Time Series Generation and Classification of MODIS Data for Land Cover Mapping

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades

der Julius-Maximilians-Universität Würzburg

vorgelegt von

René Roland Colditz

aus Würzburg

Würzburg, September 2007

Eingereicht am: 04. September 2007

1. Gutachter: Prof. Dr. Stefan Dech

2. Gutachter: Prof. Dr. Matthew C. Hansen

1. Prüfer: Prof. Dr. Stefan Dech

2. Prüfer: Prof. Dr. Jürgen Rauh

Tag der Disputation: 16. Januar 2008

Acknowledgements

Many people have helped to make this dissertation possible and supported my work during the last years. First and foremost, I am gratefully indebted to Prof. Dr. Stefan Dech (German Aerospace Center - DLR and University of Würzburg - UW), who provided the unique opportunity for a dedicated Ph.D. research with a topic almost free of choice in an inspiring environment. Besides providing these nearly perfect physical and intellectual conditions, I would like to thank Prof. Dech for his continuous support and helpful scientific advice. I am also thankful to Prof. Dr. Matthew Hansen (South Dakota State University, Brookings) for a host of trend-setting advice for my research and for being the second reviewer of my dissertation.

I am very grateful to Dr. Michael Schmidt, who provided continuous support during my Ph.D. period. As my mentor Dr. Schmidt provided several opportunities, opened doors thanks to his far-reaching connections, and gave me much helpful advice for my research and for the time beyond the Ph.D. phase.

Writing a thesis in a foreign language is a particular challenge and requires corrections and advice from several people. I gratefully acknowledge all my proof-readers, who have devoted a significant portion of their time for reading my work. I want to express my deep gratitude in particular to Prof. Dr. Detlef Busche, Dr. Christopher Conrad, and Ursula Geßner (all UW) for reading large portions of this thesis. Further proof-reading of selected parts was done by Susan Giegerich (DLR), Dr. Martin Herold (University of Jena), and Dr. Thilo Wehrmann (DLR).

Since a significant portion of my study is on South Africa, I had the opportunity to stay in this beautiful country for two months. I am very thankful to Dr. Richard Knight (University of the Western Cape - UWC, Cape Town) for his help in arranging the research stay and providing physical support as well as the contact with several counterparts. My thanks are also due to Johan Malherbe, Terry Newby, and Dawie van Zyl, (Agricultural Research Council) for providing the opportunity to visit their research facility in Pretoria. I also want to thank Prof. Dr. Bruce Hewitson and Neil MacKellar (University of Cape Town) as well as Wolfgang Lück (Council for Scientific and Industrial Research - CSIR) for fruitful discussions. Very special thanks to Gene and Russell Main (UWC) for their warm welcome and their support during my stay. Another significant visit resulting in a range of new ideas was that to the South Dakota State University (SDSU). In addition to Prof. Hansen, my thanks are due to this research group at the SDSU and Prof. Dr. David Roy in particular for their very helpful discussions and advice.

Working in a remote-sensing research group at the Department of Geography (at UW) in close relation to the German Remote Sensing Data Center (DFD at DLR) provided unique opportunities for my Ph.D. research. On the one hand, my research was integrated in the tasks and on-going projects of the DFD, and I have benefited from the knowledge of several colleagues. I want to express my gratitude to Martin Bachmann, Dr. Kurt Günther, Manfred Keil, Dr. Gerd Rücker, Dr. Günter Strunz, Dr. Thilo Wehrmann, and Dr. Klaus Wisskirchen. On the other hand, I have been able to work in a personally and spiritually pleasant academic environment, and, even more important, in a growing and very inspiring research group of fantastic colleagues. Besides science, this group also provided significant moral support, in particular during the writing phase. Therefore, I am thankful for having been a member of this remote-sensing unit. The entire current team - Dr. Christopher Conrad, Dr. Thomas Esch, Ursula Geßner, Christian Hüttich, Tanja Kraus, Dr. Tobias Landmann, Miriam Machwitz, Andreas Schenk, Gunter Schorcht, Matthias Schramm, Michael Thiel, Martin (Oggi) Wegmann, and Noellie Yao - each had their share in helping me to complete my study. Former helpful colleagues at the UW were Dr. Christoph Schultz (DLR) and Dr. Melanie Vogel (CSIR). Several students also contributed to my studies; in particular my long-term research assistant Martin Schmidt.

Very special thanks are to my family, in particular to my parents Annerose and Dr. Roland Colditz. They enabled my education and gave me their encouragement. This thesis could not have been completed without their continuous moral and financial support.

Table of Contents

Extended Abstract

Processes of the Earth's surface occur at different scales of time and intensity. Climate in particular determines the activity and seasonal development of vegetation. These dynamics are predominantly driven by temperature in the humid mid-latitudes and by the availability of water in semi-arid regions. Human activities are a modifying parameter for many ecosystems and can become the prime force in well-developed regions with an intensively managed environment. Accounting for these dynamics, i.e. seasonal dynamics of ecosystems and short- to long-term changes in land-cover composition, requires multiple measurements in time. With respect to the characterization of the Earth surface and its transformation due to global warming and human-induced global change, there is a need for appropriate data and methods to determine the activity of vegetation and the change of land cover.

Space-borne remote sensing is capable of monitoring the activity and development of vegetation as well as changes of the land surface. In many instances, satellite images are the only means to comprehensively assess the surface characteristics of large areas. A high temporal frequency of image acquisition, forming a time series of satellite data, can be employed for mapping the development of vegetation in space and time. Time series allow for detecting and assessing changes and multi-year transformation processes of high and low intensity, or even abrupt events such as fire and flooding. The operational processing of satellite data and automated information-extraction techniques are the basis for consistent and continuous long-term product generation. This provides the potential for directly using remote-sensing data and products for analyzing the land surface in relation to global warming and global change, including deforestation and land transformation.

This study aims at the development of an advanced approach to time-series generation using data-quality indicators. A second goal focuses on the application of time series for automated land-cover classification and update, using fractional cover estimates to accommodate for the comparatively coarse spatial resolution. Requirements of this study are the robustness and high accuracy of the approaches as well as the full transferability to other regions and datasets. In this respect, the developments of this study form a methodological framework, which can be filled with appropriate modules for a specific sensor and application.

The main study regions, South Africa and Germany, with very distinct environmental and structural characteristics, have served as test sites, especially for land-cover classification using time series. The study sites have been complemented with Minnesota for a special test, and Namibia for mapping vegetation types. This study has solely employed datasets of the MODIS instrument of the Terra platform operating since 2000. It provides a full global

coverage within two days and is highly suited for generating time series. The spatial resolution ranges between 250 m and 1 km. In this study land products have been utilized, including surface reflectance and NDVI.

In order to attain the first goal, time-series compilation, a stand-alone software application called TiSeG (Time Series Generator) has been developed. TiSeG evaluates the pixel-level quality indicators provided with each MODIS land product. It computes two important data-availability indicators, the number of invalid pixels and the maximum gap length. Both indices are visualized in time and space, indicating the feasibility of temporal interpolation. The level of desired data quality can be modified spatially and temporally to account for distinct environments in a larger study area and for seasonal differences. Pixels regarded as invalid are either masked or interpolated with spatial or temporal techniques.

The effectiveness of TiSeG has been demonstrated for the study region of Germany. Quality specifications ranging from lenient to strict have been used for time-series generation and have been evaluated by contextual-visual analysis and quantitative-statistical techniques. It has been found that intermediate settings perform best. Using no quality analysis has yielded an inaccurate representation of the temporal development of the surface, e.g. an NDVI-profile describing the phenological development of vegetation. On the other hand, strictest settings have resulted in an insufficient number of observations for meaningful interpolation. Therefore, it is advisable to apply the strictest setting possible, but to still retain a sufficient number of data for contextually meaningful interpolation. Expressed differently, the approach requires an appropriate balancing of data quality and data quantity.

The possibilities to evaluate time series with statistical approaches are limited. The testing of all statistical functions for assessing a time series has led to the conclusion that the strictest settings always performs best, simply due to indicating the noise component. However, as mentioned above, if the strictest quality criteria have been applied, the phenological characteristics of vegetation may be insufficiently represented. For instance, the peak of seasonality might be excluded due to a longer gap of missing observations. The evolving field of product inter-comparison seems to be the most promising alternative for evaluating the quality of time series. However, the particular characteristics of each sensor have to be considered, and a general agreement is not necessarily an indicator for correctness. Furthermore, a truly continuous validation of time series for large areas is not feasible. Even an assessment with model curves does not seem possible, due to the inherent variability of time series in time and intensity.

Considering the limitations of quantitative assessment methodologies, the visual assessment of exemplary temporal plots, together with contextual knowledge, has been found to be most appropriate for estimating the quality of time-series generation. TiSeG’s capabilities to

visualize the comparison between the original curve and the exemplary interpolation for individual pixels, a neighborhood, or small regions, have been urgently demanded. Mainly from this contextual analysis it is concluded that usually moderate quality settings perform best for generating time series. This finding has been confirmed by other studies, for instance in tropical western Africa.

The second goal, the development of an automated time-series classification procedure, has been attained by a modular design employing a supervised classification framework. Key modules of the automated classification procedure have been a multi-scale analysis and the appropriate selection of features and sample data followed by the actual classification.

The classification has been performed using an advanced decision-tree approach, which fitted into the framework of random forests and bagging (an acronym for bootstrap aggregation). However, the classification process of this study has been conducted strategically, i.e. randomization and re-use of sample data have been omitted. A soft classification approach is most suitable for better accounting for the heterogeneity of coarse-resolution cells of 500 m to 1 km. Especially the spatial heterogeneity in Germany has imposed a challenge for mapping with medium- to coarse-resolution data. Furthermore, difficulties of mapping broad transitions between classes in South Africa have been diminished by applying the concept of class-memberships for each pixel.

The multi-scale analysis has been a central module of the classification procedure. It essentially transforms a discrete high spatial resolution map into coarse-resolution cells by storing all class proportions. This fuzzy reference map serves as the baseline for a fuzzy accuracy assessment. Furthermore, homogeneous samples for training and assessing the classifier have been extracted from the multi-scale analysis. Additionally the fuzzy reference map has been used to define - from the data-perspective - a feasible classification scheme. The multi-scale analysis has been enhanced by including adjacent high-resolution pixels into the scale transformation. For a moderate increase in area, the more conservative estimate of the class proportion has yielded a clear improvement in classification accuracy of up to 10 %. If the adjacent area has become too large, accuracies have started to decrease. In this case, sample candidates insufficiently accounted for multi-modal characteristics in feature space or heterogeneous samples have been employed for training the classifier. The impact of the minimum mapping unit, inherent to many high-resolution classifications, on multi-scale analysis and image classification has also been assessed. Utilizing a dataset of Minnesota without this limitation has showed a decrease in accuracy of approximately 7 %, even for small thematic aggregations. Unfortunately, there is no alternative or possibility for correction if the minimum mapping unit had been applied initially.

As common to all supervised classification procedures, the selection of appropriate features and sample data for training a classifier has also been fundamental in this study. Time series have been transformed to metrics using simple univariate statistics, including mean and standard deviation. This significantly reduced the correlation among features, an important step for most, even advanced and robust, classifiers. However, the use of different quality levels of time series has yielded only negligible differences.

It is very important to exploit the full range of information available in the metrics record. In this respect, metrics sets of surface reflectance have performed significantly better (approximately 25 %) than using NDVI exclusively, even if a supplementary temporal segmentation has been applied. Therefore, it is more important to employ as much distinct parameters as possible, rather than generating many highly correlated derivatives of only one parameter, such as from NDVI.

Homogeneous sample candidates for training the classifier have been derived automatically by multi-scale analysis. Since decision trees have a potential bias towards classes with a higher number of training samples, an equalized sampling scheme has seemed most appropriate. However, proportional sampling as a function of the number of candidates has yielded better results of approximately 5 % in map accuracy. Therefore, stratified sampling schemes can also be recommended for decision-tree classifications, if all classes are sampled well and all modes of each class are sufficiently covered in feature space.

In order to assess the hard and soft classification results, a detailed evaluation protocol has been developed, also adopting error measures from other fields of data analysis. The fuzzy reference map in particular has provided useful information for accuracy assessment. Hard classifications have been evaluated using common accuracy statistics derived from the error matrix, which has been modified by employing the classification-certainty estimate. Fuzzy assessments have been conducted for each class and pixel separately using measures, including difference, relative and absolute mean, or RMSE.

The capabilities for classification update have been explored for both main study sites using an adoption of post-classification change detection for hard and soft classifications. Many change-detection results indicate too widespread conversions, due to uncertainties in class assignment. In this respect, change detection in this study, utilizing a highly automated classification procedure, has been challenging. For change detection using discrete classes the accompanying class-certainty estimate has been useful. It has provided the possibility of computing the change in correspondence to the uncertainty in classification. The fuzzy class estimates and hard classifications above 70 % certainty have only yielded negligible changes between consecutive years. On the other hand, clear change indicators have been derived for even temporary large-scale modifications in Germany, e.g. the severe flooding of the Elbe

River in 2002. Also, a different class composition has been shown for the exceptionally hot summer of 2003. These clear indications of negligible to subtle changes and the clear sign for known differences indirectly prove the accuracy and robustness of the classification process on the one hand, and the sensitivity of the change detection process on the other hand.

High accuracy, robustness, flexibility, and transferability have been the main requirements for the developed semi-automated to automated methodologies in this study. The transferability to other regions and datasets, in particular, has been challenging, since *a priori* knowledge must not have been included in the classification process. The regional transferability has been demonstrated for study regions of different physical-geographic settings. Although dataset transferability has not been shown directly, other studies have employed a modification of the TiSeG approach for generating an 18-year AVHRR NDVI time series for Africa. The currently implemented modules of TiSeG are suitable for MODIS products. It is planned to extend TiSeG to other advanced sensors such as MERIS and future instruments including VIIRS. The classification approach is obviously neither limited to MODIS nor to time series at all. The process is clearly data-driven, which has been demonstrated by different sets of metrics derived from time series. The selection of training data has been automated, depending on the result of multi-scale analysis and the sampling scheme. However, different modules are readily available if a high-resolution input dataset is missing. Furthermore, the fuzzy classification result enhances the interpretability of coarse-resolution maps in diverse regions. The flexibility of the classification has been demonstrated successfully by re-mapping vegetation types of Namibia, a contextually different mapping exercise.

The methods developed in this study are being and will be employed by several national and international organizations and research programs for high-quality time-series generation and accurate land-cover and vegetation mapping. In this respect, this study has a direct relation to mapping the land surface and its transformation. TiSeG, for instance, is being employed for monitoring large agricultural areas. A study on burn severity and vegetation re-growth in a Mediterranean ecosystem has employed time series produced with TiSeG. Another study employed time series of land-surface variables from TiSeG for water-balance simulation studies. In conjunction with precipitation and temperature, time series of land-surface parameters can be employed to study interdependencies, and directional patterns in multi-annual time series could be related to climate change.

The robust classification procedure allows for automated map updates and hence enables the estimation of land-cover conversion processes, including large-scale deforestation and human-induced transformation. The fuzzy class assignment in particular is important to also account for subtle changes in class composition important in heterogeneous areas, and for transition zones of ecosystems. In this respect, the classification can be readily employed for generating baseline datasets, which could be refined in a second step using regionally tuned

approaches. Therefore, the methodologies developed in this study serve for the detection of clear changes and subtle modifications of land cover in general, and vegetation in particular, using time series.

Erweiterte Zusammenfassung

Prozesse an der Erdoberfläche finden auf verschiedenen Intensitätsskalen und in unterschiedlichen Zeiträumen statt. Dabei steuert das Klima die saisonale Aktivität der Vegetation, welche in den humiden Mittelbreiten hauptsächlich durch die Temperatur bestimmt wird. In semi-ariden Gebieten hingegen ist die Verfügbarkeit von Wasser als Haupteinflussfaktor für das Pflanzenwachstum zu betrachten. Andererseits greift auch der Mensch modifizierend in das Ökosystem ein. Dies gilt insbesondere für die stark besiedelten und intensiver genutzten Räume der Erde, in denen die Umwelt nahezu ausschließlich durch den Menschen gesteuert wird. Zur Beurteilung dieser Dynamiken, sowohl der natürlichen saisonalen Muster als auch der kurz- bis langfristigen Änderungen der Landschaft, ist die Aufnahme einer Vielzahl von Messungen über eine längere Periode erforderlich. Insbesondere im Zusammenhang mit der Charakterisierung der Landoberfläche und deren Veränderung im Rahmen der Erderwärmung aber auch des wachsenden Einflusses des Menschen auf die Umwelt besteht somit ein Bedarf an geeigneten Daten und Methoden zur Bestimmung der jährlichen Aktivität von Vegetationseinheiten und der wiederholbaren Kartierung der Landoberfläche.

Die Satellitenfernerkundung ist in der Lage, durch Messung von Strahlung die Aktivität der Vegetation zu bestimmen sowie die Klassifikation der Landoberfläche abzuleiten. In vielen Fällen sind Satellitenaufnahmen die einzige Möglichkeit, große Flächen der Erde umfassend und einheitlich zu beurteilen. Dabei kann durch eine Vielzahl aufeinander folgender Aufnahmen, d.h. eine Zeitreihe aus Satellitendaten, die Entwicklung der Vegetation in Raum und Zeit beobachtet werden. Zeitreihen bieten das Potential, Veränderungen der Landoberfläche über mehrere Jahre zu dokumentieren und somit Prozesse sowohl hoher als auch niedriger Intensität abzuleiten. Neben diesen gerichteten Veränderungen können auch plötzliche Ereignisse, wie z.B. Hochwasser oder Brände, mit Zeitreihen erfasst und in Bezug auf normale Bedingungen ausgewertet werden. Insbesondere die operationelle Prozessierung der Satellitendaten und die automatisierte Ableitung von Informationen bilden die Basis für konsistente und kontinuierliche Produkte über längere Zeiträume. Somit besteht das Potential, die Ergebnisse direkt in die Erforschung der Landoberfläche und deren Veränderung, z.B. durch die Erderwärmung, Walddegradation, oder die Nutzung vormals natürlicher Flächen, einzubinden.

Diese Dissertation befasst sich mit der Entwicklung von Methoden zur Zeitreihengenerierung unter Verwendung der Qualitätsindikatoren einzelner Aufnahmen. Ein zweites Ziel der Arbeit ist die Anwendung der optimierten Zeitreihen zur automatisierten und reproduzierbaren Kartierung der Landoberfläche, wobei unscharfe Klassifikationsverfahren zur genaueren

Charakterisierung der räumlich nur grob aufgelösten Daten eingesetzt werden. Damit erfordert diese Arbeit sowohl Robustheit der eingesetzten Methoden als auch eine hohe Genauigkeit der Ergebnisse. Ebenso maßgeblich ist die Übertragbarkeit der Verfahren, einerseits auf verschiedene Regionen als auch auf verschiedene Datensätze. Daher sind die hier vorgenommenen Entwicklungen als ein Rahmen zu verstehen, der je nach Sensor oder Anwendung mit verschiedenen Modulen besetzt werden kann.

Das in dieser Arbeit eingesetzte Aufnahmesystem MODIS ermöglicht innerhalb von maximal zwei Tagen eine globale Abdeckung der Erde mit Aufnahmen in 36 spektralen Bändern. Insgesamt existieren zwei MODIS-Sensoren, die sich jeweils auf der Terra- und Aqua-Plattform befinden und Daten seit 2000 bzw. 2002 liefern. Die räumliche Auflösung beträgt 250 m bis 1 km. In dieser Arbeit wurden vorverarbeitete Komposite und abgeleitete Produkte der MODIS Land Daten (z.B. Reflexionsgrade oder NDVI) eingesetzt.

Die Hauptuntersuchungsgebiete Südafrika und Deutschland dienen insbesondere als Vergleichsregionen zur Klassifikation unterschiedlicher klimatischer Eigenschaften. Südafrika ist charakterisiert durch humide bis aride Bereiche und große Übergangszonen zwischen Landbedeckungen. Deutschland hingegen ist ein sehr heterogener Raum und stellt somit eine besondere Herausforderung für eine räumliche Auflösung zwischen 500 m und 1 km dar. Diese Hauptuntersuchungsgebiete wurden durch Minnesota und Namibia als besondere Testgebiete für bestimmte Dateneigenschaften bzw. die Eignung zur Klassifikation von Vegetationstypen ergänzt.

Zum Erreichen des ersten Zieles, der Zeitreihengenerierung, wurde das eigenständige Softwareprodukt TiSeG (Time Series Generator) entwickelt. TiSeG dient der Auswertung der Qualitätsindikatoren die mit jedem MODIS-Produkt für terrestrische Applikationen zur Verfügung gestellt werden. Dabei werden in Bezug auf die Generierung von Zeitreihen zwei Indizes der Datenverfügbarkeit ermittelt: erstens die Anzahl der ungültigen Pixel und zweitens die längste zeitliche Datenlücke. Beide Indizes werden räumlich und zeitlich dargestellt und geben so dem Bearbeiter die Information, ob mit den aktuellen Qualitätsangaben die Generierung einer sinnvollen Zeitreihe durch zeitliche Dateninterpolation möglich ist. Die Qualitätseinstellungen können sowohl zeitlich als auch räumlich angepasst werden. Eine zeitliche Änderung kann beispielsweise für bestimmte Jahreszeiten sinnvoll sein. Räumlich unterschiedliche Qualitätseinstellungen eignen sich für größere Untersuchungsgebiete mit differenzierten physisch-geographischen Charakteristika. Als ungültig betrachtete Pixel können durch einen Fehlwert maskiert oder durch zeitliche und räumliche Interpolation neu errechnet werden.

Die Auswirkungen verschiedener Qualitätseinstellungen auf die Zeitreihengenerierung mit TiSeG wurden am Beispielraum Deutschland für den NDVI aufgezeigt und sowohl inhaltlich-

visuell als auch statistisch-quantitativ ausgewertet. Tests zeigten, dass moderate Einstellungen die besten Ergebnisse liefern; ein Resultat, welches auch schon andere Studien, z.B. in Westafrika, ergeben haben. Wenn keine oder nur sehr schwache Qualitätsangaben eingestellt wurden, zeigte die Zeitreihe einen inakzeptablen Verlauf des NDVI. Zum Beispiel werden bei Wolkenbedeckung immer wieder deutlich zu niedrige Werte in die Zeitreihe übernommen, welche den phänologischen Jahresverlauf von Vegetationseinheiten nur unzureichend wiedergeben. Eine sehr rigide Einstellung der Datenqualität andererseits eliminierte zu viele Beobachtungen. Damit fehlen eine genügende Anzahl und eine hinreichende zeitliche Verteilung von Stützwerten zur thematisch sinnvollen zeitlichen Interpolation. Aus diesen Ergebnissen kann die Schlussfolgerung gezogen werden, dass immer die strengstmögliche Qualitätseinstellung gewählt werden sollte, die noch hinreichend viele Stützwerte zur sinnvollen Interpolation der Datenlücken zur Verfügung stellt. Anders formuliert, die Methode zur Zeitreihengenerierung erfordert eine Abwägung zwischen Datenqualität und hinreichender Datenquantität zur Interpolation.

Die Möglichkeiten zur Bewertung der Zeitreihen mit statistischen Methoden sind begrenzt. Statistische Verfahren, die ohne Zusatzwissen direkt auf den Zeitreihen angewendet werden, bewerten die strengsten Qualitätsangaben immer als die beste Zeitreihe. Dieses Ergebnis resultiert aus den statistischen Methoden selbst, die immer den durch die Interpolation geringeren Rauschanteil positiv bewerten. Allerdings muss eine Interpolation von Daten, wie oben schon erläutert, nicht notwendigerweise ein inhaltlich richtiges Ergebnis erzeugen. Vielmehr können möglicherweise in phänologiebeschreibenden Parametern von Vegetation, beispielsweise der NDVI, die relevanten Zeitpunkte durch Interpolation der Daten nicht korrekt erfasst werden. Zum Beispiel wird der Beginn oder das Maximum der Vegetationsperiode nur unzureichend ermittelt, wenn in dem entsprechenden Zeitraum über ein längeres Zeitintervall keine Daten mit ausreichender Qualität zur Verfügung standen.

Alternativ zur statistischen Bewertung der Zeitreihe selbst besteht derzeit ein starker Trend zu Produktvergleichen verschiedener Sensoren. Jedoch müssen auch hier Einschränkungen getroffen werden, da jedes Produkt auch von dem jeweiligen Ausgangsdatensatz und damit dem Sensor abhängt. Des Weiteren ermöglicht der Vergleich mit einem ähnlichen Produkt noch nicht den Bezug zur Realität. Es ist durchaus denkbar, dass ein Produktvergleich ein gutes Ergebnis liefert, der zeitliche Verlauf aber trotzdem nicht akkurat wiedergegeben wird, weil die Abweichung von der Realität in beiden Produkten enthalten ist. Möglichkeiten zur eigentlichen kontinuierlichen und großflächigen Validierung sind aktuell nicht gegeben und werden auch in naher Zukunft nicht zur Verfügung stehen. Auch der Einsatz von so genannten Referenzkurven, ähnlich zu Feldspektren in der hyperspektralen Fernerkundung, erscheint als keine Alternative. Dies liegt schon allein an der hohen Variabilität im zeitlichen Kurvenverlauf aber auch im unterschiedlichen Intensitätsniveau eines jeden Parameters.

Selbst in Regionen mit verhältnismäßig stabilen phänologischen Mustern, wie z.B. in Deutschland, variiert das Einsetzen der Vegetationsperiode um bis zu zwei Wochen. Auch die Intensität kann je nach Temperaturverlauf aber auch der Wasserverfügbarkeit in den Sommermonaten beträchtlich schwanken, wie Studien zum außergewöhnlich warmen Sommer 2003 deutlich gezeigt haben.

Daher wurde in dieser Dissertation die visuelle Bewertung der Zeitreihe zusammen mit dem Kontextwissen des Bearbeiters durchgeführt. TiSeG unterstützt diesen Prozess, indem es neben der Visualisierung der Datenverfügbarkeitsindizes auch den antizipierten zeitlichen Kurvenverlauf nach erfolgreicher Interpolation im Vergleich zur Originalkurve darstellen kann. Diese Beispielinterpolationen für eine Qualitätseinstellung können für jeden Pixel, eine Pixelumgebung oder eine kleine räumliche Einheit dargestellt werden. Aus dieser inhaltlich-visuellen Bewertung resultiert das Ergebnis, dass moderate Qualitätseinstellungen und nachfolgende Interpolationen die generellen Kurvencharakteristika am besten darstellen.

Eine weitere Analyse verschiedener Versionen des MODIS Vegetationsindex-Produkts zeigte klare Verbesserungen in der aktuellen Version. Durch eine Neugruppierung der Qualitätsindikatoren wurden z.T. markante Unterschiede in der Wolkenmaskierung festgestellt. Des Weiteren hat diese Analyse ergeben, dass der EVI, vermutlich auch durch die Unterstützung geeigneter Alternativalgorithmen, einen generell stabileren Kurvenverlauf darstellt, z.T. sogar wenn keine Qualitätsanalyse durchgeführt wurde.

Das zweite Ziel dieser Arbeit ist die automatische Klassifikation von Zeitreihen. Hierzu wurde ein modulares Verfahren der überwachten Klassifikation entwickelt. Aufgrund der groben räumlichen Auflösung von MODIS-Daten erschien es besonders wichtig, ein unscharfes Verfahren aufzubauen, das die Heterogenität der Klassen in vielen Räumen besser abbilden kann. Dabei wurde der Klassenanteil eines jeden Pixel ermittelt. Für das Untersuchungsgebiet Südafrika wurde erwartet, dass die weit ausgedehnten Übergangszonen zwischen natürlichen Landbedeckungseinheiten besser kartiert werden können. Im zweiten Untersuchungsgebiet Deutschland ist die Landschaft zwar deutlich stärker strukturiert, allerdings sind viele Flächen deutlich kleiner als die Auflösung eines MODIS-Pixel, was eine besonders hohe Anforderung an einen Klassifikator stellt. Hier kann die anteilige Darstellung der Landbedeckungsklassen die feine Gliederung der Landoberfläche besser repräsentieren.

Die Schlüsselmodule zur erfolgreichen Durchführung der Klassifikation waren eine Multiskalenanalyse und die geeignete Auswahl von Merkmalen und Stichproben zum Trainieren des Klassifikators. Der eigentliche Klassifikationsschritt wurde durch eine Erweiterung des Entscheidungsbaumklassifikators durchgeführt. Diese Erweiterung kann in den bestehenden Rahmen von „random forest“ und „bagging“ (ein Akronym für „bootstrap aggregation“) eingeordnet werden. Jedoch wurden die dort angewendeten Verfahren in dieser

Arbeit zu einem deutlich strategischen Vorgehen modifiziert, d.h. es wurde auf das Prinzip der Zufallsauswahl und der Wiederverwendung von Stichproben (Ziehen mit Zurücklegen) verzichtet.

Zur Beurteilung der diskreten und unscharfen Klassifikationen wurde ein detailliertes Protokoll entwickelt. Dabei hat insbesondere eine unscharfe Referenzkarte wichtige Aufgaben erfüllt. Die diskrete Klassifikation wurde mit üblichen Genauigkeitsstatistiken aus Kontingenzmatrizen bewertet. Allerdings konnte in einer Erweiterung des gewöhnlichen Beurteilungsverfahrens die Sicherheit der diskreten Klassenangabe in die Bewertung eingearbeitet werden. Somit kann je nach angestrebter Sicherheit der Klassenzuweisung ein korrespondierendes Gütemaß für die Klassifikation im Vergleich zu einer diskreten Referenzkarte ermittelt werden. Die Bewertung von unscharfen Klassifikationen ist in der Fernerkundung nicht sehr verbreitet und erforderte die Adaption von Fehlermaßen aus anderen Bereichen der Validierung. Dabei wurde insbesondere die Differenz zwischen Klassifikation und Referenz sowie daraus resultierende Abweichungsmaße, wie relativer (gerichtete Tendenz; mit Vorzeichen) und absoluter (ungerichtete Magnitude; ohne Vorzeichen) Mittelwert und Schwankung berücksichtigt. Weitere Maße, wie die mittlere quadratische Abweichung und der Korrelationskoeffizient, vervollständigten die Analyse der Genauigkeit. Die unscharfe Klassifikation wurde in zwei voneinander unabhängigen Schritten bewertet: Erstens mit Fehlermaßen für die Korrespondenz der Anteile einer jeden Klasse mit der Referenz und zweitens mit denselben Maßen für den Zusammenhang aller Klassenanteile eines jeden Pixels mit der entsprechenden Referenz.

Die Multiskalenanalyse ist ein zentrales Modul des in dieser Arbeit entwickelten Klassifikationsprozesses. Dieses Verfahren transformiert eine räumlich hoch aufgelöste Eingangsklassifikation in die gröbere Auflösung der Zeitreihen. Dabei werden nicht die Werte der Pixel als solche transformiert, wie dies in einem Resampling üblich wäre. Vielmehr werden hier die einzelnen Anteile der Klassen proportional gespeichert, wodurch ein unscharfes Bild der Klassenanteile eines grob aufgelösten Pixels in Bezug auf die vorherige hoch aufgelöste Referenz entsteht. Damit dient dieses unscharfe Bild der Klassenanteile als Referenz zur Beurteilung der Ergebnisse der Klassifikation. Basierend auf diesem unscharfen Referenzbild werden ebenfalls die potentiellen Trainings- und Testpixel ermittelt. Diese Flächen müssen entsprechend des Verfahrens zur Stichprobenerhebung (siehe weiter unten) eine hinreichend hohe Anzahl und eine hohe Homogenität aufweisen. In dieser Arbeit wurde immer ein Stichprobenumfang von mindestens 1100 Pixel pro Klasse angestrebt. Das automatische schrittweise Verfahren beginnt bei einer Homogenität von 100 % und wird bei einem unzureichenden Umfang der Stichprobe immer um 1 % erniedrigt. Ein ähnliches Prinzip wurde zum Erstellen des Klassifikationsschemas angewendet. Auch hier wurde die Homogenität der Trainingspixel betrachtet. Bei unzureichender Datenmenge wurden

inhaltlich ähnliche Klassen zusammen gefügt. Somit basieren die Klassifikationsschlüssel auf einer hinreichend großen Datenmenge sowie einer kontextuellen Beurteilung der Klassen.

Im Rahmen der Untersuchung zum Einfluss der Komponenten des Klassifikationsverfahrens wurde das Konzept der Multiskalenanalyse erweitert. Hierzu wurden nicht nur die Pixel der hoch aufgelösten Daten betrachtet, die direkt in einem grob aufgelösten Pixel liegen, sondern auch die entsprechenden Nachbarschaften. Die räumliche Vergrößerung des Analysebereiches lieferte eine konservativere Abschätzung der Homogenität der Oberfläche. Für moderate Nachbarschaftsgrößen (bis zu einem halben groben Pixel) konnte eine Verbesserung der Klassifikationsgenauigkeit bis zu 10 % festgestellt werden. Wenn jedoch der Nachbarschaftsbereich zu weit ausgedehnt wurde, nahm die Genauigkeit wieder ab. Diese Umkehr liegt an der deutlichen Limitierung des potentiellen Stichprobenumfangs zum Trainieren des Klassifikators. Einerseits wurde durch das vollautomatische Verfahren die potentielle Trainingsdatenmenge eingeschränkt, d.h. es konnten nicht mehr für alle Regionen die spezifischen Klassenausprägungen abgeleitet werden, weil die entsprechenden Flächen zu klein wurden und somit unter den Schwellenwert der Homogenität fielen. Andererseits bestand auch die Möglichkeit, dass heterogene Flächen in das Klassifikationsverfahren einbezogen werden, wenn die Anzahl homogener Kandidaten generell zu gering wurde. Eine weitere limitierende Einflussgröße auf die Multiskalenanalyse und damit die Klassifikation ist die kleinste Kartierungseinheit, ein häufig benutzter Parameter zur Reduktion des Klassifikationsrauschens in räumlich hoch aufgelösten Produkten. Zur Untersuchung des Einflusses wurde ein Datenset für Minnesota genutzt, welches diese Restriktion nicht aufweist. Es wurde selbst bei geringen Aggregationsstufen eine Verringerung der Genauigkeit um bis zu 7 % festgestellt. Allerdings lässt sich dieser in vielen räumlich hoch aufgelösten Klassifikationen inhärente negative Effekt nicht korrigieren.

Zwei weitere wichtige Einflussgrößen auf die Genauigkeit des Klassifikationsprodukts waren die geeignete Auswahl von Merkmalen und von Trainingsdaten. Diese beiden Parameter bedingen auch bei anderen Verfahren die Güte des Produkts und sind damit für vollautomatische Klassifikationen eine besondere Herausforderung. Üblicherweise werden Zeitreihen nicht direkt als Merkmale für die Klassifikation eingesetzt. Dies liegt vorrangig an der hohen zeitlichen Korrelation, die sich sowohl negativ auf die Auswahl der Merkmale zur Klassentrennung als auch auf die Robustheit des Klassifikators bei Anwendung auf den Gesamtdatensatz auswirken kann. Viele Studien haben erfolgreich so genannte Maßzahlen zur Landbedeckungsklassifikation mittels Zeitreihen eingesetzt. Dabei werden über die Zeit oder einzelne Zeitabschnitte einfache, meist univariate Statistiken (z.B. Mittelwert und Schwankung) für jeden Parameter errechnet. Allerdings führt diese Transformation auch zu kaum feststellbaren Unterschieden in der Qualität der Zeitreihen selbst. Im Zusammenhang mit früheren Studien, häufig basierend auf AVHRR-Zeitreihen mit hohem Rauschanteil,

konnte mit dieser Sensitivitätsanalyse festgestellt werden, dass die Qualität der Originaldaten keinen oder bei sehr verrauschten Zeitserien nur einen vernachlässigbar geringen Einfluss auf die Klassifikationsgüte hat, wenn die Transformation mittels der genannten simplen Maßzahlen vorgenommen wurde.

Bei den Merkmalen selbst wurden der Einfluss der zeitlichen Segmentierung und die Auswahl der Zeitreihenparameter (Reflexionsgrade, NDVI und gegebenenfalls Landoberflächentemperatur oder ausschließlich NDVI) analysiert. Es wurde festgestellt, dass die zeitliche Segmentierung für den Parameter NDVI einen Genauigkeitsgewinn von 10 % erzielt. Die Genauigkeit durch zeitliche Segmentierung wurde bei den Reflexionsgraden (z.B. 7 Bänder und NDVI) nicht erhöht. Dies resultiert aus dem sowieso höheren Informationsgehalt von Reflexionsgraden gegenüber einem einzelnen Parameter. In einem weiteren Vergleich zum ausschließlich genutzten NDVI wurde die Klassifikationsgüte um 25 % verbessert, wenn die Reflexionsgrade hinzugenommen wurden. Dieses Ergebnis zeigt sehr deutlich den Einfluss der Merkmalsauswahl auf die Genauigkeit der Klassifikation. Es kommt nicht darauf an, möglichst viele Derivate aus einem Parameter abzuleiten, weil diese Information zumeist hoch redundant ist. Es ist vielmehr von besonderer Wichtigkeit, so viel wie möglich an unterschiedlicher Information zu gewinnen, um die Ausprägung einer Klasse im Merkmalsraum optimal beschreiben zu können.

Die andere Einflussgröße, das Auswahlverfahren der Trainingsgebiete, ist zum Teil schon im Zusammenhang mit der Multiskalenanalyse besprochen worden. Ein wichtiges Kriterium neben der generellen Eignung und Verlässlichkeit von Trainingsgebieten ist die Strategie der Stichprobenziehung. Insbesondere Entscheidungsbäume gelten als besonders sensibel, wenn eine Klasse mit überproportional vielen Trainingsmengen in das Partitionierungsverfahren eingeführt wird. Durch das anteilige Auswahlverfahren, basierend auf der Menge an Stichproben pro Klasse und in Bezug zu den anderen Klassen, wird eine Überschätzung von größeren Stichproben für eine Klasse angenommen. Daher wurde in dieser Arbeit grundsätzlich die gleiche Stichprobenmenge für alle Klassen (1100 pro Klasse) abgeleitet. Testweise wurden diesem Verfahren zwei proportionale Varianten des Ziehens der Stichprobe gegenübergestellt (linear und logarithmisch), was eine skalierte Variante des generellen Verfahrens zur zufällig stratifizierten Stichprobenauswahl darstellt. Beide proportionale Auswahlverfahren lieferten überraschenderweise etwas bessere Ergebnisse, mit einem Genauigkeitsgewinn von 5 %. Allerdings muss darauf geachtet werden, dass alle Klassenausprägungen im Merkmalsraum berücksichtigt werden und die Stichprobenmenge für anteilig geringe Klassen nicht zu klein wird.

Das in dieser Arbeit entwickelte automatische Klassifikationsverfahren dient neben der genauen Bewertung einzelner Klassifikationskomponenten vor allem der wiederholten Anwendbarkeit ohne Anpassungen in regelmäßigen, z.B. jährlichen Abständen. Dazu wurde

für die beiden Untersuchungsgebiete die Klassifikation für mehrere Jahre durchgeführt. Eine Variante der Post-Klassifikations-Veränderungskartierung wurde zur Bewertung der Ergebnisse von immer zwei aufeinander folgenden Jahren angewandt. Dies weist eine besondere Schwierigkeit auf, da Veränderungsdetektionen häufig aufgrund geringer Klassifikationssicherheiten eine überproportionale Veränderung der Landoberfläche indizieren, die in der Realität zumeist nicht anzutreffen ist. Die in dieser Arbeit entwickelte Methode hat den Vorteil, dass die Veränderungskartierung auf der diskreten und auf der unscharfen Klassifikation zweier Jahre möglich ist. Dabei liefert die Klassifikationssicherheit, als ein Derivat der Transformation von unscharfen Ergebnissen zu diskreten Klassen, ein wichtiges Indiz für die Glaubwürdigkeit der Klassenzuweisung. Es kann somit der Grad der Veränderung in Abhängigkeit zur Sicherheit der Klassifikation ermittelt werden. Für Südafrika und auch für Deutschland werden ab einer Klassifikationssicherheit von 70 % kaum Veränderungen festgestellt. Damit zeigen sich hier auch indirekt die Robustheit und auch die Genauigkeit der Klassifikationen im Vergleich zweier Jahre zueinander.

Interessant ist neben der geringen Veränderung auch die Ähnlichkeit der Abhängigkeitskurven für verschiedene Jahre. Nur für Deutschland wurden in Vergleichen mit dem Jahr 2003 Abweichungen festgestellt. Eine detaillierte Analyse der Veränderungskartierung mit der unscharfen Klassifikation verdeutlichte, dass zwar keine Veränderung der Landschaft als solche eingesetzt hat, allerdings eine von anderen Jahren verschiedene Klassenzusammensetzung pro Pixel angegeben wurde. Diese Differenz beruht auf dem Einfluss des Hitzesommers 2003 auf die Ausprägung der Klassen, insbesondere Ackerland. Ebenso interessant ist die Detektion von großflächigen, wenn auch nur temporären Veränderungen. So wurde zum Beispiel das Hochwasser an der Elbe im Sommer 2003 durch eine deutlich abweichende Klassenzusammensetzung im Bezug zum Vorjahr wiedergegeben. Die auf der einen Seite hohe Koinzidenz der Klassifikationen aufeinander folgender Jahre und die auf der anderen Seite klare Indikation von subtilen und temporären Veränderungen verdeutlichen die Robustheit aber auch hinreichende Sensitivität und insgesamt die Eignung des Verfahrens.

In Bezug auf die anfangs geschilderten Anforderungen, Robustheit, Genauigkeit und Übertragbarkeit, kann an dieser Stelle festgestellt werden, dass sich die in dieser Arbeit entwickelten Methodiken als geeignet erwiesen haben. Insbesondere die Übertragbarkeit auf andere Regionen und Daten war eine große Herausforderung, da hierdurch kein zusätzliches *a priori* Wissen außer den Trainingsdaten zum überwachten Klassifizieren benutzt werden konnte. Die regionale Übertragbarkeit ist für mehrere Untersuchungsräume mit sehr unterschiedlichen physisch-geographischen Eigenschaften demonstriert worden. Obwohl TiSeG derzeit nicht direkt auf andere Datensätze außer MODIS angewendet wird, bildet die zugrunde liegende Idee der Auswertung von Qualitätsdaten zur Zeitreihengenerierung sowie

der entwickelte Rahmen der Software die Möglichkeit der Erweiterung, z.B. auf Daten von MERIS und auf das zukünftige VIIRS-Instrument. Des Weiteren ist das grundlegende Konzept der Qualitätsauswertung in einem separaten Prozessor für AVHRR NDVI Daten zu einem vollautomatischen, schrittweise interpolierenden Verfahren erweitert worden. Das in dieser Arbeit vorgestellte modulare Klassifikationsverfahren erfordert keine besonderen Eingangsdaten, wie z.B. bestimmte MODIS Zeitreihen. Dies wurde durch den Gebrauch unterschiedlicher Eingangsdaten zur Generierung der Maßzahlen bei der Sensitivitätsanalyse bestätigt. Damit ist der Prozess weder auf MODIS-Zeitserien noch auf Zeitreihen generell beschränkt. Der gesamte automatische Klassifikationsprozess ist datengesteuert, sowohl was die zu klassifizierenden Daten als auch die Trainingsdaten angeht. Die Unschärfe in den Ergebnissen ermöglicht die detaillierte Auswertung der Klassenzusammensetzung, was ein besonders wichtiger Aspekt bei grob aufgelösten Datenprodukten und deren Anpassungsfähigkeit auf andere Anwendungen ist. Diese Flexibilität auf Anwendungen außerhalb der Landbedeckungskartierung ist abschließend beispielhaft für die Re-Klassifikation von Vegetationstypen in Namibia demonstriert worden. Die für aride und semi-aride Regionen charakteristische Verteilung der Vegetationseinheiten konnte mit dem hier vorliegenden Verfahren mit sehr zufrieden stellender Genauigkeit erkannt werden. Zusätzlich zeigen die Ergebnisse deutlich die breiten Übergangszonen der Vegetationseinheiten als ein typisches Charakteristikum für semi-aride Räume, welches durch diskrete Klassifikationsergebnisse nicht aufgedeckt werden kann.

Die hier entwickelten Methoden sollen bzw. werden von verschiedenen nationalen und internationalen Organisationen, Behören und Programmen zur qualitativ hochwertigen Datenverarbeitung von Zeitreihen und Ableitung von Landbedeckung sowie Vegetationstypen eingesetzt. Damit steht diese Arbeit in direktem Bezug zur Erforschung der Landoberfläche und deren Veränderungen.

Gerade die hochgenaue Erstellung des phänologischen Verlaufs von Vegetationseinheiten durch TiSeG ist ein vielfach gefragtes Produkt, beispielsweise im Bereich des Monitorings von großflächigen Anbaugebieten und zukünftig möglicherweise auch der Abschätzung von Ernteerträgen. Durch TiSeG erstellte Zeitreihen wurden ebenfalls erfolgreich zur Beurteilung der Waldbrandintensität und der Erholung der Brandflächen in mediterranen Räumen eingesetzt. Eine weitere Anwendung der durch TiSeG produzierten Zeitreihen von Landoberflächenparametern diente der Simulation der Wasserbilanz im westlichen Afrika. Des Weiteren kann und wird TiSeG zur Analyse von Anomalien eingesetzt, die sich in Bezug auf die vorherigen bzw. nachfolgenden Jahre entweder als einmaliges Ereignis oder als gerichtete Veränderung darstellen können. Gekoppelt mit anderen Datensätzen, wie z.B. Niederschlag und Temperatur, kann ein Eindruck in die Wirkungsweise von Ökosystemen und die gegenseitigen Abhängigkeiten der einzelnen Komponenten gewonnen werden. Damit

wird mittels der großflächigen Zeitreihen nicht nur ein besseres Raum- und Prozessverständnis geschaffen, sondern auch ein direkter Beitrag zur Erforschung des Klimawandels auf die Umwelt und den Menschen geleistet.

Neben diesen subtilen Prozessen können insbesondere durch die robuste Klassifikation automatisch Veränderungen der Landoberfläche, z.B. Entwaldung oder die Nutzbarmachung vormalig naturnaher Flächen durch ein jährliches Update, erfasst werden. Insbesondere bei der Veränderungskartierung der Landoberfläche ist die grundlegende Übereinstimmung ein wichtiges Indiz für die Korrektheit beider verglichener Klassifikationen und der Eignung des Vergleichsverfahrens. Dabei ermöglicht gerade die unscharfe Kartierung ebenfalls die Erfassung geringer oder auch nur saisonaler Veränderungen der Erdoberfläche. In Bezug auf ein einzelnes Jahr kann die Unschärfe sowohl die Klassenzusammensetzung kleinräumiger Strukturen als auch weiträumige Übergangszonen natürlicher Ökosysteme hinreichend abbilden. Damit eignet sich dieses Klassifikationsverfahren zum Generieren von Basisprodukten, die durch weitere Module auf die regionalen Bedingungen und Besonderheiten angepasst und verbessert werden können. Die in dieser Arbeit entwickelten Methoden dienen somit der genauen Erstellung von Produkten zur Charakterisierung der Erdoberfläche und können beispielsweise zur Detektion sowohl deutlicher Veränderungen als auch nuancierter Modifikationen der Landbedeckung einen Beitrag leisten.

List of Figures

List of Tables

Glossary

6S	Second simulation of the satellite signal in the solar spectrum
ACTS	Automatic classification of time series
AdaBoost	Adaptive boosting algorithm
AERONET	Aerosol robotic network
AFRICOVER	Land cover classification of Africa
ANN	Artificial neural network
ANOVA	Analysis of variance
ARMA	Autoregressive moving-average
ARTMAP	Adaptive resonance theory model
ARVI	Atmospherically resistant vegetation index
ASTER	Advanced spaceborne thermal emission and reflection radiometer
ATBD	Algorithm theoretical background document
AVHRR	Advanced very high resolution radiometer
AWiFS	Advanced wide field sensor
BEAM	Software for ENVISAT data
BELMANIP	Benchmark land multi-site analysis and inter-comparison of products
BISE	Best index slope extraction
BOREAS	Boreal ecosystem - atmosphere study
BRDF	Bidirectional reflectance distribution function
C4	Collection 4 (version 4) of MODIS data
C4.5	Decision tree algorithm, successor of ID3
C5	Collection 5 (version 5) of MODIS data
C5.0	Decision tree algorithm for UNIX, successor of C4.5
CART	Classification and regression trees
CEOS	Committee on earth observation satellites
CERES	Clouds and earth’s radiant energy system
CHAID	Chi-square automatic interaction detectors
CLAVR	Clouds from AVHRR
CLC	CORINE land-cover classification
CMG	Climate modeling grid
CORINE	Coordination of information on the environment
CPU	Central processing unit
CV-MVC	Constrained view maximum value composite
CZCS	Coastal zone color scanner
DAAC	Distributed active archive center

DIVERSITAS	An international programme of biodiversity science
ENSO	El Niño Southern Oscillation
ENVISAT	Environmental satellite
EOS	Earth observing system
EOS AM-1	First EOS spacecraft (later renamed to Terra), crosses equator at 10:30am
EOS PM-1	Second EOS spacecraft (later renamed to Aqua), crosses equator at 2:30pm
ESE	Earth science enterprise
ESSP	Earth systems science partnership
ETM^+	Enhanced thematic mapper plus
EVI	Enhanced vegetation index
EVI2	Enhanced vegetation index, backup algorithm
FACT	Fast algorithm for classification trees
FIFE	First ISLSCP field experiment
FLUXNET	Global network of microclimatological towers
FPAR	Fraction of photosynthetically active radiation
FRA	Forest resources assessment
GAC	Global area coverage
GES-DISC	Goddard earth science data and information services center
GHOST	Global hierarchical observing strategy
GIMMS	Global inventory modeling and mapping studies
GIOVANNI	GES-DISC interactive online visualization and analysis infrastructure
GIS	Geographical information system
GLC2000	Global land cover classification 2000
GLCC	Global land cover characterization
GLFC	Global land cover facility
GLOBCOVER	Global land cover from MERIS
GOES	Geostationary operational environmental satellite
GOFC-GOLD	Global observation of forest and land cover dynamics
GVI	Global vegetation index
HANTS	Harmonic analysis of time series
HDF	Hierarchical data format
HRV	High resolution visible
HRVIR	High resolution visible and infrared
ID3	Iterative dichotomiser 3
IDL	Interactive data language
IFOV	Instantaneous field of view
IGBP	International geosphere-biosphere programme

IHDP	International human dimensions programme on global environmental change
IMAGE2000	Pan-European Landsat mosaic of approximately year 2000
IRS	Indian remote sensing satellite
ISLSCP	International satellite land surface climatology project
ISODATA	Iterative self-organizing data analysis technique
ITCZ	Inner-tropical convergence zone
KHAT	Kappa coefficient
LAC	Local area coverage
LAI	Leaf area index
LBA	Large-scale biosphere - atmosphere experiment in Amazonia
LCCS	Land cover classification system
LDOPE	Land data operational product evaluation
LISS	Linear imaging self-scanning systems
LP DAAC	Land processes distributed active archive center
LPV	Land product validation
LST	Land surface temperature
LTDR	Long term data record
LTER	Long term ecological research
MCD	MODIS combined product (merge of Terra and Aqua data)
MERIS	Medium resolution imaging spectrometer
MESMA	Multiple endmember spectral mixture analysis
METEOSAT	Meteorological satellite
MIR	Middle Infrared
MISR	Multi-angle imaging spectroradiometer
MLC	Maximum likelihood classification
MMU	Minimum mapping unit
MOD	(1) MODIS Terra product, (2) often also indicates MODIS products in general without reference to platform
MODLand	MODIS land product
MODIS	Moderate resolution imaging spectroradiometer
MODTRAN	Moderate resolution transmittance
MOPITT	Measurement of pollution in the troposphere
MOVAS	MODIS online visualization and analysis system
MRT	MODIS reprojection tool
MVC	Maximum value composite
MW	Microwave
MYD	MODIS Aqua product
NASA	National aeronautics and space administration

NBAR	Nadir BRDF-adjusted surface reflectance
NCEP	National center for environmental prediction
NDVI	Normalized difference vegetation index
NIR	Near infrared
NLC	National land cover classification of South Africa
NLCD	National land cover database of the United States of America
NOAA	National oceanic and atmospheric administration
NPOESS	National polar-orbiting operational environmental satellite system
NPP	(1) Net primary productivity, (2) NPOESS preparatory project
OrbView2	Satellite where SeaWiFS is mounted
PAL	Pathfinder AVHRR land
PELCOM	Pan-european land use and land cover monitoring
QA	Quality assurance
QA-SDS	Quality assurance science data set
QUEST	Quick unbiased efficient statistical tree
RAM	Random access memory
RMSE	Root mean square error
RW	Radiowave
SAFARI	Southern Africa regional science initiative
SARVI	Soil-adjusted and atmospherically resistant vegetation index
SAVI	Soil-adjusted vegetation index
SCF	Science computing facilities
SDS	Science data set
SeaWiFS	Sea-viewing wide field-of-view sensor
See5	Decision tree algorithm for Windows, successor of C4.5
SEVIRI	Spinning enhanced visible and infrared imager
SiB	Simple biosphere model
SiB2	Simple biosphere model, second version
SMA	Spectral mixture analysis
SPOT	Systeme pour l’observation de la terre
SST	Sea surface temperature
STEP	System for terrestrial ecosystem parameterization
SWIR	Short wave infrared
TIMESAT	A program for analyzing time-series of satellite sensor data
TIMESTATS	A software tool for analyzing spatial-temporal raster data archives
TIR	Thermal infrared
TiSeG	Time series generator
TM	Thematic mapper
UMD	University of Maryland

UNESCO	United nations educational, scientific, and cultural organization
USGS	United States geological survey
UTM	Universal transverse Mercator
UV	Ultraviolet radiation
VALERI	Validation of land European remote sensing instruments
VCF	Vegetation continuous fields
VEGETATION	Vegetation monitoring instrument onboard SPOT
VI	Vegetation index
VIIRS	Visible infrared imager radiometer suite
VIS	Visible light
WCRP	World climate research programme

Chapter 1

Introduction

The Earth is a complex and highly dynamic system. Exchanges of matter and energy between Earth systems are only partially understood, and quantitative measurements are limited. In order to account for processes at the Earth's surface, multiple measurements have to be recorded at various temporal scales. Processes can be rapid or slow, intense or subtle, continuous or variable, periodic or episodic. All these considerations focus on velocity, intensity, connectivity, and reoccurrence. Therefore, time, in particular the length of the record as well as the temporal resolution, is very important for process analysis. Phrased differently, processes are analyzed using time series.

Much research in Earth science has focused on endogenous geological processes. On the other hand, there are important exogenous dynamics at work at the Earth surface shaping the ecosystems and affecting the morphosphere, pedosphere, atmosphere, hydrosphere, and biosphere. For instance, atmospheric processes involving weather and climate are very important for mankind. The daily weather forecast demonstrates the movement of clouds and water vapor over the last few hours, a service supported by hypertemporal time series. In contrast to weather, climatic processes are the subject of long-term analysis ranging from seasonal studies to multi-year analysis. Projections of the future 50 to 100 years are particularly relevant to the ongoing debate on climate change. In an even larger framework of global change, the interaction of all ecosystem components and the human impact is studied and future scenarios are being modeled.

Climatic processes in particular determine seasonal dynamics of the land surface. Human activities can become a modifying variable or even the driving factor of vegetative activity in cultural landscapes. Human-induced processes comprise, for instance, deforestation, urban sprawl, land degradation, desertification, or salinization. The change of the land surface by humans, in turn, affects other ecosystem components, which may result in process-response systems.

Seasonal patterns of vegetation including green-up, plateau, senescence, and dormancy, in the following referred to as phenology, are characteristic land-surface processes in many regions. The timing, length, and intensity of the growing season and of each phenologial phase depend on the ecoregion. Climatic variability is relevant to many ecosystems, especially in semi-arid regions where precipitation is the principal determining factor for plant growth. For monitoring long-term processes, on the other hand, data of multiple years have to be analyzed. Land degradation and transformation such as plant invasions or regrowth in semi-arid ecosystems will take place over several years. Multi-year trends can be altered or interrupted by irregular events such as fire and flooding. Episodic fires, for instance, are important for the health of Mediterranean ecosystems, but can also degrade the vegetation cover if too much dry matter has accumulated.

With respect to the characterization of the Earth surface and its transformation due to global warming and global change in general, there is a need for appropriate data and methods to determine the activity of vegetation and the change of the land surface. National and international organizations urgently demand accurate, quantitative results on natural and human-induced land-change processes. Important international programs are concerned with global-change research. For instance, the International Geosphere-Biosphere Programme (IGBP) aims at describing and understanding the physical, chemical, and biological processes of the natural and human systems including their interaction and change. In this respect the IGBP focuses on the sustainability of the living Earth. Together with three other international programs, the World Climate Research Programme (WCRP) for climate change, DIVERSITAS (an international programme of biodiversity science) specifically dedicated to biodiversity and ecology research, and the International Human Dimensions Programme on global environmental change (IHDP) for the human aspects of global change, they form the Earth Systems Science Partnership (ESSP). The overarching goal of the ESSP is research on global sustainability with four issues critical for human well-being: energy and carbon cycles, food systems, water resources, and human health.

Remote sensing has the potential for monitoring the land surface and its dynamics. Often, satellites observing the Earth from space are the only means to comprehensively acquire data of large regions to be employed for assessing the state of the land surface. Furthermore, multiple measurements, forming a time series, also account for land surface dynamics. For

instance, repeated measurements from space-borne systems detect photosynthetic-active vegetation, its phenological development and variability, multi-year transformation processes, and irregular events. Physically-based objective observations of the Earth surface and highly automated data processing to value-added products are the basis of consistent, spatially and temporally continuous datasets of large areas. They provide the potential for directly using remote-sensing data and products for analyzing the land surface in relation to global warming and global change.

The following section will provide a first overview of the state-of-the-art involving time-series remote sensing, with particular emphasis on time-series generation, analysis, and their use for land-cover classification. It should be noted that more detailed background information, a thorough discussion of the literature, and important practical considerations are presented in each main chapter.

1.1 State of the Art

Since 1981 land-surface processes have been monitored by remote-sensing techniques using the AVHRR instruments onboard the NOAA platforms (de Beurs and Henebry 2004a, Justice et al. 1985). Multiple coarse-resolution land-cover classifications have been derived from AVHRR time series (Hansen et al. 2000, Loveland et al. 2000) The AVHRR data record has several limitations, though, including insufficient geometric correction, missing radiometric inter-calibration between the AVHRR instruments flown on successive platforms, orbit drifts, unsatisfactory atmospheric correction, and cloud masking (Cihlar et al. 1997, Gutman and Ignatov 1996, Roy 2000). The problems have been mitigated using several approaches, e.g. the spatial aggregation to relieve misregistration issues and the use of invariant targets for radiometric calibration. However, the most important approach to improve AVHRR data has been temporal compositing using the maximum value compositing (MVC) method (Holben 1986). It assumes that the highest value per pixel out of a specific period, often 10 days to one month, indicates the pixel with the least contamination by clouds, aerosols, water vapor, or other influencing atmospheric effects. This approach has been deemed useful for many products such as NDVI, but difficulties arise if the entire compositing period is cloudy (Cihlar et al. 1994, van Leeuwen et al. 1999).

Several time series have been generated from AVHRR data. Among global records of 1 km (Eidenshink and Faundeen, 1994), 4 km, and 8-km spatial resolution, products such as Pathfinder AVHRR Land (PAL) and Global Inventory Modeling and Mapping Studies (GIMMS) have been used for time-series studies (Townshend 1994). Initially, time series have been generated by simply stacking composites. The GIMMS time series corrects for

orbit drifts and its processing has been continued until today (Tucker et al. 2005). Often mathematical approaches such as harmonic analysis or filtering functions have been used to smooth the time series with remaining quality limitations. However, these approaches affect both, signal and noise (Chen et al. 2004a, Jakubauskas et al. 2001, Jönsson and Eklundh 2002, Roerink et al. 2000).

Nevertheless, the historical aspects have to be considered when judging the limitations of the AVHRR instruments and its non-standardized data processing. It has been the first attempt at a daily global acquisition strategy, and recent terrestrial remote-sensing programs of medium and coarse spatial-resolution sensors can be considered to be a spin-off of it. Meanwhile spanning a period of more than 25 years, the AVHRR data record is very valuable for long-term analysis, e.g. for global land-cover change studies (Hansen and DeFries 2004). Recent efforts towards long-term data records incorporate AVHRR data into a modern product generation (Gallo et al. 2005).

The advent of a new era of remote-sensing systems started in 1997 and 1998 with the launch of SeaWiFS and SPOT VEGETATION, respectively, and has been complemented with MODIS and MERIS in the new millennium. Building on lessons learned from previous programs, well calibrated sensors provide the possibility of consistent long-term observations and trend analysis, important for multiple global-change studies. Geometric correction has been clearly improved, but atmospheric correction highly depends on the number of available bands and the availability of additional data.

In this regard, MODIS offers excellent possibilities. In contrast to all other systems, MODIS' 36 spectral bands allow for a full and automated atmospheric correction and detailed cloud masking. The spatial resolution ranges between 250 m and 1 km, and global coverage is achieved within two days. The fully operational product generation and a wide range of value-added datasets for terrestrial applications following a standardized production process ensure long-term consistency of global products (Justice et al. 2002a). In addition, each product is accompanied by product-quality estimators as meta-data and at the pixel level. The pixel-level quality assurance in particular indicates the usefulness of an observation (Roy et al. 2002a). While the dataset itself usually contains values in the valid data range, the attached product-specific data-quality layer indicates possible cloud cover, shadow effects, high aerosol content, difficulties in radiometric calibration, or angular effects. Several studies have considered the quality indicators of MODIS products (Colditz et al. 2006a, Colditz et al. 2007a, Conrad et al. 2005, Erasmi et al. 2006, Lobell and Asner 2004, Lunetta et al. 2006).

An accurate characterization of the seasonal development of the land surface is in high demand. For instance, phenological studies analyze the state of vegetative activity over the year. Continuous cloud cover during a long-lasting rain season can limit the accurate temporal

characterization of the Earth surface. Colditz et al. (2006a) have shown the effects of different quality levels on MODIS NDVI of evergreen broadleaved forest and savanna in western Africa. No quality analysis or very lenient settings have resulted in a significant decrease of the NDVI during the rain season, thus not accurately representing the phenology. Excluding cloudy pixels from the analysis has revealed reasonable phenological profiles. However, applying strictest quality settings can result in missing data for the complete growing season, again yielding an inaccurate temporal course. Similar results have been shown by Colditz et al. (2007a) and Colditz et al. (2007b) for Germany and by Erasmi et al. (2006) for northeastern Brazil. Wagner et al. (2007) and Conrad et al. (2007) have illustrated the influence of the quality indicators on resulting LAI and LST time series in western Africa and Uzbekistan, respectively. Accurate MODIS NDVI time series has been generated for southern Virginia using the available MODIS quality indicators (Lunetta et al. 2006).

The large-scale analysis of the effects of El Niño Southern Oscillation (ENSO) events and droughts on vegetation activity is only one example for the application of time series. Subtle changes or modifications in phenology due to the 1997 / 1998 El Niño have been mapped for equatorial eastern and southern Africa using AVHRR data (Anyamba et al. 2002). While eastern Africa has shown an increase in biomass production, limited precipitation during El Niño has diminished plant productivity in southern Africa. Other studies on the effects of ENSO events on plant productivity have been conducted globally using SeaWiFS data (Behrenfeld et al. 2001), for Indonesian forests with MODIS data (Fuller et al. 2004), and with AVHRR for the Amazon forest Asner et al. (2000) as well as for the U.S. (Li and Kafatos 2000, Wannebo and Rosenzweig 2003). Similarly, several studies on droughts have been provided (Ji and Peters 2003, McVicar and Bierwirth 2001, Tucker et al. 1994). Reichstein et al. (2007) and Gobron et al. (2005) have employed MODIS, MERIS, SeaWiFS FPAR and flux tower data to show the dramatic impact of the drought in 2003 on Europe's vegetation, but have also indicated that most areas had quickly recovered by 2004.

Long-term change and modification processes, partially related to landscape degradation, have also been analyzed using AVHRR time series. Tucker et al. (1994) has monitored changes in the boundary between the southern Sahara and the Sahel, with its southward movement until 1984, followed by a northward trend directly related to an increase in precipitation. The change vector approach (Lambin and Strahler 1994a), a technique for change detection effectively using time series, has been applied for monitoring land-cover processes in sub-Saharan Africa (Lambin and Ehrlich 1997). Long-term analysis of vegetation index time series for rangelands in Saudi Arabia have clearly shown desertification processes (Weiss et al. 2001). Multi-year studies on the seasonal variability of phenological phases in eastern Asia have indicated a clear relation to climatic variables and land cover (Yu et al. 2003, Yu et al. 2004). Regions of temporal variability and transition with frequent changes in

seasonality have been derived from an AVHRR time series for the African continent (Colditz et al. 2006b).

Mapping and monitoring of agricultural areas and crop-yield estimation is a challenging field in which time series are analyzed in depth. Changes in phenology of agricultural regions of Kazakhstan have been extensively monitored with AVHRR data and have been attributed to changing land use following the collapse of the Soviet Union (de Beurs and Henebry 2004a). Monitoring of agricultural areas in western Africa has been conducted with AVHRR data (Fuller 1998, Tottrup and Rasmussen 2004). A recent study on soybeans in southern Brazil using MODIS LAI has highlighted the merit of the MODIS quality indicators (Rizzi et al. 2006) and MODIS EVI data have been used to map sugar cane in a neighboring region (Xavier et al. 2006). Correlations between NDVI and yield have indicated the possibility of yield estimations (Ferencz et al. 2004, Mika et al. 2002). Reeves et al. (2005) have shown that remotely sensed time series may significantly contribute to yield estimations at an early stage of the cropping season. Satellite time series have been assessed and deemed useful as an input to crop models for crops in Europe, India, and Uzbekistan (Patel et al. 2006, Roebeling et al. 2004, Shi et al. 2007).

The main driving factors of vegetation green-up and senescence are precipitation and temperature (Bonan 2002). Water availability, often directly related to precipitation and its variability, is the driving factor for most semi-arid regions. Temperature and length of day become important for the mid-latitudes (de Beurs and Henebry 2005a, Potter and Brooks 1998, Myneni et al. 1997, Nemani et al. 2003, Schultz and Halpert 1993). Several studies have been conducted on the response time between precipitation and phenological activity using satellite-based vegetation indices. In this context, a direct relationship with a lag of one to two months has been found (Camberlin et al. 2007, Colditz et al. 2007c, Los et al. 2006, Nicholson et al. 1990).

An interesting phenological study by Huete et al. (2006) has shown that the healthy Amazon rain forest has higher vegetation index values in the dry than in the wet season. This unexpected pattern was initially related to invalid data due to cloud- and water-vapor contamination in the wet season, and to aerosols due to slash and burn agriculture during the dry season. A detailed time-series analysis using only highest quality observations of MODIS vegetation index data still revealed the same patterns, also supported by ground-based measurements of carbon fluxes (Saleska et al. 2003). Furthermore, a high-quality MODIS LAI time series has indicated gradually increasing leaf area during the dry season (Myneni et al. 2007). An earlier study using a 20-year AVHRR time series has shown a rise in net primary productivity over this period, which is related to an increase in sunlight in the tropics (Nemani et al. 2003). In other words, the increase in sunlight during the dry season causes a vegetation green-up. Water availability during the dry season is not the limiting factor,

because the roots of trees may reach up to 20 m down into the soil and saprolite. On the other hand, disturbed forests and pasture with a shallow rooting system show green-up during the rainy season (Huete et al. 2006, Myneni et al. 2007). Moreover, the increase in LAI of healthy Amazon rain forests is seen as a possible trigger for the rainy season, because more leaves will lead to increased evapotranspiration accelerating convective atmospheric processes until large-scale monsoonal processes come to dominate the region (Fu and Li 2004, Li and Fu 2004). In simple terms, the Amazon rain forest itself triggers the rainy season, which is a great example of surface-atmosphere interaction.

Besides multiple applications analyzing the phenological characteristics and variability of several years, time series are useful for land-cover and land-use classification as well as rapid change detection at moderate resolution (DeFries and Townshend 1994, DeFries et al. 1995, Lunetta et al. 2006, Zhan et al. 2002). Maps of land-surface types are still one of the most requested remote-sensing products. Due to the advantages of remote sensing, land-cover datasets are derived at a comparatively low cost, in particular if the process can be automated (Bartholomé and Belward 2005, Friedl et al. 2002). However, a high level of accuracy is desired. This is particularly important for modeling, where errors will increase the uncertainty (Atkinson and Foody 2002, DeFries and Los 1999, Heuvelink 2002).

A wide range of pixel- and object-based techniques has been developed for mono- and bi-temporal image classification (Jensen 2005, Lu et al. 2004). However, new and independent information can be incorporated into classification procedures, considering the phenology of land-surface types. In turn, ignoring relevant inner-annual land surface processes by accidentally or even deliberately choosing specific single-observation data even can lead to inaccurate results (Coppin et al. 2002). For instance, the analysis of a single image during the dry or wet season falls short of distinguishing all potentially discernible vegetation types of a given area. Even a bi-temporal approach with observations in the dry and wet season inaccurately discriminates classes if the distinctive character of the target class is revealed during transitional phases only. In addition to increasing classification accuracy, temporal characteristics can be employed for the discrimination of further classes.

The transferability, reproducibility, and repeatability of most classification procedures and existing large-scale classification products to other study areas and input datasets are limited (Herold et al. 2006a). Often, classification methods and associated legends have been tied to a specific study site, a specific date, or environmental setting and cannot be applied elsewhere. Furthermore, several approaches require specific input data. If data of the same sensor as employed before is not available, or an important ancillary or supportive dataset cannot be provided, the approach has to be adjusted with potential limitations regarding the accuracy and consistency of the outputs.

Capabilities for reoccurring updates are an important objective for regional, continental, and global classifications (GCOS 2004, Herold et al. 2006a). The operational acquisition strategy, highly automated data processing and accurate product generation of many advanced instruments including MODIS allows for a regular update of land-cover datasets using time series (Friedl et al. 2002). A classification procedure should be easily transferable to other regions and datasets, i.e. specific characteristics of the study site and location should not significantly limit the results in a first-order classification. Based on this result, adjustments for regional particularities can be considered in a second step.

The Global Observation of the Forest and Land Cover Dynamics network (GOFC-GOLD) provides an excellent overview of global, continental, and regional classifications (GOFC-GOLD 2007). One of the first global land-cover datasets has been based on AVHRR data. According to UNESCO definitions (UNESCO 1973) 11 classes have been mapped using NDVI AVHRR time series of 1987 with a spatial resolution of 1° and a supervised maximum likelihood classification (MLC; DeFries and Townshend 1994). Efforts of the IGBP have resulted in a global land-cover classification with 17 classes (Belward et al. 1999, Loveland et al. 2000). This dataset has been embedded in the Global Land Cover Characterization (GLCC) data base, an international initiative for land-cover mapping (Loveland et al. 1999). The classifications are based on a 1 km AVHRR NDVI time series from 1992 to 1993 and have been generated by unsupervised classification. In addition to the IGBP legend, GLCC also provides global maps including global ecosystems classification (96 classes; Olson 1994a, Olson 1994b), USGS land use / land cover system (24 classes; Anderson et al. 1976), Simple Biosphere Model (SiB) and SiB2 (20 and 11 classes, respectively; Sellers et al. 1986, Sellers et al. 1996a), biosphere-atmosphere transfer scheme (20 classes; Dickinson et al. 1986), and vegetation life form (8 classes; Running et al. 1995). The 14 class legend of the University of Maryland (UMD), a derivative of the IGBP classification scheme (Hansen and Reed 2000), has been employed for automated classification of AVHRR time series (DeFries et al. 1998, Hansen et al. 2000). Further global classifications are provided by the International Satellite Land Surface Climatology Project (ISLSCP) program (Hall et al. 2005) and dedicated forest mapping is provided by the Forest Resources Assessment (FRA).

The Global Land Cover classification of 2000 (GLC2000) with 22 classes has been derived from a SPOT VEGETATION time series from 1999 to 2000 (Bartholomé and Belward 2005). Regional GLC2000 datasets have been generated using individual classification procedures, which have been subsequently harmonized to a global product. GLOBCOVER, the successor of GLC2000, is based on MERIS time series (Defourny et al. 2006). Many recent approaches, including GLC2000 and GLOBCOVER, follow standards of the Land Cover Classification System (LCCS; Di Gregorio 2005), the recommended world-wide system for flexible and standardized global land-cover legend generation and harmonization (Herold et al. 2006b).

Existing global land-cover classifications have been “one-time” efforts and lack regular updates. Annual updates of global land-cover data with five legends including IGBP, UMD, and vegetation life form are contained in the MODIS land-cover product (Friedl et al. 2002). Time series of MODIS surface reflectance and vegetation indices are crucial for annual classification updates. Furthermore, continuous land-surface characterizations, i.e. a fuzzy classification of tree, herbaceous, and barren, have been derived from AVHRR and MODIS time-series metrics (DeFries at al. 2000, Hansen et al. 2003).

Some global products originate from continental land-cover maps, e.g. GLC2000. Datasets for parts of Africa, i.e. AFRICOVER, are based on Landsat data, but are only available for selected countries. The pan-European CORINE land-cover project with 44 classes has also been derived from a Landsat mosaic of 1990 and 2000, with an update planned for 2006. It is attempted to automate the CORINE land-cover classification by also incorporating time series, and the potential of SPOT VEGETATION, MODIS and IRS AWiFS, have been tested (Han et al. 2004, Wehrmann 2007). Another land-cover mapping project, the Pan-European Land Use and Land Cover Monitoring (PELCOM), has used AVHRR time series for classification updates similar to CORINE, but with a reduced number of classes (Mücher et al. 2000).

Almost every developed country conducts national or regional land-cover mapping activities, which are sometimes embedded in a continental classification effort such as CORINE. As many of these classifications require high spatial-resolution input data, time series are of limited use due to their medium to coarse pixel size. Some recent studies, however, have acknowledged the advantages of time series, either for selecting appropriate dates for high-resolution image acquisition with respect to important phenological stages (Yang et al. 2001) or due to phenology’s general usefulness for class discrimination (Wehrmann 2007). National mapping activities relevant to this study are the National Land Cover classification of South Africa (NLC1995; Thompson 1996), CORINE land cover for Germany (CLC2000; Keil et al. 2005), the National Land Cover Database of the United States of America, with particular emphasis on Minnesota (NLCD2001; Homer et al. 2004) and the map of vegetation types of Namibia (Mendelsohn et al. 2002).

All existing global and regional classifications based on time series have a coarse spatial resolution of approximately 1 km. However, many land-surface structures are well beyond this spatial resolution and therefore cannot be accurately represented in discrete coarse-resolution classifications. Although it would be desirable to employ higher spatial-resolution time series, current and planned earth-observing systems will not provide satellite constellations suitable for continental or global mapping (GCOS 2004). Another solution is the combination of coarse-resolution time series with high spatial-resolution data (Borak et al. 2000). Knowledge-based approaches seem to yield reasonable results, which have been

successfully employed to distinguish specific classes using time series (Wehrmann 2007). Fractional mapping - the proportion of land-surface types is determined for every pixel - is another alternative offering more flexibility for the user.

Repeated updates of land-cover maps are in high demand as input to models and because they allow for the identification of hot spots of land-cover change. Since the spatial resolution of high temporal-resolution sensors is relatively coarse, fractional land-surface datasets appear to be most appropriate (Mather 2003). Commonly employed hard classifications assign a single class to each pixel, even if only a small portion is covered. This is particularly questionable for heterogeneous areas or broad transition zones. In contrast, fractional estimates conserve the proportion of each class for every pixel, which is used as an indicator, not a definitive class assignment. Factors determining a fuzzy classification comprise the spatial resolution of the input data, the landscape heterogeneity, and the scale of data analysis (Moody and Woodcock 1994). The propagation of the fuzziness has a direct implication on the likelihood of land-cover or land-use change, a requirement of many current projects. For example, a landscape change can be attributed if a certain threshold is exceeded (Zhan et al. 2002). Furthermore, fuzzy classifications can be transformed to discrete maps, for instance using the majority rule.

Due to the tremendous increase in data, new classification techniques have to be employed for land-surface mapping (Hansen et al. 1996). Common pixel-based approaches such as minimum-distance or maximum-likelihood classification cannot handle these large amounts of data even on computer clusters. Furthermore, it has been shown that a significant increase in dimensionality will cause decreasing classification accuracies due to the curse of dimensionality (Datcu et al. 1998). In turn, fast and robust approaches are required. Among many possibilities such as unmixing procedures and neuronal networks, decision-tree approaches have been found to be useful for remote-sensing data (Friedl and Brodley 1997, Hansen et al. 1996). According to a splitting criterion, a tree continuously divides the training data into smaller sets until a stopping rule is reached. Depending on the type of the tree, the resulting leaf will either contain a class or a value. Based on the resulting decision tree, the dataset is propagated through the tree following the rules at each node (Breiman et al. 1984).

1.2 Scientific Goals

As outlined in the previous section, time series have a high, but not fully exploited potential in remote sensing. Applications of high-quality time series in environmental research include the study of land surface dynamics and their use for land-cover classification. Land-surface processes, in particular key phenological states of vegetation, are important inputs for

environmental monitoring and agricultural studies. Furthermore, quantitative global-change research requires standardized and consistent long-term time series.

Time series are also useful for large-scale classifications and, in addition to multi-spectral data, add another dimension to improved land-surface discrimination. Global classifications in particular are directly or indirectly based on time series, and their use for continental and regional classifications has been clearly emerging over the last ten years. Operationally acquired and automatically processed datasets are needed for regular updates of land-cover maps and for the detection of hot spots of land-cover change.

In this context the study focuses on the following goals:

- A robust semi-automatic approach of time series generation
- An effective automated classification approach of fractional cover estimation for updatable land-surface mapping using time series

A software tool called Time Series Generator (TiSeG) has been developed for time-series generation. This software interprets data-quality indicators provided with each MODIS dataset. The accuracy of the time-series generation has involved detailed visual and quantitative analysis. The time-series classification builds on the principles of Classification and Regression Trees (CART; Breiman et al. 1984). Decision-tree approaches have been successfully employed for remote-sensing image classification (Friedl et al. 2002, Hansen et al. 2000, Hansen et al. 2003). The machine learning algorithm has been modified and extended to a fuzzy classifier following the ideas of random forest (Breiman et al. 2001). The classification approach has been intensively tested by a sensitivity study. Therefore, a protocol has been developed for accuracy assessment of hard and soft classifications.

The approaches developed in this thesis are independent of a specific study site, i.e. they are *a priori* not restricted to particular regional characteristics. This has been proved by four study sites in environmentally different settings and for different classification schemes. The criteria for study site selection were different climatic and vegetation conditions as well as varying heterogeneity of the landscape. While the classifications of the study sites in South Africa, Germany, and Minnesota focus on land-cover and land-use classes with different input datasets, Namibia has been selected for indicating the potential for classifying vegetation classes. The classification of time series for land-cover map updates has been conducted in accordance to regionally suitable classification schemes. The number of classes has been selected from two perspectives: (1) the thematic context and heterogeneity of the landscape (2) the feasibility, especially considering the spatial resolution of the time series.

Similarly, minor adjustments are necessary if applying techniques developed in this thesis to data of other instruments. Therefore, the approaches developed for both, time-series generation and classification, are understood as frameworks which have been tested for MODIS data. Other datasets, e.g. from MERIS and future VIIRS, have or will have similar characteristics requiring possibly a few adjustments to the software. In addition, successful attempts have been made to apply the principles to long-term AVHRR datasets.

This regional and instrument transferability is highly appreciated by many remote-sensing studies. Limitations such as reduced capabilities of class discrimination or lower accuracy if compared to regionally tuned results may exist, but are beyond the scope of this thesis. All processing steps have been developed in a modular programming fashion, which allows for a simple integration into operational processing.

1.3 Outline of Thesis

Each main chapter of this thesis contains a first section with an in-depth discussion, necessary to understand the subsequent datasets, methods, and results. The above-described goals of this thesis are approached in chapter 3 (Time Series Compilation) and chapter 4 (Automated Time Series Classification).

The second chapter comprises a discussion of the sensors suitable for time-series generation and describes the MODIS sensor and the MODIS land product generation process. Besides important background information for readers unfamiliar with the sensors, in particular with the MODIS instrument, this chapter introduces the products and study sites used in this thesis. Chapter 3 (Time Series Compilation) highlights important considerations for time-series production and describes the process of MODIS time-series generation using the developed software tool TiSeG. The analysis section illustrates the selection of different data-quality levels and their effect on time series in Germany using visual and statistical measures. The effect of changes in the quality indicators is shown for the recently released MODIS data collection 5. A last section analyzes the temporal shifts of phenological patterns in South Africa. The fourth chapter, Automated Time Series Classification, discusses several issues to be considered for classification processes, outlines the classification procedure used in this study, and analyzes the results. The analysis section contains a detailed classification assessment of South Africa and Germany. A sensitivity analysis evaluates important classification parameters for both main study sites. A third part focuses on the capabilities for land-cover updating. A final part indicates the flexibility of the classification procedure for mapping vegetation types in Namibia. The thesis concludes with a final discussion and an outlook for critical requirements and scientific developments in the near future.

Chapter 2

Sensors Datasets and Study Sites

This chapter aims at introducing important background information for the following chapters of data analysis. The first section will provide an overview of sensors or sensor constellations which can be used for time-series records. It discusses the limitations of older systems and derived datasets, presents currently available sensors, and provides a brief outlook. The next section introduces the MODIS sensor which has been used in this study. Separate sections will describe most important information of the space platforms, the sensor itself, and the MODIS land-dataset generation. Datasets used in this study will be described in more detail. The last section of this chapter introduces the study sites of this thesis, namely South Africa, Germany, Namibia, and Minnesota. It will discuss important geographical characteristics, relevant to space-borne remote sensing at medium resolution. Furthermore, land-cover datasets will be described, which have been important for training and test data generation of time-series classification.

Readers familiar with the content outlined above may skip this chapter or respective sections and continue with chapter 3 (Time Series Compilation). The chapters of data analysis, however, will require a profound knowledge of the general justifications and specifications provided in chapter 2.

2.1 Sensors Suitable for Time Series Records

A time series is defined as “a collection of observations made sequentially through time” (Chatfield 2004, p. 1). In this respect, the main sensor requirement for time-series generation is a repetitive coverage of the same area, ideally the sensors employed have a high temporal resolution to catch vegetation cover changes due to phenology. This is assured by geostationary satellites and some polar-orbiting systems. Weather satellite instruments on geostationary orbits such as SEVIRI on METEOSAT or Imagers on GOES acquire data of the same location every 15 to 30 minutes. However, the small number and placement of the spectral bands is unfavorable for terrestrial applications. Additionally, the best spatial resolution of 3 km at nadir limits the use of geostationary systems especially for land-cover studies. Further restrictions are the decreasing spatial resolution at higher latitudes, hemispheric image acquisition, and data processing focused on atmospheric and climatologic applications. The advantages of hypertemporal time series, such as accounting for BRDF effects and higher likelihood of cloud-free observations, have not been relevant to this study.

Polar-orbiting systems have been extensively used for land applications. TM or ETM^+ on Landsat and HRV or HRVIR on SPOT are specifically devoted to land-cover mapping with spectral bands at wavelengths important for land-surface studies. A constraint of these systems, however, is the limited temporal resolution. This restricts the polar orbiting systems suitable for time-series records, at least in the first place, to wide field of view instruments, such as AVHRR on the NOAA platforms, SeaWiFS on OrbView2, VEGETATION on SPOT, MERIS on ENVISAT, or MODIS onboard Terra and Aqua. All systems have full global coverage within a few days. The most important specifications and applications of the instruments and sensor systems are summarized in remote-sensing introductory literature, including Lillesand et al. (2003) and Jensen (2007).

In particular, the AVHRR instruments onboard the NOAA platforms have been widely used for terrestrial earth surface applications. Since 1981 they have been acquiring data in five spectral bands including a red and a near infrared band. The AVHRR instruments are the only long-term data source for environmental studies, for instance, of phenology (de Beurs and Henebry 2004a), land-surface variability (Anyamba et al. 2002), and land-cover mapping (Hansen and deFries 2004). AVHRR data are provided as local area coverage (LAC) with 1.1 km resolution and global area coverage (GAC) with 4 km resolution. Frequently, image composites are generated to mitigate errors in the acquisition system, to minimize the influence of atmospheric constituents such as clouds or aerosols, and to eliminate product generation difficulties (Roy et al. 2002a). The maximum value compositing (Holben 1986) has been the method of almost universal choice for AVHRR NDVI. Several global datasets have been derived (Townshend 1994), including Pathfinder AVHRR Land (PAL) at 8 km

(James and Kalluri 1994, Maiden and Greco 1994), Global Inventory Modeling and Mapping Studies (GIMMS; el Saleous et al. 2000, Tucker et al. 2005) at 8 km, a 1° dataset for climate modeling (Los et al. 1994, Sellers et al. 1994), and global vegetation index (GVI; Goward et al. 1994) with 25 km resolution. An overview of time series and approaches to time-series compilation is provided by Colditz et al. (2007a). Recently, successful attempts have been made to generate a long-term data record for numerous products spanning from 1981 into the future. The long term data record (LTDR) will combine AVHRR, MODIS, and future VIIRS (visible infrared imager radiometer suite) data from NPP and NPOESS.

Several new instruments have been launched since 1997 (launch date in parenthesis), including SeaWiFS (1997), VEGETATION (1998), MODIS (1999 and 2002) or MERIS (2002). They all complement the AVHRR data record. Disadvantages of the AVHRR instruments, such as insufficient cross-calibration, hard to estimate but known to exist deterioration of the charged coupled device, inaccurate geolocation, orbit-drifts, or inadequate and insufficient bands for a full atmospheric correction have been mitigated or eliminated. Additional and detailed meta-data serve as a means to analyze data-quality (Brockmann 2004, Roy et al. 2002a) and have been successfully used for time-series generation (Colditz et al. 2006a, Colditz et al. 2007a). Having lessons learned from AVHRR and the long-term experiments of MODIS and NPP (to be launched in 2009), the fleet of NPOESS platforms (scheduled to be launched beyond 2012) with its VIIRS instruments will continue time series for environmental applications of polar-orbiting systems (Townshend and Justice 2002).

It should be noted that wide field of view polar-orbiting instruments have large scans with a swath width between 2,000 and 2,500 km. At an orbiting altitude between 600 and 800 km this corresponds to a scan angle of approximately 50˚. At high view angles, the quality of the image data is reduced. Furthermore, the sensor systems have restrictions in spatial and sometimes spectral resolution due to limited downlink rates and data storage capabilities. These drawbacks will be solved for future missions, because recent advances in downlink and storage capacities will allow for a higher spatial and adequate spectral resolution.

An increase in the temporal resolution with high spatial-resolution datasets is also possible using pointable sensors or a sensor fleet. Pivoting instruments such as HRV on SPOT or ASTER onboard Terra, are capable to acquire a frequent coverage of a selected region. However, the focus on a particular area prohibits the view on other regions during this overpass. Hence, there is no global acquisition strategy with high temporal frequency. A high spatial resolution multi-sensor constellation for land-cover studies is currently not available. However, future missions such as Rapid Eye will provide remote-sensing data with 6.5 m spatial resolution in five spectral bands ranging from the visible blue to the near infrared, the minimum requirement for phenological monitoring and land-cover mapping. Five identical instruments will allow a high revisit time and therefore frequent coverage of a given area.

2.2 MODIS

Following the discussion of polar-orbiting sensors suitable for time series in the previous section, newer systems are preferably used, if the shorter period of data availability is deemed sufficient. Among the currently available newer instruments (e.g. MODIS, SeaWiFS, VEGETATION, and MERIS), MODIS has been selected for this study. Although SeaWiFS and MERIS have been successfully used for some terrestrial applications such as studies of El Nino effects (Behrenfeld et al. 2001) or land-cover classification (Geßner et al. 2005), their general sensor design is oriented towards oceanographic applications, and the bands are limited to the visible and near infrared wavelengths. SPOT VEGETATION focuses on terrestrial applications and has been successfully used for phenological studies (Geerken et al. 2005) and land-cover mapping at 1 km (Bartholomé and Belward 2005, Han et al. 2004). However, the small number of only four spectral bands limits the discrimination of land-cover classes. MODIS, in turn, has 36 bands, covers a wider range of the electromagnetic spectrum, and offers a higher spatial resolution.

The long-term experiment MODIS is part of the Earth Observing System (EOS) which is embedded in NASA's interdisciplinary Earth Science Enterprise (ESE; NASA 2003a, 2003b, NASA 2007). With the heritage of multiple sensors including AVHRR, TM / ETM$^+$, and CZCS, MODIS faces the challenge to suite atmospheric, oceanographic, and terrestrial needs. Different scientific fields competing against each other might appear a burden at first glance, maybe shunning added value for the user community. For instance, while a spatial resolution of 1 km is sufficient for atmospheric and ocean products, many terrestrial applications require a resolution of at least 500 m and possibly even finer. Furthermore, records of specific wavelengths are required for generating atmospheric products such as cloud temperature, ozone content, and tropospheric humidity. As such, they are of minor importance for land applications. However, a crucial requirement of the land community is a complete and accurate atmospheric correction, including scattering reduction and correction for aerosol and ozone. The atmospheric science branch provides the necessary parameters to run radiation transfer models for accurate surface-reflectance retrieval. Accurate sub-pixel cloud identification is also essential for land studies. Altogether, after seven years in operation, the MODIS experiment is considered a successful enterprise and joint work between different science communities.

2.2.1 Platforms and Instrument

Identical MODIS instruments are mounted on the Terra (formerly known as EOS AM-1) and Aqua (EOS PM-1) platforms, launched on December 18th, 1999 and on May 4th, 2002, respectively (Barnes et al. 1998, Barnes et al. 2003). While Terra operates on a descending orbit with a mean equatorial crossing time at 10:30 am (Justice et al. 2002a), Aqua runs in an ascending mode and crosses the equator at 1:30 pm (Parkinson 2003). Applications in tropical regions are limited, because cloud convection in the late morning and early afternoon equatorial crossing time prevents the view onto the ground. Although the orbit characteristics of both platforms are rather similar this study will focus on Terra, because of a longer availability of the data. Information on the Aqua platform and derived products can be found in the EOS data products handbook volume 2 (Parkinson and Greenstone 2000) and in a special edition of IEEE Transactions on Geoscience and Remote Sensing (Parkinson et al. 2003).

Terra operates at an altitude of 705 km in a circular, sun-synchronous, near-polar orbit with 98.1˚ inclination (Justice et al. 2002a, King et al. 2004). The repeat cycle at the same ground track is 16 days, with an orbit period of 98.9 minutes. The satellite weights 5,190 kg with a length of 6.8 m and a diameter of 3.5 m and a power of 2,530 watts. Terra already exceeded its designed lifetime of five years (Xiong and Barnes 2006). Besides MODIS, four other instruments are onboard Terra including ASTER, CERES, MISR, and MOPITT. An overview on Terra and its instruments can be obtained from the EOS data products handbook volume 1 (King et al. 2004) and from a special edition of IEEE Transactions on Geoscience and Remote Sensing (Kaufmann et al. 1998).

The MODIS instrument itself weights 250 kg, spans 1 × 1.6 × 1 m, and requires 225 watts (Barnes et al. 1998). The scan angle of 55˚ yields a swath width of 2,340 km. Due to its wide swath the instrument provides daily global coverage above 30˚ latitude and two day coverage in the tropics. MODIS' 36 bands range from the VIS to the TIR in the electromagnetic spectrum (for wavelength see Table A.1). Depending on the wavelength the spatial resolution ranges from 250 m to 1 km at nadir (Guenther et al. 2002, Justice et al. 2002a). Selected reflective and emitting bands suitable for land applications such as vegetation indices, land cover, or biogeophysical parameters are listed in Table 2.1. The radiometric resolution of all bands is 12 bit.

The whiskbroom scanning system consists of a set of linear detector arrays sweeping across track. Depending on the spatial resolution of 250 m, 500 m, or 1 km, each scan comprises 10, 20, or 40 rows with 1,354, 2,708, or 5,416 observations, respectively. While the ground-projected instantaneous field of view (IFOV) at nadir matches the nominal resolution as mentioned above, at the edge of a scan at 55˚ it is 2.0 and 4.8 times larger along and across

Table 2.1: Selected MODIS bands and their spatial resolution, spectral range, and major applications for land studies.

Band	Abbreviation	Spatial resolution [m]	Wavelength [nm]	Major application for land surface mapping
1	VIS red	250	0.620 - 0.670	Chlorophyll, land cover mapping
2	NIR	250	0.841 - 0.876	Chlorophyll, land cover mapping
3	VIS blue	500	0.459 - 0.479	Soil vegetation difference, suspended matter in water
4	VIS green	500	0.545 - 0.565	Green vegetation
5	SWIR1	500	1.230 - 1.250	Leaf canopy differences
6	SWIR2	500	1.628 - 1.652	Vegetation mapping
7	SWIR3	500	2.105 - 2.155	Soil / ground differences
20	MIR1	1,000	3.660 - 3.840	LST
21	MIR2	1,000	3.929 - 3.989	Forest fire and volcanoes
22	MIR3	1,000	3.929 - 3.989	LST
23	MIR4	1,000	4.020 - 4.080	LST
29	TIR1	1,000	8.400 - 8.700	LST
31	TIR2	1,000	10.780 - 11.280	Forest fires, volcanoes, LST
32	TIR3	1,000	11.770 - 12.270	Forest fires, volcanoes, LST

Note: Divisions of wavelengths are provided in Table A.1. A list of all bands and detailed characteristics is provided in Barnes et al. (1998) and Guenther et al. (2002).

track, respectively (Masuoka et al. 1998, Wolfe et al. 1998, Wolfe et al. 2002). The MODIS sensor and orbit design with a mirror angular velocity of 2.127 rad/sec and a forward velocity of 7.5 km/sec assures that each scan adjoins at nadir. However, with increasing scan angle, consecutive scans intersect considerably. At an angle of 24˚ the overlap is 10 %, increasing up to 50 % at the edge of the scan. Due to this so-called "bow-tie" effect (Figure 2.1) the same object may be sensed three times at the scan edge, which is particularly apparent at distinct boundaries such as a shoreline. This effect is corrected in the course of gridding for the MODIS standard product suite (see 2.2.2 MODIS Land Datasets; Wolfe et al. 1998). Further information on MODIS and the developed land products can be obtained from the EOS data products handbook volume 1 (King et al. 2004), from a special issue of Remote Sensing of Environment (Justice and Townshend 2002), or from Justice et al. (1998).

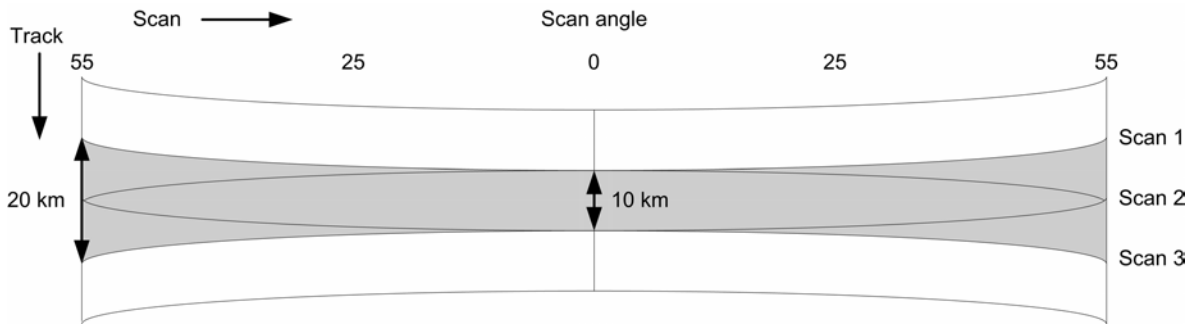


Figure 2.1: Ground projection of three consecutive MODIS scans illustrating the bow-tie effect (Wolfe et al. 1998, Wolfe et al. 2002).
Note: The area in gray indicates the second scan and shows the growing overlap of consecutive scan with increasing view angle.

2.2.2 MODIS Land Datasets

The first part of this section will focus on the various processing levels of MODIS data and explains the data generation process. The following sections will highlight general characteristics and algorithms of datasets employed in this study (Table 2.2). These sections also introduce the formats and ancillary data supplied as science dataset (SDS) with each composite.

Table 2.2: List of MODLand products used in this study.

Parameter	Product	Spatial resolution [m]	Temporal resolution [days]	Level	TiSeG
Surface reflectance	MOD09A1	500	8	3	x
LST / emissivity	MOD11A2	1,000	8	3	x
Land cover dynamics	MOD12Q2	1,000	Bi-yearly	3	
Vegetation index	MOD13A1	500	16	3	x
Nadir BRDF-adjusted reflectance	MOD43B4	1,000	16	3	x

Note: **A complete list of all MODLand products is provided in Table A.2. Column TiSeG indicates products which can be processed with the Time Series Generator (see chapter 3).**

Data Production and Levels of MODIS Data

The product generation of MODIS imagery follows a well-defined procedure with several levels. The following sections will mostly focus on MODIS land (MODLand) products (Table 2.3 and Figure 2.2, see also Table A.2; Justice et al. 1998, Justice et al. 2002a, King et al. 2004, Masuoka et al. 1998, Roy et al. 2002a, Wolfe et al. 1998). Products are provided to the user community in the hierarchical data format (HDF; Masuoka et al. 1998). Each product is abbreviated with MOD, MYD, and MCD to indicate Terra, Aqua, or combined Terra / Aqua products, respectively. However, MOD is also often used to specify MODIS products in general and will be used in this sense unless noted differently. A combination of numbers and characters follows the three capital letter identifier indicating the specifics of the product.

Level 1A data (MOD01; Masuoka et al. 1998, Nishihama et al. 1997) are derived by splitting an overpass into five minute granules covering approximately 2,340 × 2,030 km across and along track, respectively, which equals 288 files per day (Justice et al. 2002a). Geolocation parameters and radiometric calibration coefficients are provided along with quality indicators and other instrument data. During nighttime, only band 20 to 36 (MIR and TIR) excluding band 26 are processed. Level 1A data serve as input for MOD02 (also called Level 1B; Barbieri et al. 1997, Masuoka et al. 1998) and for a separate geolocation dataset including elevation as well as sun and view angles (MOD03; Nishihama et al. 1997, Wolfe et al. 2002).

Table 2.3: Processing levels of MODIS data according to Justice et al. (1998), Wolfe et al. (1998), Roy et al. (2002a), and King et al. (2004).

Level	Variables	Processing stage	Example of MODLand data
Level 0	Raw data	Reconstructed unprocessed instrument / payload data with communication artifacts removed	
Level 1A	Digital counts	Radiometrically and geometrically calibrated	
Level 1B	Radiance	Calibrated radiances at sensor	Calibrated radiance
Level 2	Geophysical	Geophysical variables	LST, snow cover, thermal anomalies / fire
Level 2G	Gridded geophysical	Mapped / gridded level 2 data to a uniform map	Surface reflectance
Level 3	Value-added	Derived from (gridded) geophysical variables, composited to uniform time intervals	Vegetation indices, land cover
Level 4	Modeled	Modeled products, result from lower level and ancillary data	LAI / FPAR, NPP, VCF

MOD02 contains the calibrated and geolocated at-sensor radiance. Both, MOD02 and MOD03 are inputs for all downstream processes, and errors in these datasets would affect all higher level MODIS products (King et al. 2004, Wolfe et al. 2002).

The MODLand Level 2 product generation requires some atmosphere products. For instance, the surface-reflectance data (MOD09) are derived for all non-cloudy pixels, identified with the cloud mask (MOD35; Ackerman et al. 2006) and with band 26 for thin cirrus clouds. It is corrected for aerosols (MOD04; Remer et al. 2006), water vapor (MOD05; Gao and Kaufman 1998), and ozone (MOD07; Seemann et al. 2006). In addition, previous BRDF coefficients (MOD43; Schaaf et al. 2002, Strahler et al. 1999a) are used to account for surface-atmosphere coupling (Vermote et al. 1997a, Vermote and Vermeulen 1999, Vermote et al. 2002). The surface reflectance is a major input for almost all Level 3 and 4 MODLand products. Other MODLand products generated at level 2 are land surface temperature and emissivity (MOD11; Wan 1999, Wan et al. 2002), thermal anomalies / fire (MOD14; Justice et al. 2002b, Kaufmann and Justice 1998), and cryosphere datasets for snow and sea ice cover (MOD10 and MOD29; Hall et al. 2001, Hall et al. 2002).

Subsequently, the so-called Level 2G datasets for MODLand products are generated by gridding (Masuoka et al. 1998). In particular surface-reflectance data (MOD09; Vermote et al. 1997a, Vermote and Vermeulen 1999, Vermote et al. 2002) are provided to the user community in Level 2G (Wolfe et al. 1998, Wolfe et al. 2002) where all swaths of one day are reprojected and combined in a gridded dataset. It is important to note that all granules are still represented; hence there is no compositing in a classical sense. Accurate geolocation and gridding is tremendously important for time-series analysis, because a spatial mismatch of adjacent pixels would lead to erroneous results (Justice et al. 1998, Justice et al. 2002a, Roy 2000, Wolfe et al. 1998, Wolfe et al. 2002). Additional Level 2G products provide angular information of the view and the sun (MODMG; Nishihama et al. 1997, Wolfe et al. 1998). So-

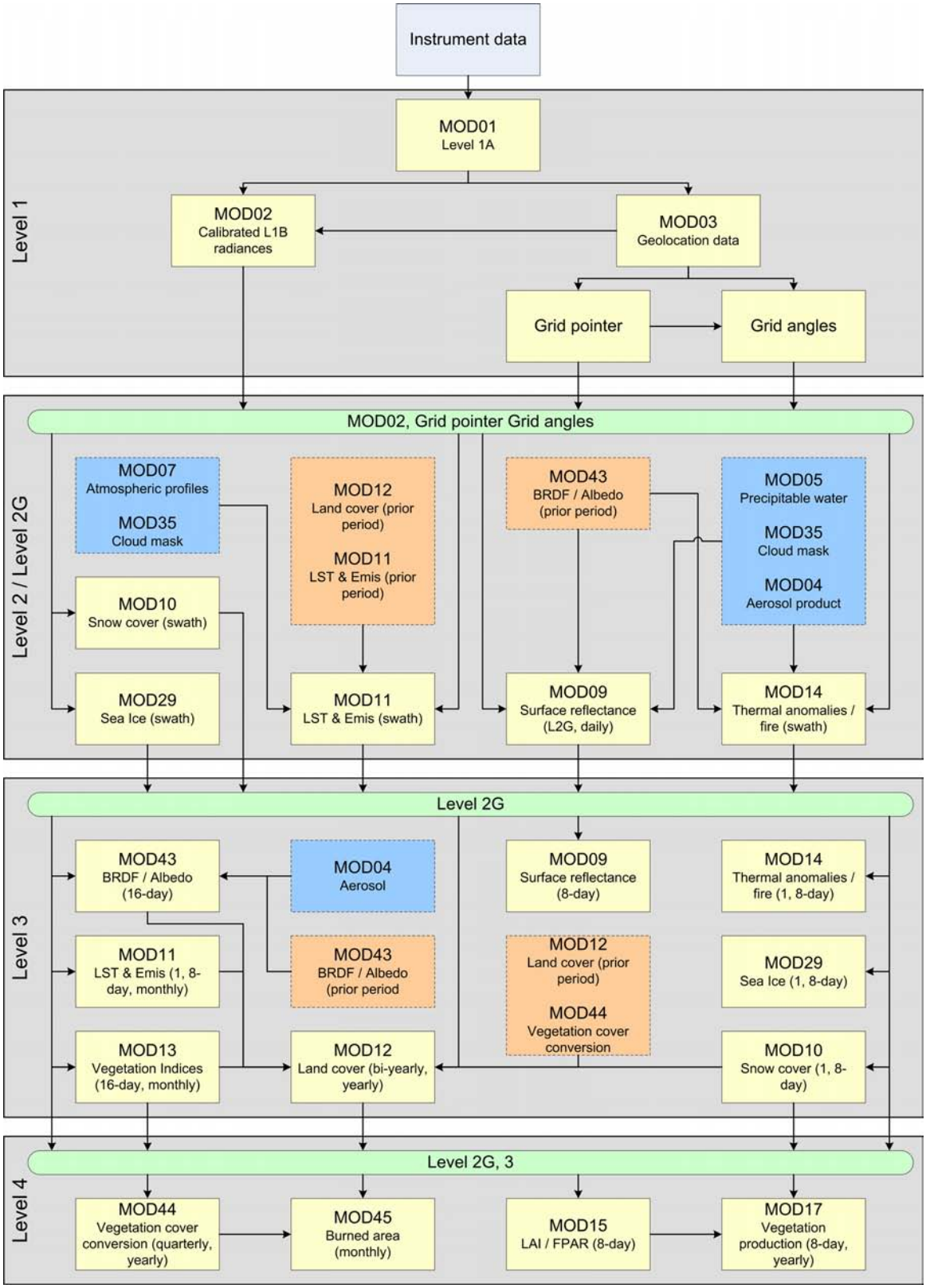


Figure 2.2: **Generalized flowchart of MODLand processing according to Justice et al. (1998), Justice et al. (2002a), Roy et al. (2002a), and King et al. (2004). For a complete list of MODLand products see Table A.2.**

Note: **Blue boxes indicate MODIS atmosphere products. Brown boxes specify previously generated MODLand datasets; standard settings had been used in the beginning of production. MOD10 (snow cover) and MOD29 (sea ice) are not strictly considered MODLand data.**

called pointer files (MODPT; Nishihama et al. 1997, Wolfe et al. 1998) reference the original granule, line, and sample. Besides the granule information, the pointer product stores sub-pixel offsets accounting for the MODIS point spread function, pointers to corresponding coarser resolutions, and observation coverage. Therefore Level 2G datasets can be seen as a link between Level 1 and 2 swath data in granules and Level 3 and 4 gridded products, referred to as tiles (Masuoka et al. 1998, Wolfe et al. 1998). The tiles of Level 2G, 3, and 4 are non-overlapping and have an extent of approximately 1,200 × 1,200 km, which corresponds to an area of $10 \times 10^{\circ}$ at the equator. 326 out of 460 global tiles contain land pixels (Justice et al. 2002a). Using the angular information and pointer files, the above mentioned bow-tie effect is mainly removed during the geometrical transformation to the Sinusoidal projection (Figure A.1).

Level 3 MODLand products are derived from geophysical variables and can also be termed value-added datasets. As such, Level 3 products are characterized by spatially and temporally uniform composites in tile format (Figure A.1) of 8 days, 16 days, 1 month, quarterly, or yearly datasets (Justice et al. 2002a). In addition to MOD09, MOD11, and MOD14, land cover and land-cover dynamics (MOD12; Friedl et al. 2002, Strahler et al. 1999b, Zhang et al. 2003), vegetation indices (MOD13; Huete et al. 1999, Huete et al. 2002), and BRDF / albedo (MOD43; Schaaf et al. 2002, Strahler et al. 1999a) complement the Level 3 MODLand suite. The compositing approaches vary depending on the product (Huete et al. 2002, Schaaf et al. 2002, van Leeuwen et al. 1997, van Leeuwen et al. 1999, Wolfe et al. 1998), but take advantage of the Level 2G MOD09 data and prefer cloud-free, near-nadir observations. Rolling compositing, i.e. datasets of 16 days of observation are generated every eighth day, is performed for some products such as collection 5 MOD43 data (for data versions see below). Other products, of collection 5 (e.g. MOD13), are processed in phased production between MODIS from Terra and Aqua to effectively provide a vegetation index time series with datasets every eight days (Didan and Huete 2006).

Modeled Level 4 datasets are derived from lower level data and also employ ancillary datasets. In particular the MODLand Level 4 data comprise biophysical parameters such as the leaf area index and the fraction of photosynthetically active radiation (MOD15; Knyazikhin et al. 1998a, Knyazikhin et al. 1998b, Knyazikhin et al. 1999, Myneni et al. 2002) or the net photosynthesis (MOD17; Running et al. 1999a). Vegetation cover conversion and vegetation continuous fields (MOD44; Hansen et al. 2002a, Hansen et al. 2002b, Hansen et al. 2003, Townshend et al. 1999, Zhan et al. 2000, Zhan et al. 2002) and burnt area (MOD45, not yet released; Justice et al. 2006, Roy et al. 2002b) are also considered modeled Level 4 datasets (King et al. 2004) because they require additional information (e.g. climate datasets) during their production.

Since the launch of MODIS onboard Terra several data versions, also called collections, have been processed and distributed. While science evolves during the years, up-to-date scientific results are included in the irregularly updated data versions (Justice et al. 2002a). The initiation of a new collection results in a reprocessing of the entire MODIS archive, and earlier versions are deleted after completion of processing. Collection 1 was generated following the launch of 2000. The first reprocessing, termed collection 3 (collection 2 data were never produced), was initiated in December 2001 and also comprised Aqua data since May 2002 (Justice et al. 2002a). Reprocessing to collection 4 (C4) was initiated in December 2002, and is currently the most complete collection of MODIS data spanning from the beginning in 2000 to the end of 2006. Since 2007, collection 5 data (C5) processing has been started. The entire MODIS archive will be reprocessed to collection 5 by September 2008 (MODIS 2007a). Unless noted differently, this study is based on collection 4 data.

MODIS Surface Reflectance (MOD09)

The correct retrieval of the Earth's surface reflectance is the baseline for accurate time-series analysis and for almost all higher level datasets, including land cover (MOD12), vegetation indices (MOD13), or LAI (MOD15). The approach builds on the heritage of AVHRR and SeaWiFS and takes advantage of new sensor developments (Vermote et al. 2002). The surface-reflectance product (MOD09) is computed for all non-cloudy pixels that passed the Level 1B quality control. The at-sensor radiance (MOD02, Level 1B) is transformed into ground reflectance, which would have been measured without atmospheric scattering and absorption (Vermote et al. 1997a). A simplified 6S radiation transfer model is used to perform a full atmospheric correction (Vermote et al. 1997b, Vermote and Vermeulen 1999). If available, the algorithm employs MODIS-derived atmospheric products to account for atmospheric scattering, surface pressure, water vapor, ozone, and other gasses. Otherwise, modeled climatology data from the National Center for Environmental Prediction (NCEP) are used as backup data. Aerosol optical thickness, the most critical parameter for visible and near infrared bands, is solely based on MODIS data and is computed in the atmospheric product suite. Furthermore, the algorithm corrects for coupling between surface and atmospheric BRDF and the adjacency effect caused by the atmospheric point spread function (Vermote et al. 1997a). Validation is performed at selected homogeneous sites with Landsat 7ETM$^+$ data. The analysis revealed good agreement (Vermote et al. 2002), with typically not more than 5% deviation (Liang et al. 2002). Current efforts towards validation stage 2 (Morisette et al. 2002) show 94 % and 86 % agreement for the red and near infrared band, respectively.

Surface-reflectance data are provided daily as Level 2G and as 8-day composites in 250 m, 500 m and 0.05° (C5 only) resolution. Table A.3 lists the SDS provided in each 8-day

composite of 500 m surface-reflectance data (MOD09A1) used in this study. Besides the reflective bands stored in seven two-dimensional SDS extensive pixel-level angular and quality information is provided. Details of the quality flags are shown in Table A.4 and Table A.5 for band specific quality issues and atmospheric / surface state, respectively.

MODIS Land Surface Temperature and Emissivity (MOD11)

Similar to the surface reflectance, land surface temperature is a key parameter, e.g. for the land-cover product (MOD12) and thermal anomaly and fire product (MOD14). LST products are required for several climatic, hydrological, ecological, and biogeochemical studies (Justice et al. 1998). While LST is the measured temperature (in Kelvin) at the top of the Earth's surface at the contact with the atmosphere, the emissivity (unitless) is the ratio of the radiant flux exiting the surface and a blackbody at the same temperature (Jensen 2007). While MIR MODIS bands (20, 21, 22, and 23) are also relevant to land applications such as fire mapping, TIR bands (9, 31, and 32) are primarily used for emissivity and LST mapping (Petitcolin and Vermote 2002).

Two algorithms are used for LST retrieval of non-cloudy pixels, a common split window approach for 1 km data (Becker and Li 1990, Wan and Dozier 1996) and a newly developed day / night method for coarse spatial-resolution products (Wan and Li 1997). Depending on the land cover derived from the MOD12 product the split-window algorithm computes the LST from the linear difference of the brightness temperature between bands 31 and 32. The algorithm accounts for view angle, water vapor, and atmospheric lower boundary temperature. Similarly, band emissivities are estimated from the IGBP classification of the MOD12 product and BRDF and emissivity modeling (Snyder et al. 1998, Snyder and Wan 1998). The new physically-based approach computes band-averaged emissivities and day and night temperatures of 4.6 km spatial-resolution cells and climate modeling grids (CMG). The approach is based on day / night pairs of MODIS MIR and TIR bands. This approach employs MODTRAN-derived look-up tables for efficient processing and also accounts for atmospheric uncertainties such as day / night near surface air temperature and water vapor (Wan and Li 1997, Wan et al. 2002, Wan et al. 2004). Validation of LST is difficult and occurred on homogeneous surfaces and different cover types during several field campaigns. It revealed differences of less than 1 K for temperatures between 263 K and 323 K (Wan et al. 2002, Wan et al. 2004).

MOD11 products are provided as Level 2 granules and Level 3 tiles with daily, 8-day, and monthly compositing periods. Furthermore, Level 3 LST and emissivity is derived at 1 km and 4.6 km spatial resolution and in global CMGs of 0.05˚. This study used 8-day composites

of averaged day and night LST with 1 km resolution (MOD11A2). Additional pixel-level information such as data-quality and retrieval error, view zenith angle, time of observation, and the number of clear sky conditions is available (see also Table A.6). Table A.7 lists the quality specifications of MOD11A2 data.

MODIS Land Cover and Land Cover Dynamics (MOD12)

The MODIS land cover and land-cover dynamics product (MOD12) retrieves land-cover classes and phenology information (Friedl et al. 2002, Zhang et al. 2003). However, only the dynamics product has been used in this study (MOD12Q2; Zhang et al. 2003), which is based on the EVI (see below) and computed from nadir BRDF-adjusted surface-reflectance data (see below). The dynamics product indicates phenologically relevant days including the beginning of green-up, maturity, senescence, and dormancy. Following the identification of periods with increasing and decreasing vegetation index, EVI values are fitted to piecewise logistic functions. Next, rate of changes in the curvature are computed, and points of local minima and maxima indicate the phenological transitions (Zhang et al. 2003). The algorithm is capable to estimate up to two full phenological cycles per year. A rigorous quantitative validation of the vegetation dynamics product has not yet been published. First visual inspections of the dataset corresponded with the expected pattern, e.g. for the mid-latitudes, where the timing of green-up is fairly well known (Zhang et al. 2003).

The annual MOD12Q2 product with 1 km spatial resolution requires two years of data for processing. It is generated in rolling composites of six months to also account for peak phenologies of the southern hemisphere. The product contains the phenological dates in three-dimensional SDS, where the third dimension stores a potential second phenological cycle (see Table A.8). Additional information such as the value of minimum and maximum EVI, the integrated EVI over the growing season, and preliminary quality estimates are provided. However, the quality information of this product is still incomplete and has not been used for data processing in this study.

MODIS Vegetation Index Product (MOD13)

Since the beginning of terrestrial remote sensing, vegetation indices have been used to characterize the land surface and the study of land-surface dynamics (Justice et al. 1985, Townshend and Justice 1986, Tucker 1979). The widely used NDVI, a normalized difference between the red and near infrared, is computed to continue long-term time series of the AVHRR instruments. Other vegetation indices were developed, for instance, accounting for

sparse vegetation (SAVI; Huete 1988), atmospheric interferences (ARVI; Kaufman and Tanre 1992), or the combination of both (SARVI; Huete et al. 1994). Due to additional visible bands of MODIS, the NDVI was complemented with the EVI, a novel index which also incorporates the blue and red surface reflectance (Huete et al. 1997, Huete et al. 2002). Compositing of vegetation indices is most important for time-series analysis and can considerably improve the image itself. Depending on the number of cloud-free observations, different compositing methods are applied. BRDF compositing (van Leeuwen et al. 1999) is used for more than five cloud-free observations per pixel, the constrained view maximum value compositing (CV-MVC; Huete et al. 2002) for one to five good pixels, simple vegetation index computation for one good observation, and the MVC method (Holben, 1986) if all observations have limitations. Vegetation index uncertainties are estimated to be lower than 2% for the dynamic range (Miura et al. 2000). The comparison of the MOD13 products for several ecosystems and periods showed good correspondence to other physical and biophysical parameters including surface reflectance and LAI. The MODIS NDVI showed a much higher dynamic range in vegetated areas due to a narrower spectral range of the red and NIR bands compared to the AVHRR instruments (Huete et al. 2002). Compared with NDVI, EVI shows a better stability, is less susceptible to atmospheric disturbances, and does not saturate in high biomass regions.

MOD13 data are provided as 16-day and monthly composites and with spatial resolutions of 250 m, 500 m, 1 km, and as CMG with 0.05˚. This study employed 500 m datasets with a 16-day compositing period (MOD13A1). The product consists of eleven two-dimensional SDS (Table A.9) containing the indices, extensive quality estimates (Table A.10), critical ancillary data such as view and sun zenith angles, and surface reflectance. Exemplarily, C4 and C5 data of the MOD13A1 product are compared in this study. The C5 product contains twelve two-dimensional SDS (Table A.11) and a rearranged quality assurance science dataset (QA-SDS; Table A.12) as well as a newly introduced reliability SDS (Table A.13; Colditz et al. 2007b, Didan and Huete 2006).

MODIS Aerosol / BRDF Product (MOD43)

Surface albedo, BRDF parameters, and nadir BRDF-adjusted surface reflectance (NBAR) data are provided in the MOD43 product suite (Schaaf et al. 2002). For wide-field-of-view sensing instruments such as MODIS, BRDF-corrected datasets are particularly important for consistent annual time series and long-term data records over several years. The MODIS BRDF / albedo algorithm uses the semi-empirical Ross-Thick-Li-Sparse reciprocal kernel (Lucht et al. 2000, Wanner et al. 1997) accounting for volume scattering (Ross 1981), surface scattering, and shadowing (Li and Strahler 1992). If at least seven cloud-free observations

passed initial tests a full inversion is attempted, otherwise a magnitude inversion is computed (Strugnell and Lucht 2001). Predominant snow cover during the compositing period is considered by the algorithm. In order to process the NBAR data, the BRDF parameters are applied to atmospherically corrected surface-reflectance data of the MOD09 product with a mean solar zenith angle of the compositing period. This dataset mimics Landsat TM / ETM^+ observations but on a coarser spatial resolution and is ideal for time-series analysis and advanced classification approaches (Schaaf et al. 2002). Comparisons of NBAR data with corrected Landsat $7ETM^+$ imagery show very good agreements with a deviation of less than 1% (Liang et al. 2002). Moreover, temporal plots exhibit a very smooth and consistent profile, useful for time-series analysis (Schaaf et al. 2002).

MOD43 data are provided as 16-day composites at 500 m (C5 only), 1 km, and as CMG with 0.05˚ spatial resolution. This study employed the 1 km MOD43B4 NBAR product. It consists of two SDS with three dimensions each (Table A.14). The first SDS stores seven bands of nadir BRDF-adjusted surface-reflectance data, and the second SDS contains extensive quality information of surface state and band dependent issues (Table A.15 and Table A.16).

2.3 Study Sites and Land Cover Datasets

Altogether, four study sites of different characteristics have been selected for this study, namely the Republic of South Africa, the Republic of Namibia, the Federal Republic of Germany, and Minnesota in the United States of America (for approximate location in MODIS global tile system see Figure A.1). The rationale for selecting these widely distributed sites is the aim to cover diverse environmental settings and different reference datasets as an input for classification. It should be noted, however, that South Africa and Germany are the main sites, whereas Namibia and Minnesota complement the study. The description of the environmental characteristics relevant to remote sensing will be followed by an illustration and brief explanation of the land-cover datasets, which are used as training and validation data in this study.

2.3.1 Physical Setting

This section does not attempt to provide a thorough characterization of the regional geography of each study site covering all physical and anthropogenic systems, because many of those complexes cannot be observed from space. Table 2.4 contains an overview of the most important characteristics with emphasis on features that can be acquired using satellites. In the

Table 2.4: Most important geographical characteristics of study regions with relevance to coarse resolution satellite imagery.

	South Africa, coastal lowlands and Drakensberg escarpment	South Africa, remaining country	Namibia	Germany	Minnesota
Elevation [m]	0 – 3,500	1,200 – 2,000	1,000 – 1,500	0 – 1,000	200 - 300
Topography	Mountainous	Flat	Mountainous	Flat to mountainous	Flat
Climatic region (Köppen)	Cf	BW, BS, Cs, Cw	BW, BS	Cf	Df
Precipitation rate	High	Low to moderate	Low	Moderate	Moderate
Precipitation variability	Low	High	High	Low	Low
Temperature	Moderate	High	High	Moderate	Moderate
Seasonality	Stable	Variable	Variable	Stable	Stable
Vegetation characterization	Forest	Desert to woodland	Desert to woodland	Deciduous forest	Deciduous forest
Landscape structure	Moderate	Low	Low	High	High
Patch size	Moderate	Large	Large	Small	Moderate
Patchiness	Moderate	Low	Low	High	Moderate
Transitions	Transitions and boundaries	Transitions	Transitions	Boundaries	Boundaries

following paragraphs, in particular topography, climate and its effect on vegetation, and the structure and heterogeneity of the landscape are described.

Topography is an important characteristic of the landscape. It influences the land use, patch size, and phenology. Furthermore, satellite data are often distorted in mountainous terrain where shadow also becomes a significant issue. The flat lowlands of Minnesota have a mean elevation of 370 m with few peaks up to 700 m. While the northern lowlands of Germany are flat, central Germany is characterized by several mountain ranges with peaks between 800 m and up to 1,500 m and deeply incised valleys. The southern part consists of escarpments and wide plateaus between 300 m and 500 m with higher uplands at the eastern and western edge, but increases in elevation up to 1,000 m towards the Alps. The low topography in the interior with an elevation ranging between 900 m and 1,300 m typifies large portions of Namibia. Along the escarpment in the western part, peaks rise up to 2,500 m. The terrain is even steeper and higher in South Africa. While extensive parts of the interior increasing from 1,000 m in the West to 1,500 m in the East are comparatively flat, the encircling escarpment rises well above 2,000 m and peaks at 3,400 m in the Drakensberg region. Coastal lowlands are narrow except for the eastern portion at the border to Mozambique.

Climatic characteristics substantially determine the state of the earth surface and in particular contribute to the development of vegetation. According to the Köppen climate classification, Minnesota belongs to the snow climate regions with sufficient precipitation during all months

(Df). Germany and the eastern part of South Africa including the Drakensberg escarpment are located in the moderately warm zone with precipitation during the entire year (Cf). A small portion of the northern provinces of South Africa is characterized by a distinct dry period during wintertime (Cw). The western part of the Cape Floristic Region exhibits a Mediterranean climate with winter rain and a dry summer period (Cs). The interior of South Africa and entire Namibia belongs to the dryland climate with desert and steppe characteristics (BW and BS).

Temperature and precipitation are regarded the most influencing climatic variables on vegetation activity. In contrast to temperate regions of mid-latitudes where temperature determines the state of vegetation, the precipitation rate and variability is most important for semi-arid ecosystems in the sub-tropics (Bonan 2002). In Minnesota and Germany, mainly temperature guides the phenological states of vegetation. The inter-annual phenological variability is rather low. A clear green-up in spring, a long plateau phase during summer, and a rapid senescence in fall caused by rapidly decreasing temperatures and frost at night characterize natural vegetation units. Climatic variables, driving the phenological development of vegetation in Germany, are slightly modified by aspects such as increasingly continental climate towards the South, elevation, or rain shadow effects (Glaser et al. 2007). In contrast to Germany, the higher range and extremes in temperature are typical for Minnesota because of its continental location (Bailey 2002). Except for higher elevations, the temperature in both, Namibia and South Africa is appropriate for year-round vegetation growth. Only the increasing continental climate towards the interior with a higher range in temperature causes several days of frost, which is unsuitable for certain vegetation types such as Mopane woodlands in Namibia (Mendelsohn et al. 2002). However, the limited water availability and the insufficient moisture storage capacities of soils cause seasonal patterns. A long dry period, increasing with distance from the ocean and amplified by the rain shadowing effect of the encircling escarpment with frequent inversions during winter, characterizes eastern South Africa. It is interrupted by a rain season in summer (van der Merwe 1994). Descending air masses, trade winds, and the cold Benguela current offshore western South Africa and Namibia keep moist air masses away from the western regions. Northern Namibia is influenced by heavy rains from the southward moving inner-tropical convergence zone (ITCZ) between November and March (Mendelsohn et al. 2002). During wintertime, cold fronts cause rain in the Cape Region (van der Merwe 1994).

In addition to the lack of precipitation, a considerable inter-annual variability of rain with regard to timing, intensity, and duration exhibits further stress on the vegetation in South Africa and Namibia. Climate variability is mainly attributed to the El Niño Southern Oscillation (ENSO) phenomenon and the variation of sea surface temperature (SST) in the southern Atlantic and Indian Ocean (Tyson and Preston-Whyte 2000, Vogel and O’Brien

2003). Their influence on vegetation for the 1997 / 1998 ENSO event has been demonstrated by Anyamba et al. (2002). The scarcity and variability of precipitation increases to the northern and western parts of South Africa, whereas the northern part of Namibia receives more precipitation from the southward moving ITCZ. Vegetation growth is highly dependent on rain occurrence, and species have adapted to the variability of precipitation. Local thunderstorms can initiate vegetation growth of annuals, whereas in a distance of some kilometers no green-up is observed. Furthermore, the intensity of precipitation can be enormous, and areas become literally flooded (Mendelsohn et al. 2002). Regions with annual vegetation growth and lower variability are found along the southern and eastern coast extending northward along the escarpment of South Africa. A combination of westward moving tropical cyclones, frontal systems, and orographic effects at the escarpment ensures sufficient water for vegetation growth during summer (Vogel 2003), but the timing may vary approximately one month.

Although not directly related to climate, important modifiers such as fires or the water storage capacity of the soil should not be disregarded by analyzing the vegetation activity of semi-arid regions. Fire can dramatically influence the environment and leads to an alteration of landscape characteristics for a particular year. With regard to remote sensing fire drastically affects the phenology, and therefore the temporal course of, for instance, vegetation indices.

A characterization of the landscape structure, i.e. patch size, patchiness, and transition, is very important with respect to the coarse spatial resolution of MODIS, the structure and heterogeneity of the land surface is revealed in satellite images. The detection of these features, however, is directly related to the spectral and especially spatial resolution of the satellite data, as structures smaller than the pixel size may not be represented. Observations of Germany from space indicate a highly structured, even dissected landscape with many patches beyond the resolution of 500 m. Almost all of the country is thoroughly shaped by more than 80 Mio citizens, and agriculture clearly dominates with an area proportion of 53 %. Furthermore, 30 % of the country is covered by managed forests and 13 % by built-up areas (Glaser et al. 2007). Apart from some small exceptions, natural landscapes are absent. Located in the U.S. Midwest, Minnesota is characterized by a well-structured landscape, but compared with Germany the average size of patches is considerably larger. A spatial resolution of 500 m and even 1 km is considered suitable for most areas. The southern part is intensively used for agriculture, and forest dominates in the Northeast (McKnight 1992, Paterson 1994). South Africa and Namibia exhibit a less to non-structured landscape (Mendelsohn et al. 2002, van der Merwe 1994). Broad ecotones are clearly visible in satellite images. Many of these low biomass ecoregions, e.g. the Karroo in the western interior of South Africa and southern Namibia, are vulnerable due to high inter-annual variations attributed to the above described climatic variability (Dean and Milton 1999). Droughts

during El Niño events have dramatic influences on the natural ecosystems and on agricultural areas of dryland farming in the so-called bread basket of central and northern South Africa (Dilley 2003). It has been argued that inappropriate land-use practices have irreversibly changed the landscape of these sensitive areas. However, these statements of Acocks (1988) that - driven by overgrazing - semi-desert landscapes such as the Karroo are expanding into surrounding grasslands, has meanwhile been questioned (Hoffman and Ashwell 2001). Some areas in South Africa, e.g. the Cape, the greater Johannesburg region, areas in Kwazulu Natal and agricultural regions in the central part, show a clearly structured landscape. However, for the majority of South Africa and Namibia a resolution of 1 km is deemed sufficient. Only urban areas and parts of the Eastern Cape and KwaZulu Natal exhibit patches clearly smaller than 1 km.

In conclusion, there are two distinct settings. Germany and Minnesota expose a highly structured landscape in the mid-latitudes where phenology is driven by a distinct seasonal increase and decrease in temperature. However, these two study regions are different with regard to the size of patches. Many landscape features in Germany are much smaller than the employed MODIS resolution of 500 m, but 1 km cells are regarded sufficient for Minnesota. On the other hand, South Africa and Namibia reveal a less structured landscape with large portions of natural or semi-natural ecosystems and broad transition zones where 1 km spatial resolution seems suitable. Both are located in a sub-tropical humid to largely semi-arid and arid climate. Vegetation dynamics are driven by highly variable precipitation occurrence and intensity.

2.3.2 Land Cover Datasets

This section describes the land-cover datasets of the four study regions. The land-cover data have been recoded to legends appropriate for the spatial resolution of MODIS data. In this respect, the resulting datasets are employed as training or test data for time-series classification or accuracy assessment.

The National Land Cover classification of the Republic of South Africa from 1994-95 (NLC1995) contains 31 land-cover and land-use classes. The classification was derived from single date Landsat 5TM datasets by visual image interpretation and a mapping scale of 1:250,000. A minimum mapping unit (MMU) of 25 ha was applied, which should not significantly limit the results when considering the landscape characteristics for the clear majority of the country. Following recommendations of Thompson (1996), a hierarchical classification scheme with twelve classes at level 1 and 23 classes at level 2 was used. A third level is flexible and depends on the definition of the individual user. An estimation of the

classification accuracy for the NLC1995 has not been available. It should be noted that the classification of 1994 / 1995 has been updated with multi-temporal Landsat 7ETM$^+$ images of 2000, called NLC2000. The legend consists of 49 classes, which are also in accordance with the definitions of Thompson (1996) and convertible into the 31 class scheme of the NLC1995. The mapping scale of the NLC2000 is 1:50,000, and a MMU of 1-2 ha was used. However, the NLC2000 was not available at the time of data processing.

For this study the original classification scheme of the NLC1995 with 31 classes has been recoded to 14 classes (Figure 2.3). The recode scheme is provided in Table A.17. Recoding has been necessary to obtain sufficiently large training sample sets for every class and to combine classes which cannot be distinguished by multi-spectral and multi-temporal approaches, e.g. detailed urban sub-classes. In this study, NLC1995 land-cover data of South Africa are projected in Albers equal area (Table A.18).

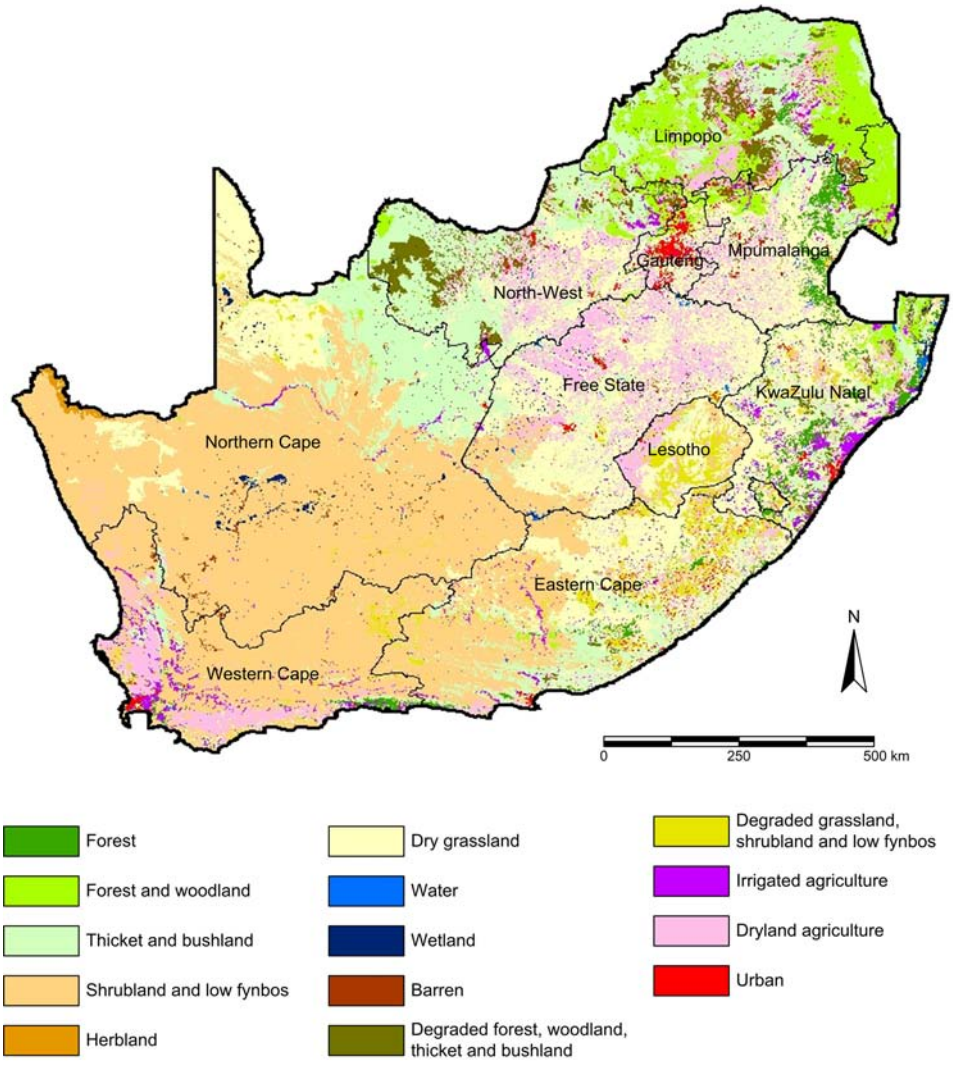


Figure 2.3: Land cover map of South Africa.
Note: The original NLC1995 has been recoded to 14 classes. The recode scheme is provided in Table A.17. All data sets of South Africa are Albers equal area (Table A.18). The scale bar, north arrow, and legend are not shown for all other figures.

The CORINE land-cover classification is a pan-European effort for consistent land-cover mapping. The first classification, CORINE land cover 1990 (CLC1990), was based on Landsat TM data, mainly of the year 1990. Landsat 7 ETM^+ images recorded between 1999 and 2001 were used to generate the European Landsat mosaic IMAGE2000, which is used as baseline for CLC2000. In order to generate a spatially and spectrally continuous dataset, IMAGE2000 generation involved a radiometric and geometric correction (De Lima 2005). The legend of Germany comprises 36 out of 44 classes of the European dataset. CLC mapping follows a strict hierarchy of three levels and considers land-cover as well as land-use classes. The classification was mainly accomplished by visual image interpretation and only supported by digital image processing (Keil et al. 2005). The MMU of 25 ha limits the identification of many small patches in Germany. The overall classification accuracy of CLC2000 for Germany is 85 %. Approaches to automated and semi-automated CLC update for Germany have been developed (Wehrmann 2007). There are on-going efforts to update CLC for 2005 / 2006 using IRS-P6 LISS-3 and AWiFS data.

For this study the CORINE classification scheme has been recoded to 14 land-use and land-cover classes (Figure 2.4). In particular, urban classes as well as some scattered small-patch land-cover types have been combined. The recode scheme varies between all hierarchical levels (Table A.19). It should be noted, however, that class combinations such as “beach”, “barren”, “sparse vegetation”, and “glacier” to a general category such as “barren” also lead to higher internal class variability. The projection of CLC2000 land cover of Germany is UTM (Table A.18).

The multi-temporal, multi-spectral, and multi-source National Land Cover Database of 2001 (NLCD2001) of the United States updates the NLCD1992 classification. In contrast to the NLC1995 and the CLC2000 datasets, no MMU has been defined; hence the original spatial resolution of 30 m is preserved. The effects of no, small and large MMUs on the classification of time series has been assessed with the NLCD2001. For NLCD2001 processing, Landsat images were selected during phenological key stages such as green-up, peak, and senescence. The appropriate timing was determined by a previous analysis of AVHRR time series (Yang et al. 2001). The NLCD2001 database contains radiometrically calibrated as well as geometrically and atmospherically corrected Landsat 5TM and Landsat $7ETM^+$ imagery. The database was complemented with ancillary datasets, including a digital elevation model and its derivatives, a Tasseled Cap transformation, estimations of imperviousness and tree cover, a delineation of the country into 66 homogeneous mapping zones, and GIS data such as roads, population density, and city lights (Homer et al. 2004). Classification trees using the C5.0 algorithm (Quinlan 1993) were used to generate decision rules and derive 27 land-cover classes. Class-specific accuracies are between 73 % and 77 %. (Homer et al. 2004).

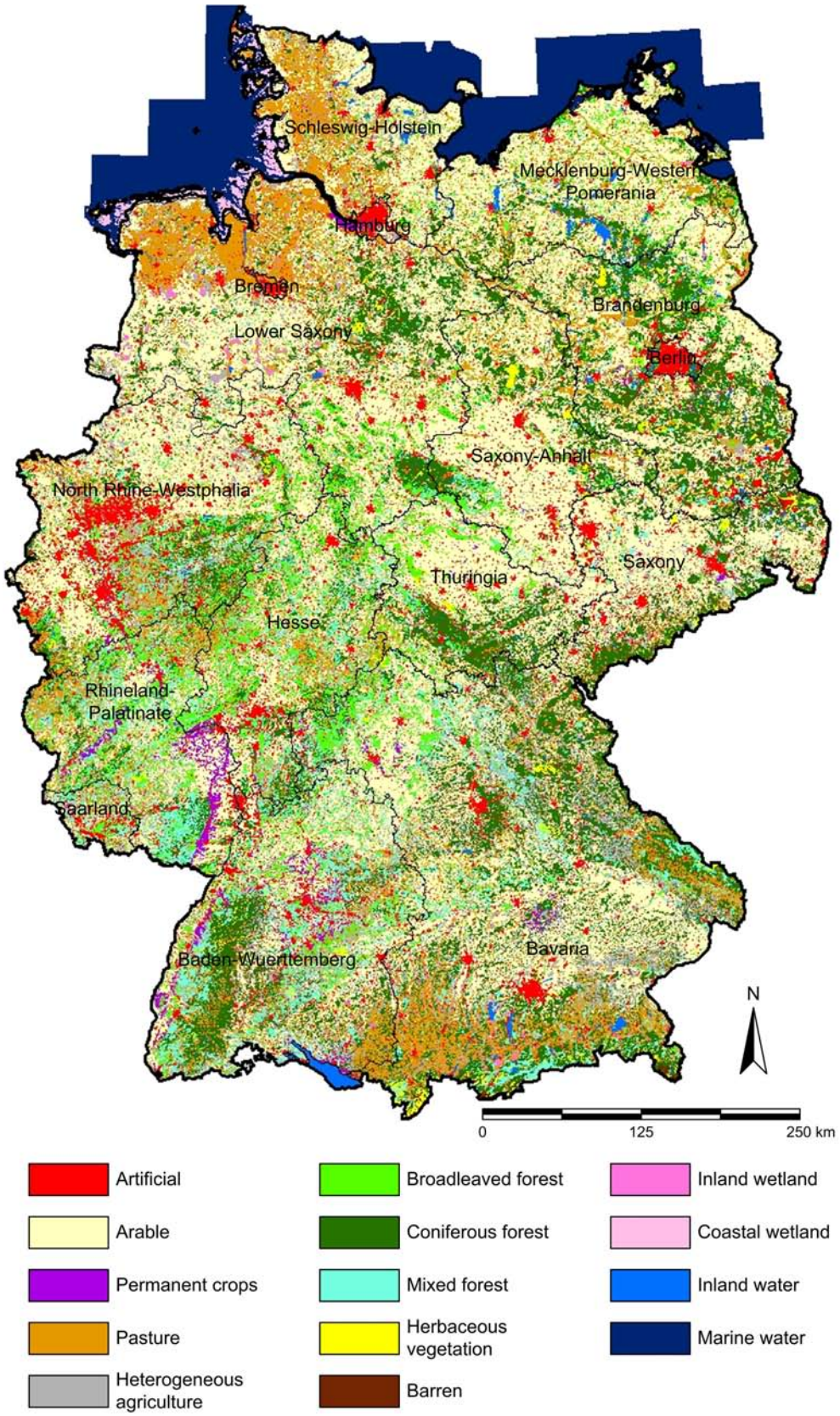


Figure 2.4: Land cover map of Germany.
Note: The original CLC2000 has been recoded to 14 classes. The recode scheme is provided in Table A.19. All data sets of Germany are UTM zone 32 (Table A.18). The scale bar, north arrow, and legend are not shown for all other figures.

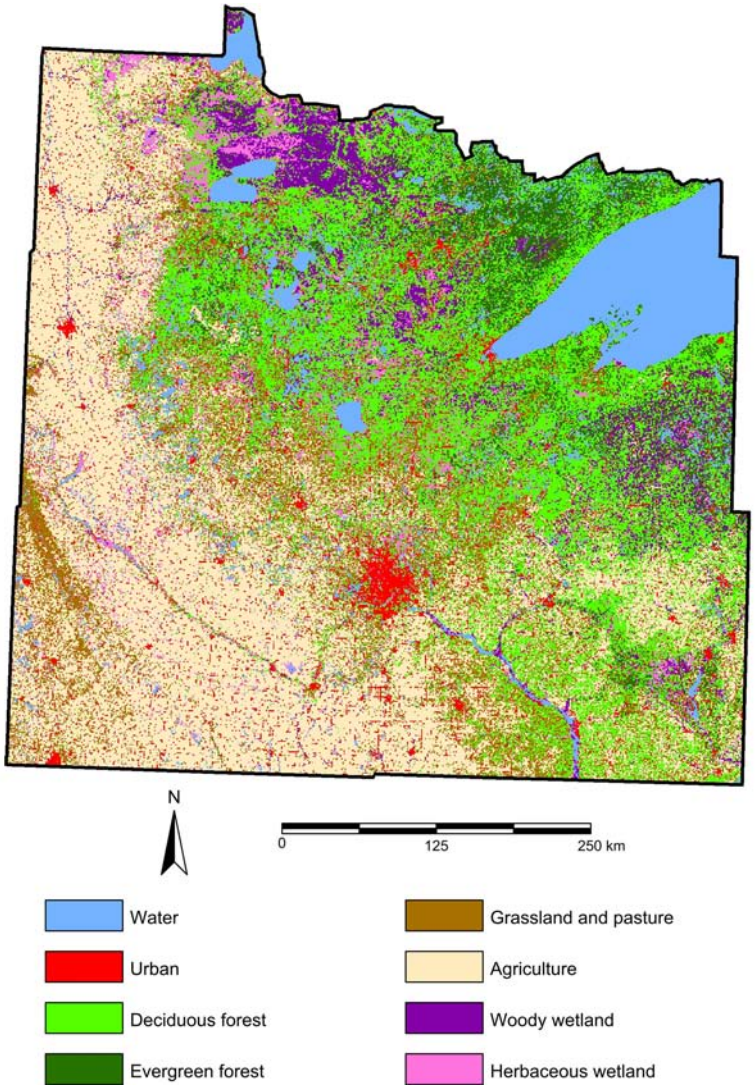


Figure 2.5: Land cover map of Minnesota.
Note: The original NLCD2001 has been recoded to 8 classes. The recode scheme is provided in Table A.20. A very small portion in the easternmost and westernmost edge of the country was not mapped to limit data processing. All data sets of Minnesota are UTM zone 15 (Table A.18). The scale bar, north arrow, and legend are not shown for all other figures.

For the state of Minnesota 15 out of a total of 21 land-cover classes for the entire USA (27 including wetlands in coastal regions) were mapped. The classification scheme of this study has required a recode to eight classes (Figure 2.5). The recode from the original NLCD2001 classification scheme is shown in Table A.20. The projection of the NLCD2001 land-cover data of Minnesota is Albers equal area (Table A.18).

An existing vegetation map of Namibia with a total of 29 vegetation types has been used as a base-line dataset (Mendelsohn et al. 2002). This GIS dataset is not a land-cover map in a strict sense but is meant to demonstrate the potential of the classification procedure if applied to other types of land-surface data. The map was originally compiled from multiple vegetation maps including Giess (1971), Mendelsohn and Roberts (1997), Mendelsohn et al. (2000), Mendelsohn and el Obeid (2003), and Strohbach (2001).

Altogether 23 vegetation types have been remapped excluding the Caprivi and some small and scattered classes, which are not detectable with 1000 m pixel size (see Figure 2.6). The classification scheme is provided in Table A.21. The projection of the vegetation types of Namibia is Albers equal area (Table A.18).

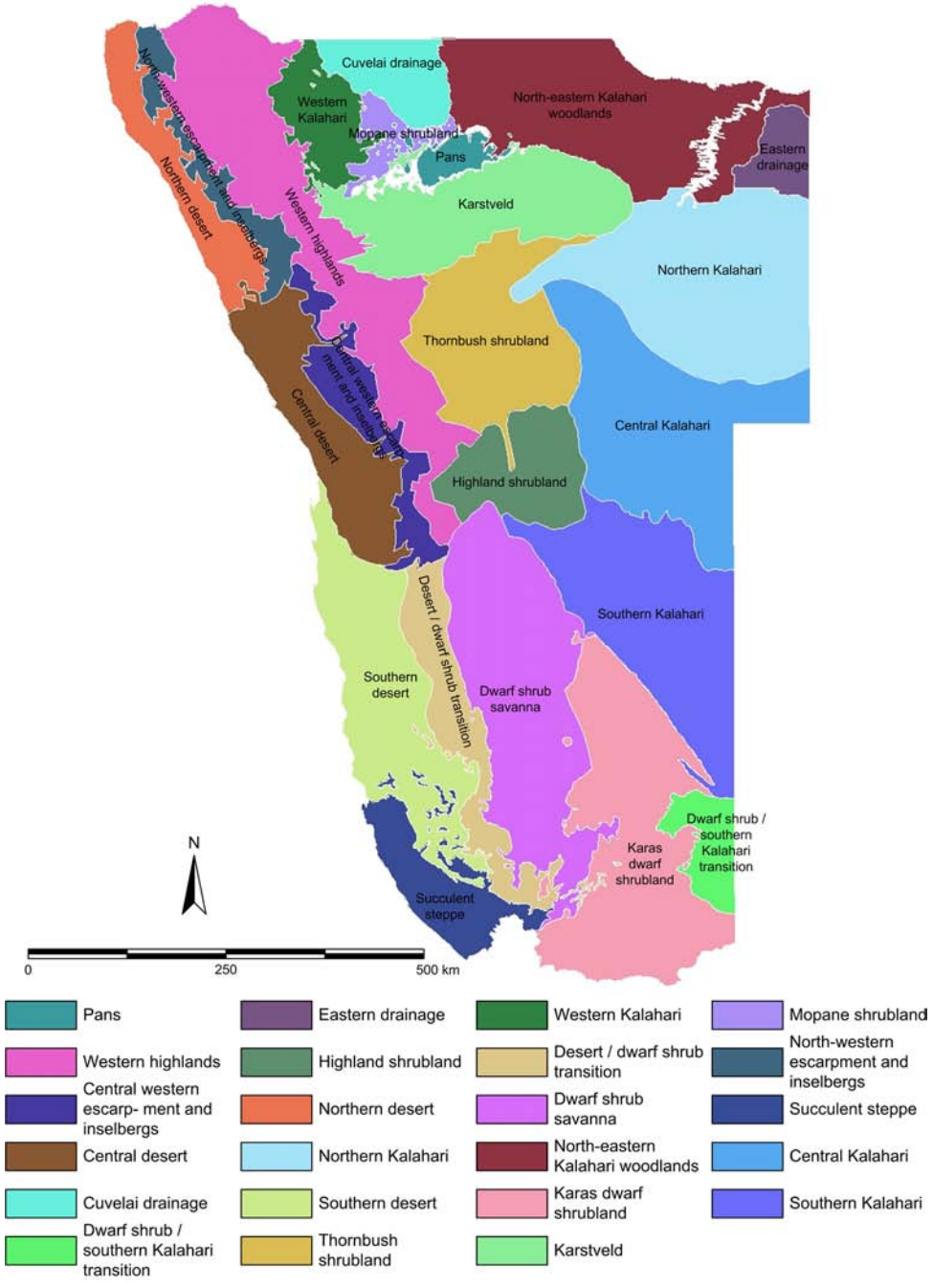


Figure 2.6: Vegetation types of Namibia.
Note: The original vegetation map has been reduced to 23 classes, excluding the Caprivi areas and spatially scattered vegetation types. The complete list of classes is provided in Table A.21. All data sets of Namibia are Albers equal area (Table A.18). The scale bar, north arrow, and legend are not shown for all other figures.

Chapter 3

Time Series Compilation

While the first section reviews approaches of time-series generation and highlights several important prerequisites, the second main part of this chapter will introduce an advanced methodology of time-series compilation for MODIS data and will explain TiSeG's functionalities, software design, and performance of critical processes. The third section will focus on the analysis of MODIS time series. In a first study the suitability of temporal interpolation algorithms of TiSeG will be analyzed. The next part in the analysis compares improved and original time series of Germany. Thirdly, differences in collection 4 (C4) and collection 5 (C5) data of the MOD13 dataset will be demonstrated for Germany. The use of time series for phenological studies will be illustrated in a fourth part by analyzing multiple annual time series of South Africa. This case study indicates temporal shifts in seasonality due to a high variability of precipitation of this semi-arid environment.

It should be noted that the focus of this chapter is on the analysis of temporal patterns. Usually images are evaluated with regard to their high spatial autocorrelation or their spectral characteristics. This study, however, aims at analyzing the temporal consistency of image time series and concentrates on the characteristics of each pixel as a vector in time.

3.1 Theoretical Background

The generation of accurate time series is a critical, if not the most critical, step in time-series processing of remote-sensing data. A considerable level of noise and multiple sources of errors require data analysis before or during time-series generation. Besides sensor artifacts, cloud cover and other atmospheric gases and aerosols are major sources of errors in multi-spectral data. Time series of surface parameters such as vegetation indices require the removal or interpolation of those factors.

Time-series generation can be considered as an additional preprocessing requirement, because the result, the time series, is the object to be analyzed. Similarly to other fields of remote sensing, inaccurate preprocessing, for instance improper geographical alignment (Roy 2000) or incomplete atmospheric correction (Cihlar et al. 1997), will significantly alter the result of the analysis and probably lead to low-quality or erroneous findings.

However, studies on error analysis of observations in a time series are scarce. Much of the statistical and applied literature on time series, e.g. in economics, business, and natural and social sciences, focuses on stochastic models of time series and transfer functions for forecasting (Box et al. 1994, Chatfield 2004). Many studies in these fields do not require data analysis before time-series generation. If conducted, error analysis of the input data is usually limited to a general outlier analysis. Other approaches are often also not feasible, because there is no ancillary information for quality control of single observations or measurements.

Remotely sensed data, and in particular datasets from the MODIS sensor, contain extensive additional information such as data-quality indicators and angular datasets at the pixel level. In some MODIS products, e.g. MOD09, MOD13 and MOD43, invalid pixels are not indicated by fill values but contain “actual” data regardless of their quality. Only the ancillary data layers indicate the validity and level of accuracy. It has been concluded that the use of the ancillary dataset is required for accurate data analysis of single observations and time-series generation (Roy et al. 2002a). Due to the focus of the statistical literature already explained, there is a significant lack of knowledge and a need for approaches how to generate time series if additional indicators of data quality are available.

The following sections of the theoretical background address several important issues. First, important considerations on data interpolation with respect to the context of the data are discussed. Next, three broad groups of time-series generation are reviewed. The final section discusses possibilities of time-series comparison and validation strategies.

3.1.1 General Considerations on Interpolation

An important consideration in data analysis is the handling of data gaps. Many subsequent approaches of data analysis require a consistent dataset without invalid values in space and time. Therefore, suitable approaches of data interpolation are required. In general, interpolation can be applied spatially for each image separately or temporally for each pixel vector. Its success highly depends on the size of the spatial or temporal gap between valid observations.

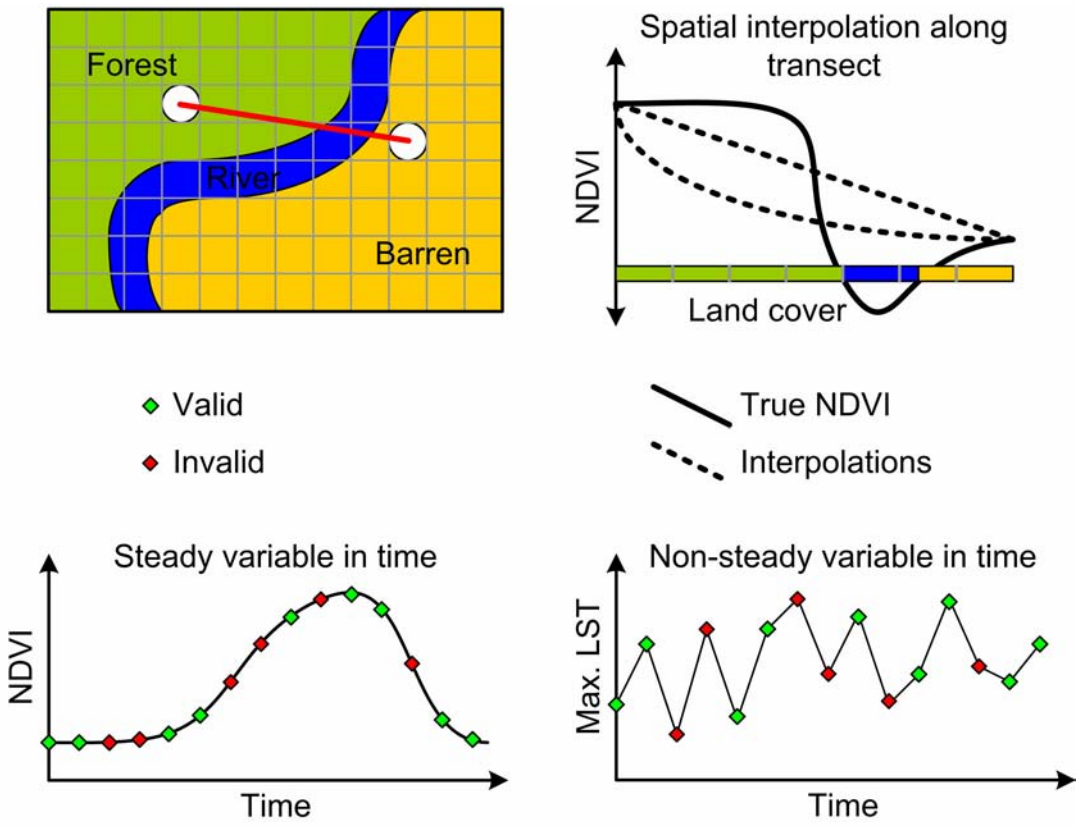


Figure 3.1: Simplified examples for unsuitable spatial interpolation (top) and temporally steady (bottom left) and unsteady (bottom right) time series.
Note: The red line in the upper left indicates the profile depicted in the upper right.

The appropriate selection of a spatial or temporal interpolation should be conducted with regard to the context of the data (see Figure 3.1). For instance, the vegetation index of a known land-cover type, e.g. of forest, can be assumed to first increase steadily and then decrease, with no simultaneous land-cover change. The phenological development of a natural land surface is not assumed to alter dramatically during a short period. The resulting curve can also be considered steady in a mathematical sense, and a temporal interpolation would be appropriate (Figure 3.1 bottom left). On the other hand, a spatial interpolation would yield erroneous results due to the lack of spatial coherence in the landscape. For instance, the interpolation along the profile at the top of Figure 3.1 would, according to the selected interpolation algorithm, follow one of the dashed lines. It is impossible to spatially interpolate

correctly without additional knowledge such as land cover. However, one goal of this study is time-series generation for land-cover mapping.

Land surface temperature (LST), however, changes dramatically during the day and over several days and depends on several variables such as incoming solar radiation, wind speed, surface roughness, air temperature, and precipitation. In mathematical terms, LST is clearly unsteady in time (Figure 3.1 bottom right). Depending on the study, it might be more suitable to apply a spatial interpolation to each particular image, assuming spatial coherence between the valid data, because LST is not assumed to change dramatically. If these assumptions are not valid, observations of insufficient quality have to be masked. It should also be noted that the interpolation of LST depends on the observation period and possibly the merging of several acquisitions to one image, in the following termed composite. A composite integrates several observations by a defined function, such as average, sum, or maximum. A time series of mean LST composites may be considered steady, because the averaging procedure will have leveled the peaks or lows of single observations. In this case temporal interpolation becomes an option. On the other hand, maximum-value compositing of LST would not yield any meaningful temporal interpolation. In contrast, LST measurements acquired consistently every couple of minutes by geostationary satellites can also be interpolated temporally, as the daily cycle of LST can be approximated with sufficient confidence. Thus, temporal steadiness depends on the parameter to be studied, the temporal resolution of image acquisition, and the selected compositing approach.

It can be concluded that steady observations in time should be interpolated by temporal methods. Temporally unsteady data should be filled with a defined flag value or, if appropriate, interpolated by spatial approaches. Complex combinations of spatial and temporal interpolations are available in several statistical software packages. These approaches, however, have not been sufficiently studied with regard to the special requirements of remote-sensing time series.

3.1.2 Approaches to Time Series Generation

Approaches to time-series generation of remotely sensed data for terrestrial applications can be assigned to three broad categories: (1) compositing and stacking, (2) filtering, and (3) use of ancillary data. A brief overview of approaches to time-series generation is provided in Colditz et al. (2007a).

Compositing and Stacking

The most common and most frequently applied approach to time-series generation is maximum value compositing (MVC; Holben 1986) and stacking of multi-day composites to a time series. The rationale for MVC is that all disturbing influences will lead to lower values than good observations. The assumption holds true for atmospheric artifacts and clouds in particular (Cihlar et al. 1997, Dech et al. 1998). For each pixel separately, the observation with the highest value is selected and stored in the composite. MVC has mainly been applied to NDVI and LST data of the AVHRR sensor (Cihlar et al. 1994, Roy 1997). Many studies have been based on MVC time series, including the derivation of biophysical parameters (Dech et al. 1998, Sellers et al. 1996a), land-cover mapping (DeFries et al. 1995, Hansen and DeFries 2004, Loveland et al. 2000, Moody and Strahler 1994), phenological analysis (de Beurs and Henebry 2004a, de Beurs and Henebry 2005a, Huete et al. 2006, Kasischke and French 1997), and change detection (Lambin and Strahler 1994b, Lambin and Ehrlich 1997, Zhan et al. 2002).

Although the MVC method tends to select cloud-free pixels, the chosen selection does not necessarily prefer near-nadir observations (van Leeuwen et al. 1997). Several studies have shown that off-nadir pixels may have higher NDVI values and will be considerably contaminated by atmospheric particles or clouds (Cihlar et al. 1994, Cihlar et al. 1997, Goward et al. 1991, Moody and Strahler 1994). For PAL processing, CLAVR (Clouds from AVHRR) flags, indicating cloud contamination of AVHRR data (Stowe et al. 1991), have not been used for 10-day NDVI MVC composites. This has led to considerably contaminated time series in tropical areas (Colditz et al. 2006b, Gutman and Ignatov 1996). The remaining atmospheric effects can be reduced by making use of longer compositing periods, e.g. bi-weekly or monthly instead of the common 10-day composites, although thereby limiting the temporal information content of the time series (Goward et al. 1994, Los et al. 1994, Reed et al. 1994). Advances with regard to high scan angles have been made by the MODIS vegetation index product (Huete et al. 2002, van Leeuwen et al. 1997). The constrained view maximum value compositing (CV-MVC) method, for instance, considers the two highest vegetation index values and selects the observation with the lowest view angle. Another difficulty in compositing is the selection of the most appropriate observation while ensuring spatial continuity, i.e. the selection of the actual value varies for neighboring pixels within the compositing period. A contextual approach ensuring spatial continuity is approached in the current C5 production of MODIS vegetation index products (Didan and Huete 2006).

It has been concluded that MVC works best with atmospherically uncorrected data (Cihlar et al. 1994, van Leeuwen et al. 1997). Despite the limitation in atmospheric correction and sensor drift of multiple AVHRR instruments, considerable improvements have been made to generate NDVI time series of 10-day composites such as the GIMMS data record (el Saleous

et al. 2000, Tucker et al. 2005). However, atmospheric correction is a prerequisite for recent developments to ensure intercalibration between multiple sensors and consistency of long-term data records (Justice et al. 2002a). With regard to surface anisotropy and bidirectional reflectance, MVC is not appropriate, as the algorithm tends to select distorted and radiometrically less accurate pixels at off-nadir locations (Cihlar et al. 1994, van Leeuven et al. 1999). Advanced methods are preferable for corrected high-quality composites, including CV-MVC or BRDF compositing for MODIS (Huete et al. 2002, van Leeuwen et al. 1999) or averaging for SPOT VEGETATION composites (Duchemin and Maisongrande 2002). Due to its robustness, MVC is often employed in most advanced compositing approaches as the ultimate backup (Huete et al. 2002).

Filtering

Filtering functions are another possibility for generating time series from single observations or previously generated composites. The filtering functions include simple boxcar and adaptive moving-average filters (Reed et al. 1994), forward-estimation processes assuming steady time-series characteristics (Viovy et al. 1992), Savitzky-Golay filtering (Chen et al. 2004a), Fourier filtering (Jakubauskas et al. 2001, Roerink et al. 2000, Sellers et al. 1994), and fitting to models (de Beurs and Henebry 2005b, Jönsson and Eklundh 2002, Piwowar and LeDrew 2002). Usually the data are assumed to be sampled at equal intervals, but some approaches such as model fitting can also handle non-equidistant sampled data.

A statistical framework for time-series analysis has been presented by de Beurs and Henebry (2005b) combining different tests for discontinuities, trends, and seasonality and quadratic models. The approach has been tested for PAL NDVI data in Central Asian steppe and desert environments and selected areas of the northern high latitudes (de Beurs and Henebry 2004a, de Beurs and Henebry 2004b, de Beurs and Henebry 2005a).

The so-called best index slope extraction (BISE) approach has been suggested by Viovy et al. (1992), whenever no sudden changes are assumed. This is a meaningful criterion for vegetation indices, excluding rapid changes such as fire, flooding, or land-cover conversion. The approach starts a forward search from one good observation, accepting only higher values, but not above a previously defined threshold. Decreases are only acceptable if they are followed by gradual regrowth. Rejected data are interpolated and the search starts again from the next good observation. The BISE approach has been modified and used for PAL NDVI studies in Australia (Lovell and Graetz 2001).

Harmonic analysis, based on the Fourier transformation, is used for both, time-series analysis and time-series generation. The Fourier transform decomposes a function into a continuous

spectrum of its frequency components. The basic idea of harmonic analysis is to approximate a periodic function as the sum of sine und cosine functions. Thus harmonic analysis is meant to separate signal from noise, where the signal is represented by the low frequencies, hence the first couple of Fourier coefficients. Only those coefficients are used for the inverse Fourier transformation, and a series of summed-up sine and cosine waves and an additive term model a complex curve (Jakubauskas et al. 2001). The result is similar to low-pass filtering. Advanced methodologies such as the Harmonic Analysis of Time Series (HANTS) algorithm combine harmonic analysis and least-squares fitting, where the time series is iteratively fitted to the harmonic curve (Roerink et al. 2000). Values outside of previously defined thresholds are replaced by estimates of the inverted Fourier transformation. Harmonic analysis has been successfully applied for generating cloud-free time series and for studying the climate impact (Roerink et al. 2000, Roerink et al. 2003), for describing phenologies and temporal shifts (Menenti et al. 1993, Moody and Johnson 2001, Sellers et al. 1996b), and for crop classification (Jakubauskas et al. 2002).

A step-wise variant of the harmonic analysis has been successfully employed for generating smoothed time series, and to fill temporal as well as spatial data gaps (Bittner et al. 1994, Wuest and Bittner 2006). In contrast to other harmonic-analysis approaches, this variant does not require the availability of all observations, but interpolates missing values. A sine function is matched to a time series, using least-squares fitting, and the residuals are calculated. Another sine function is fitted to the residuals until some previously defined termination criteria are reached, e.g. the number of sine functions to be fitted. To improve the results, the so-called all-step method is used, where previous best fits are optimized during the computation of the current best fitting sine function (Wuest and Bittner 2006). The step-wise approach has been used for temperature and ozone data (Bittner et al. 1994, Wuest and Bittner 2006).

While harmonic analysis seems to be an appropriate methodology for smoothing, fitting, and even interpolation, the symmetrical characteristics of the sine and cosine waves can be problematic with respect to asymmetrical phenological curves (Chen et al. 2004a, Jönsson and Eklundh 2002). While simple moving average filters are often regarded as being too simple, Savitzky-Golay filtering, with the focus on the so-called upper envelope, is deemed sufficient for SPOT VEGETATION MVC data (Chen et al. 2004a). Jönsson and Eklundh (2002) have used least-squares fitting of asymmetric Gaussian functions to remove the considerable noise content of PAL NDVI time series, but additional knowledge, such as the CLAVR clouds index, also improved the results. A multitude of filtering functions has been provided with TIMESAT, a time-series tool for satellite data (Jönsson and Eklundh 2004). Time-series analysis with autoregressive moving-average (ARMA) models (Box et al. 1994) has been conducted by Piwowar and LeDrew (2002).

Ancillary Data

The third general category of time-series generation methods evaluates ancillary information such as data-quality indicators. If the data quality of a given observation is too low, the pixel is excluded from the analysis and can be interpolated by temporal or spatial interpolation (see above). The respective quality indicators are either derived from the data itself, or by analyzing its consistency and cross-checking with other datasets.

Selected quality indicators such as the CLAVR flags (Stowe et al. 1991) or angular and sensor information (James and Kalluri 1994, Kidwell 1998) have provided even for AVHRR time series such as PAL. However, they have rarely been used for compositing or time-series generation (Gutman and Ignatov 1996, Jönsson and Eklundh 2002). An attempt to improve PAL NDVI time series from 1982 to 2000 using CLAVR and angular datasets has been made by Colditz et al. (2006b) for Africa.

New sensor developments comprise extensive quality information and make considerable efforts to ensure correct sensor calibration as a prerequisite for long-term data records. For instance, multiple surface and atmosphere states can be visualized with the BEAM software package for data from the ENVISAT platform (Brockmann 2004, Fomferra and Brockmann 2005).

The unique concept of quality assurance and product validation of MODIS data allows the generation of science-quality time series (Justice et al. 2002a). During MODIS processing numerous quality datasets are generated and stored for downstream data analysis. GIOVANNI, an interactive online visualization and analysis tool at the GES-DISC, analyzes multi-source satellite data. The MODIS sub-system MOVAS (MODIS Online Visualization and Analysis System) generates maps and plots of MODIS atmosphere data of aerosol, cloud, or water vapor (Leptoukh et al. 2005).

The product-dependent data quality of each MODLand dataset is stored as meta-data, and in a separate quality assurance science dataset (QA-SDS) at the pixel level. As already noted, it is highly recommended to evaluate the quality indicators of each dataset, in particular for science-quality time series (Roy et al. 2002a). Linear temporal interpolation has been applied to remove low-quality pixels for subsequent cropland analysis of MODIS NDVI time series (Lobell and Asner 2004). Similarly, Lunetta et al. (2006) have flagged invalid MODIS NDVI data from 2002 to 2006 and filled data gaps with a discrete Fourier transformation for land-cover change detection in the eastern U.S. Other studies of time-series generation using the MODLand QA-SDS have been conducted by Colditz et al. (2005), Conrad et al. (2004), Conrad et al. (2007), Erasmi et al. (2006), Landmann et al. (2005), Neteler (2005), and Wagner et al. (2007). A bundle of code called LDOPE (Land Data Operational Product

Evaluation) has been provided for reading and extracting the data quality, but handling is rather complicated (LDOPE 2007). Therefore, a user-friendly environment of quality analysis and time-series generation of gridded MODLand products (see Table A.2), called Time Series Generator (TiSeG), has been developed (Colditz et al. 2006a, Colditz et al. 2007a, Conrad et al. 2005).

There are possibilities to combine two or more time-series compilation approaches, e.g. quality control of single observations, before applying filtering functions. For instance, TiSeG evaluates the quality indicators of composites and interpolates the resulting invalid data. Optionally, curve smoothing approaches such as harmonic analysis are provided by the TiSeG software.

3.1.3 Time Series Comparison and Validation

Efforts towards regional and global comparison and validation strategies are currently evolving (Baret et al. 2006, Brown et al. 2006) and are supported by the CEOS working group on calibration and validation, and in particular by the land-product validation sub-group.

Common statistics such as the RMSE and the correlation coefficient as well as histogram analysis and scatter plots are often used to assess differences between NDVI time series from AVHRR (GIMMS), MODIS, SeaWiFS and SPOT VEGETATION (Brown et al. 2006). Attempts towards advanced techniques of time-series comparison between PAL, GIMMS, and SPOT VEGETATION have been made by Fensholt at al. (2006) using regression approaches and accounting for differences in the dynamic range. A comparison between AVHRR and MODIS NDVI time series has been performed by Huete et al. (2002). A detailed study in the conterminous U.S. has derived simple land-cover-dependent linear regressions to correct for differences between AVHRR and MODIS NDVI time series (Gallo et al. 2005). Another inter-comparison study has been conducted for LAI data from SPOT VEGETATION and MODIS (Weiss et al. 2007). The above-employed approaches are suitable for comparisons, but neither time series can be considered to be correct. In the present study three new quantitative strategies will be introduced to assess the plausibility of a time series.

With regard to validation it should be noted, though, that it is impossible to validate time series rigorously for large areas. Therefore it is common practice to compare selected observations only if extensive additional measurements are being taken during a dedicated field campaign. Basically, there are two approaches: (1) spatial validation during intensive field campaigns for selected regions and periods or (2) long-term analysis at permanent but widely scattered field sites. Examples of extensive initiatives and field campaigns, also

beyond satellite remote sensing, are, for instance, the Southern Africa Regional Science Initiative (SAFARI; Privette and Roy 2005), the Large-scale Biosphere - Atmosphere Experiment in Amazonia (LBA; Ferreira et al. 2003), BigFoot (Cohen et al. 2003), the System for Terrestrial Ecosystem Parameterization (STEP; Muchoney et al. 1999), and the Validation of Land European Remote Sensing Instruments (VALERI; Baret et al. 2007). Furthermore, CEOS Land Product Validation (LPV) has initiated several, often permanent, core sites for product validation. Permanent global networks for scattered long-term validation are, for instance, the Long Term Ecological Research (LTER), the Aerosol Robotic Network (AERONET), or the FLUXNET stations (Holben et al. 1999, Running et al. 1999b). A current CEOS initiative, BELMANIP (Benchmark Land Multi-site Analysis and Inter-comparison of Products) , specifically focuses on the current bias of validation sites in the northern mid-latitudes (Baret et al. 2006).

Extensive validation has been carried out for MODIS products following the global hierarchical observing strategy (GHOST; Morisette et al. 2002). The MODLand product validation occurs at EOS land-validation core sites and at MODLand product sites, often connected to global validation networks. Several campaigns in the framework of biophysical studies such as BigFoot, the First ISLSCP Field Experiment (FIFE), and the Boreal Ecosystem - Atmosphere Study (BOREAS) contributed to product assessment and validation (Cohen and Justice 1999). The extensive SAFARI ground campaign also allowed the validation of biophysical time series such as FPAR and LAI (Huemmrich et al. 2005). Product-specific field campaigns were conducted to validate surface reflectance (Liang et al. 2002, Vermote et al. 2002), LAI and FPAR (Privette et al. 2002), and BDRF and albedo (Liang et al. 2002, Schaaf et al. 2002).

3.2 Time Series Generation of MODIS Data: TiSeG

The Time Series Generator, abbreviated with TiSeG, analyzes the quality assessment science dataset (QA-SDS), which is provided with each MODIS dataset. The flow chart in Figure 3.2 outlines the processes of TiSeG in a generalized and simplified manner. The user-interactive software interprets the pixel-level quality information and computes two critical indices, the number of invalid pixels and the maximum gap length. The quality settings can be adapted in space and time, and the analyst weights data quality against data-quantity. The final selection of quality settings depends on the analyst. As noted in previous sections, the goal of the study, the parameter to be generated, and the subsequent approaches to time-series analysis are important considerations. Overviews on TiSeG are provided in Colditz et al. (2006a) and Conrad et al. (2005). An in-depth description is published in Colditz et al. (2007a).

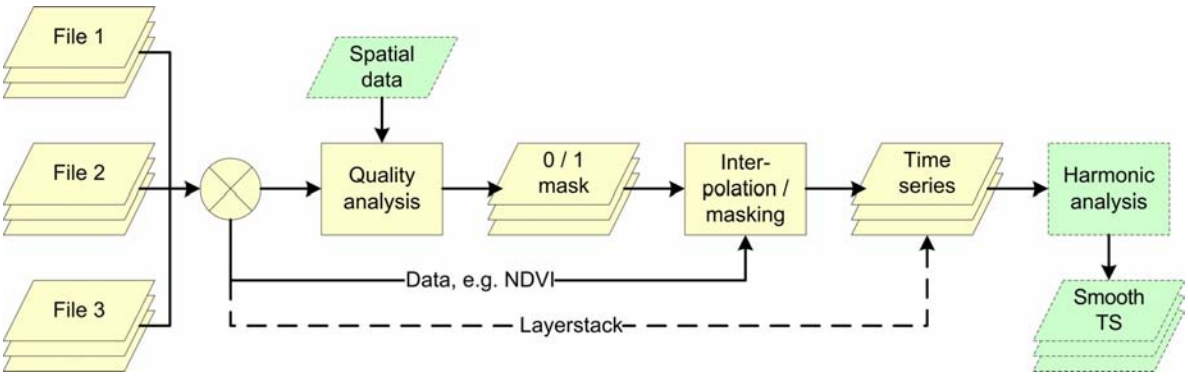


Figure 3.2: Generalized flow chart of TiSeG processes.
Note: Green boxes indicate optional data sets and processes.

The current version of the software exclusively focuses on MODLand products of the Terra and Aqua platform as well as combined products. Suitable datasets for time series have daily, 8-day, 16-day, and monthly temporal resolution (see Table A.2). The underlying principle, however, is formulated more generally, can be and partly has been applied already to ocean and atmosphere MODIS products as well as to other sensors such as MERIS, SPOT, or AVHRR.

The following sections briefly describe the QA-SDS and data availability indices and explain the functionalities of the TiSeG software package. Further sections focus on the software design and highlight the performance of computationally intensive operations.

3.2.1 The Quality Assessment Science Data Set

This section only highlights the general outline of the MODIS quality assessment procedure and considerations of the QA-SDS important for time-series generation (for details on quality settings see section 2.2.2 MODIS Land Datasets and respective tables in the appendix). Much effort has been made to ensure highest quality standards of MODIS data with regard to the spatial connectivity of individual datasets and the stability of the signal for long-term analysis. Besides a science team dedicated to quality assurance, research groups responsible for individual MODIS products have also been integrated in the quality assessment process and product validation (Morisette, et al. 2002, Roy et al. 2002a).

Roy et al. (2002a) list several sources of error during processing, which can be grouped into instrument errors, interaction with the atmosphere, and product generation. Starting from Level 1 to modeled products in Level 4, the data quality is assessed and propagated in an appropriate aggregated structure to downstream products. The LDOPE facility (LDOPE 2007) has been established by the NASA to assess the accuracy of all land products and is supported by individual science computing facilities (SCF) for each MODLand product. However, it is

only possible to apply tests to selected products, where it is attempted to investigate the data quality of approximately 10% of all daily observations (Roy et al. 2002a). Both, LDOPE and SCF ensure high data quality by visual inspection and a number of operational checks, e.g. using time series of summary statistics for globally distributed regions. Global browse images indicating erroneous data are generated every six hours and are posted in the Internet (LDOPE 2007). They are reduced in size by either computing totals or mean values of land-surface parameters and allow a rapid synoptic data evaluation. Secondly, a QA database of the LDOPE team can be used to query for quality issues in the meta-data. Results are displayed as text, graphs, scatter plots, or global images. Other assessments are made with multi-year time-series plots to assess instrument characteristics and calibration issues such as sensor degradation. Plots of specific tiles and long-term test sites are provided in the Internet (LDOPE 2007, Roy et al. 2002a).

General and product-specific quality information is provided for the user as meta-data and at the pixel level. The meta-data indicate quality issues for the entire spatial extent of the dataset and contain the most informative quality indicators at seven levels, which are important for data ordering. The per-pixel quality, the QA-SDS, can be used to assess the quality of each grid cell (Roy et al. 2002a). Its design depends on the product (see numerous tables in the appendix for products used in this study) and is described in the respective algorithm theoretical background documents (ATBD). Only the first two bits, the so-called mandatory flag with four quality indicators, are available for all MODLand products. This unique concept of product-specific pixel-level quality information provides full information and maximum flexibility, leaving the ultimate decision of sufficient data quality to the user.

The level of required data quality always depends on the subsequent analysis. Hereby, data quality is related to the level of uncertainty contained in the data and propagated to the results, which, in turn, has a high influence on their accuracy (Atkinson and Foody, 2002). With regard to time series the temporal consistency during the year and for multiple years is of utmost importance (Roy et al. 2002a). Intra-annual variations of a vegetation index profile, which cannot be attributed to actual changes at the earth surface, are quality issues and clearly limit the usefulness of the data. For example, cloud cover and many other atmospheric particles have a substantial influence on the signal and need to be either corrected or at least indicated. It should be noted at this stage that, for instance, clouds are often not visible in the actual dataset, e.g. a NDVI or a LAI. The additional QA-SDS, however, indicates cloud cover for the affected pixels. Therefore, the use of the per-pixel quality attached to each MODLand product is highly recommended for accurate data analysis (Roy et al. 2002a).

3.2.2 Data Availability Indices

Two critical indices are derived from the data cube, which indicate the reliability of the temporal interpolation of each pixel. The number of invalid pixels is the sum of all unacceptable composites in time and shows their usefulness according to the quality settings for the entire length of the time series. The maximum gap length retrieves the longest consecutive period of invalid data and is a good indicator of the reliability of a possible temporal interpolation. Considering the example from above, this index will be high if a strict setting is applied for a period with moderate quality issues, and the remaining valid data may not be sufficient for subsequent temporal interpolation.

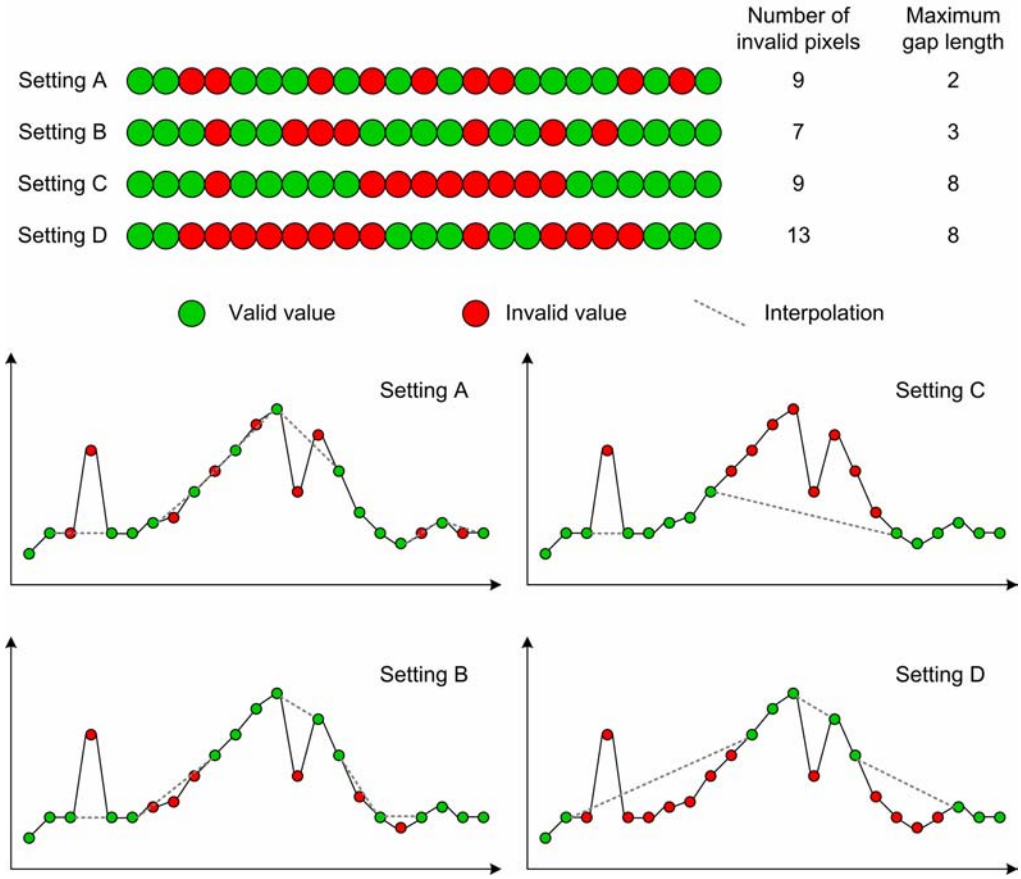


Figure 3.3: Differences in quality indices (top) and hypothetical interpolations (bottom).
Note: The upper part illustrates the differences in data availability indices number of invalid pixels and maximum gap length for four quality settings. The lower part depicts identical hypothetical uni-modal curves and applies the above settings with valid (green) and invalid (red) data. Dashed lines indicate the resulting plots if linear temporal interpolation would be used.

Figure 3.3 indicates the difference and importance of both indices with respect to the original and interpolated curve. Each example consists of 23 observations, i.e. it resembles an annual time series of 16-day composites such as vegetation index data. The lower portion of Figure 3.3 indicates the effects of linear temporal interpolation on a hypothetical uni-modal curve. Nine invalid pixels were retrieved for setting (A) but with a small maximum gap length. The

slight differences of setting (B) yielded fewer invalid pixels but a longer gap length. The interpolation of the invalid data has small effects on the resulting curve. Both outliers at the beginning and in the middle of the curve have been effectively eliminated, conserving the general characteristics of the curve. The difference between the interpolations of both curves is insignificant. Even though the number of invalid pixels of the stricter setting (C) is identical to setting (A), the longest consecutive gap with eight composites is considerably higher. Regarding the general curve characteristics, the results of the linear interpolation are dramatically different. The rising as well as the falling limb, representing green-up and senescence, are removed. and the curve characteristic is altered to a different surface type, e.g. barren. The very strict setting (D) yielded 13 invalid composites with a longest gap of eight. Although the original curve has been altered by the interpolation, the general characteristics remain with a rising and falling limb. The rising phase is extended by the interpolation, but the comparatively gentle increase may not be influenced by a robust subsequent statistical analysis. Compared with setting (C), the interpolation of setting (D) is still deemed meaningful.

3.2.3 Software Functionalities

In the following, the functionalities of the TiSeG software will be introduced in a technical fashion (Figure 3.4). In general, three blocks are distinguished: import, quality analysis, and processing. Most TiSeG applications require the use of all three steps in this sequence. Exceptions are layer stacking, where no quality analysis is needed, and harmonic analysis.

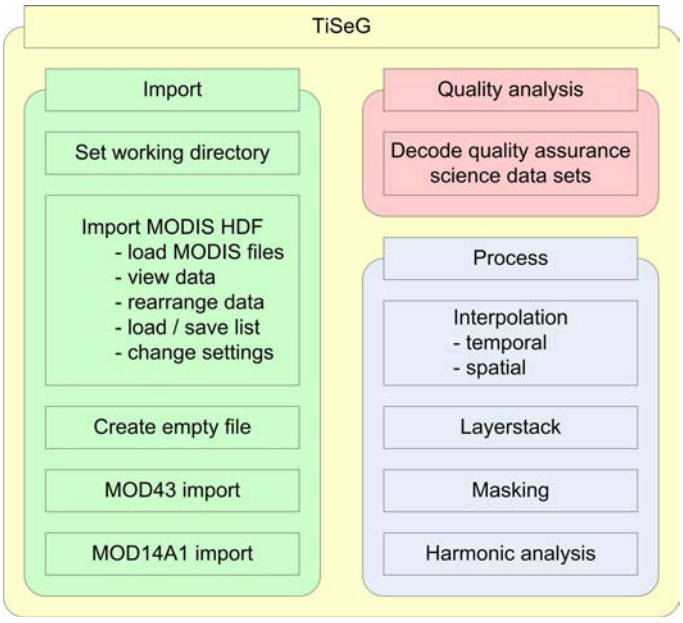


Figure 3.4: Overview of TiSeG functionalities.

Import

All inputs and outputs of TiSeG are in the hierarchical data format (HDF). The format and filename structure of MODLand files, as downloaded from the distributed active archive center (DAAC), is checked for consistency in products, tile, and version. This minimizes mistakes of the analyst and omits meaningless time series. The user is enabled to modify the consistency settings for special applications. In addition, the import function orders files temporarily, but manual temporal rearrangement is also possible. Identical file sequences to be loaded multiple times can be stored in a separate file for rapid reloading. Lastly, SDS of loaded files can be viewed for quick visual checks and interpretations.

Sometimes MODLand products are not provided for particular days or periods, for example if the system was recalibrated. In order to generate equidistant time series, a dummy file needs to be generated. The “create empty file” option opens an existing MODLand file, retrieves the fill values of all SDS from the meta-data, and makes a new file. Some MODLand products, e.g. MOD14A1 or all MOD43 products of C4, contain three- or four-dimensional SDS. TiSeG, however, only handles two-dimensional SDS internally; the more-dimensional SDS must be split into single layers. Optionally, MODIS reprojection tool (MRT) can be employed for this operation. A working directory can be defined to quickly read and write data during processing.

Quality Analysis

The quality-analysis menu is the core of TiSeG (Figure 3.5). The aim is to decode the QA-SDS and to display the data availability according to user-defined quality settings. The appearance of the graphical display is identical for all MODLand products of Table A.2. Only the product-specific quality settings depicted on the left-hand side are adapted for each product. The selected quality settings are displayed in a table at the center, where the sequentially ordered input datasets are shown in each row. As mentioned before, the analyst is enabled to modify the quality settings in time and space. Each input dataset can be analyzed with a different quality setting by assigning the user-defined settings from the left-hand side to a particular cell in the table. For instance, while strict settings are applied during seasons with generally good atmospheric conditions, quality settings during a long rainy season may require more lenient settings to keep some observations for interpolation. Optionally, the large extent of the MODIS datasets may require separating the image into spatial units with different quality settings. An additional layer with discrete units and the same extent as the time series is loaded, and the spatial units are displayed as columns in the quality table. Similar to temporally varying quality settings, the user assigns the quality settings to each

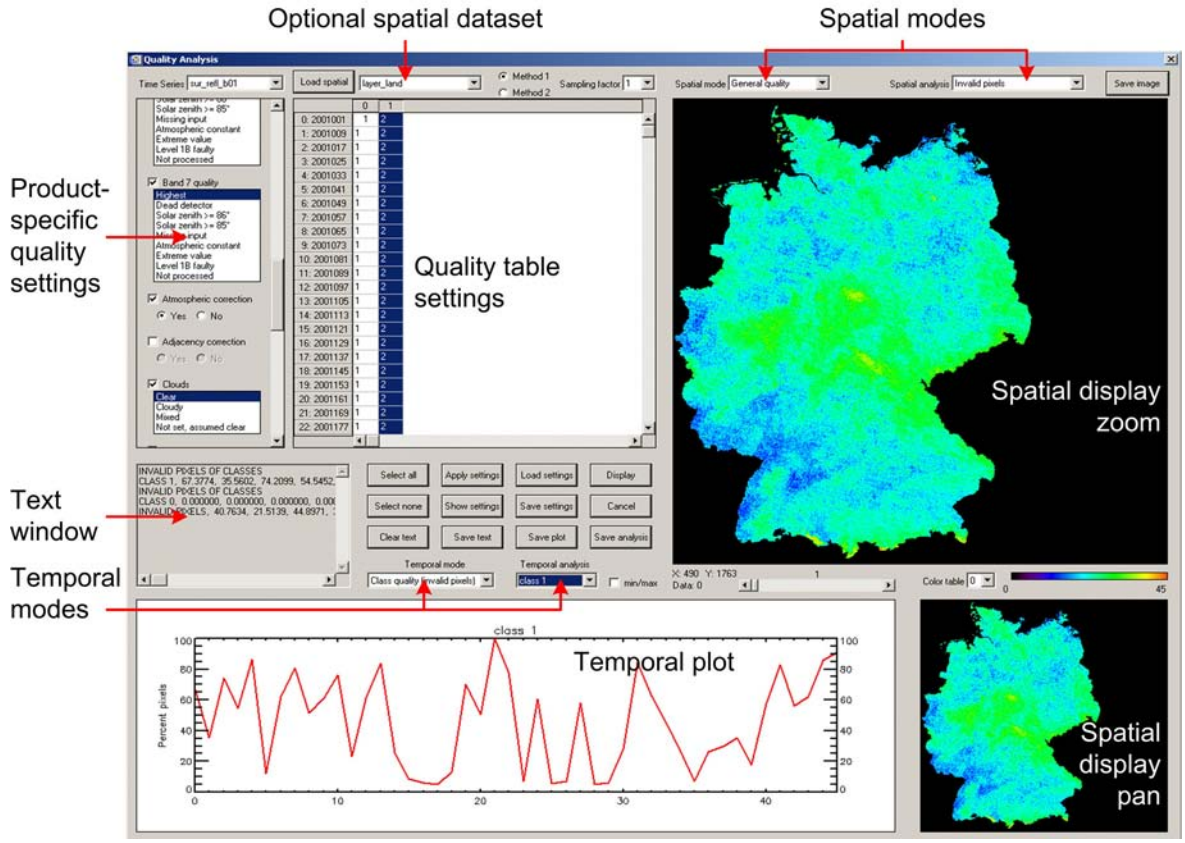


Figure 3.5: Annotated screenshot of quality-analysis graphical user interface of TiSeG.

cell. Multiple settings are displayed by different numbers, and the actual quality setting is retrieved by clicking on a particular number. Altogether the analyst has maximum flexibility and is allowed to modify quality settings temporally and spatially.

Several MODIS products contain multiple contextual SDS. The vegetation index product (MOD13), for instance, contains NDVI and EVI (enhanced vegetation index), and therefore the dataset of interest is selected from a droplist in the upper left corner. Two methods of quality analysis are available, having the same result but very different performance (see below). In order to accelerate data analysis, a sampling factor can be applied. The result of the quality analysis indicates if the particular pixel of the input dataset is valid. It is then stored in an intermediate HDF file called “data cube”, where separate SDSs indicate each input file.

Both quality indices, the number of invalid pixels and the maximum gap length, are displayed spatially and temporally (see Figure 3.5). While the temporal plots indicate the proportional data availability in percent, the spatial display depicts the quality per pixel. The window in the lower right can be used for panning, and a scroll bar for zooming. In addition to the quality indices, the data quality of each input dataset, the parameter of interest (e.g. NDVI), and the optionally loaded dataset for spatially varying quality settings can be loaded. In addition to temporal plots of the quality indices, sample interpolations for selected pixels or pixel neighborhoods or mean interpolations for spatial units can be depicted. The cumulative plot of

the maximum gap length indicates the required maximum gap to interpolate a specific proportion of the data, or *vice versa*. All values are also shown in a small text window. The text as well as the spatial and temporal displays can be exported for further analysis, visualization in other software packages, and demonstration.

Processing

The process menu generates the time series by either simple stacking or the use of the data cube, the result of the quality-analysis. Furthermore, post-processing options such as harmonic analysis are included to smooth already existing time series. Layer stacking is a fast approach and has been widely used in the past (see 3.1 Theoretical Background). The approach does not consider the quality data and thus cannot be recommended for MODIS time series of science-quality (Justice and Townshend 2002, Roy et al. 2002a).

Employing the masking function, the previous analysis of the QA-SDS is used, and invalid data are flagged with a user-defined value. This approach is useful for generating time series of temporally unsteady parameters such as maximum daily LST, where interpolation is not meaningful (see also 3.1 Theoretical Background). The use of this function also depends on subsequent algorithms and their ability to handle non-equidistant time series, i.e. data with gaps.

Similar to masking, interpolation considers the result of the quality analysis, but interpolates the invalid data with temporal or spatial approaches. Spatial interpolation approaches use distance functions such as nearest neighbor, inverse distance, and kriging. They perform solely on single SDS. Generally, temporal interpolation is preferable due to the aim of generating a time series. Temporal interpolation is solely performed at a pixel as a vector in time (see Figure 3.3). Each of the temporal interpolation approaches - linear, spline, and polynomial - requires additional parameters. The interpolation parameter “maximum gap length” can be used to omit interpolation of gaps longer than specified and saves a user-defined fill value at these locations instead. Furthermore, the parameter “good values” omits the interpolation of vectors with very few good values regardless of their location in time. It can be considered the converse to number of invalid pixels. Either the vector is filled with fill values or the original vector remains in the new time series. These parameters leave maximum flexibility to the analyst. Polynomial interpolation can be either performed using a user-defined degree of polynomials, or a best-fit polynomial retrieved with the standard-error criterion. The chosen polynomial is provided in a further dataset.

In addition, spline and polynomial interpolation can be confined to the beginning and end of the time series. The optional parameter “set first / last” places the first good observation at the

end of the time series and the last good observation at the beginning. This parameter is inherently required for linear temporal interpolation. Otherwise the curve is derived by the spline and polynomial function. Due to the lowest confidence of correct time-series generation at the beginning and end of the period, it is recommended to add a so-called shoulder dataset. These additional layers extend the time series into the past and future, and therefore shift the higher uncertainty to periods out of interest. These additional time-series layers will be deleted following interpolation.

Harmonic analysis in the TiSeG software package is used to smooth time series and provide some parameters which can be used for curve characterization. The underlying principle has been described in 3.1 Theoretical Background, and this explanation will only highlight the necessary parameters and outputs. The number of harmonics and proportion of cumulatively explained variance are two criteria to derive meaningfully smoothed time series. While the first parameter specifies the maximum number of harmonics to be used for smoothing, the latter terminates the process if the cumulative explanation of the variance of single harmonics compared to the original curve exceeds the previously specified percentage. The output comprises the smoothed time series, the number of harmonics, and explained variance needed to terminate the process. Parameters such as amplitude, phase, and explained variance of single harmonics can be stored additionally. These parameters can be used for further curve characterization and data analysis.

3.2.4 Software Design

A three-layer software architecture with a data, application, and graphical layer has been employed for TiSeG (Figure 3.6). The data layer comprises the input and output MODIS data and time series, respectively. A meta-data file system and an optional spatial dataset for quality analysis are provided. The application layer contains the functionalities to read, write, and calculate the necessary operations of the quality analysis and data processing in TiSeG. It

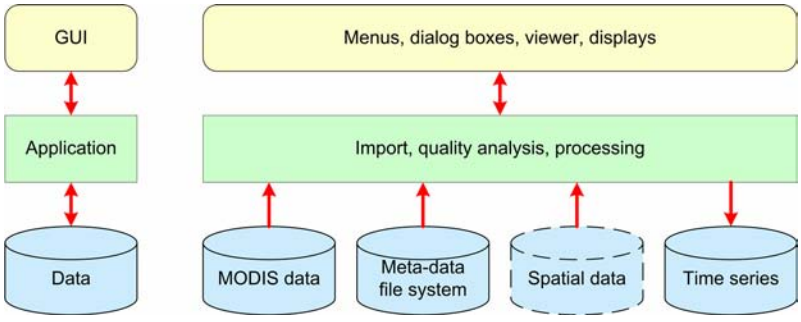


Figure 3.6: Generalization of three-layer software architecture of TiSeG.
Note: GUI…Graphical User Interface. The dashed frame indicates the optional use of the spatial data.

consists of a bundle of IDL (Interactive Data Language; IDL 2007) routines coded according to the iterative programming paradigm, which is deemed most useful for data processing. The graphical user interface, also generated in IDL, allows for interaction with the analyst. The user provides parameters, selects operations, and starts routines in the application layer which may access the data layer. Results of the operations are returned to the graphical user interface or are stored in files on the hard drive. The modular design allows quick and easy maintainability.

TiSeG requires a minimum screen size of 1200 × 1024 pixels. A processing machine with a CPU of 2 GHz, 2 GB RAM, and at least 10 GB of free disk space are recommended. Currently, Windows XP or higher is required for TiSeG, but porting to other operating systems is planned. Furthermore, the freely available IDL virtual machine (IDL 2007) is needed, which allows to distribute TiSeG in a precompiled version to a wide user community.

3.2.5 Software Performance

The computationally most expensive operation in TiSeG is quality analysis. The pixel-level quality data, stored as decimal numbers in a separate SDS, are compared to the binary semantic of user-defined quality settings. In addition, this conversion and comparison is the most important operation and must be executed for each user-defined quality setting to display data availability or save the data cube for subsequent masking or interpolation. Slow processing of this interactive operation would clearly limit the applicability and effectiveness of TiSeG. Other operations such as interpolation or harmonic analysis are also computationally expensive, but are not used interactively and are executed only once during time-series generation.

The obvious solution of the quality analysis is a numerical conversion from decimal to binary and a comparison of the binary pattern with the user-defined quality settings. Although decimal-to-binary conversion only requires standard CPU operations such as looping, division, modulo, and subtraction, the process must be executed for each pixel separately. Furthermore, the comparison of the binary pattern is computationally expensive. For instance, the user may not be interested in the snow cover bit provided in the data quality, which needs to be considered in the comparison routine.

The obvious solution, subsequently termed simple forward approach, translates the quality settings of the analyst to its binary pattern as depicted in Figure 3.7. Unspecified quality settings are indicated by an entire column of (x). Next, the decimal numbers of the QA-SDS are transformed to binary for each pixel and the binary patterns are compared for all user-

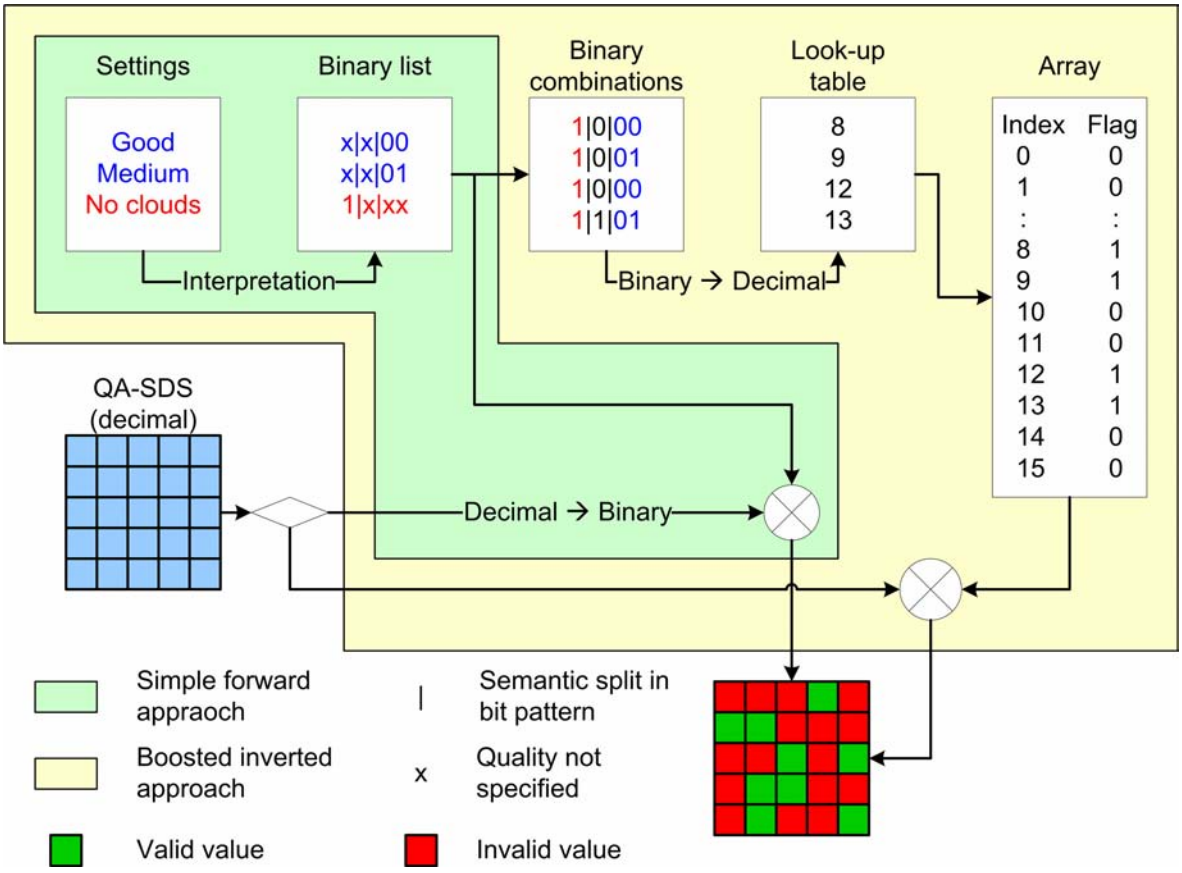


Figure 3.7: **Detailed flow chart of the simplified forward and boosted inverted approach of quality analysis, the computationally most expensive process in TiSeG.**

Note: **Three hypothetical settings are specified and decoded to their corresponding binary lists. The simple forward approach applies a decimal-to-binary conversion to all pixels of the QA-SDS (forward) and compares the elements to the specified values in the binary list (simple). The decimal-to-binary conversion and the stepwise comparison for each pixel are inefficient for large data sets. Conversely, the boosted inverted approach interprets the binary list and fills unspecified settings with all possible binary combinations (bit 2 in this example). Next all combinations are converted to decimal (inversion). A sorted list could be used for direct comparisons with the QA-SDS using binary search. However, the process is boosted by generating an array with the length of all possible decimal values, e.g. 16 in case of four-bit data. Elements at indices with a possible quality setting are set to 1, which enables a simplified and tremendously fast comparison (direct addressing) for all pixels of the QA-SDS at once.**

defined quality settings. Empirical tests for different MODIS products with more or less complicated QA-SDS have been performed on a DELL with a 3 GHz dual Intel Xeon CPU and 2 GB RAM. The computation times in seconds are shown in Table 3.1 and clearly state that the simple forward approach is impractical for an interactive analysis. Even reducing the dataset to selected samples, e.g. every fifth pixel, will not lead to sufficiently fast computations.

Therefore the problem of quality analysis has been inverted, i.e. all binary combinations are generated, converted to decimal numbers, and stored in a run-time look-up table. At this stage a sorted table could be used for comparison with the QA-SDS using binary search. However, the comparison process would be necessary for each pixel. Therefore processing is further

Table 3.1: Computation time of quality analysis using simple forward or boosted inverted algorithm (Figure 3.7) for selected MODIS products.

Product	Bits	Extent [rows x samples]	Composites	Simple forward algorithm [sec]	Boosted inverted algorithm [sec]
MOD09Q1	16	4,800 x 4,800	46	35,021.5	215.3
MOD11A2	8	1,200 x 1,200	46	1,697.0	11.9
MOD13A2	16	1,200 x 1,200	23	1,186.2	7.0
MOD13A1	16	2,400 x 2,400	23	4,759.7	24.0
MOD13Q1	16	4,800 x 4,800	23	18,797.7	114.6
MOD15A2	8	1,200 x 1,200	46	1,875.5	16.0

Note: The test was performed on tile h18v03 (Figure A.1), using setting mandatory flag "good" for one year of data. The processing machine was a DELL with a dual Intel XEON CPU, 3GHz, and 2GB RAM.

accelerated by generating an array with flags indicating whether this decimal number is a possible setting. Thereby the comparison is reduced to the simple test as to whether the array at an index of the QA-SDS decimal value is flagged or not. The process is termed boosted inverted approach and is also outlined in Figure 3.7. The direct addressing saves enormous computation time, because once the array has been generated, no further numerical operation is needed. Specific routines allow a check with the entire QA-SDS simultaneously. Furthermore the process may be easily split and performed on multiple CPUs. Exceptions for very extensive quality specifications, e.g. for MOD09 or MOD43, are also considered where multiple arrays are generated and compared. Table 3.1 lists the computation time of the boosted inverted approach. It is obvious that this fast computation is preferable and that TiSeG would not be applicable without this decisive invention.

3.3 Time Series Analysis

The following sections will focus on the analysis of time series. The methodologies of TiSeG as outlined above will be applied to MODIS products covering Germany and South Africa respectively. Vegetation indices are particularly well suited for analysis, because the depicted phenologies can also be compared visually to expected patterns of vegetation development.

Four issues will be discussed in the subsequent sections: the impact of different interpolation functions, the analysis of different quality settings, the comparison of different MODIS data collections, and the analysis of phenological variability. In the first section, different temporal interpolators of TiSeG will be exemplarily applied to two temporal profiles. The analysis shows the potential to interpolate varying lengths and types of data gaps. In the following sections, different quality settings will be employed for time-series generation of MOD09 data

of Germany. The use of data availability indices of TiSeG will be analyzed, followed by multiple qualitative assessments and a visual analysis of different time series. The comparison between C4 and the currently released C5 data of the MOD13 product will be shown for Germany in a third section. The study reveals substantial differences in MODIS versions, which have to be considered in time-series generation. The fourth section will present an applied study on temporal shifts in phenology between consecutive years. The study site of South Africa has been chosen, because vegetation development there is highly dependent on the considerable precipitation variability from year to year, typical for semi-arid environments. The annual EVI time series of the MOD13 product comprises the time from 2000 / 2001 to 2005 / 2006.

3.3.1 Analysis of Temporal Interpolation Functions

This section will demonstrate the performance of temporal interpolation algorithms of TiSeG for selected profiles. In this respect, this section illustrates already discussed issues of the hypothetical example using actual data (see Figure 3.3 and 3.2.2 Data Availability Indices). A first section will briefly describe the dataset generation followed by a detailed analysis of different settings. A final section draws first conclusions and justifies the use of the linear temporal interpolation technique, which has been used throughout this study

Data generation

Two temporal NDVI plots, derived from an 8-day MO09A1 time series of Germany, have been extracted. The curves exhibit a uni- to a slight bi-modal characteristic, which is typical for agricultural land use in Germany. Both profiles show an increase during springtime and a marked peak in summer. The rapid senescence is followed by a steady increase in late fall and during winter for the second profile, most likely indicating the green-up of winter crops. The high dynamic range, for the second profile even during a short period, is a particular challenge for temporal interpolation algorithms. Many other land-cover types indicate a moderate temporal dynamic; hence the differences between the interpolation algorithms would be less clear.

Several data gaps of varying length and at multiple positions in the temporal course have been artificially created; hence, they are not related to the data quality of MODIS. Next, interpolation functions of TiSeG have been employed to fill gaps. Besides linear interpolation, a spline function has been applied in two modes to illustrate the difference of confining the function at the beginning and end of the time series (“set first / last” see 3.2.3 Software

Functionalities). Three polynomial functions have been applied in the unconfined mode but with different polynomial degrees. The best-fitting polynomial employed the standard error measure for obtaining the optimal degree (3.2.3 Software Functionalities).

Analysis

Figure 3.8 illustrates both profiles and the results of the interpolations for three different sets of invalid data. The position of invalid observations is shown in red at the bottom of each graph. With regard to the data availability measures as such, the number of invalid pixels (maximum gap length) is 27 (9) for Setting 1, 28 (5) for Setting 2, and 25 (7) for Setting 3. Although the difference among all settings is not very high, the impact of the interpolation on the resulting time series also depends on the local temporal dynamics. Setting 1, for instance, mimics a long data gap during the summer, i.e. during the peak of phenology. It is indicated that none of the methods is capable to represent the peak of Profile 1 adequately. The best-fitting polynomial achieves the best-fitting results. Spline, for instance, does not show satisfactory results, because the temporal course of the available neighboring values before and after the data gap does not indicate high slopes. However, the ranking of most suitable interpolation techniques reverses for Profile 2. Here the polynomial with 15 degrees fits best, followed by spline and linear interpolation.

The picture is rather different for Setting 2, where all functions except for the third-degree polynomial sufficiently represent the expected temporal course of both profiles. However, there are marked discrepancies at the beginning and the end, due to missing data at the margins of the time series. Shoulder datasets, which extend time series into previous and following periods and therefore mitigate the uncertainty, have not been applied. The impact of confined and unconfined methodologies is clearly shown. The “set first / last” option places the first good value at the end, and the last good value at the beginning. Although fitting the time series to these values is suitable for many applications, in particular for producing annual time series, the error to be introduced can also be higher as exemplarily shown for the end of Profile 2. Nevertheless, the conservative use of the “set first / last” option seems appropriate in most cases, especially if shoulder data are not available.

Setting 3 depicts marked differences among the interpolators for the senescence of Profile 1. Since the slight drop during the peak period is regarded as valid, no temporal interpolation function is capable to represent the extended peak due to missing vertexes. Furthermore, it is usually expected that short data gaps can be easily interpolated with temporal functions, an assumption which is true for most cases. However, even a short gap can be difficult to

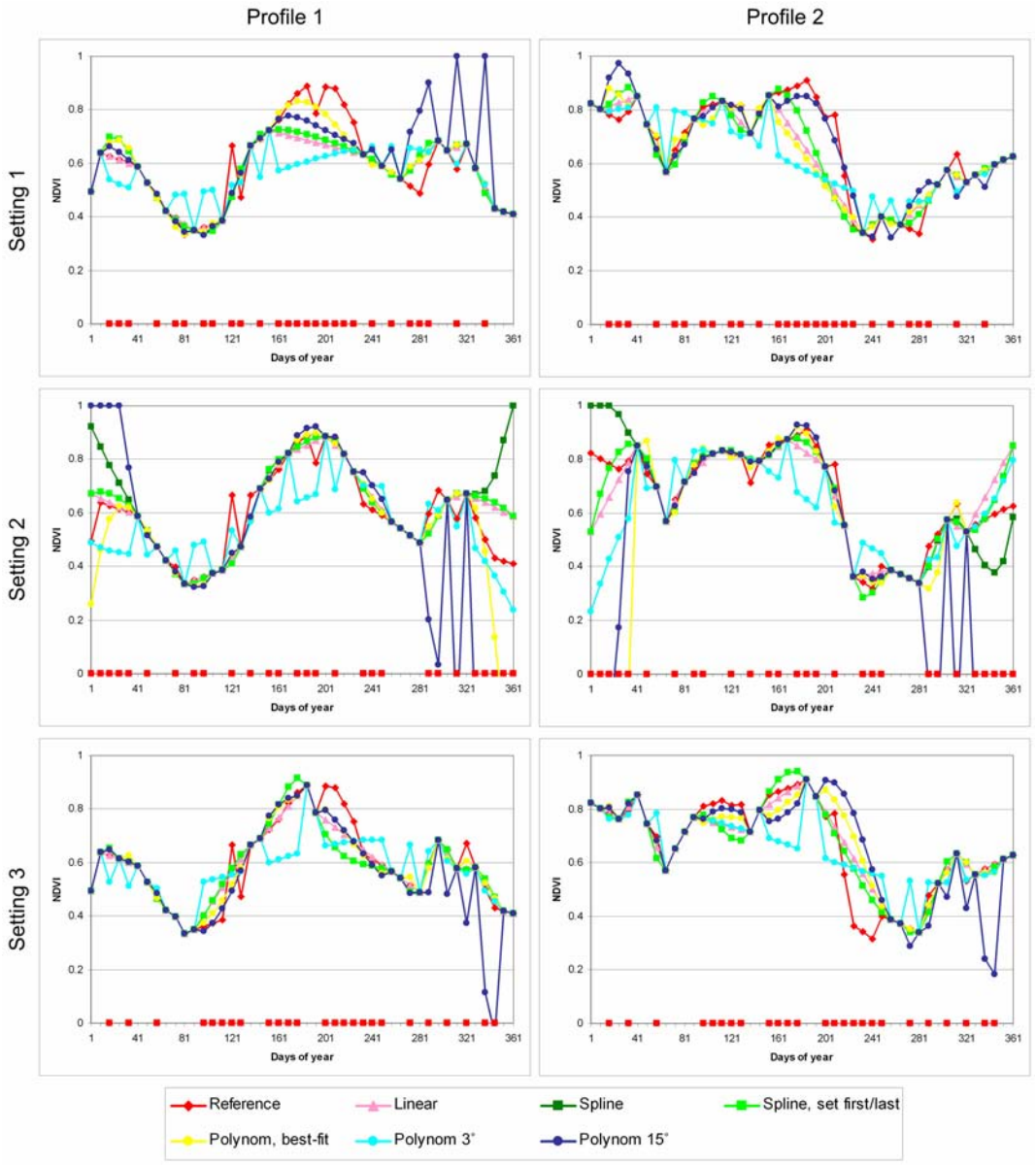


Figure 3.8: **Examples for the interpolation of data gaps with temporal interpolation algorithms implemented in TiSeG.**

Note: **The data gaps of different settings are indicated at the bottom of each graph. The significant discrepancies of some algorithms at the beginning and the end of Setting 2 result from the unconstrained interpolations (not using the "set first / last" option). Both spline methods yield identical results for Setting 1 and Setting 3.**

interpolate, as exemplarily shown for the peak in phenology of Profile 2. The interpolation functions draw rather different patterns, where spline and linear are most meaningful.

Comparing the course of temporal interpolation functions with regard to the reference, the specification of polynomial degrees clearly appears inappropriate. The use of three degrees cannot accommodate for the high temporal dynamics during a short period and 15 degree polynomials introduce noise patterns. However, the best-fit polynomial yields a good approximation of the expected pattern. There is no significant difference between linear and spline.

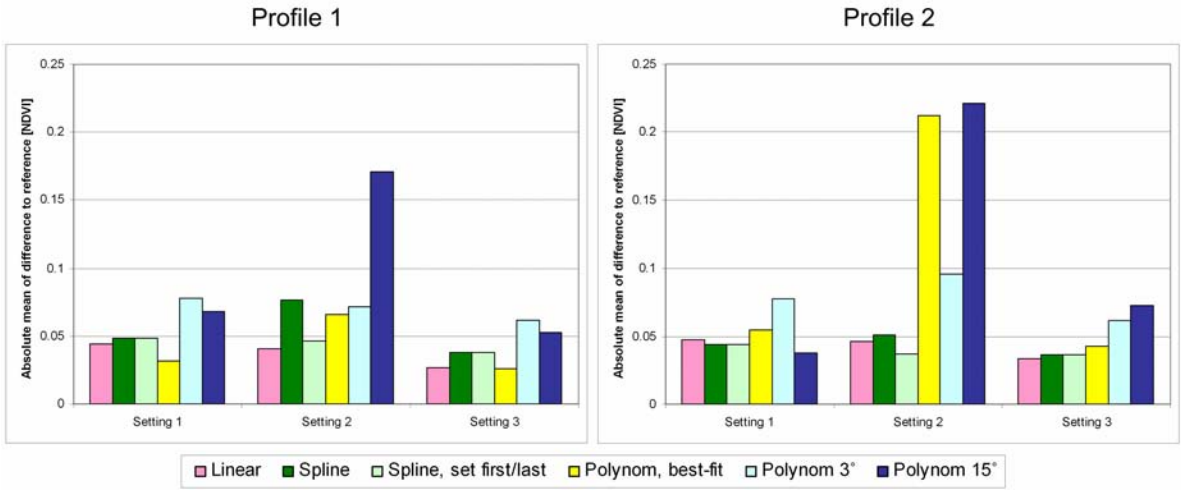


Figure 3.9: Absolute mean of difference to the reference graph for temporal interpolation algorithms of TiSeG.
Note: For specifics of the temporal course see Figure 3.8. The markedly higher values for some algorithms of Setting 2 result from the unconstrained interpolations (not using the "set first / last" option) at the beginning and the end of the time series.

A computation of the mean absolute difference to the reference curve is shown in Figure 3.9. It should be considered that differences to the reference have to be expected, since the interpolators are meant to fill gaps of invalid data, e.g. the peak and drop during the green-up of Profile 1, which have to result in some discrepancy. Marked differences are shown for unconfined techniques of Setting 2, i.e. the uncertainty at the beginning and the end of the time series. Also, the unconstrained spline method shows differing results. For the other settings, third and 15 degree polynomials mostly indicate highest differences. The discrepancy between all other techniques, i.e. linear, spline, and bets-fit polynomial is low.

Discussion and Implications

This first assessment of interpolation functions has indicated significant differences among temporal interpolation techniques implemented in TiSeG. It has illustrated that data availability indicators are critical, providing as a first picture of the feasibility of interpolation. However, it is highly recommended that the analyst applies example interpolations to determine the contextual feasibility of interpolation, i.e. the correct characterization of the temporal course. If no interpolation seems to appropriately represent the expected temporal pattern, the quality settings of the respective season can be modified. Of course, this can incorporate low-quality observations into the time series, but could also lead to more meaningful temporal plots.

It is generally recommended to employ the constrained methodology, i.e. the use of the "set first / last" option, also for algorithms which could be applied in the unconstrained mode

(spline and polynomials). The option must be employed for linear interpolation. Obviously, the use of shoulder datasets, extending the time series into the previous and following period will shift most of the uncertainty to periods not of current interest. Therefore, the use of shoulder data is recommended whenever possible.

With regard to the most appropriate algorithms, there is no definite result. The visual inspection of the temporal course as well as the difference analysis did not yield a clear ranking. Throughout this study, linear temporal interpolation has been employed due to several reasons. Clearly, it is the fastest algorithm to be applied. Another advantage is its inherent bounding of interpolated data between the highest and lowest values of vertexes. Due to the solely statistical manner of other interpolation algorithms, it is possible to create values beyond the allowed data range. TiSeG automatically rejects these values by replacement with the maximum or minimum. This has been shown for the defined polynomial functions, but such cases can also occur if using best-fit polynomials and spline. Linear interpolation, in contrast, only yields a connection between the highest and lowest vertex from the original dataset. On the other hand, the interpolation of longer gaps will yield data on a straight line linking two reference values, which is usually considered an unlikely characterization of the reality. However, because real data are missing, a curve is not necessarily the better estimator.

3.3.2 Comparison between Improved and Original Time Series

This section will demonstrate the suitability of TiSeG for time-series generation. The MOD09 product has been employed for calculating the vegetation index NDVI for displaying the state and development of vegetation activity in Germany. The main goal of this section, however, is the assessment of data quality settings with respect to data availability indices of TiSeG. The effects on the resulting time series will be evaluated by quantitative measures and visual inspection. The first section will describe the time-series generation process. The data availability indices will be evaluated in space and time, followed by a detailed temporal analysis of selected regions. In a final section, the resulting time series will be statistically assessed.

MOD09 Time Series Generation

A time series of 8-day 500m surface-reflectance data (MOD09A1; Vermote et al. 2002; 2.2.2 MODIS Land Datasets) of collection 4 was generated for 2001. Germany is covered by tiles h18v03 and h18v04 (Figure A.1), which were mosaiced, reprojected to UTM zone 32, and subset to the country extent using MRT (Table A.18). Ten shoulder datasets were used for

time-series generation (3.2.3 Software Functionalities). Furthermore, missing day 169 was filled with a dummy file to ensure temporally equidistant time-series data.

The products were imported to TiSeG and quality analysis was applied. Besides the mandatory flag, the MOD09 QA-SDS contains a multitude of band-specific quality indicators (Table A.4). An additional SDS provides quality indicators of atmospheric and surface characteristics (Table A.5). In addition to a simple layer stack, five time series with consideration of the quality flags were generated for all surface-reflectance bands. Details of the quality settings can be obtained from Table 3.2. All invalid data were interpolated using linear temporal interpolation. Next, the NDVI was calculated from the VIS red and NIR band.

Table 3.2: Quality settings of MOD09A1 (8-day, 500m surface reflectance) time series of Germany for 2001.

Name	Settings for band 1 to 7	Athmospheric correction	Clouds	Cloud shadow	Aerosol	Cirrus	Cloud mask	Snow / ice	Snow mask
H7	Highest								
H7-CS-S	Highest	Yes	Clear	No		None	Clear	No	No
H7-C-S	Highest	Yes	Clear			None / small		No	No
H7-CS	Highest	Yes	Clear	No		None	Clear		
H7-CS-S-A	Highest	Yes	Clear	No	Low	None	Clear	No	No

Note: Only selected settings are shown in this table. For a complete list of QA-SDS specifications see Table A.4 and Table A.5.

While the lenient setting H7 only excludes pixels with substantial instrument issues, such as striping, bad detectors, or difficulty in Level 1B generation, all other settings also consider additional quality flags indicating cloud cover, shadow effects, snow on the ground, or successful atmospheric correction. Moderate setting H7-C-S excludes composites of cloud cover and snow cover. Strict setting H7-CS-S also excludes shadow effects and evaluates an additional cloud mask of the MOD09 product. Identical quality specifications were used for setting H7-CS, but snow cover was not regarded as invalid. Setting H7-CS-S-A excludes pixels with other than low aerosol content. Aerosol is a critical parameter for accurate atmospheric correction, but highly variable and difficult to retrieve. It should be noted that the aerosol content has been retrieved inconsistently in time, thus sometimes climatology information is used instead. However, the study area was not affected in 2001.

Spatial and Temporal Analysis of the Data Quality for Germany

The number of invalid pixels (see also 3.2.3 Software Functionalities and Figure 3.3) in the temporal plot of Figure 3.10 (left) indicates substantial differences in data availability for the selected quality settings (see Table 3.2). With regard to seasonality, late spring, summer, and

early fall reveal a higher data presence than for the rest of the year. However, even during the summer season an 8-day compositing period is often not sufficiently long to eliminate cloud cover and other quality issues. Day 169 was not available from the DAAC and is therefore indicated by 100 % invalid data. It is important to note that 16-day composites, e.g. of MOD13 vegetation index data, are deemed sufficiently long for cloud-free observation during the summer, as shown by Colditz et al. (2007a). During the winter period low data availability usually results from a combination of cloud cover and snow.

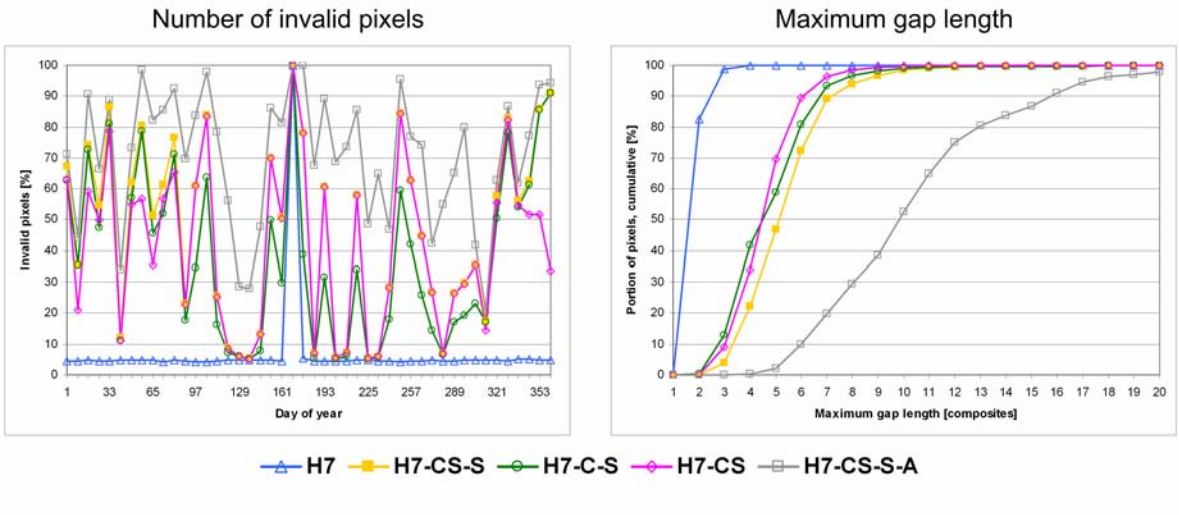


Figure 3.10: Temporal quality analysis of Germany.
Note: For specifics of the quality settings see Table 3.2.

With regard to the quality settings, lenient setting H7 only excludes few pixels with the above-mentioned quality limitations and does not show a temporal characteristic. During the winter period moderate setting H7-C-S and strict setting H7-CS-S follow a similar trend, with a 5 to 10 % absolute difference in data availability. During the summer period the difference increases to approximately 25 %. The setting H7-CS does not regard snow cover as invalid and shows a clearly different pattern compared to H7-CS-S and H7-C-S during the winter, e.g. for days 9, 17, 57, 65, and 345 to 361. For the rest of the year H7-CS is identical to strict setting H7-CS-S, due to similar cloud settings. Data availability between the setting H7-CS and moderate setting H7-C-S, having trends similar to strict setting H7-CS-S during wintertime, reverse at days 89 and 337, indicating the beginning and end of the summer period, respectively. Aerosol setting H7-CS-S-A does not show any particular seasonal pattern and often identifies remarkably more pixels as invalid than all other settings, especially during the summer.

The right-hand side of Figure 3.10 shows plots of the maximum gap length (see also 3.2.3 Software Functionalities and Figure 3.3). This graph starts with one gap for all settings due to the dummy dataset for day 169. The cumulative presentation of the maximum gap length indicates the percentage of data that can be interpolated at a specific data gap, and *vice versa*.

Choosing lenient setting H7, 80 % of all invalid data are interpolated at a longest gap of two and 99 % at a gap of three. On the other hand, 90 % of all invalid pixels derived with strict setting H7-CS-S require an interpolation of up to seven consecutive composites. This corresponds to 56 days or almost two months without data. The 90 % level of aerosol setting H7-CS-S-A is reached by a gap length of 16 composites, or more than four months without data. In comparison to setting H7-CS-S, H7-CS requires the interpolation of one composite less at longer data gaps and indicates temporary snow cover during the winter period. With regard to the gap length, plots of moderate setting H7-C-S and snow setting H7-CS reverse, where the moderate setting interpolates less data quantities at longer data gaps.

Figure 3.11 depicts the number of invalid pixels and maximum gap length for the quality settings of Table 3.2 spatially. Lenient setting H7 does not show a particular spatial pattern. Moderate and strict settings H7-C-S and H7-CS-S depict the influence of upland areas in the center and the Alps. The likelihood of cloud cover and snow increases with elevation. Setting H7-CS, with the same cloud specification as the strictest setting H7-CS-S, highlights that northern Germany is more frequently covered by clouds than the southern part. Compared with the moderate setting H7-C-S, the influence of the additional cloud mask of the MOD09 team is shown spatially for northern Germany. Furthermore, setting H7-CS did not account for the snow cover in southern Bavaria. Aerosol setting H7-CS-S-A indicates a high number of invalid pixels for the entire country. The gap length is also remarkably higher than compared to all other settings, but except for an area north of Berlin, northern Germany seems to have longer gaps than the South. Considering the number of invalid pixels, setting H7-CS with a large extent of low-quality observations, seems stricter than moderate setting H7-C-S. The maximum gap images reverse this first impression and indicate that especially mountainous areas in central Germany cause considerable difficulties due to additional snow cover during winter.

Figure 3.12 shows a comparison of selected days between the original datasets without quality analysis (layer stack) and an improved time series of the strict quality setting H7-CS-S. The lower row depicts the absolute difference between both datasets. Issues such as NDVI saturation, susceptibility to atmospheric influences, and the interpolation of snowy pixels are discussed below. On day 33 substantial areas throughout Germany were obscured by cloud cover, and upland areas were covered by snow as indicated in the layer-stack. The strict settings of the quality-analyzed image interpolated the atmospheric interference and the snow cover. According to the temporal quality analysis (Figure 3.10), 86 % of Germany was regarded invalid for day 33. The remaining areas with lower NDVI indicate urban centers such as Berlin, Hamburg, Munich, the Ruhr district in North-Rhine Westphalia, the greater Frankfurt industrial district, open strip mines in southern Brandenburg, and some barren fields along the Danube in Bavaria. Similar spatial patterns were also found by Colditz et al. (2007a)

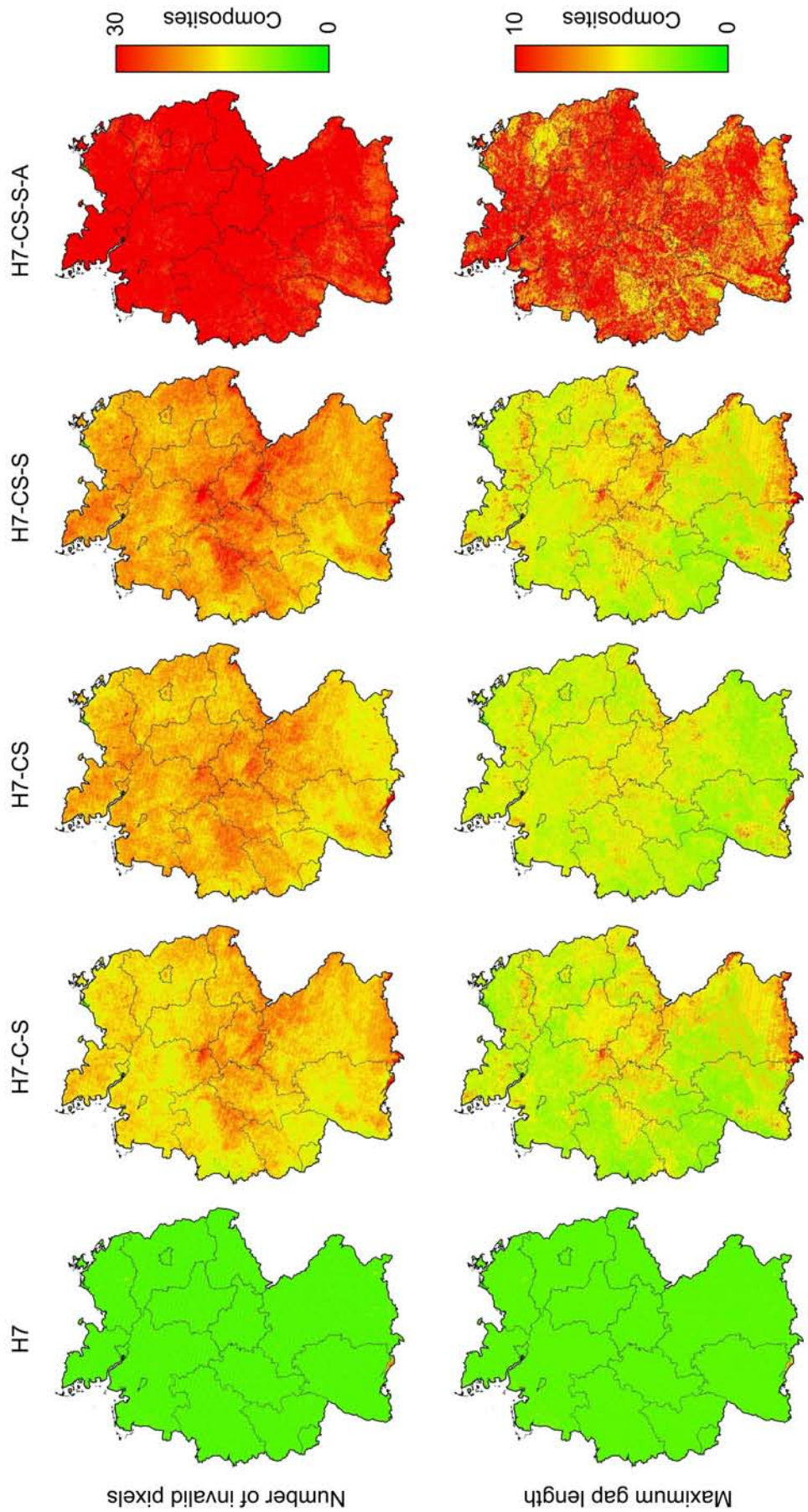


Figure 3.11: **Spatial quality analysis of Germany.**
Note: **Upper row: number of invalid pixels with values greater 30 displayed in red. Lower row: maximum gap length with values greater 10 displayed in red. For specifics of the quality settings see Table 3.2.**

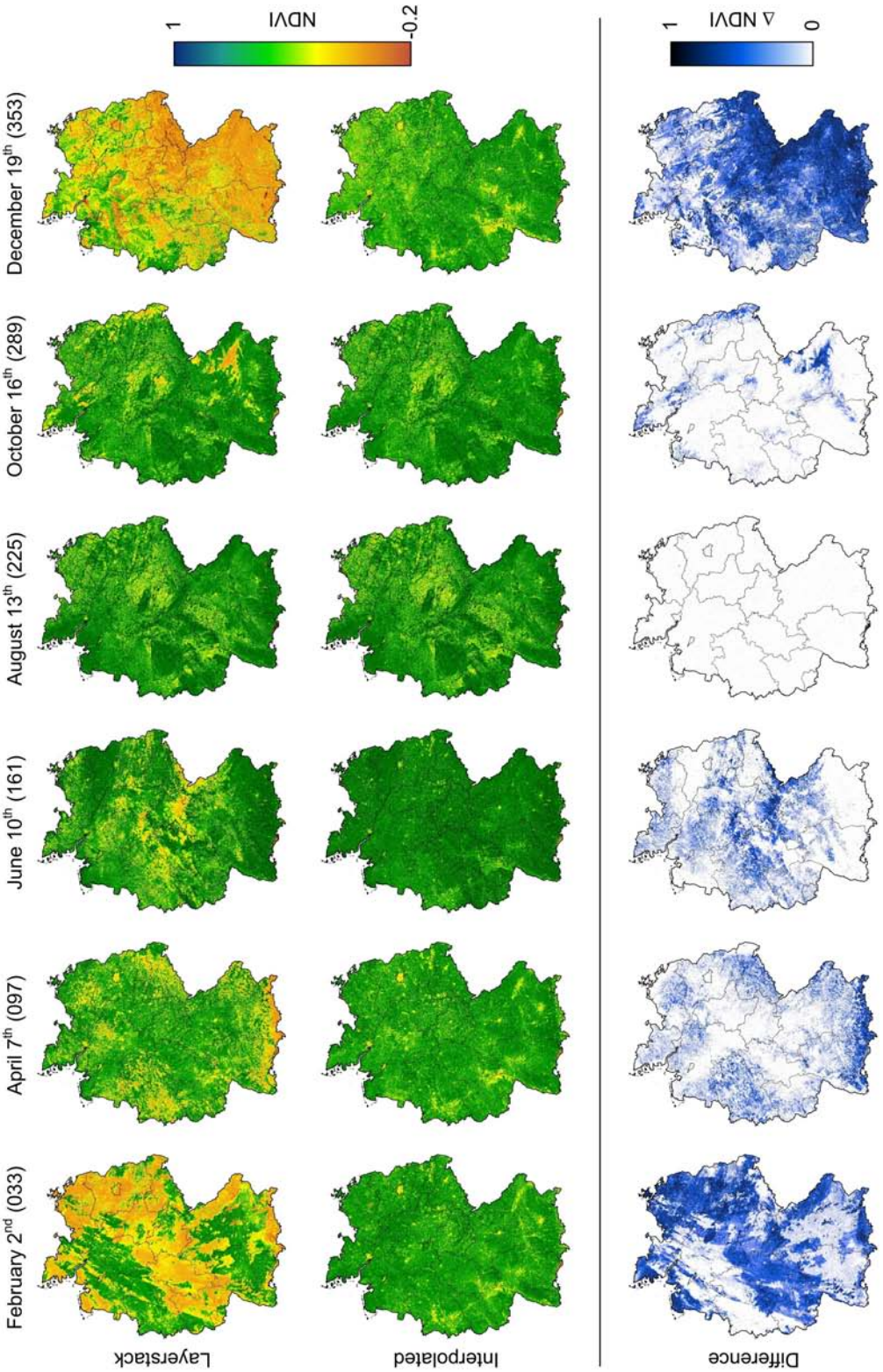


Figure 3.12: **Time series of Germany with selected composites of 2001.**
Note: **Upper row: layer stack without quality analysis. Middle row: linearly temporally interpolated time series with setting H7-CS-S (Table 3.2). Bottom row: absolute difference in NDVI between layer stack and interpolated time series. The color stretch of the NDVI aims at distinguishing the difference between layer stack and interpolation and therefore cannot represent the dynamics of the interpolation itself.**

using the EVI. The layer-stack of day 97 shows the expected response of NDVI for large parts of Germany where green-up usually starts in March. Areas with low NDVI in eastern, western, and southern Germany were obscured by clouds. Similarly, the spots of low NDVI in central Germany were covered by clouds on day 161. The interpolated image of June also shows the green-up in urban areas. Furthermore, a very high NDVI in June indicates the plateau-phase of the agricultural growing season (see below). In comparison, harvesting of large fields started in August, indicated by a lower NDVI in eastern Germany. The layer stack and interpolated image of day 225 are almost identical, with 5% invalid data due to MODIS detector failures. The October image shows prominent differences in southern Bavaria due to cloud cover. Clouds and snow caused dramatic differences between the layer stack and interpolation in late December.

Temporal Analysis of the Data Quality for Selected Regions

In order to account for regional differences in the temporal information, selected land-cover units of natural regions of Germany were analyzed. An average temporal plot of the quality settings for entire Germany would not show a meaningful result, because temporal curves vary considerably between regions and land-cover types. Selected natural regions of Germany (Meynen and Schmithüsen 1953) were employed for temporal analysis, spanning from the lowlands in northern Germany over prominent uplands in the middle to the Alpine Foreland in the southern part (Figure A.2). In addition to regional analysis, land-cover types according to the CORINE classification were considered (Keil et al. 2005, see also 2.3 Study Sites and Land Cover Datasets). Four dominant land-cover types were analyzed: arable land, pasture, broadleaf, and coniferous forest. Relative and absolute area-statistics of selected land-cover types are provided in Table A.22.

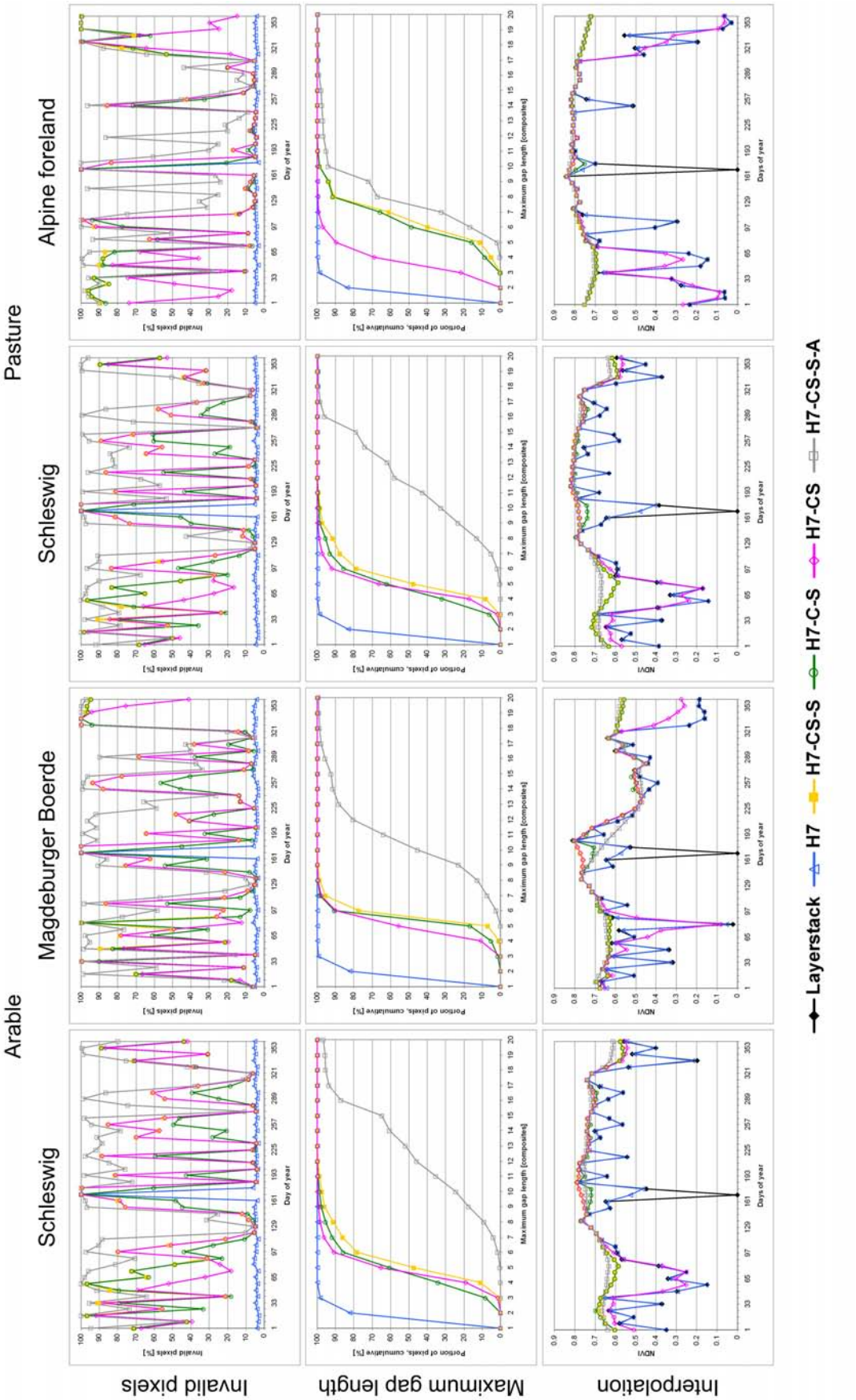
Figure 3.13 shows plots of the number of invalid pixels, maximum gap, and average interpolation for the quality settings of Table 3.2. Two natural regions were selected for each dominant land-cover type. The general characteristic of the number of invalid pixels and maximum gap length regarding the ranking of quality settings is similar to plots for Germany (Figure 3.10). The number of invalid pixels increases from lenient setting H7 to moderate setting H7-C-S and strict setting H7-CS-S. Snow setting H7-CS is identical to the strict setting for the summer months, but retains snowy pixels during the winter period. Interestingly, the beginning and end of snow cover seem to be very consistent for all selected regions. Except for the Harz and the Alpine Foreland, the maximum gap follows already described patterns. For both exceptions H7-CS does not traverse H7-C-S. The aerosol setting H7-CS-S-A indicates a north-to-south gradient. While aerosol content is high in the northern

regions causing a long gap length, plots of the southern sites indicate fewer aerosol particles with a shorter gap of consecutively missing observations.

Land-cover class “arable” was analyzed for two regions in northern Germany. While the Magdeburger Boerde is the most intensively used agricultural area in Germany with fertile loess soils and very large fields, Schleswig is characterized by fertile to medium-fertile ground-moraine soils and much smaller fields. Differences in the invalid pixels and data gap are related to cloud cover, with less interference in the Magdeburger Boerde due to the rain-shadow effect of the Harz Mountains. Both plots show a high aerosol content for the whole year, which manifests in a very long maximum gap length. The interpolated average NDVI of Schleswig merely describes a pasture profile with a long growing season from 81 to day 321. NDVI profiles of the Magdeburger Boerde indicate the expected pattern with a gentle green-up, almost no plateau phase, and a prominent harvesting period starting in July. Choosing setting H7-CS-S-A, both regions reveal the effect of too small a number of pixels for interpolation. In Schleswig, the dormancy and plateau phase are overestimated and underestimated, respectively. Even more obvious, the peak in phenology and the rapid senescence phase of arable areas in the Magdeburger Boerde is not correctly represented in setting H7-CS-S-A, because no valid pixels were mapped for the peak growing period. This corresponds with described effects as illustrated in Figure 3.3 and has been previously shown by Colditz et al. (2006a) for savanna and evergreen forests in western Africa. The number of available pixels is insufficient, which indicates that strictest settings with only highest quality data may not always be the best selection for interpolation and representation of an accurate temporal profile.

The comparison of pasture lands between northern and southern Germany depicts substantial differences. While the rather continental Alpine Foreland reveals very good atmospheric conditions during the summer, interrupted by two prominent cloud periods and brief aerosol inferences, maritime Schleswig in the north is characterized by changing weather conditions. Schleswig shows a characteristic pasture profile with a gentle green-up, a long plateau phase, and a rapid senescence period. The average NDVI of pasture in the Alpine Foreland shows almost no phenology. Rains during the spring and fall and afternoon thunderstorms in the summer are one reason. Furthermore, grasslands in the Alpine Foreland are usually not plowed and exclusively used for cattle grazing, which is the major source of income for farmers. Agriculture and crop rotation is not common in this region. The third reason is the interpolation of persistent snow during wintertime.

The potential natural vegetation type, broadleaf forest, is not very prominent in the highly managed landscape of Germany. Broadleaved forests dominate the lower portion of the Harz Mountains along the southeastern rim, at a similar elevation as in the Taunus Mountains. Most interesting is the temporal NDVI profile, which is very similar for both regions.


Arable
Pasture
Schleswig
Magdeburger Boerde
Schleswig
Alpine foreland
Invalid pixels
Maximum gap length
Interpolation
Layerstack
H7
H7-CS-S
H7-C-S
H7-CS
H7-CS-S-A

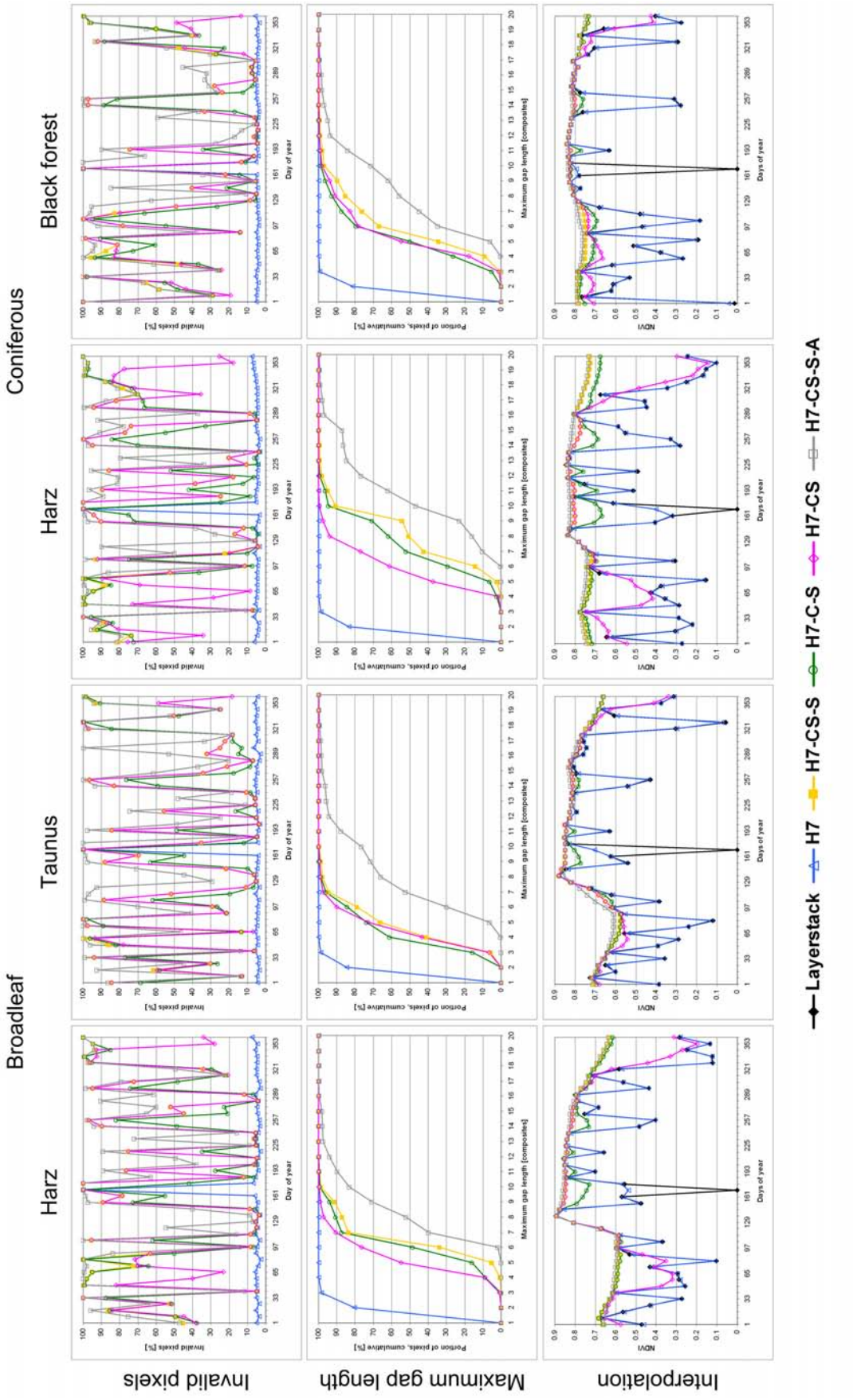


Figure 3.13: Temporal quality analysis and average interpolation of selected natural regions and CORINE land cover classes (Figure A.2).
Note: Upper row: number of invalid pixels. Middle row: maximum gap length. Bottom row: average of linear temporal interpolation. For specifics of the quality settings see Table 3.2.

Following a long phase of dormancy with no leaves, rapid green-up starts by the end of March (day 89) in the Taunus, and an even steeper increase in NDVI by the end of April (day 113) in the Harz. A long plateau phase with average NDVIs between 0.8 and 0.9 lasts until mid-October (day 289), while the NDVI constantly decreases during the summer, possibly related to drought. The gentle senescence indicates defoliation in fall. A second effect regarding quality settings is noticeable in particular for broadleaf as well as needle-leaf classes in the Harz. For instance, between days 145 and 185 or 241 and 281, setting H7-C-S yields dramatically lower values, due to the ignored cloud shadow flag. The H7-C-S NDVI profile does not describe the expected phenological pattern and indicates the importance of accurate cloud masking, including shadow effects.

Coniferous stands dominate several uplands of Germany. The Harz is the most prominent upland in northern Germany, densely covered by coniferous forests, in particular at its steep western side. Coniferous trees also dominate the steep western slopes of the Black Forest in southwestern Germany. Despite the higher elevation of the Black Forest, quality setting H7-CS (not excluding snow pixels) has a higher influence on the maximum gap length and the resulting curves in the Harz, e.g. from days 49 to 81. In other words, in 2001 snow played a more important role in the Harz than in the Black Forest. Furthermore, the more continental Black Forest is indicated by more available pixels during the summer, when the Harz is often obscured by clouds. Both regions indicate slight differences of the aerosol setting H7-CS-S-A. However, the very rigorous setting seems to improve the results, eliminating minor variations. Although the forests are dominated by needle leaf trees such as pine, spruce, and some firs, the mean NDVI profiles indicate some phenological development due to neighborhood effects and green-up of the ground layer in sparser stands. Furthermore, sprouting needles contribute to the slight increase in NDVI in springtime, which gradually decreases during the summer months.

A few issues should be addressed to put the NDVI plots into the right perspective. As already shown in the interpolations of Figure 3.12, the NDVI is noticeably high for large parts of Germany. The interpolation average for the entire country is higher than 0.6, even for the winter composites when snow covered pixels are interpolated. The NDVI is a sensitive vegetation index, i.e. even a low green vegetation activity affects the signal. On the other hand, the signal saturates quickly for high biomass regions (Huete et al. 2002). Saturation does not seem to occur for large areas, but regional average NDVIs above 0.9 indicate a local saturation during the summer. Other vegetation indices such as EVI are specifically designed for high biomass regions. A 2001 EVI time series of Germany has been employed by Colditz et al. (2007a), showing similar temporal plots, but lower absolute values and a higher dynamic range. Another reason for small low NDVI areas in Germany is the patchiness of the

landscape. There are few homogeneous areas dominated by a distinct land-cover or surface type. The usually mixed signal almost always contains some vegetation response.

Another issue is the interpolation of snowy surfaces. It may be argued whether a snow cover can be interpolated or not, and the ultimate decision always depends on the goal of the study. The period of snow cover on the ground is important for calculating annual gross primary productivity in physical units or computing gaseous exchanges between soil and atmosphere. On the other hand, except for Polar Regions, snow cover is an ephemeral phenomenon. It can considerably vary in space and time and prohibits a comprehensive and comparative representation of annual phenologies. Thus, land-cover mapping with annual time series requires the removal and interpolation of these dramatic episodic effects.

Quality Assessment of Time Series using Statistical Methods

As already noted in 3.1 Theoretical Background, it is impossible to validate time series for large areas. Point measurements for specific observations or during field campaigns can be used to calibrate datasets and to estimate their accuracy. Furthermore, several statistics are used to compare different time series. Steady time series can also be assessed with regard to their inherent consistency.

Three approaches of time-series consistency checks are used: (1) the average standard deviation of regional time series for particular land-cover classes, (2) total differences between the initial and locally boxcar-filtered time series, and (3) temporal autocorrelation. It should be noted that all of the approaches are only useful for steady time series.

The temporal mean of all standard deviations σ was computed for natural regions and land-cover classes following Equation 3.1:

$$\overline{\sigma(R,LC)} = \frac{\sum_{i=1}^{N} \sigma(R,LC)_i}{N}, \qquad \text{Equation 3.1}$$

R...region
LC...land cover type
N...length of time series

The standard deviation is calculated for all composites. Next, the mean of all temporal standard deviations is computed. The underlying idea of this assessment is that time series of the same land-cover type of a small region should have a low variability. This approach

highlights the spatial correspondence of time series for homogeneous areas while considering geographical differences.

The bar-plot (Figure 3.14) indicates a very consistent ranking among time series. The data analysis focuses on the relative differences among quality settings for one land-cover type and region. Layer stack with no quality analysis and lenient setting H7 feature the highest variability. Snow setting H7-CS shows varying patterns which correspond to the geographical location. For instance, the snow-prone Alpine Foreland depicts a high standard deviation of setting H7-CS compared to all other quality specifications of this region. The relative difference between H7-CS and other settings is by far highest for Alpine Foreland, which indicates the dramatic effects of snow cover on NDVI time series in this region. Small differences are revealed for strict setting H7-CS-S and aerosol setting H7-CS-S-A.

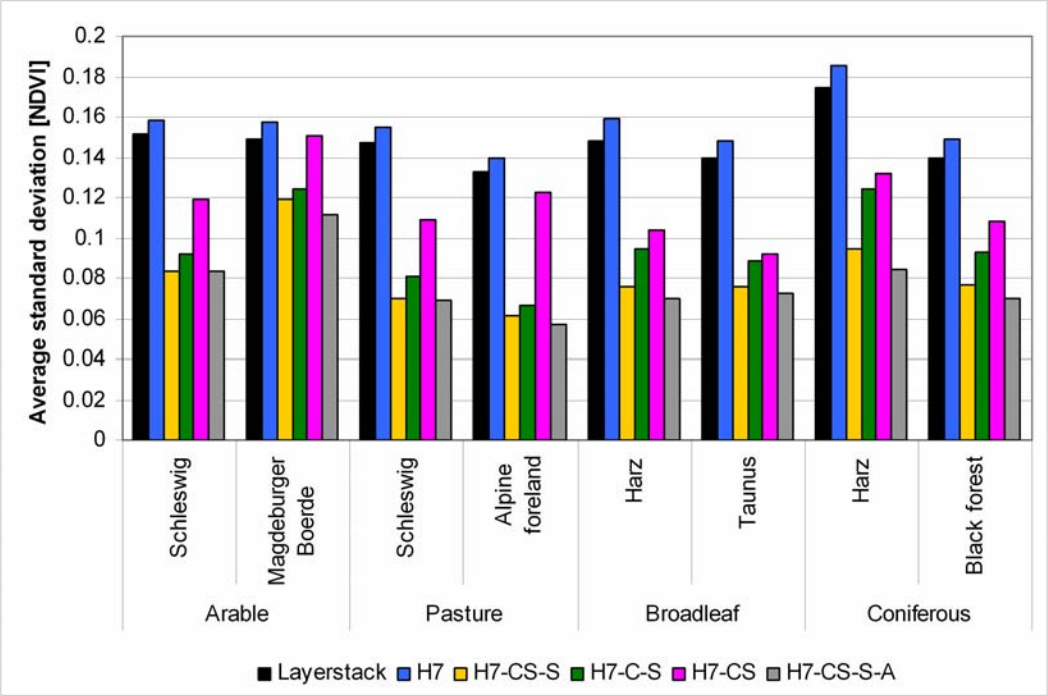


Figure 3.14: Bar plot of mean standard deviation of selected natural regions and CORINE land cover classes (Figure A.2).
Note: For specifics of the quality settings see Table 3.2.

Absolute differences of quality settings among the regions and land-cover types can only be explained with contextual knowledge and the relation to topography, geographical location, and exposition. For example, the variability of dense coniferous forest of the Harz Mountains is higher due to its location on the steep western slopes with more cloud cover and generally higher elevation. Broadleaved forest of this upland is located along the rim of the lower eastern portion. Interestingly, arable areas of the Magdeburger Boerde show a higher standard deviation than Schleswig, potentially due to winter and summer crops, where both are considered agriculture with varying NDVI values.

The second assessment approach analyzes the local noise of time series. This approach is computed for each pixel separately, i.e. independent of regional or land-cover dependencies. In order to assess local deviations, a time series is smoothed using a moving average filter of a previously defined size without weights. The average absolute difference between the smoothed and the original profile is computed.

$$smooth_i = \frac{\sum_{k=i-F/2}^{i+F/2} original_k}{F}, \qquad \text{Equation 3.2}$$

$F \ldots filter\ size$
$N \ldots length\ of\ time\ series$

Equation 3.2 shows the moving average filter with filter size *F*. Equation 3.3 calculates the mean absolute difference between the smoothed and the original curve.

$$\overline{difference} = \frac{\sum_{i=0}^{N} \left| smooth_i - original_i \right|}{N} \qquad \text{Equation 3.3}$$

In order to account for local differences, filter width F was set to five, i.e. the previous and following two composites. Larger filters would provide a more gentle profile of the smoothed time series with higher differences. Figure 3.15 shows the mean differences in NDVI for the different data-quality settings, including the layer stack for Germany. The figure reveals the expected patterns, with high local differences for lenient quality settings and very low differences for strict settings. The result was to be expected because data gaps were interpolated by linear temporal interpolation. Other temporal interpolation approaches such as spline or polynomial would cause higher local differences also for strict settings, but the general ranking among quality settings would be similar. However, some regional differences are obvious, in particular for layer stack and lenient setting H7, where urban areas and some agricultural regions have considerably lower mean differences in NDVI. This pattern is caused by a lower dynamic range of the NDVI signal in less vegetated areas, which leads to a minor impact of cloud presence. It also shows in the comparison of time series between the layer stack and setting H7-CS-S (Figure 3.12). Snow setting H7-CS indicates regional differences for snow-prone uplands. A reverse image is depicted for strict setting H7-CS-S, with lower values for uplands than for lowlands, due to the interpolation of data gaps in wintertime. Compared with the strict setting H7-CS-S, moderate setting H7-C-S indicates a higher mean difference due to insufficient cloud masking, except for the Alpine Foreland with less cloud cover.

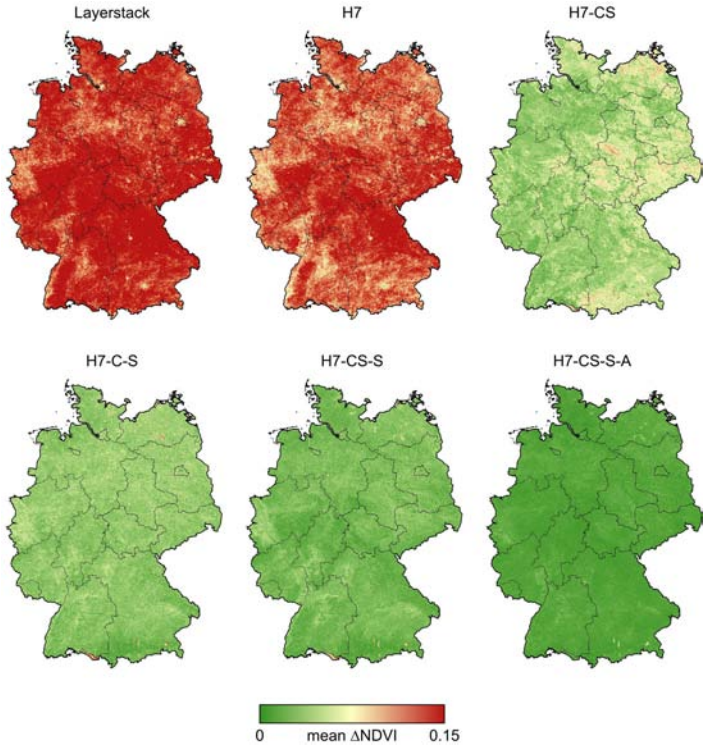


Figure 3.15: Spatial display of average absolute differences between locally smoothed and reference time series.
Note: The smoothing algorithm was a boxcar filter of window size five. For specifics of the quality settings see Table 3.2.

A third assessment was made using temporal autocorrelation. This approach analyzes the temporal connectivity between observations. Or expressed in statistical terms, the stochastic dependence of subsequent composites is assessed. As noted earlier, a NDVI time series is not expected to change dramatically from one observation to the next. Exceptions such as rapid land-cover changes due to extensive flooding or fire are unlikely for Germany. It should be noted that connectivity is not related to stationarity (Chatfield 2004). Stationarity is rarely tested in applied statistics. In this study it is assumed to have mildly stationary time series, a plausible assumption following from the visual analysis of temporal plots.

$$r_k = \frac{\sum_{i=0}^{N-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=0}^{N}(x_i - \bar{x})^2}$$ Equation 3.4

r...autocorrelation coeffcient
k...lag
N...length of time series
z...confidence
s...standard error

Equation 3.4 shows the computation of the autocorrelation coefficient *r* at lag *k*. It is common practice to compute the mean for the entire time series and apply it to all lags despite the decrease of the samples (Chatfield 2004). The temporal connectivity is analyzed using the significance of *r* at lag *k* (Equation 3.5) and the required standard error *s* (Equation 3.6). All *z* above 1.96 are significant, computed for a two-tailed Gaussian test with 95% confidence.

$$z_k = \frac{r_k}{s_k} \qquad \text{Equation 3.5}$$

$$s_k = \frac{\sqrt{1 + 2 \cdot \sum_{j=1}^{k-1} (r_j)^2}}{\sqrt{N}} \qquad \text{Equation 3.6}$$

The connectivity is expressed by the highest significant lag *k*. In other words, starting at lag one (eight days), the significance is computed and tested until it becomes insignificant for the first time. The percentage of pixels in Germany which becomes insignificant at a specific lag *k* is shown in Figure 3.16 for all quality settings. The cumulative plot depicts a rapidly increasing proportion of lenient settings at short gaps, while the proportion of pixels for strict settings occurs at longer gaps. This indicates that lenient settings or no quality analysis do not contribute to a high temporal connectivity.

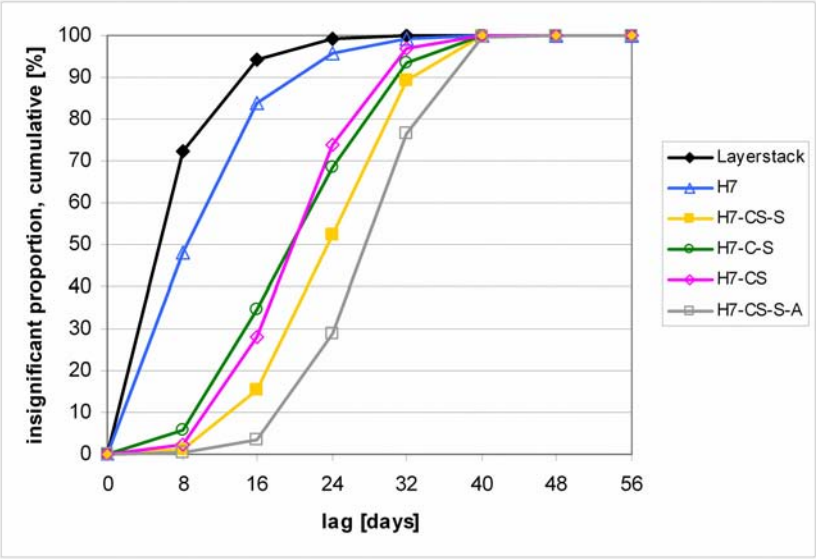


Figure 3.16: Cumulative plot of the proportion of pixels becoming insignificant with increasing lag.
Note: For specifics of the quality settings see Table 3.2.

The last significant lag is displayed spatially in Figure 3.17 for all quality settings. A longer lag indicates a better temporal connectivity. This analysis reveals a ranking similar to the average local difference analysis (see above), with a low temporal stability for lenient settings. In contrast to the mean differences using boxcar filters, this analysis highlights some

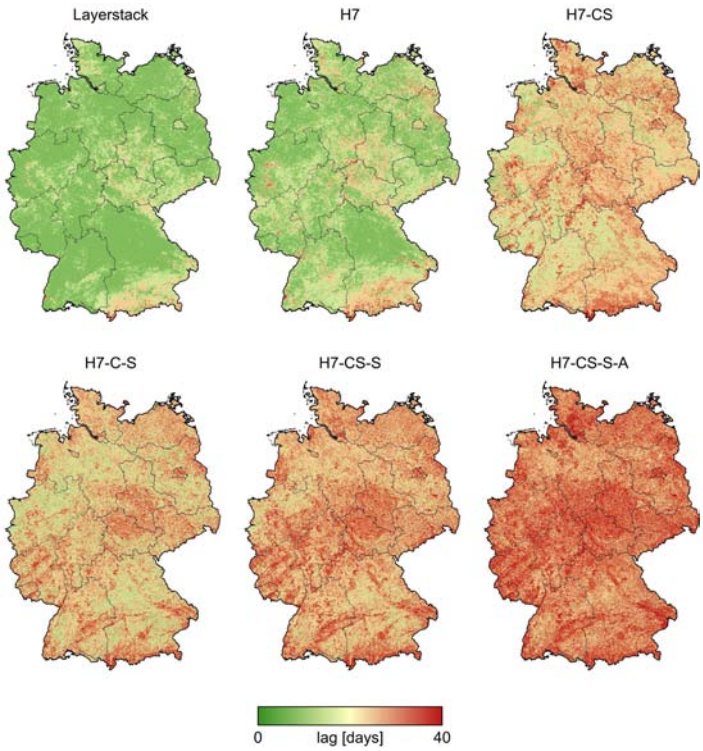


Figure 3.17: Spatial display of maximum significant lag.
Note: The lag indicates the last significant autocorrelation coefficient (95 % confidence, two-tailed). For specifics of the quality settings see Table 3.2. For specifics of the quality settings see Table 3.2.

distinctions among land-cover types. Vegetation with a clear phenology shows longer lags in some uplands and agriculture areas for moderate to strict settings. Coniferous areas, as in the Black Forest or Harz Mountains, however, do not indicate particularly high lags due to its small changes in NDVI intensity during the growing season.

Discussion and Implications

In general terms, all qualitative assessments indicate that stricter settings are usually better, having either a low standard deviation for small regions and similar land-cover type, smaller local differences compared to smoothed curves, and a better temporal connectivity. Partially, this results from the methods used for assessment. Another reason is the way the data have been processed and interpolated with linear temporal interpolation. Although, for instance, temporal interpolation with spline or polynomial functions would indicate slightly less temporal connectivity, they still show that strict settings are best.

This conclusion may appear to be correct at first glance, but does it withstand scrutiny? Temporal analysis of average time plots (Figure 3.13) for agriculture and pasture lands in northern Germany indicates that a very strict setting, such as H7-CS-S-A, excludes too many data for a contextually meaningful interpolation. In particular, arable land in the Magdeburger

Boerde is incorrectly interpolated because of missing vertexes during the short plateau phase. Similarly, this has been found true by Colditz et al. (2006a) for savanna and evergreen broadleaf forest in western Africa and by Colditz et al. (2007a) for Germany using the EVI time series. Therefore, it may be concluded that optimal time-series generation requires balancing between high data quality and sufficient data quantity.

It is neither possible to provide a quantitative measure of correctness nor a number of data to be available, because data quality varies spatially and temporally. Statistical measures solely based on the time series cannot account for contextual correctness. However, what other approaches are applicable for assessment ? Some comparison and validation strategies have been outlined in section 3.1.3 Time Series Comparison and Validation. The inter-comparison of datasets may seem useful at first glance, but a high similarity does not necessarily indicate a correct time series. For instance, Huete et al. (2002) indicated a generally good agreement in the temporal course between AVHRR and MODIS vegetation index data. However, MODIS data showed clearly higher values and a higher dynamic range, which has been mainly attributed to different wavelength and the improved capabilities of MODIS to correct for water vapor. Another inter-comparison of multiple time series concluded that SPOT VEGETATION may not correctly represent temporal surface conditions for cloudy areas in Africa, because it significantly differed from all others (Fensholt et al. 2006). However, it is not clear if all other time series are correct simply because of their correspondence.

Furthermore, almost all sensors have an equatorial crossing time in the late morning or even early afternoon, atmospheric conditions and cloud cover in particular influence or prevent the observation of the surface, especially in the tropics. An inter-comparison may result in a high dissimilarity, which may be due to rapidly changing atmospheric conditions during the acquisitions period and differences in data processing. However, it is unclear which dataset correctly represents the temporal course of the Earth surface. The issue becomes even more serious for frequently cloudy areas, much longer than a compositing interval.

Time-series validation with ground data would, of course, be the most appropriate method. Since a thorough validation in space and time is obviously not possible, the comparison with model curves seems to be an alternative. Analysis of synthetic curves has shown an impact of TiSeG's temporal interpolators on different length of data gaps at varying positions in a uni-modal annual time series. However, very large sets of model curves would be necessary, because time series shift in time and intensity, and are highly dependent on the region, surface state, and, of course, timing. Therefore, in reality, additional knowledge is necessary, including land cover and land use, precipitation, data on water storage in the soil, and fire. However, it not possible to provide such a large set of curves, and even if this set would be available, it seems impossible to select the best-fitting course for correction. The ideas of hyperspectral remote sensing with a large set of endmember spectra cannot be applied to

temporal remote sensing. Hyperspectral remote sensing is based on the physical characteristics of the electromagnetic spectrum, which is constant for all circumstances, disregarding influences such as soil moisture and atmospheric interferences in this discussion. The temporal course of a land surface, in contrast, is highly variable.

In this regard, it is most appropriate to evaluate time series from a thematic point of view. Analyzing the temporal course with expected patterns by visual inspection is recommended for most applications. Of course, the analyst should have a profound knowledge of the study area.

It is also important to question the correctness of automatically and globally generated MODIS quality indicators. For instance, sometimes the MODIS QA-SDS may indicate snow or ice cover in desert regions. These surfaces are mainly located in desert regions and correspond with salt pans. This issue was prevalent in early versions of MODIS data and has been successfully corrected in collection 4. Still, remaining artifacts indicate the limits of operational production and ask for the careful consideration of quality indicators also using contextual knowledge during data analysis.

It also needs to be considered that the quality of the data depends on the algorithms used for assessment. Therefore, changes in upstream data processing or alterations in the sensor, including degradation, electronic, or mechanical problems, have a significant impact on rendered data quality and significantly influence the time series. This issue will be addressed in the next section, showing the impact of updated MODIS data on time series.

3.3.3 Differences between Collection 4 and 5

While the previous analysis has specifically focused on the evaluation of different quality settings of time series, this section will elucidate the main differences between two data collections of the MOD13 product. As already outlined in 2.2.2 MODIS Land Datasets, MODIS data are irregularly updated to include new scientific knowledge and improvements in product generation and accuracy. Each MODIS data version / collection requires a full reprocessing of the entire data archive for all levels. In September 2006, reprocessing from C4 to C5 has started for all MODLand products.

The differences in data quality between C4 and C5 MOD13 products will be described in the first section. A brief explanation of the time-series generation process will be followed by a discussion of the main differences between NDVI and EVI time series of C4 and C5. The content of this section has also been published in Colditz et al. (2007b).

Changes in Data Collections for the MOD13 Product

Substantial changes in science and structure have been applied to C5 data of the MOD13 product and are documented in Colditz et al. (2007b) and Didan and Huete (2006). It is not recommended to combine both versions for a reliable data analysis (Didan and Huete 2006). Scientific modifications have been applied to cloud and aerosol retrieval, another backup method of EVI retrieval, and an improved compositing method. Aerosol and cloud masking have already been changed in upstream processing of atmospheric products and surface reflectance retrieval in C5. In MOD13 processing, filtering functions have been adjusted and strengthened, e.g. for pixels adjacent to clouds. Compositing of these pixels have been changed to the simple but robust MVC method. In C4 the SAVI (Huete 1988) has been used as EVI backup for pixels with cloud cover, snow and ice surfaces, or if the blue band was out of range. In C5 a newly developed equation, termed EVI2 (Equation 3.7), is used for better spatial and temporal continuity with the standard EVI.

$$EVI2 = 2.5\frac{\rho_{NIR} - \rho_{RED}}{1 + \rho_{NIR} + \rho_{RED}} \qquad \text{Equation 3.7}$$

ρNIR...surface reflectance of the near infrared band
ρRED...surface reflectance of the red band

Compositing with regard to the view angle (CV-MVC, Huete et al. 2002) has been changed to ensure better spatial stability. In C4 the two highest vegetation index values with less than 30% deviation were nominated and the observation with the lowest view angle was selected for each pixel separately. This led to spatially varying vegetation index values within the 16-day compositing period. C5 processing is adjusted with a stricter deviation of not more than 10%, and a contextual temporal selection with regard to neighboring pixels to ensure better spatial continuity.

Structural changes have comprised a different arrangement of SDS, modifications of the QA-SDS, and phased production between Terra and Aqua. Due to negligible differences in QA-SDS of NDVI and EVI, C5 only contains a single QA-SDS valid for both indices. SDS, containing the day of vegetation index value selection and a so-called reliability dataset, have been added, where the latter also becomes a critical data-quality indicator. Since the BRDF compositing approach (van Leeuwen et al. 1999) has never become operational in C4, this flag has been excluded in the QA-SDS of C5, and the full three-bit land and water mask is now included instead. An 8-day shift between Terra and Aqua MOD13 production effectively generates a time series with 8-day intervals from 16-day vegetation index composites, if data of both platforms are combined. Furthermore, C5 data are compressed and contain additional meta-data parameters.

Time Series Generation

C4 and C5 data of the 16-day 500 m vegetation index product (MOD13A1) of Germany have been processed for one year, starting on day 49 (February 18th) in 2000. The beginning in February was due to the currently limited availability of C5 data; just this period has been available at the time of writing. Time series had to be generated without shoulder datasets. Using MRT, original composites of tile h18v03 and h18v04 (Figure A.1) were mosaiced, reprojected to UTM zone 32, and subset to the extent of Germany (Table A.18).

Table 3.3: Quality settings of C4 and C5 MOD13A1 (16-day, 500 m vegetation indices) time series of Germany for 2000 / 2001.

Setting	Usefulness index	Mixed clouds	Snow / ice	Shadow
C-S-S		No	No	No
UI3-C-S-S	Perfect - acceptable	No	No	No
UI5	Perfect - intermediate			

Note: Only the selected settings are shown in this table. For a complete list of QA-SDS specifications see Table A.10 for C4 and Table A.12 for C5. The reliability SDS with settings good and marginal data (Table A.13) is used in addition to C5 QA-SDS settings indicated by C5+R.

In addition to a simple layer stack, NDVI and EVI time series were generated with three quality settings (Table 3.3). For detailed QA-SDS descriptions of C4 see Table A.10 and Table A.12 for C5. While C5 data only contain one QA-SDS for both vegetation indices, NDVI and EVI data of C4 were retrieved from their respective QA-SDS. The settings can be ranked lenient (UI5), moderate (C-S-S), and strict (UI3-C-S-S). Furthermore, C5 contains the newly introduced reliability SDS with additional indicators for clouds and snow / ice surfaces (Table A.13). The selection of good and marginal quality from the reliability SDS, excluding clouds and snow / ice pixels, is abbreviated by C5+R.

Time Series Comparison between Collection 4 and 5 MOD13 Products

While the general characteristics of the lenient, moderate, and strict quality setting with low to high numbers of invalid pixels are well represented, Figure 3.18 also depicts substantial differences with regard to data collection and the consideration of the reliability SDS. It is shown for all settings that C5 data exclude more data from analysis than C4 products. The only exception is lenient setting UI5 for days 81, 193, and 17. The additional consideration of the reliability layer masks even more data as invalid. Particular differences in data collections are shown for days 177 to 209 and 305 to 1. With regard to the maximum gap length, the differences between C4 and C5 seem to be less relevant for lenient quality setting UI5 and increase towards strict settings such as UI3-C-S-S. In contrast, the additional use of the

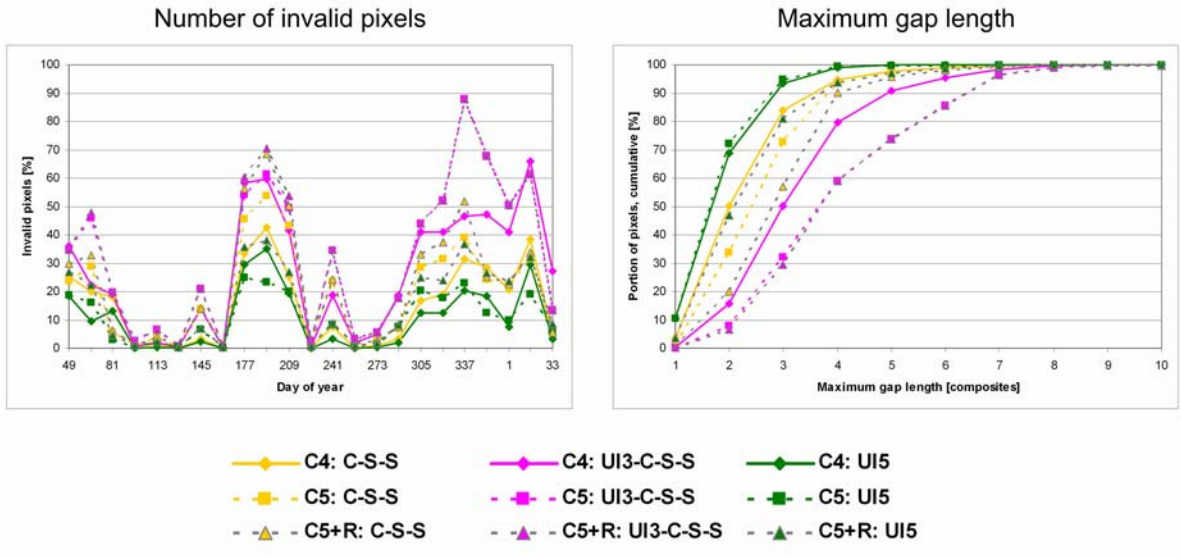


Figure 3.18: Temporal quality analysis of C4, C5, and C5+R for Germany.
Note: For specifics of the quality settings see Table 3.3.

reliability SDS makes almost no difference for strict settings, while lenient setting UI5 shows an increase of approximately one composite. The influence of the reliability SDS on the time series is therefore relevant for lenient settings, which may still contain potentially invalid data.

A spatial view on the number of invalid pixels highlights the regional differences for specific quality settings and the respective data collection (Figure 3.19). Lenient setting UI5 shows almost no increase in invalid pixels from C4 to C5, but clearly different patterns between C5 and C5+R (the additional use of the reliability SDS). The strict setting UI3-C-S-S depicts a reverse sequence. It also excludes large areas in northern Germany, mainly due to frequent cloud cover. Most upland areas in central Germany and the Alps have less valid data. It is surprising that the Black Forest upland in southwestern Germany, with peaks up to 1,500 m a.s.l., shows a low number of invalid pixels. These good conditions are attributed to the rain-shadow effect of the Vosges Mountains at the western side of the Rhine valley in France, with substantially more cloud cover.

Average interpolations including the layer stack of NDVI and EVI have been computed for natural regions Schleswig, with land cover “pasture”, and the Harz Mountains, with land cover “coniferous” (Figure 3.20). For the location of the natural regions and land cover see Figure A.2. NDVI plots of Schleswig indicate clear differences between C4 and C5. Even the layer stack reveals less dramatic decreases for composites of days 209 and 353 in the new data collection. Furthermore, the dynamic range seems to have increased. While the lenient and moderate settings reveal a characteristic profile, strict setting HI3-C-C-S excluded too many observations during the dormancy period, resulting in a meaningless curve. This effect of too strict quality settings and an insufficient quantity for interpolation has also been shown for arable land use in the Madgeburger Boerde (see 3.3.2 Comparison between Improved and

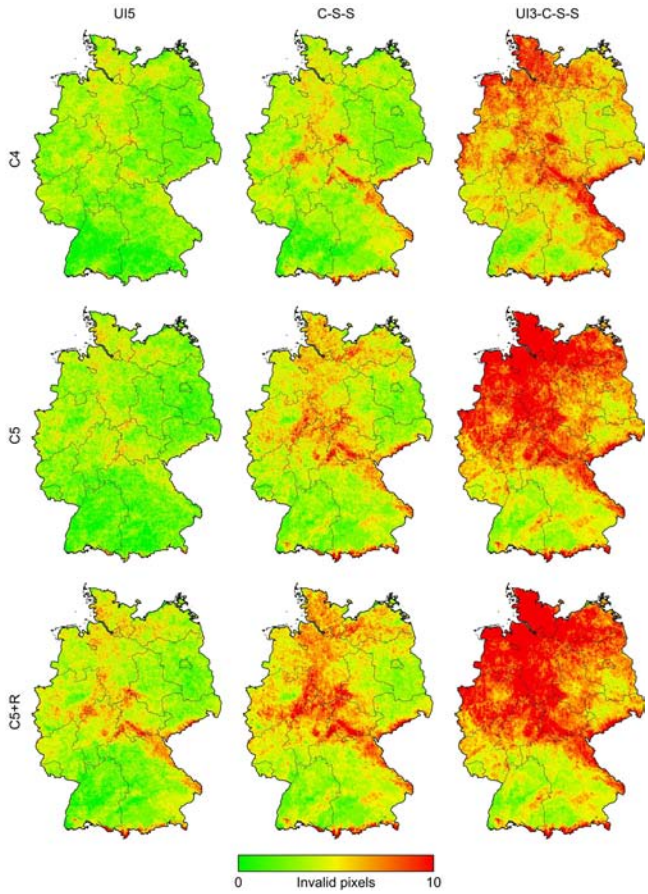


Figure 3.19: Spatial quality analysis of C4, C5, and C5+R for Germany.
Note: Upper row: collection 4 data. Middle row: collection 5 data without reliability SDS. Bottom row: collection 5 data with reliability SDS. Values greater 10 are displayed in red. For specifics of the quality settings see Table 3.3.

Original Time Series). Plots of the Harz Mountains depict a clear drop in C5 NDVI for days 177 to 209 for all quality settings. This decrease is due to a persistent cloud cover during the period. While in C5 the mixed cloud flag only masks partially cloudy pixels, the same flag was also used to mask completely cloudy pixels in C4. The additional use of the newly introduced reliability SDS is required to exclude all cloudy pixels and to generate expected phenological plots.

EVI time series show a slightly higher dynamic range and a remarkably lower intensity. This indicates the usefulness of the EVI for high biomass regions (Huete et al. 2002). Furthermore, EVI plots are smoother and do not indicate substantial differences among quality settings. The long plateau phase of Schleswig as shown in the NDVI data merely becomes a gentle decrease in EVI, without a clear senescence phase. Similar patterns are shown for the Harz Mountains. It is obvious, however, that the EVI backup, either the SAVI in C4 or EVI2 in C5, limit the influence of cloud-obscured pixels on the quality of the data. Comparing the cloud period of days 177 to 209 for the Harz, the use of SAVI caused slightly lower EVI values in C4, while EVI2 in C5 even computed higher values than the interpolation and a smoother curve during wintertime.

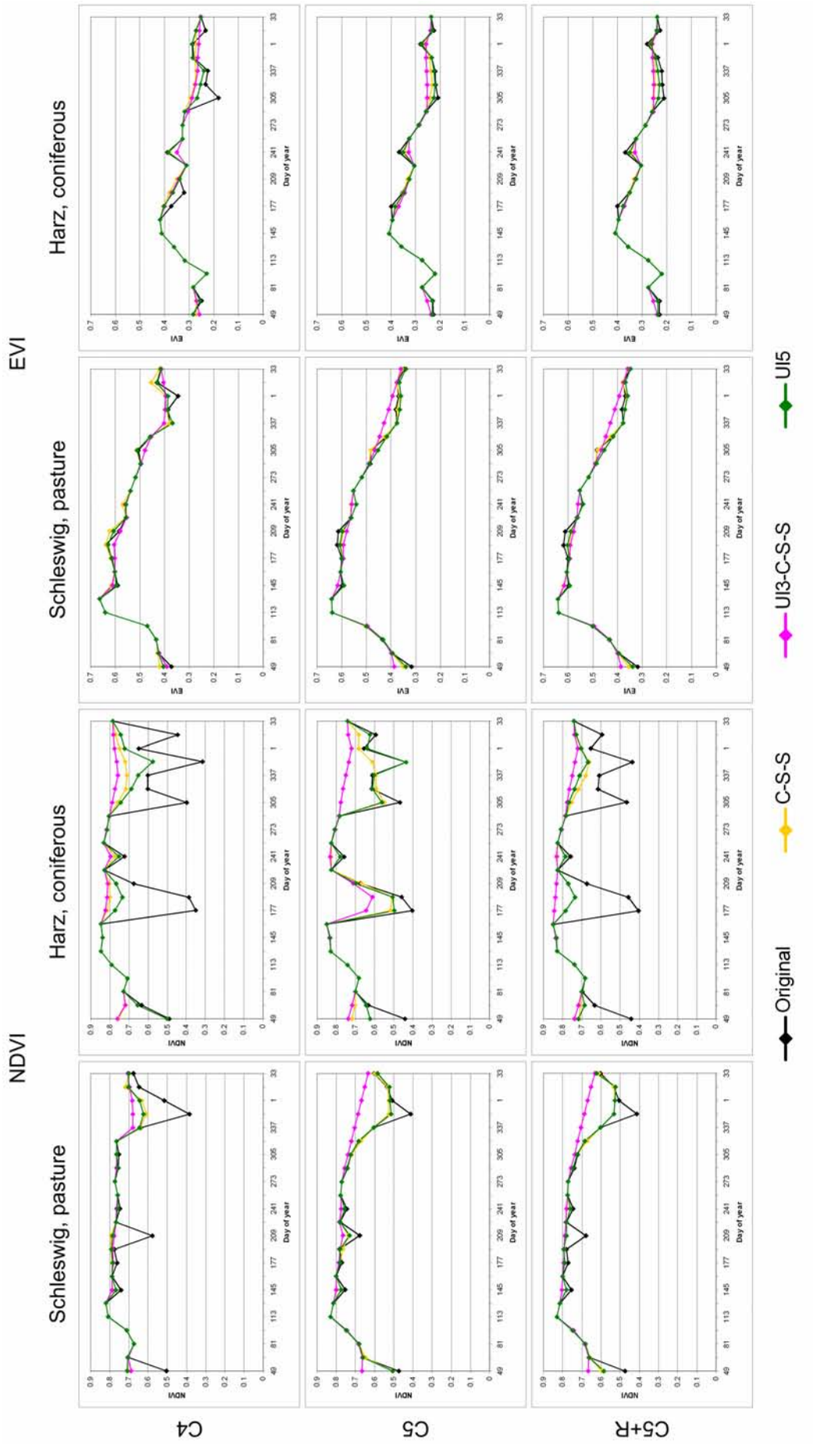


Figure 3.20: **Average interpolation of C4, C5, and C5+R for selected natural regions and CORINE land cover classes (see Figure A.2).**
Note: **Upper row: collection 4 data. Middle row: collection 5 data without reliability SDS. Bottom row: collection 5 data with reliability SDS. For specifics of the quality settings see Table 3.3.**

Discussion and Implications

It has been shown that the scientific and structural changes in C5 have a substantial influence on both vegetation indices with regard to their intensity and their quality. Both collections cannot be used interchangeably or together. This has implications on the on-going research of long-term data records spanning from several AVHRR instruments to MODIS and future VIIRS. Methods are needed for generating continuous and consistent time series.

The time-series comparison between both collections shows remarkable differences, and the new version of MODIS data generates a more realistic phenological curve. However, the consideration of the newly introduced reliability SDS is highly recommended for C5, in particular to exclude all cloudy composites from further analysis. Furthermore, EVI time series are smoother in comparison to NDVI plots. The EVI time series appears more stable even for low quality data or a simple layer stack, most likely due to suitable backup algorithms, if the EVI cannot be computed directly. While the impact of the quality analysis and the reliability SDS of C5 is clearly necessary for NDVI time series, the importance for the EVI is relatively low.

3.3.4 Regional Time Series Analysis of South Africa

The previous analysis sections specifically addressed issues of data quality and interpolation, either with respect to multiple QA-SDS or to different data collections. This final section on time-series analysis focuses on a multi-year analysis of temporal shifts of vegetation phenologies. The study region of South Africa was deemed useful, because seasonal patterns are highly correlated with precipitation, which, in turn, is highly variable in this semi-arid environment. The aim of this case study is to demonstrate how time series can be employed for the analysis of a particular region.

A first part will provide the specifics of the time-series production followed by an explanation of approaches used for time-series analysis. In particular, a baseline characterization of seasonality patterns (non-modal, uni-modal, bi-modal) is necessary for a thorough assessment of temporal differences among consecutive years. The results will be discussed in a final section. Parts of the approaches used for time-series analysis are also discussed in Colditz et al. (2005), Colditz et al. (2006b). The results of this analysis have been published in Colditz et al. (2007c), with an additional comparison with precipitation rate estimates.

EVI Time Series Generation for South Africa

The enhanced vegetation index with 500 m spatial resolution and a 16-day compositing period (MOD13A1) has been used for time-series generation. In contrast to NDVI, the EVI is preferred due to a generally smoother profile, a higher dynamic range, and because it has been found to better mimic the real vegetation phenology (see 3.3.3 Differences between Collection 4 and 5, Colditz et al. 2007b, Huete et al. 2002). Using MRT, four MODIS tiles (h19v11, h19v12, h20v11, h20v12; Figure A.1) were mosaiced, reprojected to Albers equal area, and confined to the country of South Africa (Table A.18). Annual time series were generated using TiSeG, where each year started on day 177 (June 26th), to account for phenological phasing on the southern hemisphere for large parts of the study region. Although the data quality is generally very good, remaining clouds, shadow effects, or pixels with a usefulness index below fair (see Table A.10) were eliminated and interpolated with linear temporal interpolation. This quality setting seemed most appropriate for the following analysis. Its minor influence of the quality analysis and subsequent interpolation of invalid pixels mainly corrected cloudy pixels over mountain ranges along the escarpment.

Approaches to Multiyear Time Series Analysis

The approaches to multi-year time-series analysis comprise the characterization of seasonality and the analysis of temporal shifts between two consecutive years. Five major types of seasonality have been identified. Non-modal profiles with no phenological development are related to either deserts or evergreen forests with a low or a high EVI intensity, respectively. The uni-modal type of seasonality shows clear phenological phases such as green-up, plateau, and senescence and includes land-cover types such as grasslands, dryland agriculture, and deciduous woodlands. This type of seasonality is further distinguished by the peak in summer- or wintertime to account for the winter-rain climate in the Cape Floristic Region. Bi-modal profiles characterize areas which are influenced by winter and summer rain climate, likely to be expected along the southern coast. Furthermore, irrigated agriculture and riparian vegetation show this type of seasonality, although the temporal development is not coupled with the seasonality in precipitation.

Harmonic analysis has been used to categorize each pixel into five types of seasonality. In particular the explained variance of each harmonic is a useful indicator (Colditz et al. 2006b). Uni-modal is attributed, if the first harmonic explains most of the common variance between the original and the transformation. Uni-modal curves were further divided into winter and summer peak according to the timing of the maximum. Uni-modal curves with maximum EVI from April to September were assigned to winter-rain, whereas peaks from October to March

were considered representative for summer-rain environments in the larger eastern portion of the country. Bi-modal is selected if the second harmonic, indicating two waves, had the highest explanation. In order to map non-modal pixels, two approaches have been necessary. First, the annual dynamic range was computed. If the range was lower than 0.1, excluding the two highest and lowest values to account for outliers, the pixel was assigned to non-modal. This approach selected shrublands and semi-desert landscapes with low EVI values. Secondly, the sum of all explained variances of harmonics greater than two was used. Usually higher harmonics are regarded to indicate noise. On the other hand, a high short-term variability is typical for wetlands due to periodic or episodic inundations.

In order to exclude comparisons between consecutive years with dramatic changes in the phenological characteristics, the type of seasonality must be identical for both years. No further analysis was also made for non-modal seasonality, because the temporal cross-correlation could be largest at any lag. If both prerequisites were fulfilled, the temporal cross-correlation coefficient was computed.

$$r_k = \begin{cases} \dfrac{\sum_{i=1}^{N-|k|}(x_{i-|k|} - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} & \text{for } k < 0 \\ \dfrac{\sum_{i=1}^{N-k}(x_i - \bar{x})(y_{i+k} - \bar{y})}{\sigma_x \sigma_y} & \text{for } k \geq 0 \end{cases}$$

Equation 3.8

r...temporal cross correlation coefficient
k...lag
N...length of time series

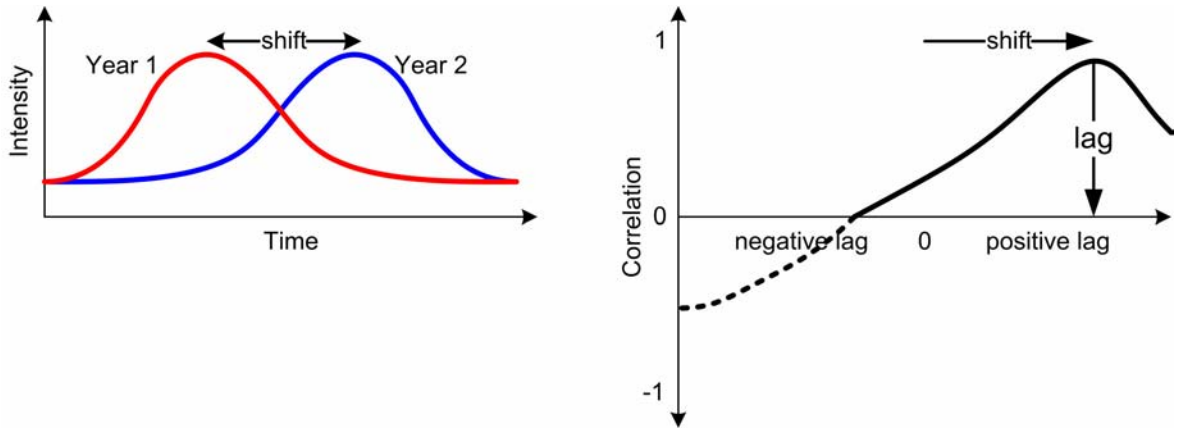


Figure 3.21: Computation of lag with highest cross-correlation coefficient.
Note: Two time series of subsequent years are compared (left). Year 1 is shifted in time. The highest correlation coefficient is derived by a shift to the right, i.e. at a positive lag. Shifting year 2 would result in the same lag, but inverse sign.

The temporal cross-correlation coefficient *r* (Equation 3.8) indicates the relative correspondence of the time series, with lag *k* between two years stressing the temporal aspect of both curves (Colditz et al. 2005). No conclusions can be drawn to absolute similarity, i.e. the intensity of the EVI for both years. In this respect, the cross-correlation coefficient is more appropriate for analyzing temporal shifts than the RMSE, which also accounts for the absolute difference between time series. The lag *k* of the highest temporal cross-correlation coefficient indicates the temporal shift in phenology between the EVI time series of two consecutive years (Figure 3.21).

Analysis of Phenological Characterization and Temporal Shifts

Figure 3.22 shows the types of seasonality for three years as a map and corresponding temporal EVI profiles. It should be noted that both approaches, used to distinguish non-modal phenologies from vegetation growth, are depicted. All profiles show a remarkable variability in intensity and timing among the years, except for the non-modal curve with a low dynamic range. While the variability of the beginning and intensity of the rainy season is one reason, spatially smaller and locally varying seasonality also caused some changes in the temporal profiles. Besides the inherently higher dynamic range of the non-modal class, the changing spatial extent among the years, e.g. in the Western Cape, Eastern Cape, and the northern provinces, also contributes to some intra-annual variability in the temporal profile.

Similarly, the bi-modal seasonality type shows substantial changes in intensity for the first peak, and slight temporal shifts of the second mode. It should be noted that the bi-modal seasonality is mainly located along the main rivers, Limpopo, Vaal, and Orange in the interior of the country, and the class is spatially stable for these areas. Besides riparian vegetation, irrigated agriculture dominates these areas which are dependent on the hydrological regime and land management and are therefore decoupled from climatic dependencies. Furthermore, a near-surface groundwater table in the Free State with agricultural land use leads to a stable bi-modal seasonality. Bi-modal seasonality along the southern coast of the Eastern Cape in 2005 / 2006 visualizes the uncertainty of seasonal classification for some areas. In addition to orographic effects, this region is influenced by both rainy seasons. The intensity and duration of the rainy seasons and the potential vegetation dry-down in the dry season can assign this area to any type of seasonality, dependent on the prevailing regime in the respective year.

Uni-modal seasonality with EVI maximum in winter is mainly confined to the Cape Floristic Region in the southwestern part of South Africa. It was found in western and eastern Africa that vegetation reaches maximum greenness with a one-to-two month shift following the rainy season (Nicholson et al. 1990). Similar patterns are revealed for South Africa (Colditz et al.

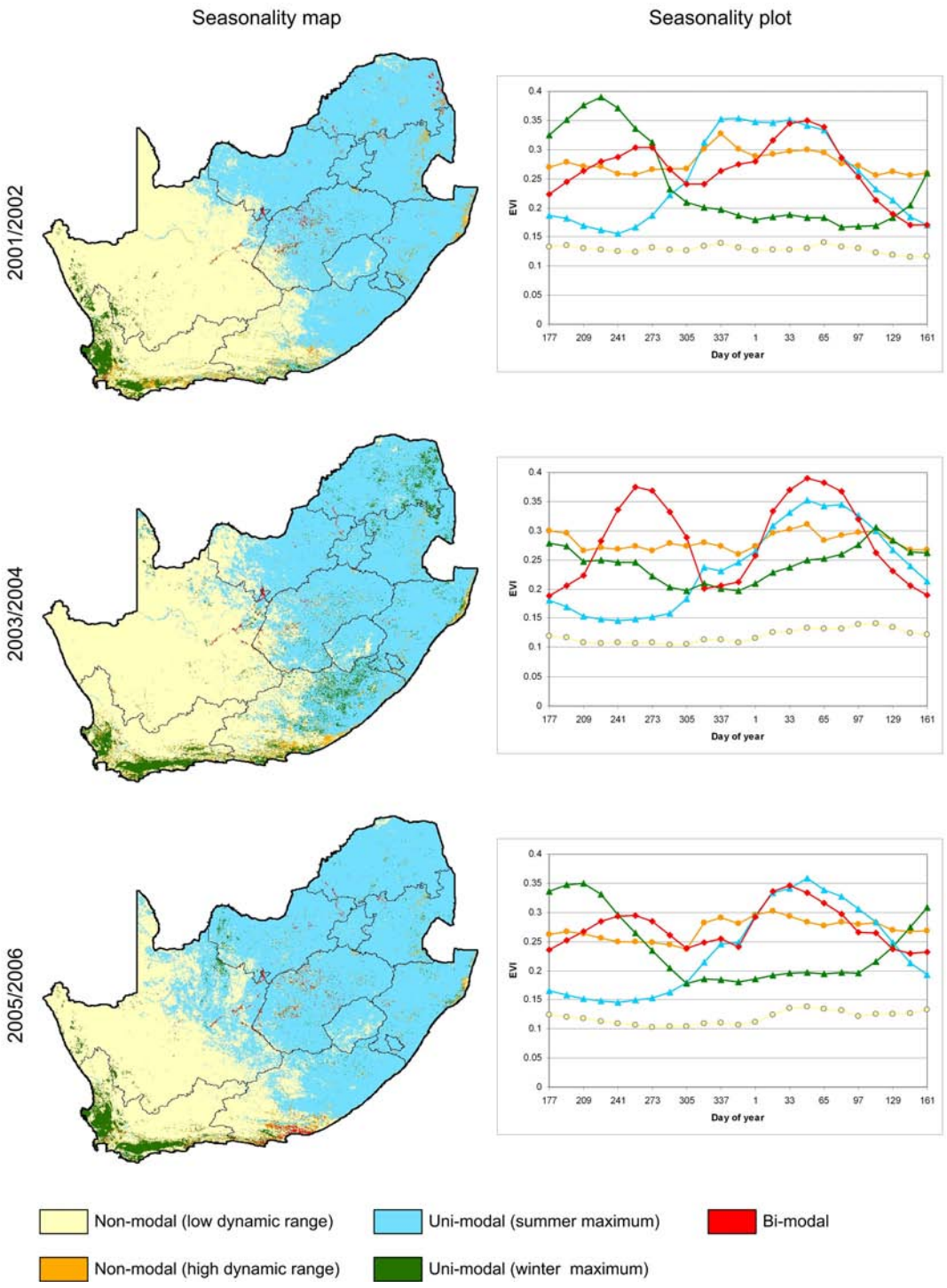


Figure 3.22: Types of seasonality in a map and average EVI plot for South Africa for years 2001 / 2002, 2003 / 2004, and 2005 / 2006.

2007c). While this expectation holds for most years, a different pattern is shown in the temporal plot for season 2003 / 2004, which assigns substantial parts in the Eastern Cape and Limpopo Province to the winter-rain climate. This assignment is due to a high temporal variability in rainfall and to the approach by which seasonality was characterized (see above). It has been argued in Colditz et al. (2007c), that the maximum of EVI integrals over one- to two-month periods would potentially yield more stable and accurate results to distinguish

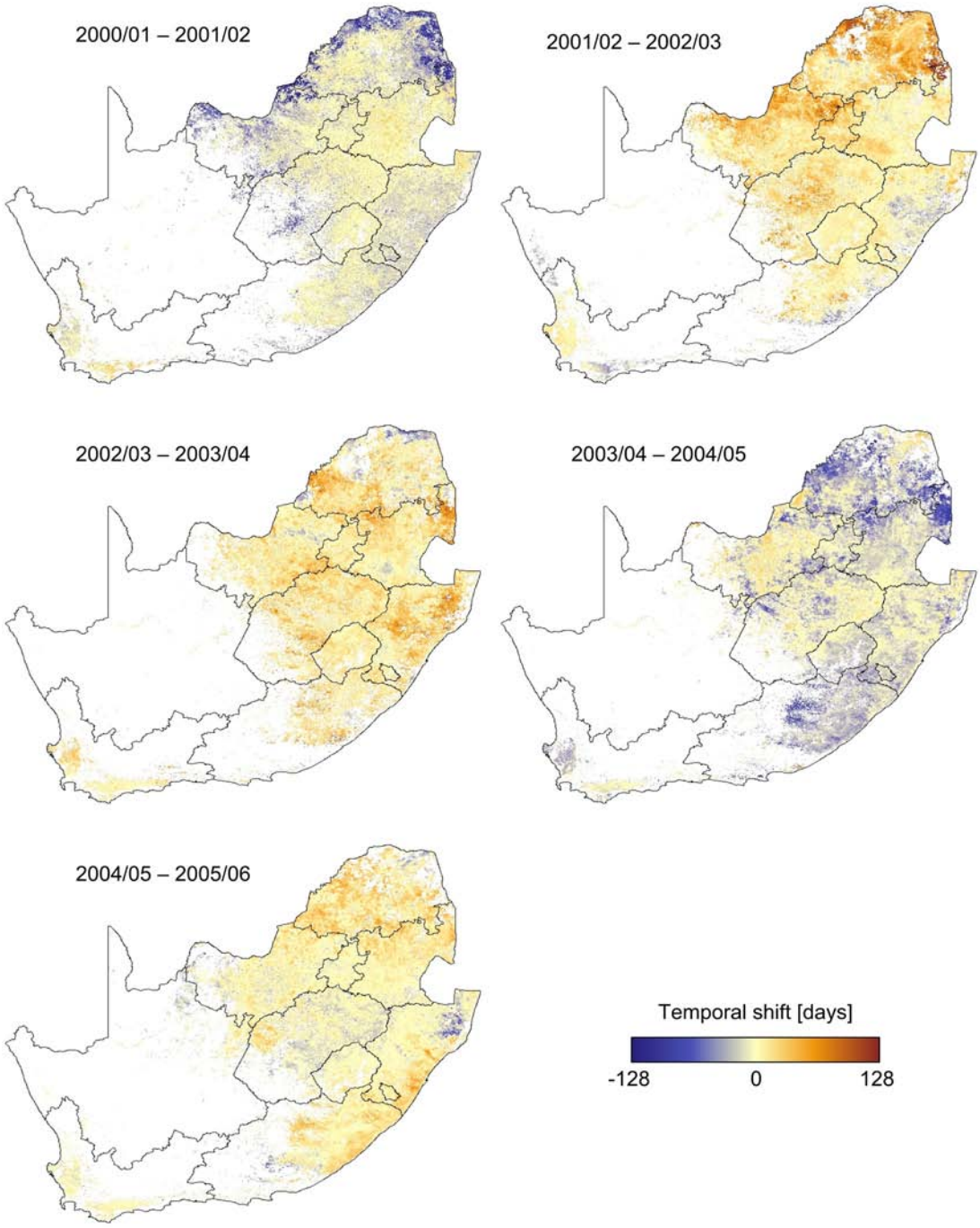


Figure 3.23: Temporal shifts between consecutive years for South Africa using an EVI time series.
Note: The highest cross-correlation coefficient at a specific lag indicates the temporal shift. Shifts are only computed for pixels belonging to the same type of seasonality for both years.

summer and winter-rain uni-modal curves. Though, further research is needed to account for a correct classification of temporally varying seasonality. The uni-modal seasonality with its vegetation peak during summer is rather stable for the large eastern portion, including the plateau, escarpment, and some coastal areas. However, the boundary to the west varies spatially, depending on the green-up intensity along the wide shrubland-bushland ecotone in the eastern Karroo (Dean and Milton 1999).

Figure 3.23 depicts the temporal shifts in seasonality by comparing two consecutive years, e.g. the year 2000 / 2001 with 2001 / 2002. Alternations in the inter-annual variability of South Africa's ecosystems are represented in variations of the timing of phenological cycles. White areas cannot be analyzed because of non-modal seasonality or different types of seasonality in the respective years. A general characteristic of this analysis is that seasonal temporal shifts of one comparison are often compensated for with a reverse in the next year. This is particularly evident for regions in the northern part of the country with a remarkable temporal variation. Moderate shifts are depicted for Kwazulu Natal, where summer rain and orographic effects contribute to variable seasonal phasing. Interestingly, a visual comparison to the NLC1995 of South Africa indicates a coincidence between high temporal variability and forest, woodland, and thicket / bushland land-cover types. No correspondence was found between areas affected by fire. Agricultural areas in the Free State also show moderate shifts due to land management issues. Besides these varying areas, the winter-rain region in the Western Cape and large parts of Lesotho, Free State, and Mpumalanga are rather stable. For instance, only a minor temporal variation is indicated for the Western Cape between 2002 / 2003-2003 / 2004 with an immediate reverse shift in the following season

Discussion and Implications

The accurate characterization of seasonality patterns is very important for the meaningful analysis of temporal shifts. Essentially, this prerequisite prevents from comparing apples and oranges. However, hard classification can be inappropriate for transition zones. This limits the study to rather conservative estimates of areas where the type of seasonality is very clear. A soft estimate will be attempted in future analysis.

The previous analysis merely focused on annual integrals of NDVI and comparisons among different areas. The time-series generation with TiSeG produces an improved time series suitable for advanced analysis techniques. This novel approach of a truly temporal analysis of phenological shifts concludes that a high temporal variability corresponds with high biomass areas in natural and semi-natural environments. Furthermore, agricultural land uses depict some variability, depending on rainfall occurrence and land management. However, no trends in temporal shift can be concluded. The patterns merely indicate a high temporal variability of the phenological phasing. The correspondence between vegetation seasonality and a time series of precipitation-rate estimates, the driving factor of vegetation activity in this environment, was established by Colditz et al. (2007c). It has been shown that for most parts vegetation growth lags approximately one to two months behind the main precipitation period. This is in agreement with earlier findings by Nicholson et al. (1990) for western and eastern Africa.

Chapter 4

Automated Time Series Classification

This chapter will focus on the classification of time series. Especially the high dimensionality of the data requires a suitable classification procedure. Because one of the aims of this study is the transferability of the approaches to other regions and datasets without adjustments, a highly automated and robust classifier is required. However, not only the classifier itself, but the entire classification framework is important for accurate results. Furthermore, the map accuracy has to be viewed from different perspectives to provide a comprehensive conclusion of appropriate classification strategies.

Section 4.1 Theoretical Background will discuss several basic considerations of data classification. Specific sections will focus on approaches previously used in time-series classification, a description of decision-tree generation as the selected classifier for this study, and techniques for accuracy assessment. The second section (4.2 Classification Procedure) will provide a detailed description of the methodology developed in this study. In the last main section, the analysis of the time-series classifications will be shown for four study regions (4.3 Time Series Classification Analysis). Separate sections will contain a thorough description of classification and accuracy for South Africa and Germany, a sensitivity study of classification parameters, the capabilities of the approach to updating land-cover classifications, and the application to other datasets such as vegetation maps.

4.1 Theoretical Background

Satellite images consist of complex measurements of Earth's reflectance and emission recorded at multiple wavelengths and stored in multi-dimensional datasets. In order to obtain usable data, e.g. for decision making, information extraction techniques are highly demanded which transform data to useful, contextual information (Jensen 2005). Image classification, a field in pattern recognition, is one possible technique of thematic extraction. Commonly, image classification is understood as the transformation of continuous multi-dimensional measurements into discrete classes. The number and detail of classes depend on the analyst and respective application. There is a wide field of methods which often originate from other sciences, such as statistics, probability or information theory, or data mining.

In remote sensing, image classification is the most demanded, most investigated, and most applied operation. Several other fields necessitate classifications of satellite image data. For instance, climate modeling requires coarse-resolution land-cover characterizations. Similarly, biodiversity modeling or carbon and nitrogen budgeting models need gridded land-surface characterizations. In contrast to land cover, land use is required by many regional studies. Urban mapping requires high spatial resolution data, and the classification of crops or minerals employs hyperspectral imagery. Change-detection studies require land-surface classification at all spatial scales.

Land-cover and land-use mapping clearly dominate the field of image classification of satellite data, while mapping of vegetation types or surface water types plays a minor role. Land cover is defined by the properties of the land surface, such as vegetation, soil, topography, water, and human structures. In contrast, land use focuses on the exploitation and treatment of the land by humans (Lambin et al. 2006). For instance, grassland is a defined land-cover type, but does not indicate how this land is used, e.g. for ranges, natural meadows, or golf courses. While land cover can be mapped quite accurately, land use is often only detectable with additional data or contextual knowledge. Advanced approaches generate contextual information from satellite data (Wehrmann 2007). For instance, a vegetated area, usually defined as "open woodland", is assigned to the land-use class "parkland" if the surrounding areas are mapped as class "urban". Several classifications, including this study, mix land cover and land use in the classification scheme.

In essence, all supervised classification procedures follow the general framework depicted in Figure 4.1. As input, some kind of data is necessary, which will be transformed into classes. In particular, supervised classifications require sample data, i.e. data where the classes are known. Often, sample data are split into training and validation sets which originate from the same distribution but are independent. Therefore, they must not mix during the classification

process. For the known training data, signatures are derived from the image. A signature is essentially a vector of data properties for each class. Multiple vectors for each class describe its variability and can indicate multiple distributions in feature space. It is highly important to acquire signatures of all frequency distributions, in order to accurately map a class throughout the entire dataset. The signatures are ingested into a classification algorithm for decision making and subsequently applied to the dataset. Finally, the classification is assessed using the independent test data.

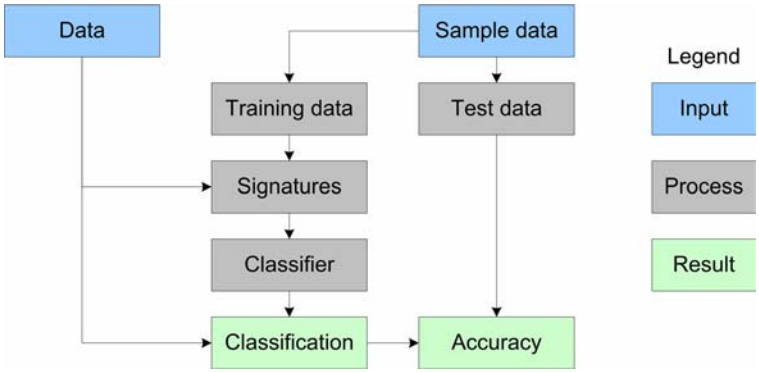


Figure 4.1: General process flow of a supervised classifier.

There are three important considerations of image classification which should be discussed beforehand: (1) pixel- vs. object-based classification, (2) hard vs. fuzzy class assignment, (3) unsupervised vs. supervised approaches. Pixel-based image classification has been applied for a long time to almost all datasets. Object-based classification techniques have become appropriate since the recent advent of high and highest spatial resolution image acquisition. Objects are built before conducting the actual classification by grouping pixels to rather homogeneous segments. In a second step a classifier assigns a thematic class to each object. Although these rather recent and still evolving approaches generate promising results, their current usefulness and application aim at higher spatial resolution data (30 m or less), because the generation of image segments requires the detection of structures in the dataset. Since time series are currently only available at medium to coarse spatial resolution (250 m and coarser), pixel-based approaches have been used in this study.

Although many users and applications demand a discrete or hard classification, it can be appropriate to generate thematic output with fractional estimates of each class. The latter result is also known as soft or fuzzy classification and is particularly suitable for heterogeneous surfaces. Heterogeneity is related to the spatial resolution of the data, the landscape heterogeneity, and the scale of analysis (Mather 2003, Moody and Woodcock 1994). Even though this example is clearly exaggerated, images with 1 cm spatial resolution are suitable for capturing the flowers, lawn, pavement, terrace, and swimming pool of a backyard. On the other hand, a resolution of 10 m of the same surface contains a mixed signal of several surface types. Although no spatial reference is made where a specific class is

located in this pixel, fuzzy classifications provide a better surface cover estimate. Furthermore, each fractional estimate can be transformed into a hard classification, e.g. by using a majority rule (“the winner takes all”) for the class with the highest proportion of membership.

Variants of soft classification are available for almost all classification techniques. Terms used for fuzzy estimates of each class range from fraction to proportion to membership, and will be used interchangeably in this study. Sometimes fuzzy estimates are indexed between zero and one, but can similarly be expressed in percentages between 0 % and 100 %. If not inherent to the classifier (as in this study), a constraint usually ensures that all class proportions add up to 1 or 100 %.

With regard to time series and their inherently coarser spatial resolution, fractional cover mapping seems to be particularly suitable, because many pixels at 500 m to 1 km belong to more than one land-cover class. This issue has also been increasingly addressed by the global and regional modeling community, and state-of-the-art products such as Vegetation Continuous Fields (VCF) satisfy the requirement (Hansen et al. 2000).

The aforementioned third consideration is related to a previously defined classification scheme or legend, and the availability of training data. Supervised classification can be applied if *a priori* knowledge is available. Typically, spatially located training sites are available from fieldwork or other datasets, and some kind of pattern recognition approach generates a classification by extrapolation to all unknown areas. If no training data or even the classification scheme is known, unsupervised classification can be employed to divide the data into statistically similar regions. This requires some kind of similarity distance measure, e.g. the Euclidian distance.

The next section will specifically introduce appropriate approaches to time-series classification, and concludes that widely accepted decision trees are most suitable as a baseline classifier. The basic principles and several issues of classification trees will be outlined. Third, approaches to classification assessment will be introduced. The final section indicates the lack of suitable methods for fuzzy classification accuracy assessment.

4.1.1 Classification Procedures for Time Series

In order to provide a sound background on classification of high-dimensional data, this section reviews several classification procedures which have been applied to time series in previous studies. The sections comprise unsupervised approaches, statistical classifiers, linear spectral unmixing, artificial neural networks, and decision trees. It discusses the advantages

and disadvantages with particular emphasis on time-series classification, hence high-dimensional input datasets and fractional cover estimation for large regions. Readers familiar with the above mentioned methods may proceed with 4.1.2 Background on Decision Trees. Table 4.1 provides a brief overview of the classification approaches.

Table 4.1: Advantages and disadvantages of classification algorithms suitable for time series.

Method	Advantages	Disadvantages	Applications
Unsupervised approaches	easy to generate, regionally detailed mapping	manual assignment to thematic classes, potential of global inconsistency	IGBP (Loveland et al. 2000)
Statistical approaches	well established techniques	requires normally distributed data, unsuitable for data with high dimensionality	Global land cover (DeFries and Townshend 1994, DeFries et al. 1995)
Linear spectral analysis	derives truly fractional estimates, very accurate for regional studies	cannot distinguish many classes, not applicable for automated large-scale and global mapping	Global tree cover (DeFries et al. 1999, DeFries et al. 2000)
Artificial neural networks	automated, robust, accounts for multimodal frequency distributions	black box, classification process difficult to interpret, fine tuning required	Global land cover (Gopal et al. 1999)
Decision trees	automated, robust, accounts for multimodal frequency distributions, easy biophysical interpretation of classification process	bias to larger sample sets, pruning required	Global land and tree cover (DeFries et al. 1998, Hansen et al. 2000, Hansen et al. 2003, Friedl et al. 2002)

Unsupervised Classification

Unsupervised classification often employs techniques of cluster analysis. The techniques are grouped in hierarchical and partitioning approaches. Hierarchical methods do not require a previous definition of the number of classes and are further divided in agglomerative (bottom-up or ascending) and divisive (top-down or descending) algorithms. Agglomerative techniques initially consider each element as one cluster and continuously merge clusters. This approach is computationally expensive and therefore not suitable for remotely sensed data with high dimensionality. Divisive algorithms regard the entire dataset as one cluster and split it into more homogeneous groups (Richards and Jia 2006). Suitable splitting and stopping rules are difficult to define for hierarchical algorithms.

Partitioning approaches, for instance the well known k-means algorithm, require the previous definition of the number of clusters to be generated. Following an initialization of all cluster centers, each pixel is assigned to the cluster with the closest distance. Next, new cluster centers are computed by the mean of all pixels belonging to this cluster. The process is repeated until a stopping criterion, e.g. a maximum number of iterations of a convergence threshold, is met (Richards and Jia 2006). A variant of k-means, ISODATA, merges and splits

clusters according to heuristics and additional parameters (Jensen 2005). A soft variant for unsupervised classification is, for instance, the fuzzy c-means classifier (Bezdek 1981). Besides statistical techniques, there are also unsupervised variants of other classification techniques, e.g. self organizing maps, the unsupervised variant of artificial neural networks (ANN; Mather 2003).

An advanced approach, suitable for automatic classification of time series (ACTS), combines the advantages of partitioning and divisive hierarchical algorithms (Viovy 2000). The algorithm does not require a defined number of classes, but starts from one cluster (divisive hierarchical). It reassigns pixels to clusters during the hierarchical process (partitioning) and is comparatively fast. Starting from a parent cluster, the dataset is recursively split in two clusters, whereas a first guess of the cluster centers is made using the difference of the lower and upper standard deviation from the cluster mean. Next, each pixel is assigned to the new clusters, and cluster centers are rearranged until some stability is reached. The splitting is rejected and the parent cluster is regarded as stabilized by *a posteriori* testing the relevance of the split.

All unsupervised classifications lack thematic context of the resulting clusters. In other words, the approaches explore the statistical dissimilarity of the data and result in (mostly) discrete clusters. Contextual knowledge supplied by the analyst is needed to assign a thematic class *a posteriori*. Although standard unsupervised classification approaches have been used for the global land-cover classification of time series (IGBP, GLCC; Loveland et al. 2000), this process can hardly be automated. Besides data exploration, hybrid approaches combine the advantages of unsupervised and supervised classification (Crews-Meyer et al. 2004). In this study, though, the availability of training data for image classification is assumed. Therefore supervised approaches seem to be more appropriate to automated land-cover mapping and will be explored in the following sections.

Supervised Classification

Supervised approaches range from simple threshold techniques to complex machine learning algorithms. Advanced approaches also incorporate additional data and contextual knowledge. The main requirement for supervised approaches is *a priori* training data, also known as learning sets in other scientific fields. They should be representative for all aspects of the dataset to be classified (Mather 2003). A classification scheme is inherently related to training data, but may be adjustable. Often, hierarchical classification schemes are useful for combining classes for other purposes *a posteriori* or applying subsequent splitting from general to detailed classes.

An advanced methodology for classification schemes is the land-cover classification system (LCCS), a standard which focuses on global transferability and harmonization of land-cover products (Di Gregorio 2005, Herold et al. 2006a). LCCS has not been used in this classification, because regional land-cover datasets were an important input to derive the set of classes and define training areas and fuzzy reference maps by multi-scale analysis. The focus of this study was not extensive, field-based sampling and class definition, but rather the development of an approach to rapid, fractional land-cover classification for regular updates. Therefore, the approach builds on existing classification schemes, which do not necessarily correspond to LCCS.

In this context, the following sections provide an overview of suitable techniques, and briefly discuss their advantages and disadvantages with a particular focus on time series. This excludes some artificial intelligence models, e.g. linear discriminant analysis and support vector machines, because they are computationally too expensive or immature for large-scale applications of high-dimensional datasets.

The goal of all supervised algorithms is the definition of appropriate thresholds for class discrimination or fractional estimates in fuzzy classification. In case of a hard classification, a simple threshold definition by the analyst is easiest. Of course, this approach is not feasible for automated and repeated processing. Furthermore, thresholds are difficult to define manually for high-dimensional data such as time series.

Statistical Approaches

Simple statistical algorithms use distance measures such as Euclidian and Mahalanobis distance. While the first simply assigns a pixel to the closest class-center by the mean of the training data, the latter also considers the variance. This concept can be extended to probability models, e.g. maximum likelihood classification (MLC) or Bayesian classifiers, using *a priori* probabilities for each class (Jensen 2005). A fuzzy MLC has been developed by Wang (1990).

The high dimensionality and the comparatively slow processing make statistical and stochastic approaches unsuitable for time-series classification. In particular the curse of dimensionality limits the applicability of these techniques to multi-spectral images (Datcu et al. 1998). The high correlation of time series would considerably limit the results of most statistical approaches. Furthermore, most statistical classification approaches require normally distributed data, a rare case in almost all datasets. Difficulties with non-normal distributions increase for the mapping of large areas, where one class consists of multi-modal frequency distributions, leading to different cluster centers (Friedl et al. 2002).

One of the first global land-cover maps was generated with AVHRR NDVI time series at 1˚ resolution using MLC (DeFries and Townshend 1994). The significant difficulty with correlation was partly compensated by deriving metrics. Metrics are basic statistics such as mean or standard deviation derived from a high-dimensional dataset. Another global land-cover map from 8 km AVHRR spectral data and NDVI metrics was produced with MLC (DeFries et al. 1995). A comparison between MLC and a classification-tree approach was conducted for a global land-cover classification using AVHRR data (Hansen et al. 1996). The study showed marginally better accuracies for the tree approach. However, the clear structure of the tree was advantageous for biophysical interpretation. Today, recent time-series classifications are conducted with a much higher number of features, where statistical approaches tend to include also less relevant or even corrupting information into the classification process. This effect has been proven by Friedl et al. (1999) and Gopal et al. (1999) for the global 1˚ spatial resolution AVHRR NDVI dataset. The studies concluded clearly higher accuracies for boosted classification trees (up to 96 %) and ANN (85 %) than for MLC (78 %).

Linear Spectral Unmixing

Linear spectral unmixing is a truly fractional classification approach. It has been extensively applied to hyperspectral data, and often to multi-spectral images. The basic idea of the so-called linear spectral mixture analysis (SMA) is that each pixel consists of a representative sum of fractional endmember spectra. The model is solved by an over-determined set of linear equations, i.e. the number of classes must not exceed the dimensionality, minimizing the error by least-squares fitting. An endmember is the purest spectral signature of one cover type. Endmembers are selected from an existing library, field spectroscopy, or appropriate pixels in the image. Often, three surface types, e.g. green vegetation, barren land, and water (or shadow) are selected (Roberts et al. 1993). Advanced models such as multiple endmember spectral mixture analysis (MESMA; Dennison and Roberts 2003a, Roberts et al. 1998) consider multiple endmembers among these three surface types.

Linear spectral unmixing is widely applied to hyperspectral imagery (Asner and Heidebrecht 2003, Roberts et al. 1993, Roberts et al. 2003). In addition, it has been successfully employed for fractional mapping of coarse-resolution data. For instance, Kressler and Steinnocher (1999) used SMA for land-cover change studies with AVHRR data. High accuracy fractional snow cover estimates have been mapped for forests in Norway using MODIS (Vikhamar and Solberg 2003). Multi-spectral MODIS images were also employed for unmixing of North African landforms (Ballantine et al. 2005).

Since unmixing is based on the assumption that endmembers indicate the extremes in feature space, most applications are limited to single day images. Dennison and Roberts (2003b) demonstrate the effects of vegetation phenology on endmember selection, and showed substantial effects on MESMA results. Some studies applied linear spectral unmixing to time series (Lobell and Asner 2004). Although the results are promising, several issues limit temporal unmixing. For instance, one class can exhibit a high intra-annual variability in the temporal development. Multi-year analysis is limited due to different vegetation growth patterns, depending on the climatic variability and land management. Furthermore, the temporal development of classes can be too similar, which makes unmixing an inappropriate technique (Lobell and Asner 2004).

In some studies, global continuous fields of tree, herbaceous, and barren have been mapped using SMA (DeFries et al. 1999, DeFries et al. 2000). In contrast to Lobell and Asner (2004), large-scale and global classifications are based on time-series metrics which limit the influence of temporal variability. However, linear spectral unmixing has several disadvantages, because the process can only be partly automated. For instance, unmixing has to be adjusted to regions and surface cover types. Furthermore, endmember selection has to be most accurate, and is unlikely to cover the extremes in all dimensions for a large area. This usually limits linear spectral unmixing to small-scale applications, which, if carried out with necessary diligence, are very accurate. As robust alternatives, machine learning techniques such as trees and artificial neural networks have been suggested for global land-cover and continuous fields mapping (DeFries et al. 1997; DeFries et al. 1998, Friedl et al. 2000, Friedl et al. 2002, Gopal et al. 1999, Hansen et al. 2000, Hansen et al. 2002b).

Artificial Neural Networks

During the last 15 years, artificial neural networks (ANN) have been increasingly employed for image classification. This technique borrows from the human thinking process, where interconnected neurons weight incoming information to attain a specific goal. In terms of image classification, the input is a multi-dimensional training dataset, and the goal is a set of classes. The network usually consists of three layers: input, hidden, and output. Sigmoid activation functions are frequently used for a transfer to other neurons and indicate the non-linearity of the approach (Richards and Jia 2006). Learning is a critical operation, and back-propagation is the most common technique for weight modification. The error between the actual and target output is computed, and weights are adjusted, e.g. by using the gradient descent technique (Mather 2003). This process is iterated until the error is below a specific threshold.

ANNs have been successfully applied to all kinds of satellite data (Atkinson and Tatnall 1997). Many studies have stated that ANNs yield higher accuracies compared to statistical approaches (Atkinson and Tatnall 1997, Foody et al. 1995, Gopal et al. 1999, Murthy et al. 2003). As for all supervised classification approaches, the accuracy of the resulting map is highly dependent on the size and characteristics of the training and test data, as well as the dimensionality of the input data for feature separation. However, the number of hidden layers and neurons does not have a significant impact on map accuracy (Foody and Arora 1997).

ANNs have several advantages compared to classical approaches, e.g. its non-parametric fashion, the computation of one neuron simulating a multivariate linear regression, adaptive and objective learning, tolerance to some noise in the training data, and flexibility in input data types (Jensen 2005, Mather 2003). However, disadvantages of ANNs comprise substantial training time and a complex architecture, which yield a black box. The frequently employed gradient descent technique must not lead to a global minimum (Mather 2003). Furthermore, random initial weights significantly influence the classification result (Skidmore et al. 1997). Depending on the number of neurons in the hidden layer and the number of iterations, ANNs have a potential for under- and over-training, which can yield too general or too inflexible structures, respectively. This becomes evident if the network is applied to the dataset to be classified (Atkinson and Tatnall 1997, Mather 2003).

Fuzzy ANNs have also been applied to coarse-resolution AVHRR data (Atkinson et al. 1997) and time series (Gopal et al. 1999). In particular the fuzzy Adaptive Resonance Theory models (ARTMAP, Carpenter et al. 1992) have been used for fractional classifications (Carpenter et al. 1997, Carpenter et al. 1999, Muchoney and Strahler 2002, Pax-Lenney et al. 2001). Adaptive resonance theory is a highly sophisticated ANN with extensive learning capacities. Fuzzy ARTMAP was anticipated as a potential classifier for the annual global MODIS land-cover product (Borak and Strahler 1999, Gopal et al. 1999). The eventual selection of a tree approach as MODIS land-cover classifier was made for several practical reasons and yielded similar classification accuracies (Friedl et al. 1999, Friedl et al. 2002).

Decision Trees

The idea behind tree approaches is splitting the data into homogeneous units. While the root contains the full variety of classes or groups, passing data through the tree will reduce the heterogeneity until, ideally, a homogeneous class or group is reached. The simplest possible decision-tree generation is the definition of thresholds by the user, often also called an expert system classifier (Jensen 2005). However, there is a wide variety of automated approaches to generate trees (discussed below), because decision trees are a well established technique in

machine learning. Trees are robust to noisy data and learn decision making expressions from the training data (Mitchell 1997). One basic distinction regards the output. While trees with discrete, nominal results are termed classification trees, so-called regression trees split data into a set of continuous values.

Decision-tree approaches are well established in remote sensing. Applications to multi-spectral data showed good performance compared to standard statistical techniques such as MLC (Rogan et al. 2002). Comparisons to other machine learning approaches such as ANN resulted in similar accuracies (Pal 2006). Burnt area mapping in northern Australia using SPOT VEGETATION data yielded good accuracies (Stroppiana et al. 2003). The use of decision trees in hyperspectral applications has been demonstrated for agricultural fields (Goel et al. 2003, Yang et al. 2003). A thorough comparison between decision trees, MLC, and ANN for land-cover classification indicated the superior performance of tree approaches for multi-spectral data (Pal and Mather 2003).

Decision trees have been extensively applied to time series, in particular for large scale and global mapping. In order to reduce the dimensionality and data noise, metrics such as mean, standard deviation, minimum, maximum, or range are commonly computed (DeFries et al. 1995, DeFries et al. 1998, Hansen et al. 2000, Reed et al. 1994). Large-scale continental to global studies have demonstrated the capabilities and high accuracies of decision trees for MODIS and AVHRR time series (DeFries et al. 1997, DeFries et al. 1998, Friedl et al. 2002, Hansen et al. 2003, McIver and Friedl 2001). A distinction should be made between maps with discrete classes using classification trees, e.g. from 1 and 8 km AVHRR (DeFries et al. 1998, Hansen et al. 2000) or MODIS (Friedl et al. 2002) and continuous outputs. Forest cover percentage, for instance, is often derived by regression trees using 1 and 8 km AVHRR (DeFries et al. 1997, Hansen et al. 2002b) and MODIS (Hansen et al. 2003). Applications of forest cover fractions are important for global-change research and carbon budgeting (Hansen and DeFries 2004), and can be useful for global forest inventories such as FRA. Land-cover maps, produced by decision trees, are employed for estimating biophysical variables such as LAI, FPAR, and NPP (Chen et al. 2004b, Lotsch et al. 2003). The decision-tree algorithms seemed most appropriate for automated updating of global land-cover maps from annual MODIS time series (Friedl and Brodley 1997, Friedl et al. 1999, Friedl et al. 2002, McIver and Friedl 2001, McIver and Friedl 2002). Regional applications are provided by Conrad (2006) for crop classification in an irrigation system in Uzbekistan.

Because a considerable majority of the literature regards decision-tree approaches as a good, if not the best approach to automated large-scale mapping of time series, a tree approach has been selected in this study. Therefore, the next section provides an in-depth description and important considerations when using decision trees.

4.1.2 Background on Decision Trees

The following paragraphs provide important background information on the decision-tree generation process. First, some general advantages and disadvantages of decision trees will be discussed. Next, a simplified example will show how decision-tree generation recursively splits data into more homogeneous groups. It should be noted that the entire discussion focuses on binary classification trees following the Classification and Regression Tree (CART) principle. Important considerations on data sampling, splitting rules, and the termination of tree generation will be explained. Another section will briefly highlight milestones in decision-tree algorithm development and tree methods different from CART. Finally, state-of-the-art meta-approaches, often used together with tree classifiers, will be introduced.

A General Note on Decision Trees

This section aims at general issues of decision-tree generation, highlights the advantages and disadvantages, and discusses concerns of the tree architecture. Beforehand a few terms should be introduced (Figure 4.2). In many ways, trees in computer science are similar to trees in nature. However, the tree is always depicted upside down, i.e. the origin, the root of the tree, is at the top. According to a splitting rule, the dataset is divided into more homogeneous groups. Each split contains a condition of the form “if variable is lower than threshold-value, then propagate to left sub-tree, otherwise to right sub-tree.” In the case of Figure 4.2 the dataset is always divided into two subsets, a special case known as binary decision tree. Splits are also known as nodes or inner nodes. A connection to another node is called a branch or link. Ultimately, splitting terminates at a leaf, also denoted as a terminal node. Each leaf contains a result, which, depending on the type of tree, is either a class or a value. Therefore, nodes of the tree describe a set of questions, being “yes or no” questions in the case of binary trees, which eventually yield a result.

Two types of decision trees are known, with different outputs. Leaves of classification trees contain discrete classes, while values are stored in the terminal nodes of regression trees. Besides this external difference, the decision-tree generation process of regression trees usually employs some kind of variance measure. However, the structure and concept between both types is rather similar, i.e. dividing the dataset into more homogeneous (classification tree) or less varying (regression tree) groups. In the following, the discussion will concentrate on the classification-tree concept, as employed in this study.

In contrast to several statistical classification procedures, trees do not have any prerequisites on the distribution of the data. This non-parametric characteristic is particularly important for classifying data with high dimensionality. Secondly, several leaves can lead to the same target class, which indicates multiple cluster centers in feature space. This consideration of multi-modal frequency distributions of each class is especially important for classifications of large areas, because each thematic class can have multiple characteristics. In particular semi-arid environments, being highly dependent on precipitation varying considerably in time and space, require classifiers which can handle multi-modal distributions in feature space.

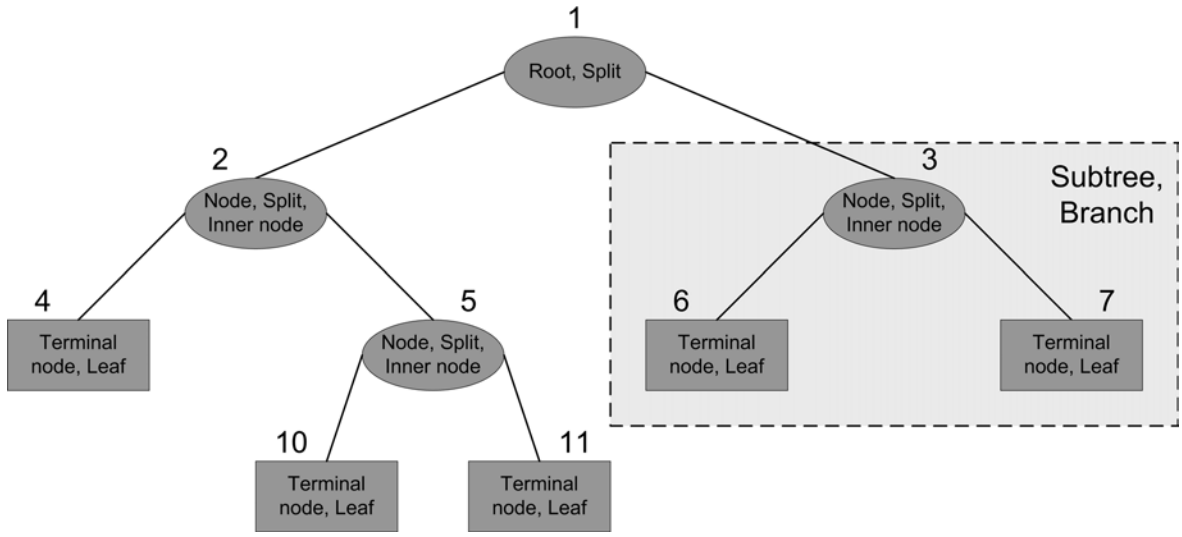


Figure 4.2: Structure and elements of a decision tree.
Note: The number indicates the identifier for each node and leaf.

Another advantage is the inherent hierarchy in decision making, depicted in the dendrogram, i.e. the structure of a tree as outlined in Figure 4.2. This often allows for a biophysical interpretation of the classification process. For instance, when using NDVI data for land-cover classification, the root often contains a split which separates vegetation from non-vegetated areas. The following hierarchies divide non-vegetated areas in water, barren, and urban classes. The other sub-tree splits vegetated units into regions with permanently green cover and agricultural areas with alternating barren and vegetated surfaces.

Trees predict the class, membership, or value by recursive partitioning of the training data into more homogeneous groups. In other words, training data are used to generate a decision tree, which becomes the decision making process and is applied to the entire dataset. Therefore, the training data should be selected carefully, just as for every other supervised classifier. Strategies for training data sampling are discussed below. A potential disadvantage of trees concerns classes with more samples. These classes can become overemphasized in a decision tree, because each split computes the probability according to the samples at its node (Hansen et al. 2000).

The principle of decision trees, recursive division of data into more homogeneous subsets, always computes the local optimum solution at each split, i.e. looking only one step ahead (Duda et al. 2001). In other words, at each split the best variable and the best threshold value are selected to divide the dataset. However, if a split decision is made, it cannot be rejected at a lower hierarchy, even if a lower level split indicates that another splitting rule of the ancestors might result in a better solution (Mitchell 1997). In computer science this local optimum solution is known as the greedy principle (Duda et al. 2001). Although the best overall solution might not be reached with greedy algorithms, strongly misleading decisions are rarely made. Furthermore, backtracking algorithms, which revise made decisions and lead to the overall optimum, are computationally extremely expensive, and therefore not feasible for most applications. In addition, the overall optimum solution might be too constrained to the training data and the tree becomes not useful when applied to the actual dataset.

Trees allow simple navigation and inherently contain a hierarchy of the decision making process. Although the level at which a decision is made does not seem to be important, the order is still relevant, because a previous split may not direct the data to a particular node. The connecting node at the preceding level is often referred to as the parent node or ancestor. Connected nodes or leaves of the subsequent lower hierarchy are called sons or descendants. In case of binary trees, addressing nodes between levels is very easy (Figure 4.2). Let the root be denoted as node 1 and its sons node 2 and 3. The left son, node 2, splits into node 4 and 5. Hence, the descendants of the next level are simply retrieved by:

$$\begin{aligned} left\ son &= 2 \cdot node\ number \\ right\ son &= 2 \cdot node\ number + 1 \end{aligned} \qquad \text{Equation 4.1}$$

Addressing in the reverse direction is computed with inverse operations. Furthermore, the level in the hierarchy is computed by

$$level = \lfloor \log_2 (node\ number) \rfloor \qquad \text{Equation 4.2}$$

A Simple Decision Tree Example

A simple example illustrates how trees split heterogeneous data into homogeneous units. Figure 4.3 shows a hypothetical two variables (feature, metric), three class example. The increasing homogeneity is shown by the reduction in impurity, where $\log_2$ Entropy is used as the impurity measure.

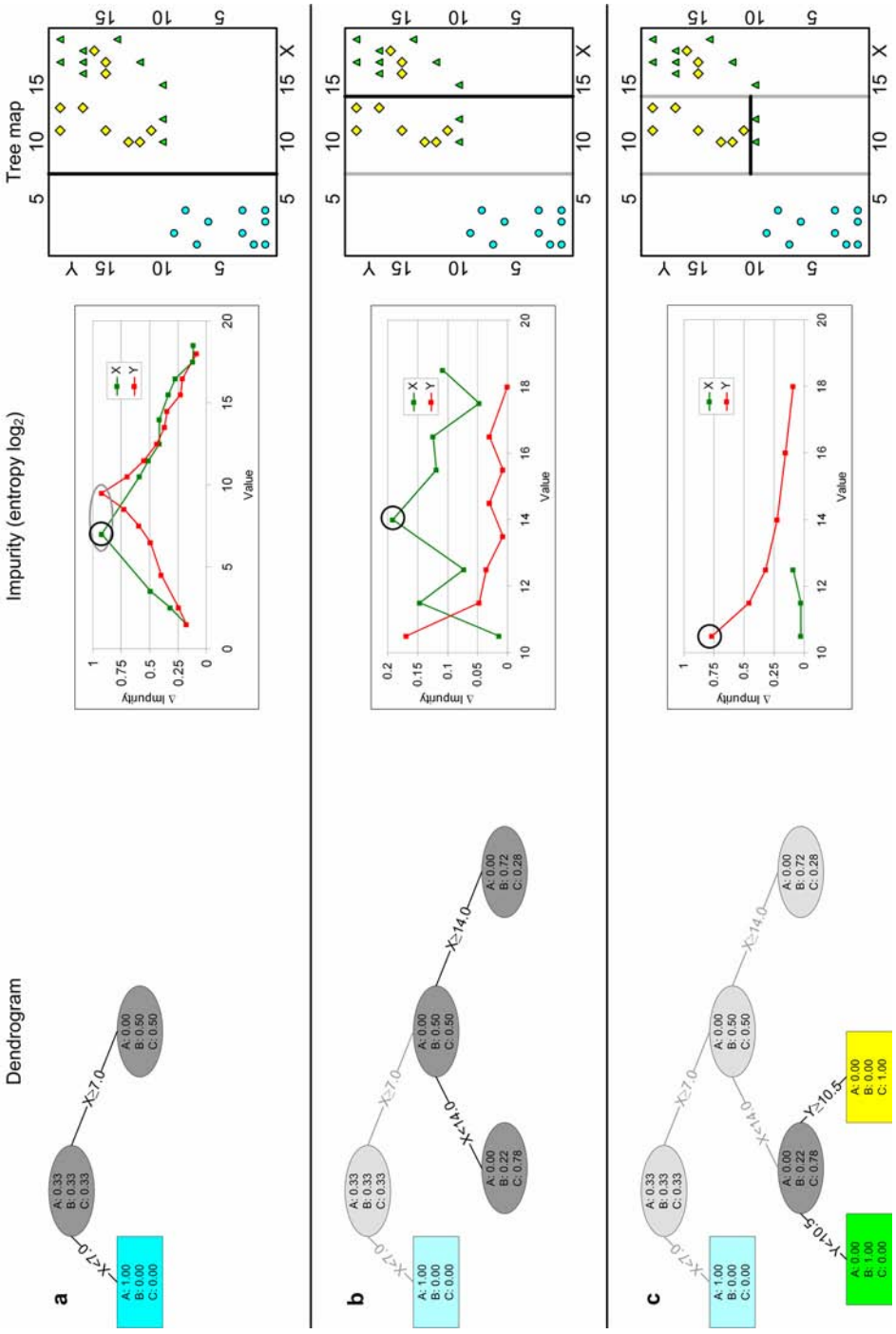
The distribution of the data in two-dimensional feature space between variables *X* and *Y* is depicted in the tree map. Classes *A* (cyan circles), *B* (green triangles), and *C* (yellow

diamonds) have ten samples each. This results in a class probability of 0.33 at the root (Figure 4.3a). The first split at the root separates all samples of class *A* from all other samples. The split variable *X* with threshold value 7 is also depicted in the tree map. However, the plot of impurities indicates identical optima for both variables, and *X* is merely selected, because it was analyzed first during processing. If *Y* would have been analyzed first, it would split the data without implications for the tree. This indicates, though, that the impurity analyzes the separation of classes, and does not consider their actual values in feature space. Looking at the tree map, splitting with variable *X* at threshold 7.0 appears much more confident than with *Y* and threshold 9.5, because the data are well separated in feature space. But again, the actual location in feature space has not been analyzed by the partitioning process. The arbitrary selection of splits with identical reduction in impurity can have implications if the tree is applied to non-training data.

Since the left node of the root split only contains samples of class *A*, this node becomes a leaf with a probability of 1.00 for class *A*. All samples with *X* greater than 7.0 are contained in the right node. Because samples of class *A* are absent, the probability of class *B* and *C* increases to 0.50. Hence, the class probability is always computed using the samples of the present node.

The next split partitions this node again using variable *X* and a threshold value of 14.0 (Figure 4.3b). The split separates many class *C* samples (left) from a majority of class *B* samples (right). The impurity plot indicates that a split of *Y* and threshold 10.5 was almost as good, which would have resulted in a leaf with three samples of class *B*. Even a split of *X* with threshold value 11.5 yielded a local optimum, and would have separated five samples of class *C* and one of class *B*. Both resulting nodes are mixed and need to be split recursively.

The left node is simply divided with variable *Y* and threshold 10.5, separating class *B* (left) from all samples of class *C* (right) (Figure 4.3c). Both nodes are homogeneous and become leaves. Figure 4.3d depicts the partition of the right branch, where variable *Y* with threshold value 16.5 was found to be optimal. This separates five samples of class *B*, making a leaf, from a mixture of class *B* and *C*. The impurity plot also indicates a local optimum with *Y* equal to 14.5. However, this split would have separated only three samples of class *B*. The lower absolute value of the reduction in impurity (0.3) is not related to the location in feature space, but to the number of samples to be separated into more homogeneous subsets. The mixed node with probabilities of 0.50 for class *B* and *C* is split into homogeneous leaves (Figure 4.3e)


Dendrogram
Impurity (entropy log2)
Tree map
a
b
c
X<7.0
X≥7.0
X<14.0
X≥14.0
Y<10.5
Y≥10.5
A: 1.00 B: 0.00 C: 0.00
A: 0.33 B: 0.33 C: 0.33
A: 0.00 B: 0.50 C: 0.50
A: 0.00 B: 0.22 C: 0.78
A: 0.00 B: 0.72 C: 0.28
A: 0.00 B: 1.00 C: 0.00
A: 0.00 B: 0.00 C: 1.00
X
Y
Value
Δ Impurity

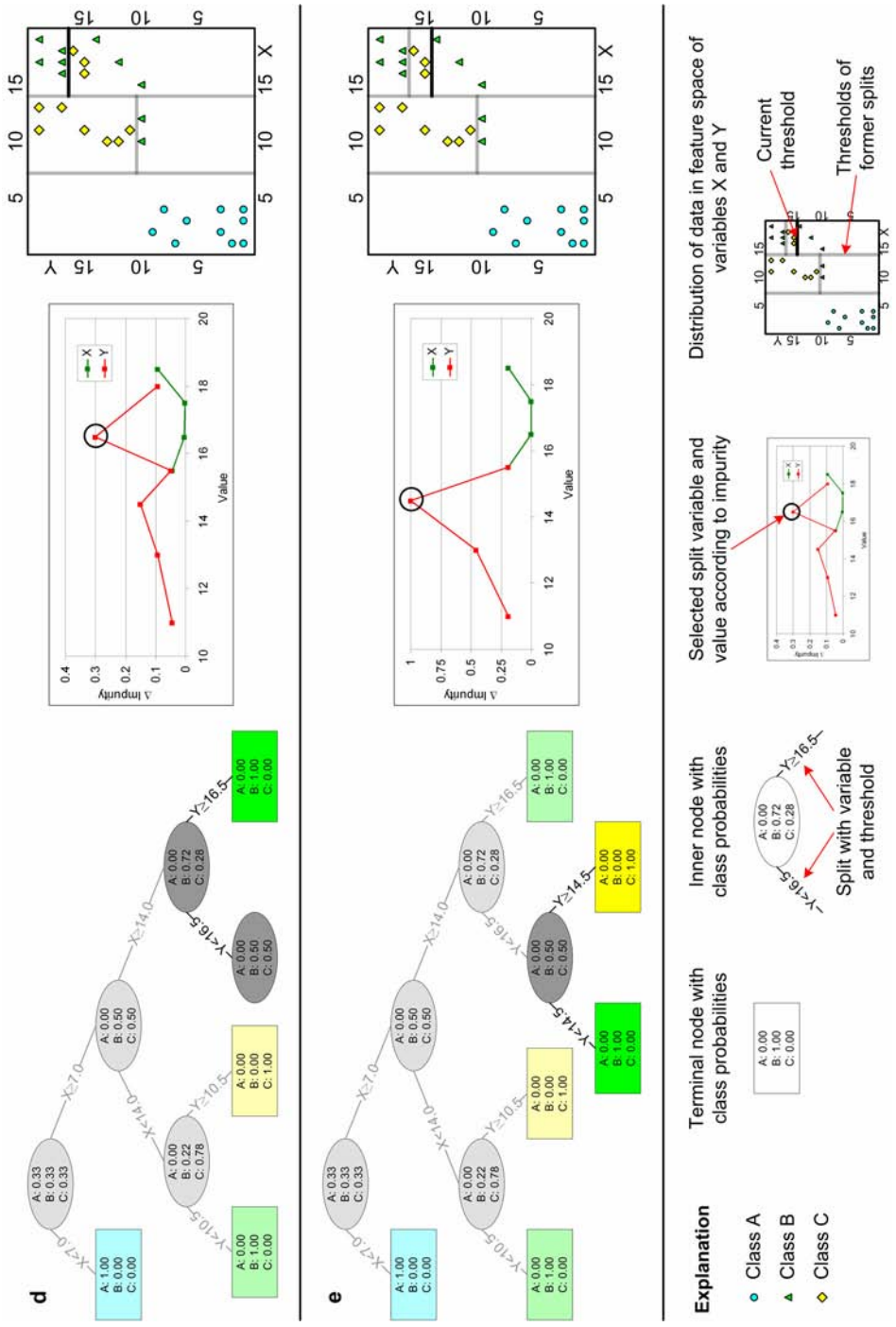


Figure 4.3: **Simplified example of the recursive partition of sample data to generate a classification tree.**
Note: **In this example, the two dimensional feature space (variables X and Y) with samples for three classes (A - cyan dots, B - green triangles, C - yellow diamonds) is depicted on the right (tree-map). The impurity of each variable using entropy $\log_2$ is shown in the middle. The variable and threshold value of the highest impurity (indicated by a black circle) is selected as splitting criterion. The resulting split is illustrated in the Dendrogram (left) and by the black line in the tree-map (right). The process is recursively applied (steps a - e) and terminated as soon as the leaves are homogeneous.**

Sampling

Any discussion of sampling has many aspects and depends on the selected classification procedure (Foody et al. 1995). For instance, some classification algorithms, including decision trees, tend to overestimate classes with a higher number of samples. Four issues are addressed in the following paragraphs. First, it is important to consider the database on which the sampling is based. Secondly, the spatial distribution of samples has to be considered in order to obtain data from all possible surface properties. Furthermore, the between-class sample size is an important consideration. A fourth issue regards splitting an overall sample set into different subsets, either to allow for training and testing, or to enable advanced approaches such as repeated and refined classification with independent samples.

Sample data are very important for image classification purposes. Often, samples refer to both, training and test data. Since they are derived at the same time, they belong to the same distribution. Field surveys are the preferable strategy for sampling. However, constraints in time and cost limit fieldwork to classification of small areas. Alternatively, other images and secondary datasets are useful for deriving sample data for large study regions. For instance, ancillary data such as other global maps have been used by DeFries and Townshend (1994) to generate a training and test set for a global classification of AVHRR data. Interpretations of high spatial resolution imagery have been used as training data for more recent global classifications (DeFries et al. 1998, Hansen et al. 2003). The STEP database has been developed for global land-cover classification (Muchoney et al. 1999), and recent efforts of CEOS-BELMANIP acknowledge the need for global validation sites (Baret et al. 2006).

However, not only the data on which sampling is based, but also the spatial sampling scheme has an influence on the classification. Data sampling in one region is most likely not representative for other areas. The issue is inherently connected to multi-modal frequency distributions of classes. For instance, grasslands in semi-arid regions have different phenological cycles, because green-up is dependent on the onset and intensity of precipitation. When mapping larger areas, grassland can expose different clusters in feature space, which have to be represented in the training data. Therefore, it is critical that samples be spatially distributed throughout the study region. This issue is addressed for global classifications, e.g. the STEP database (Muchoney et al. 1999), or for interpretations of high spatial resolution images (DeFries et al. 1998). It might be argued whether sampling needs to be equally distributed in space, which, in turn, can result in skewed distributions among classes (see below). However, clear undersampling of a unique region will most likely lead to insufficient representation in the classification process, and eventually to a lower map accuracy.

The third issue, the sample size of each class, is important for classification trees, since it is often inherently related to the spatial distribution of samples (see above; Foody et al. 1995). It

has been noted before that trees can be biased towards classes with larger samples sets. Generally, there are three strategies: random, random stratified, and equalized sampling (Jensen 2005). Random sampling might appear to be the most honest approach, but it underestimates classes with smaller spatial cover. Random stratified sampling ensures a minimum number of samples for each class, and equalized sampling derives the same number of samples for each class. Considering the potential bias of trees towards classes with larger samples sets, equalized sampling seems appropriate.

In addition to the spatial distribution of samples and weighted samples size per class, sub-sampling into separate sets is necessary for training, testing, and potentially pruning. Of course, the properties of the overall sample set should be maintained, i.e. each sub-sample set should be spatially distributed with similar intra-class relationships. One strategy involves the division of the entire sample set into two or three subsets. For instance, Hansen and DeFries (2004) used 50% of the samples for training and the other half for pruning. However, subsets for testing and pruning can be smaller than the actual training set. Another global land-cover classification employed independent sets of 60%, 20%, and 20% for training, pruning, and testing, respectively (DeFries et al. 1998).

Some applications such as random forest (see below) require splitting the overall sample set into *k* subsets. If the overall sample set is large, *k* sub-samples are drawn, again ensuring the properties of the overall sample set. However, deriving independent sub-sample sets is often impossible due to scarcity of the overall set. In this case, *k*-fold cross-validation becomes an alternative. The entire sample set is split into *k* subsets, where *k* should not be too high (e.g. set to 10). Next, *k*-1 subsets are merged and used for training, while the other, independent set is used for testing or pruning (the procedure is known as: “leave one out cross-validation”). The merging process is iterated with differing remaining subsets for testing. The idea underlying cross-validation assumes that the different iterations are independent, which is actually not true, because *k*-2 subsets are always used for training two subsequent iterations. Still, *k*-fold cross-validation is the only way to strategically sub-sample small overall sets, and will be discussed later in comparison with bootstrapping and advanced meta-strategies.

Splitting

The following notations are necessary for the next paragraphs. For detailed definitions see Equation A.1- Equation A.7 in the appendix.

i...impurity
s...split
t...node
j...class
l...left subtree
r...right subtree
$p(t)$*...probability that a sample belongs to node t*
$p(j \mid t)$*...probability that a sample which is in node t belongs to class j*

The goal of each split is to reduce impurity. Expressed in other words, the sons of each node should contain a more homogeneous set of samples. Ideally, the set becomes completely pure, i.e. all samples belong to only one class (Figure 4.3a). This terminates the recursive partitioning process and makes a leaf (see next section for terminating decision trees). However, this does not imply that each split necessarily separates a homogeneous portion of data, as indicated in Figure 4.3b, but considers the overall reduction in impurity for this particular node.

Expressed in more mathematical terms, the reduction in impurity is computed with the following equation:

$$\Delta i(s,t) = i(t) - p_l(t) \cdot i(t_l) - p_r(t) \cdot i(t_r) \qquad \text{Equation 4.3}$$

where $\Delta i(s,t)$ is calculated for each potential split. A potential split is estimated by selecting one variable from feature space and sorting all existing values. Moving from low to high values, the first potential threshold is placed between the lowest and the second lowest value and computing Equation 4.3. Next, Equation 4.3 is calculated with a threshold set between the second and third lowest values. The computation is repeated for all potential threshold values and variables. The highest $\Delta i(s,t)$ is selected and indicates the split with the corresponding variable and threshold value. Expressed in mathematical terms, the reduction in impurity is maximized.

Several impurity measures exist in the literature on information theory. Maybe the most common measure is the Entropy, in other fields also known as Shannon index (Breiman et al. 1984, Duda et al. 2001, Mitchell 1997, Quinlan 1993) computed by:

$$i(t) = -\sum_j p(j \mid t) \log[p(j \mid t)] \qquad \text{Equation 4.4}$$

The choice of the logarithm function is free for all impurity computations, i.e. there are no implications for the selection of the splitting rule between *ln*, log_2, or log_x functions. However, the impurity value itself, as plotted in Figure 4.3, becomes different. As shown in Equation 4.3, the impurity needs to be computed for each potential split with a new assemblage of the data in the left and right sub-tree. A similar measure, the Deviance (Clark and Pergibon 1992, Venables and Ripley 2002), is frequently employed for splitting remote-sensing data (DeFries et al. 1998, Hansen et al. 2000).

Another widely applied impurity function is the Gini impurity (Breiman et al. 1984, Venables and Ripley 2002):

$$i(t) = 1 - \sum_j p(j \mid t)^2 , \qquad \text{Equation 4.5}$$

The Gini impurity is a generalization of the two-class variance impurity, and sometimes multiplied with a constant 0.5 (Duda et al. 2001). A third commonly used impurity measure for multi-class binary splits is the Towing criterion (Breiman et al. 1984, Duda et al. 2001):

$$\Delta i(s,t) = \frac{p_l(t) \cdot p_r(t)}{4} \left[\left| \sum_j p(j \mid t_l) - p(j \mid t_r) \right| \right]^2 \qquad \text{Equation 4.6}$$

It has to be noted that towing directly computes the reduction in impurity, i.e. Equation 4.3 is not needed. This criterion attempts to split a set of data into two super-classes. The split is the best division where the super-classes are most dissimilar. At the top of the tree, Twoing attempts to group large numbers of classes which are similar, often resulting in equally sized groups. At the bottom the criterion separates single classes (Breiman et al. 1984, Duda et al. 2001).

Similar to the Entropy, Towing results in fairly balanced trees. In contrast, the Gini tends to separate homogeneous groups close to the root, i.e. a sub-tree with few but rather homogeneous data, and another branch with all other, heterogeneous samples. This often leads to unbalanced trees and can require more computation time, because recursive partitioning has to be applied to larger samples sets at lower hierarchies.

Stopping and Pruning

The splitting process as described above has to be terminated. Obviously, the tree generation ends with one sample per node. This ultimate termination, however, is neither necessary nor useful and often leads to lower classification accuracies if applied to an actual dataset. There are two basic strategies for the determination of an appropriate tree size: stopping and pruning. While stopping the spitting process occurs during tree generation, pruning requires grown trees which are cut. Often, stopping is seen as a top-down and pruning as a bottom-up approach, albeit pruning does not necessarily follow typical bottom-up processes. Another view regards stopping as an *a priori* technique performed during tree growth, while pruning is *a posteriori* (Duda et al. 2001).

Stopping follows specific sets of rules. Besides the obvious stopping at one sample per node, three stopping rules are commonly used for decision trees; further stopping rules are shown in Duda et al. (2001):

- Pure nodes
- Minimum samples
- Low impurities

It is, of course, not necessary to grow each tree to leaves with one sample each. Instead, splitting is terminated if a resulting branch is pure, i.e. this node only contains samples of one class. In this case the node is defined as a leaf and identifies this class with a membership of 100 %. This pure node stopping rule is applied to almost all decision trees and is also illustrated in Figure 4.3.

The minimum sample rule requests an input from the analyst of how many samples are necessary for further splitting. Usually, two values can be provided, minimum samples per node and minimum samples per leaf. If the samples in a node are lower than the specified value, tree growth is stopped, even if the branch is not pure. The resulting node is defined as a leaf, and the class membership is stored proportionally with the likeliest class stored for discrete outputs (see Equation A.8). Similarly, a rule for minimum samples per leaf stops the tree growth at a particular branch, if the new branch contains fewer samples than specified. It is obvious that the minimum node samples must be at least twice as many as the minimum leaf samples. Sometimes the rule of minimum samples per leaf is implemented slightly differently, and ensures that each leaf contains a minimum number of samples. This implementation, however, shifts the split to a less optimal reduction in impurity.

Tree growth is stopped if the maximum reduction of impurity becomes lower then a user defined cut-off value. This criterion is highly sensitive to the impurity measure and distribution of samples in each node (see Figure 4.3). As an alternative, a ratio indicating the deviation from the root node is frequently used, e.g. tree growth is stopped when a fraction of a node impurity value is lower than the root impurity. This idea can be extended to parent nodes or other ancestors.

Pruning is an alternative for generating right-sized trees. Again, a decision tree which fully fits to the training data is not flexible enough if used for classification. Therefore, a more general, hence shallower tree seems more useful and often leads to higher classification accuracies. Pruning aims at finding this optimal tree. There is a multitude of opportunities for pruning, where all approaches have pros and cons. All techniques require previously grown trees which extend beyond the optimal size. However, it is not necessary to derive fully grown trees, i.e. often conservative stopping can be applied. The idea of pruning is to cut insignificant of even error-prone sub-trees, defining new leaves for a shallower tree. Here, two common strategies are briefly described:

- Cost complexity (Breiman et al. 1984)
- Error-based pruning (Quinlan 1993)

Notations (for additional computations see also Equation A.9 - Equation A.11 in appendix):

$M(T)$...*misclassification error for tree* T
α...*complexity*
$\tilde{T}$...*number of leaves*

The general idea of cost complexity pruning is to find a rooted sub-tree which is not too complex and has no significant misclassification costs. This follows the basic computation (Breiman et al. 1984):

$$M_\alpha(T) = M(T) + \alpha \cdot \tilde{T} \qquad \text{Equation 4.7}$$

Usually, several values of α are tested and the remaining number of leaves is plotted against an error indicator or tree impurity (for computing error indicators see Equation A.9 - Equation A.11; Clark and Pergibon 1992, Venables and Ripley 2002). The misclassification error, estimating the error rate from the training data, is a first estimate of tree accuracy. It is modified by the number of leaves and the complexity factor α. Of course, it will be best for a fully grown tree, but usually shows only a slightly increasing error if less significant branches are cut. Typically, there is a point, where the error drastically increases, indicating the suitable value of α. Instead of reusing training data, error measures from secondary and independent

pruning datasets or cross-validation frequently provide more accurate estimates of α (Venables and Ripley 2002). In this sense, cost complexity pruning is a top-down approach. Finally, the decision tree is cut to its optimal size.

In contrast to the use of a distinct pruning dataset for cost complexity pruning, error-based pruning only considers the training data. It computes the error probability from the upper confidence limit of the binomial distribution (Quinlan 1993). If the sum of predicted error in the leaves exceeds the predicted error rate in the parent node, the tree is pruned and the parent node becomes a leaf. The process is repeated until the summed predicted error in the leaves becomes lower than in the parent node. Therefore error-based pruning is a true bottom-up approach, i.e. it compares inversely recursively the predicted sum of errors at the leaves to the connected parent node.

It should be noted that the appropriate size of a tree is highly important for classification accuracy. Stopping and pruning aim at finding this tree. However, *a priori* settings of stopping rules are considered insufficient in many studies. It is likely to over- or under-size the tree. Furthermore, stopping rules apply general thresholds and might only be suitable for some sub-trees. Therefore stopping is regarded as a suitable option for conservative termination of tree growth, but not for finding the optimal tree. The optimal tree should be specific to the training data dataset, but still general enough for the classification of the dataset. Pruning, conversely, focuses on finding optimal trees (Duda et al. 2001). Many studies have shown the importance of pruning compared to unpruned examples. It is not just a simple cutting of unimportant or less critical splits, making the tree structure simpler; it also can remove error-prone sub-trees if applied to real data. The lack of consensus on pruning approaches should not prevent analysts from carrying out this important operation.

Other Decision Tree Algorithms

Many ideas of decision-tree algorithms originate from the field of machine learning, embedded in the science of artificial intelligence (Russell and Norvig 2003). Some other algorithms were developed by statisticians, but the general idea is quite similar: recursive splitting of a heterogeneous dataset into more homogeneous units. Because trees are also used for data mining applications, almost all approaches are capable of handling large datasets. Although the algorithms have been developed for a wide variety of data and applications, few to no adjustments are required for remote-sensing data. Besides the principles of CART, which have been explained above, the following paragraphs provide a brief overview of other algorithms and highlight the main differences.

One alternative is the CHAID algorithm (chi-square automatic interaction detectors; Kass 1980). The basic distinction to CART and other decision-tree approaches is the inherent stop before the tree becomes too large, i.e. pruning is not necessary. Furthermore, CHAID operates on categorical variables, where metric-scale variables are previously transformed into groups. The splitting is computed using Pearson's χ^2-test for independence. The variable with highest χ^2-distance is selected, which can lead to trees with more than two branches. The approach is computationally expensive.

Computational complexity can be problematic for larger datasets. CART or similar approaches have to test every variable for each potential threshold value. The Fast Algorithm for Classification Trees (FACT) is an appropriate alternative, where ANOVA is used to estimate the best variable, and linear discriminant analysis selects the best threshold value (Loh and Vanichsetakul 1988). This result in multi-way splits, i.e. the number of branches equals the number of classes. A derivative, the Quick, Unbiased, Efficient, Statistical Tree (QUEST), aims at combining the computational efficiency of FACT with binary splits, unbiased variable selection processes, and optional pruning using CART-based techniques (Loh and Shih 1997). Comparisons of tree accuracy among CART, FACT, and QUEST indicate no clear superiority (Loh and Vanichsetakul 1988, Loh and Shih 1997).

Techniques originating from machine learning science comprise the Iterative Dichotomiser 3 (ID3) algorithm and its successors C4.5 and C5.0 / See5 (Quinlan 1993). Instead of using the reduction of impurity according to Equation 4.3 such as in CART, ID3 employs the gain criterion as a measure of the effectiveness of a split. This criterion has been extended to the gain ratio to allow multi-way splits (Duda et al. 2001, Mitchell 1997). ID3 itself has been designed for nominal data, but C4.5 is also suitable for continuous variables with binary splits similar to CART. While CART handles missing data by surrogate splits, i.e. choosing the variable with the second best reduction in impurity, C4.5 computes estimates according to the relative frequency of the available data generating fractional training data. Another feature included in C4.5 is the conversion of a decision tree to a set of rules (Quinlan 1993). However, the decision tree is not simply interpreted for each leaf, i.e. having as many rules as leaves in the tree, but the rules are also analyzed for redundant conditions. Besides simpler interpretation, rules also mitigate the hierarchical interpretation inherent in decision trees, where nodes near the leaf tend to have a higher priority than at the bottom when applied for decision making (Mitchell 1997). However, rules can be disadvantageous with noisy data (Duda et al. 2001). Advancements of the successor of C4.5, C5.0 (UNIX / Linux) and See5 (Windows), comprise the speed, memory, compactness of trees, winnowing, and support for boosting.

Meta Strategies for Classification

The following paragraphs highlight some advanced strategies which can be used together with decision-tree generation. In this regard they can be seen as meta-approaches, which often increase the accuracy compared to individual classifiers. Generally, any classification algorithm could be used in conjunction with the techniques described below, but the high correlation of time series and the high dimensionality, even if using metrics, limits the individual classifier to non-parametric methods including decision trees.

Here, state-of-the-art principles such as bagging, random sub-spaces, random forest, and boosting are introduced. In some sense, bagging and random sub-spaces are regarded as bootstrapping, which borrows from the idea of cross-validation with less restrictions. Bootstrapping derives subsets of data with replacement, i.e. data can be repeatedly used in different subsets (Duda et al. 2001). The idea behind bootstrap techniques is that errors are compensated by averaging over- and under-estimations. Even though the integrated result will never be worse than an individual decision-tree classification, there is no mathematical proof that the results are necessarily better than a single classifier (Bishop 2006). Empirical studies, however, indicate the use of bootstrapping for increasing classification accuracies and the improved stability of the classification results in particular.

Bagging, a compound derived from bootstrap aggregation, generates multiple sets of sample data (Breiman 1996). It is known that a single classifier can be unstable, which is also true for decision trees, where the greedy algorithm always generates local optimal splits. A change in training data modifies the splitting rule and yields a different classification. Bagging generates several decision trees for each training set. To compute a global result, either the decision trees are integrated or their individual results are merged.

While bagging splits the samples into a distinct number of subsets, the idea of random sub-spaces aims at selecting different features (Ho 1998). This technique generates many decision trees by randomly selecting different sets of variables from an overall set. Although the variables were not selected randomly, Conrad (2006) employed the sub-space approach to remote-sensing data. In this regard the method takes advantage of highly dimensional data. Just like bagging, learning based on subsets is carried out in parallel - a plus when processing time is an issue. A general view on bagging and random sub-spaces regards both techniques as bootstrapping, either on the samples or on the variables.

Applying the above ideas, bagging and random sub-spaces, to decision trees they belong to the category of random forests (Breiman 2001), also known as random decision forests (Ho 1995). According to the definition, a random forest consists of a collection of tree-structured classifiers. Every classifier is independent of every other one (i.e. random) but originates from

the same distribution of data (being a subset) and leads to only one class (Breiman 2001). This opens the door to further randomness and merely defines a broad framework. As for bagging and random sub-spaces, a combination of the individual classifiers selects the likeliest class.

Although not necessarily fitting into bootstrapping approaches, boosting also aims at improving individual classifiers by combining several weak learners. If joined, they result in higher accuracy. Boosting, however, adds a new classifier to existing ones by reducing the classification errors of previous iterations. While bagging is performed parallel on individual training sets, boosting can only be carried out sequentially (Bishop 2006). The adaptive boosting algorithm (AdaBoost) (Freund and Schapire 1997, Freund and Schapire 1999) is one of the most employed boosting approaches and has been successfully used for time-series classification of AVHRR (Friedl et al. 1999) and MODIS (Friedl et al. 2002) data. The basic idea of AdaBoost is the iterative employment of a training set, where misclassified data receive a higher weight and will therefore be preferred in the next classification. This weight can be either directly assigned to the data or sub-sampling according to the previous weighting. The overall result is achieved by summing the individual classifiers with their relative importance (Freund and Schapire 1997, Freund and Schapire 1999).

There is a multitude of sometimes contradictory literature on comparisons of the above-described techniques. Bauer and Kohavi (1999) and Dietterich (2000) state that bagging is superior to boosting if applied to classifications with noisy patterns. In many other cases, boosting seems to perform better (Bauer and Kohavi 1999, Quinlan 1996). Ho (2002) compared bagging and random sub-spaces and found that for smaller sets bagging is superior, while highly dimensional sets are better classified with random sub-spaces. However, there are also cases where neither technique performed better than a single classifier, e.g. for highly fractional class boundaries in feature space, which would require numerous splits for accurate discrimination. In another study the random sub-space method was almost always superior to bagging, boosting, and individual trees (Ho 1998). Furthermore, Breiman (2001) stated that random forests are comparable with boosting approaches.

4.1.3 Background on Accuracy Assessment

A thorough assessment of the classification procedure is an integral part of this study. Furthermore, a sensitivity study aims at determining the most important classification parameters. Both necessitate a robust and detailed accuracy assessment protocol.

The assessment of hard and soft classification requires different approaches. The issue of appropriate techniques for soft classification assessment is only partly addressed in the literature, but will be comprehensively exploited in this study. While statistics for hard classification are well investigated and widely employed, there is a remarkable lack of measures for soft classification accuracies (Foody 1996). In this respect, fuzziness is related to uncertainty regarding map accuracy (Atkinson and Foody 2002). This section aims at reviewing selected existing techniques. A different view on fuzzy classification assessment will be described in 4.2.7 Accuracy Assessment, which also contains a description of issues related to the reference data.

A common accuracy assessment of discrete classifications employs statistical measures derived from the contingency table, also known as the error or confusion matrix (see Figure A.3 and Equation A.12 to Equation A.15). The table fulfils the special case of a quadratic matrix, i.e. it has the form of $n \times n$, with the same categories at the rows and column for classified and reference data, respectively. Class-specific error measures such as user's or producer's accuracy and general estimates such as overall accuracy or KHAT (kappa coefficient) are conventionally calculated (Congalton and Green 1999, Liu et al. 2007, Rosenfield and Fitzpatrick-Lins 1986). Foody (2002) and Stehman and Czaplewski (1998) have addressed several issues related to accuracy assessment with confusion matrices.

The easiest fuzzy classification assessment is a simple transformation of the fuzzy map into a hard classification, e.g. using the majority rule. Next, common contingency matrix statistics can be computed. This transformation, however, is inappropriate to account for the error in the actual fuzzy map. A considerable amount of information is lost by the transformation to discrete classes. Especially mixed pixels cannot be assessed with sufficient confidence, and are likely to be misclassified (Foody 1996).

A step towards fuzzy assessment using confusion matrix statistics has been suggested by Gopal and Woodcock (1994) and Woodcock and Gopal (2000). The assessment considers the uncertainty in the reference data and evaluates the magnitude, nature, frequency, and source of the error. However, the approach cannot account for fuzziness in both datasets but requires a hard classification.

The evaluation of a soft classification with the error matrix imposes several difficulties. Because an error matrix only considers two states, correct or incorrect class assignment, the common view does not suffice the need for fuzzy set estimates (Foody 1996). Even analyzing one fuzzy class estimate by grouping the membership into discrete ranges imposes serious limitations. For instance, fuzzy estimates of one class ranging from 0 to 100 % can be grouped into 20 units with 5 % intervals. A comparison with fuzzy ground truth data or a soft reference map will always yield the highest errors at merely artificial breaks of actually

continuous data, i.e. at the 5 % threshold. The magnitude of mismatch depends on the mean difference between fuzzy classification and reference, and the error distribution in the range of memberships. In this respect, accuracy calculation has to account for off-diagonal cells within the mean error range as correctly classified data. However, no study has been found on this view of the analysis of the error matrix. Considering the difficulties caused by the transformation of continuous values into discrete groups, another approach to fuzzy set classification assessment is required.

Since the contingency table is fairly unsuited to fuzzy classification assessment, other measures have to be employed to account for uncertainty. Uncertainty can be split into vagueness and ambiguity. The first is associated with impreciseness in dividing classes, while the latter is connected to one-to-many decisions. Distance measures such as Euclidian distance seem to be appropriate alternatives for fuzzy classification assessment (Klir and Folger 1988). Foody and Arora (1996) have employed mean Euclidian and Manhattan distance for fuzzy membership assessment. Another suggested measure of uncertainty is Entropy (similar to Equation 4.4), which has been utilized by McIver and Friedl (2001). It expresses the support of mutually exclusive classes, but can be disadvantageous if reference and classification have partial class membership. In this case, Cross-Entropy has been suggested (Foody 1996).

4.2 Classification Procedure

This main section will introduce the classification procedure of this study. First, the general framework will be laid out, followed by a detailed explanation of all modules. An integral part of this study is the comprehensive assessment of classification accuracy. Because a multi-scale analysis allows for generating fuzzy reference data, fuzzy accuracy assessment can be conducted. The last section will introduce a classification assessment protocol with an innovative view on fuzzy map assessment.

As noted before, essentially every supervised classification procedure, including this thesis, follows the general outline as depicted in Figure 4.1. The process for time-series classification of this study is shown in Figure 4.4. This approach is capable of mapping land-surface types when training data exist. The general term land-surface types has been chosen, because not only land cover (and land use to a lesser extent) but also vegetation types or other surface characteristics can be mapped; the terms will be used interchangeably in the following sections. The approach is flexible and single modules can be substituted by other techniques. For instance, the classification technique is not restricted to classification trees. Furthermore, training data from other sources can be ingested into the classification process. In this respect,

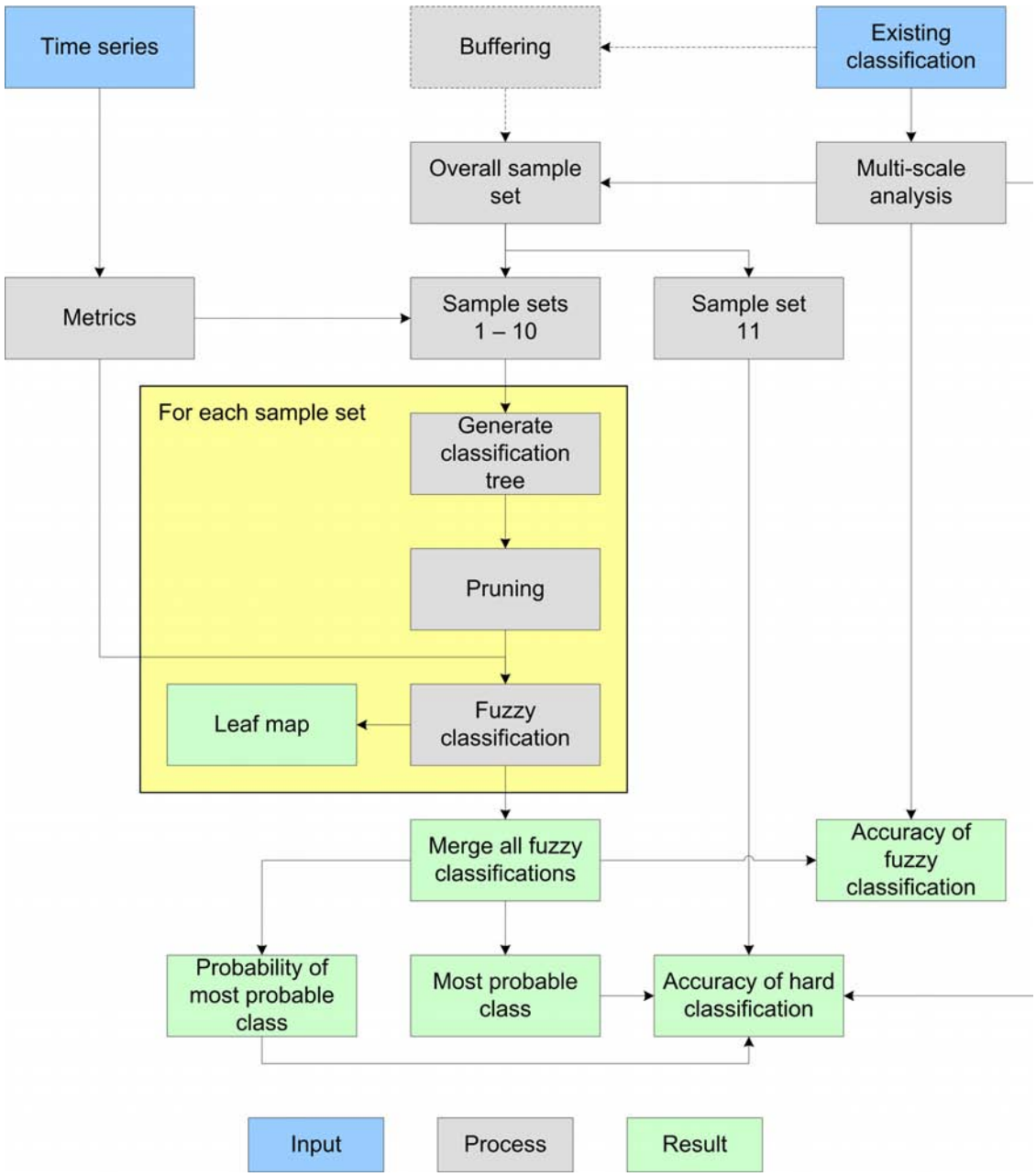


Figure 4.4: Classification procedure developed in this study.
Note: The processes are the modules of the classification (partly also intermediate results) and are discussed in detail in the text. The yellow box indicates that the enclosed processes and the resulting leaf map are generated for each individual decision tree. The results are shown in more detail in Figure 4.10.

the approach is merely a framework. The single modules developed in this study have seemed most useful to accomplish the goal of time-series classification. Some selections of modules have been made from previous studies, which have presented promising results in the scientific literature. Others have been made by the author and tested against several alternatives. The sensitivity study (4.3.2 Sensitivity Analysis of the Classification Procedure) following this main section will present the advantages and disadvantages as well as suitable parameters; hence, justifying the decisions made in specific modules.

Time series are transformed into sets of metrics, an approach which has been found suitable by almost all large-scale time-series classification studies (Figure 4.4). Sample definition, on the other hand, is more difficult. Because it is impossible to collect field samples for large areas, ancillary datasets have to be employed. In the case of land-cover studies, existing high-resolution land-cover maps are used as input. In order to select homogeneous areas for accurate training and testing, a multi-scale analysis is utilized. However, high spatial resolution data do not exist for the vegetation types of Namibia. In this case a buffer is generated along class boundaries to reduce mismatches in ecotones. Mathematical approaches ensured correct sampling of appropriate locations. For these locations the properties of all metrics are analyzed and stored in an overall sample set. In this study, this set is split into eleven subsets, where ten are used for training and the remainder for accuracy assessment. It should be noted that the number of eleven subsets could be altered depending on specific characteristics of the study area, e.g. the availability of samples. The training sets are connected with the metrics set to extract class signatures. The term signature is used interchangeably with training samples throughout this study. The approach is related to bagging, however, without replacement by bootstrapping. Ten classification trees are grown from each sample set and pruned subsequently. The resulting decision trees are used for classification. However, the class of each leaf is not important in this study. On the one hand, the identifier of each leaf is stored, since the terminal nodes indicate rather homogeneous groups. This explorative method can be interpreted similar to unsupervised classification results, and an overlay with reference data may indicate sub-units and new features. On the other hand, the membership of each class is exploited, which can also be termed a proportion or fractional class estimate, as these terms are used interchangeably in this study. This provides a first fuzzy class estimator. Next, all classifications are merged into a single dataset, again storing a fuzzy classification. This entire fuzzy classification is compared with the multi-scale input data using fuzzy error measures. Furthermore, the fuzzy classification is transformed into a crisp classification, storing the class with the highest membership and the membership itself. Hard classification accuracy was assessed with conventional statistics derived from the error matrix. In the following sections, all important processes will be explained in more detail.

4.2.1 Metrics

The derivation of metrics is an important prerequisite for classifying high-dimensional and highly correlated data. Metrics compute statistics (e.g. mean or standard deviation) or indices (e.g. ratios between bands or different periods of the time series) from complex datasets. This markedly reduces the feature space and can even derive a more meaningful, a less correlated,

and a more robust dataset for classification. An option for temporal segmentation of a time series and its hypothetical impact is illustrated in a second section.

Statistics for Metrics

There is a consensus in remote-sensing image classification that time series are transformed into a set of metrics, which become the baseline dataset for image classification. This transformation was a critical step in the past, where noisy AVHRR time series were used. However, even recent studies using MODIS data employ, at least partly, sets of metrics instead of the raw time series (DeFries et al. 1998, Friedl et al. 2002, Hansen et al. 2003). Besides the reduction of data noise, time series exhibit significant temporal autocorrelation, which prohibits the use of many classifiers (e.g. statistical methods) and also imposes difficulties to advanced techniques such as decision trees. This redundancy in feature space is clearly reduced by the calculation of metrics. A third reason for sets of metrics is the derivation of value-added information, e.g. greenness during the peak of a growing season in contrast to the dormant period. In other words, appropriately computed metrics represent the information content sufficiently well by reducing data noise (DeFries et al. 1998).

Reed et al. (1994) has suggested twelve metrics to characterize the phenological dynamics of the earth surface. DeFries et al. (1995) has investigated several simple and complicated metrics sets for time-series classification. This study using MLC has revealed that simple metrics increase the accuracy up to 10% compared to raw time-series data. Another study using time series of several reflective bands concluded that simple metrics such as mean, maximum, minimum, and amplitude are most suitable for decision trees (DeFries et al. 1998). Therefore, this study employs simple metrics including mean, maximum value, minimum value, standard deviation, and range. This set of metrics has been computed for all parameters, i.e. seven reflective bands and NDVI. Semiarid study regions have been complemented with time series of day and night LST. For some sites the number of harmonics used to explain 80% of the original variance (maximum 3) and the explained variance of the corresponding number of harmonics has been added.

A general distinction should be made between intensity and temporal metrics. Intensity metrics, as employed in this study, are derived by a statistic computed over time, for example the annual mean of the NDVI. In contrast, temporal metrics indicate a particular day when the temporal plot significantly changes. For instance, considering the NDVI, the phenological development of vegetation can be described by four defined stages: increase, plateau, senescence, and dormancy. These four stages are provided by the MODIS vegetation

phenology product (MOD12Q2). The temporal metric set can be extended by several other measures, e.g. length of the growing period or number of phenological cycles.

Most remote-sensing studies only employ intensity metrics. Initial tests of this study solely employed temporal metrics, but resulted in low classification accuracies due to considerable noise and insufficient class separability. This effect results from the above described high temporal auto-correlation of time series. Also, the combination of temporal and intensity metrics in decision trees showed that, if ever, temporal metrics were used for splitting at very low hierarchies in the tree, and therefore had a low influence on the classification.

Temporal Segmentation

Although the timing does not seem particularly useful as a defined metric of time series, it can be supportive for deriving additional intensity measures with new information content. Many studies employ sets of annual metrics, but some also incorporate ancillary information such as climate data. In addition to annual metrics, for instance, Hansen et al. (2000) use temperature data and derive metrics only using the four warmest months. The same study also considers the eight greenest months in order to eliminate snow cover for global mapping. A strategic temporal segmentation was conducted by Conrad (2006), where a year is also divided into static segments of halves and thirds. Next, a set of statistics is computed for each period, which results in a more detailed representation of the temporal dynamics. Figure 4.5 indicates the gain in information when using temporally segmented time series for a hypothetical example. Making two segments of an annual time series, some differences are revealed in the mean and standard deviation. An additional splitting into three periods highlights the seasonal patterns of this uni-modal curve. Due to this clear increase in new information, all metrics sets were derived from annual, 6-month, and 4-month time series. Furthermore, if, for instance, the year was segmented into three partitions for metrics generation, then the dataset for classification also comprised half-year and annual metrics.

An option for adaptive temporal segmentation provides the MODIS phenology product, which indicates the onset of green-up, plateau, senescence, and dormancy. The starting dates of phenological stages during the year were used to extract the beginning of vegetation increase, maximum, decrease, and minimum. Figure 4.5 (bottom) indicates the gain in information using this type of segmentation, e.g. with high mean and low standard deviations during maximum and high standard deviations during transitional seasons. Several considerations had to be accounted for, e.g. the year does not always start with an increase. Furthermore, the product contains up to two vegetation cycles, where the respective periods (e.g. the minimum in the hypothetical example of Figure 4.5) were merged into one.

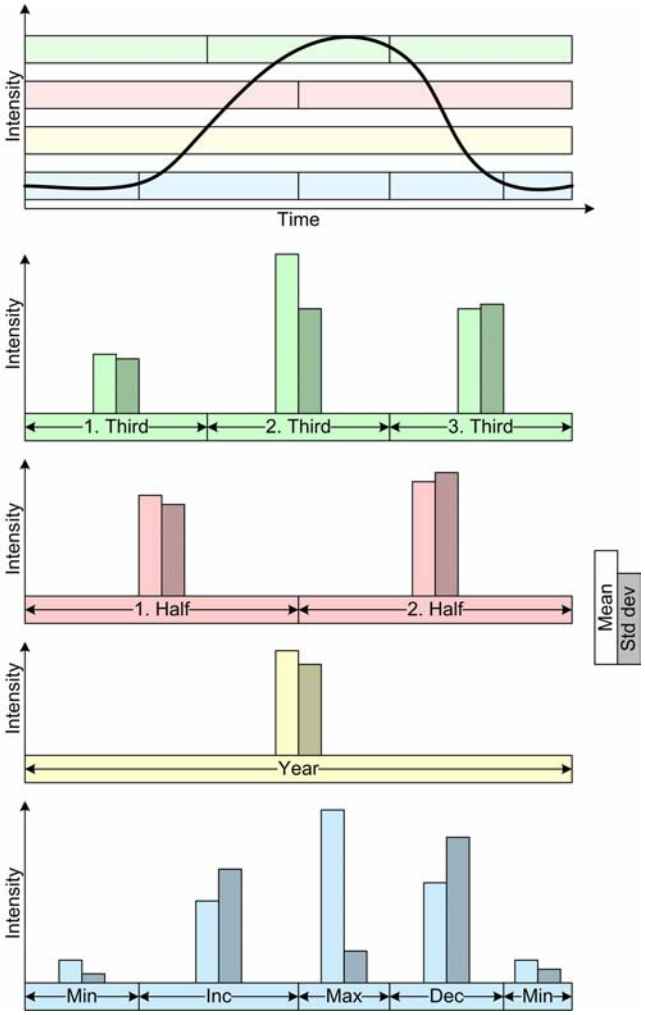


Figure 4.5: **Example for temporal segmentation and computation of selected metrics using a hypothetical uni-modal curve.**

Note: **Two metrics, mean and standard deviation, are computed over the temporal course of the each segment.. Temporal segmentation is conducted in two ways. The green, red, and yellow bars indicate partitioning the time series in static thirds, halves, or using the entire year without consideration of the dynamics. This process follows Conrad et al. (2007). The segmentation represented by the blue bar follows relevant phenological points such as dormancy (min), green-up (inc), plateau (max), and senescence (dec). These points are defined using a phenological dataset (e.g. MOD12Q2; Zhang et al. 2003; see 2.2.2 MODIS Land Datasets).**

Sometimes, a specific stage had not been derived. For example, a day of increase is followed by decrease, because the algorithms could not find a clear point for the beginning of the maximum phase. In this case, the beginning of the maximum was set between the onset of increase and decrease. Lastly, the product contains some errors, e.g. the day of decrease was mapped earlier than the beginning of maximum. For MOD12Q2 data of 2001 the proportion of errors was 0.077 % for Germany and 0.719 % for South Africa. The errors were corrected, removing invalid dates and defining new starting dates between meaningful phenological onsets. Following these corrections, metric sets of four segments were derived, including increase, maximum, decrease, and minimum. This adaptive temporal segmentation was tested against fixed segments (Figure 4.5; Conrad 2006).

4.2.2 Multi Scale Analysis of High Resolution Datasets

The goal of this module is calculation of the internal homogeneity of a coarse-resolution pixel using high spatial resolution data. The computation is useful for selecting potential samples which primarily belong to a single class. Similar approaches have been successfully employed in other studies (Colditz et al. 2005, DeFries et al. 1995, DeFries et al. 1997, DeFries et al. 1998, Hansen et al. 2000, Walz et al. 2007). Together with sampling (see 4.2.4 Sample Generation and Subsetting), it was also used for deriving the classification schemes from a data point-of-view. Furthermore, it allows the generation of a fuzzy class estimate at the coarse spatial resolution, which can be employed for fuzzy accuracy assessment. The following sections illustrate the basic concept and an option for advanced analysis.

Basic Concept

Figure 4.6 depicts the general idea for homogeneity computation. The chart shows the percentage of classes red, green, and blue, corresponding to a hypothetical spatial arrangement within one coarse-resolution cell. The ratio between high- and low-resolution data is simplified to 1:10 throughout this study. It is evident that Example (A) is very mixed in the proportion of classes and spatial distribution. Example (B) shows the same class proportions in the chart, but the spatial distribution of high-resolution data is clearly

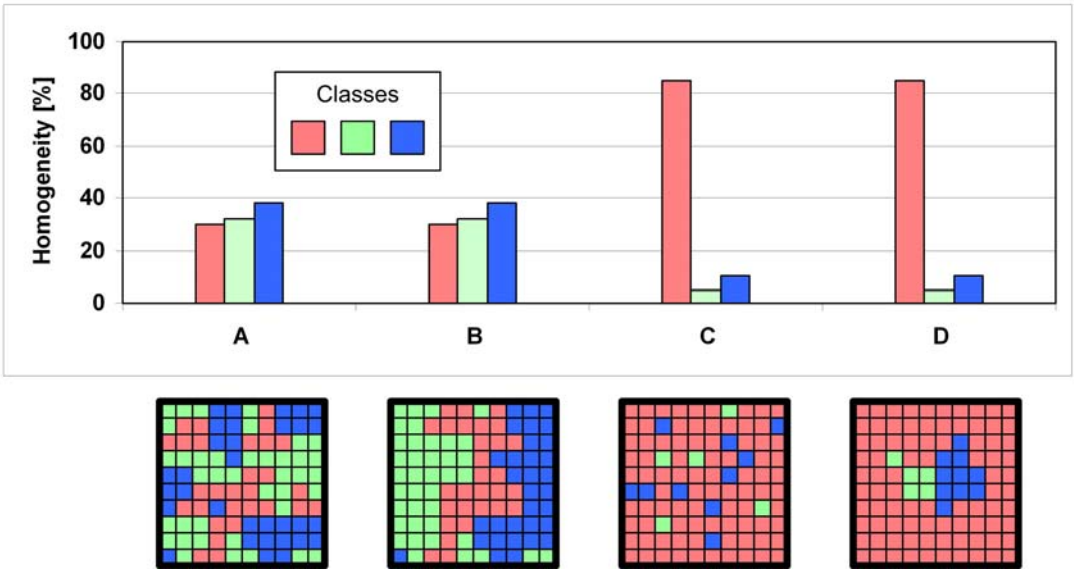


Figure 4.6: Simplified examples of multi-scale analysis.
Note: The grids at the bottom depict the assemblage of three classes in a high resolution datasets superimposed by one coarse resolution pixels (ratio 1:10). The bar charts depict the class proportions in each coarse resolution cell, the so-called homogeneity. A high homogeneity, i.e. one class dominates the coarse resolution cell, indicates candidate pixels for training and testing the classifier.

structured. Pixels belonging to the same class are connected. Expressed more analytically, the number of patches decreased from 17 in Example (A) to seven in Example (B). The noisy patterns in Example (A) can, for instance, indicate a transition zone; in this study to be expected in South Africa. Example (B), in contrast, indicates a well-structured but heterogeneous landscape with small patches, a typical characteristic of the study site Germany. No difference was made for computing homogeneity for sample data extraction.

Examples (C) and (D) of Figure 4.6 depict a clear proportional majority of the red class. This coarse-resolution cell could be regarded as homogeneous and potentially become a training or test pixel. However, there is a distinctive arrangement of the high spatial resolution data. While Example (C) shows classification noise of the green and blue classes, Example (D) depicts a connected pattern. Pavements, for instance, are known to cause spectral blurring even if only a small proportion of the sample is covered by this surface type. Dampening effects, i.e. a decrease in NIR, is known for small, water-filled ponds. Although a connected spatial arrangement of high-resolution data will clearly result in different reflectance values, no difference has been made in this study. For most study regions, the minimum mapping unit does not allow for analysis of these small-scale structures (see below).

Enlarging to Adjacent Areas

It seemed more important to account for a more conservative selection procedure, also considering potential inaccuracies in the spatial alignment of the datasets, i.e. the coarse-resolution MODIS data and the high spatial resolution land-cover data. Co-registration has also been found important by other studies which combined high and coarse-resolution data (DeFries et al. 1997, DeFries et al. 1998). It has been shown that the geometric accuracy of MODIS is very good, with reference inaccuracies lower than 50 m (Wolfe et al. 2002). Since many classifications are based on Landsat data, spatial autocorrelation with selected Landsat 7ETM$^+$ imagery showed generally good agreement. The best average correlation coefficient indicates a spatial offset of clearly less than one MODIS pixel. Similar studies with low offsets were conducted comparing MODIS and ASTER, i.e. data from the same platform and overpass (Conrad 2006). Despite the generally good agreement between the datasets, even an offset of a half MODIS cell can result in clearly misleading sample selections. Another difficulty with MODIS data is related to widely varying angles, which can have severe implications on the dataset. The bow-tie effect, for instance, has been described in 2.2.1 Platforms and Instruments. The gridded composite itself will not indicate this limitation, but algorithms can only partly compensate for distorted data at the edge of a scan. Therefore, a more conservative selection process is needed.

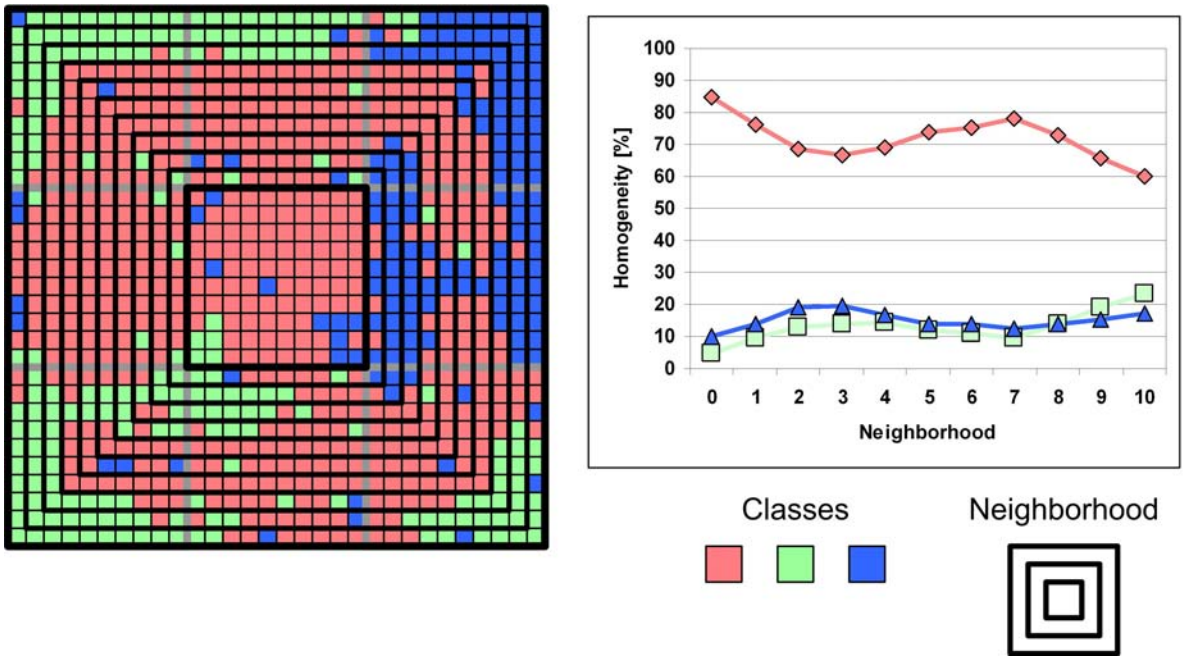


Figure 4.7: **Simplified example of multi-scale analysis with extended neighborhoods.**
Note: **The field of high resolution pixels on the left-hand side is superimposed by a coarse resolution cell (bold black square in the center). This cell is continuously enlarged to include neighboring high resolution pixels into the multi-scale analysis. The respective plot of class proportion (homogeneity) with increasing neighborhood is shown on the right-hand side.**

Spatial correlation and subsequent geometric coregistration between MODIS time series and a discrete high spatial resolution classification is not possible. Therefore, in addition to the MODIS pixel itself, a neighborhood surrounding each cell was analyzed. Figure 4.7 depicts the procedure developed in this study for a hypothetical example (ratio 1:10), where the graph plots the homogeneity of all classes with increasing neighborhood. The class homogeneity is essentially the class proportion of a respective area. Homogeneity computation without considering the neighborhood results in identical class proportions as in examples *C* and *D* in Figure 4.6. However, the red proportion dramatically decreases for small neighborhoods in favor of classes blue and green. Class red proportion increase for neighborhood sizes five to seven, almost reaching 80% homogeneity. However, class green benefits from extensive spatial cover for large neighborhoods. This conservative approach clearly limits the amount of potential sample data. For instance, at a ratio 1:10, a neighborhood of five maps an area of four coarse-resolution cells, and a surrounding of ten maps nine coarse cells. Therefore, homogeneity computation is a function of landscape heterogeneity, size of coarse-resolution cells, ratio between high and coarse spatial resolution data, high-resolution classification accuracy, minimum mapping unit, and neighborhood. In order to consider the limiting and potentially misleading influence of the minimum mapping unit (MMU), the study is complemented with one land-cover dataset without a MMU. The recently released NLCD2001 of the USA is analyzed for the impact of several MMUs on multi-scale analysis and classification accuracy.

In essence, the resulting homogeneity image mimics an estimate of class proportion at coarse spatial resolution. Viewed from this perspective, it is similar to a fuzzy classification and as such it is used as reference for fuzzy accuracy assessment. Details and constraints on this view are discussed in section 4.2.7 Accuracy Assessment. However, the fuzzy class estimate can also be transformed to a discrete map using the majority rule, which selects the class with the highest membership. The corresponding membership value accompanies the hard classification and indicates the certainty of discrete class assignment (for details see also 4.2.6 Classification and Merge).

4.2.3 Buffering of Coarse Spatial Resolution Datasets

Multi-scale analysis is the preferable option to prepare ancillary data for sample data identification. However, if high spatial resolution reference data are not available, buffering along class boundaries is an alternative. The reason for buffering lies in the increasing likelihood of incorrect class assignment along the boundary of classes. Boundary uncertainty is a well-known phenomenon in GIS, where the confidence of class assignment increases with distance from boundaries (Foody 2001, Klein et al. 1998, Middelkoop 1990). Colditz et al. (2006c) quantitatively analyzed the edge / core effects on accuracy statistics for multi-resolution data. In this study, buffering was employed for remapping vegetation types in Namibia. Sampling was prohibited within a 10 km buffer along class boundaries.

4.2.4 Sample Generation and Subsetting

Different issues of sampling strategies with regard to spatial and between-class distribution have already been discussed in 4.1.2 Background on Decision Trees. Due to the design of this study, a minimum requirement of 1,100 samples per class has been defined. In a first step, a procedure evaluates which level of homogeneity of the multi-resolution analysis is required to achieve 1,100 samples for each class. Of course, completely homogeneous pixels are most desirable, but some classes with smaller proportions or small patch sizes might also be mapped with lower than 100 % homogeneity. Homogeneity up to 80 % seemed appropriate for sampling. For some classes, however, the classification scheme has to be adjusted, too, e.g. small classes were merged into other, contextually suitable classes. This combination of classes is partly reflected in the classification recode schemes of the study sites (Table A.17, Table A.19, Table A.20, and Table A.21). In other words, there is an interactive step between designing the classification scheme and multi-resolution analysis to ensure sufficiently large sample sets for each class.

If fewer than 1,100 samples are available for a particular class, the homogeneity is reduced by one percent until the number of samples reaches the required threshold. In the case of more samples, even at homogeneity of 100 %, the samples are spatially distributed using a sample interval strategy.

$$interval = \frac{samples}{1100} \qquad \text{Equation 4.8}$$

This strategy ensures that all areas are equally sampled according to the available potential training data. In essence, exactly 1,100 samples were collected for each class. This equalized sampling is believed to prevent decision trees from overestimating classes with larger sample sets.

In order to test the influence of sampling on classification accuracy, the standard procedure of equalized sampling was complemented with a random stratified sampling using simple linear scaling between an upper and a lower threshold and a logarithmic scaling approach. Both linear and logarithmic sampling were constrained to 1,100 and 5,500 samples for classes with the least and most potential samples. Scaling was performed using a standard minimum-maximum scaling function (*i…class*):

$$scaled_i = \frac{(samples_i - \min(samples)) \cdot (5500 - 1100)}{\max(samples) - \min(samples)} + 1100 \qquad \text{Equation 4.9}$$

For logarithmic sampling the available sample sets were previously transformed using the natural logarithm.

Figure 4.8 indicates the distribution of samples along the curve. For linear sampling, the class with most samples will be sampled with 5,500 samples, while the smallest contains 1,100 samples, by definition. The samples of all other classes depend on the difference between the samples of the particular class and the class with the most samples, represented by the slope of a straight line. If, for instance, one class contains many more samples than all the others, the sets of all other classes will be comparatively small (but always above 1,100). This issue is relieved by the logarithmic transformation. However, the logarithmic curve for sampling depends on the difference between potential minimum and maximum samples due to scaling with Equation 4.9 following the logarithm computation, which is exemplarily depicted for selected differences.

Next, the sample sets are strategically split into equally sized subsets. In this study eleven subsets were generated; ten for growing decision trees and one for assessment. Again, the subsets are distributed equally in space to limit the potential bias of regionally overrepresented sets.

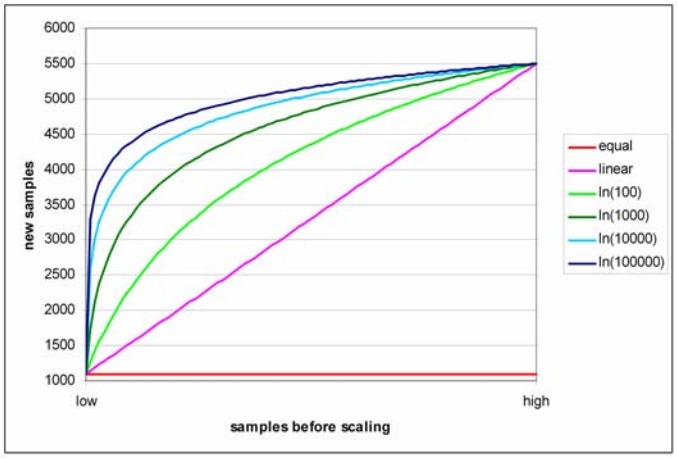


Figure 4.8: **Graph of equalized, linear, and sample size dependent logarithmic sampling schemes.**
Note: **Equalized sampling draws 1,100 samples for each class. Linear and logarithmic sampling draw between 1,100 samples for the class with the smallest set of candidates and 5,500 samples for the largest set. For linear sampling, the distribution of all other class sets is scaled on the straight line, yielding an unbalanced distribution if one class predominates. The logarithmic function mitigates this effect by a previous transformation using the natural logarithm, but yields different curves depending on the maximum set of candidates.**

Altogether, it can be stated that the entire sampling procedure aims at equally-sized and spatially equally distributed sample sets to limit between-class or regional bias. Furthermore, sampling is carried out strategically as a well-defined process. Random sampling was not considered appropriate, because results will likely not be reproducible and the gain or loss in accuracy cannot be controlled. Even sample subsetting was not performed with conventional bootstrapping procedures due to the inherent randomness, and the potential of bias. With classical bootstrapping the same sample can be used in different sets for decision-tree generation or assessment. In this regard, the approach chosen in this study does not employ cross-validation or randomization. In a strict sense, even bagging (bootstrapping with replacement) has not been chosen. Therefore, the term random forest, indicating multiple decision trees generated by bootstrapping, has to be redefined to “strategic forest” in this study.

4.2.5 Generation and Pruning Classification Trees

Growing decision trees follows the principle of CART (Breiman et al. 1984), and has been outlined in the respective sections in 4.1.2 Background on Decision Trees. For splitting, the Entropy with natural logarithm was used. This main splitting criterion was tested against Gini and Towing in the sensitivity study. In total ten decision trees were generated for each time-series classification. Tree growth was terminated at homogeneous nodes. Furthermore, stopping rules terminated the tree when a node contains less than ten samples.

Several pruning strategies have been suggested for generating right-sized decision trees (Breiman et al. 1984, Pal and Mather 2003, Quinlan 1993). Most studies conclude that pruning increases the accuracy when the decision tree is applied to non-training data. However, no pruning strategy seems superior. Therefore, a new pruning approach has been developed, comparing the resubstitution error *R*, weighted by the number of elements *e* between parent and sum of child nodes.

$$R_e = R \cdot e \qquad \text{Equation 4.10}$$

The computation of the resubstitution error *R* of a node *t* over all classes *j* is similar to the misclassification error at a particular leaf (see also Equation A.4 and Equation A.9).

$$R(t) = 1 - \max_{j \in \{1 \ldots J\}} (p(j \mid t)) \qquad \text{Equation 4.11}$$

Two critical statistics were derived, ratio and deviation.

$$ratio = \frac{R_e(parent)}{R_e(children)} \qquad \text{Equation 4.12}$$

$$deviation = \frac{R_e(children)}{R_e(parent)} \qquad \text{Equation 4.13}$$

Ratio and deviation indicate the upper and lower bound where no further splitting seems appropriate. The tree is pruned to the position of the parent node, if the ratio is greater than six or deviation lower than 0.975. In either case, the sub-tree below the parent node is deleted and the parent node is redefined to a leaf.

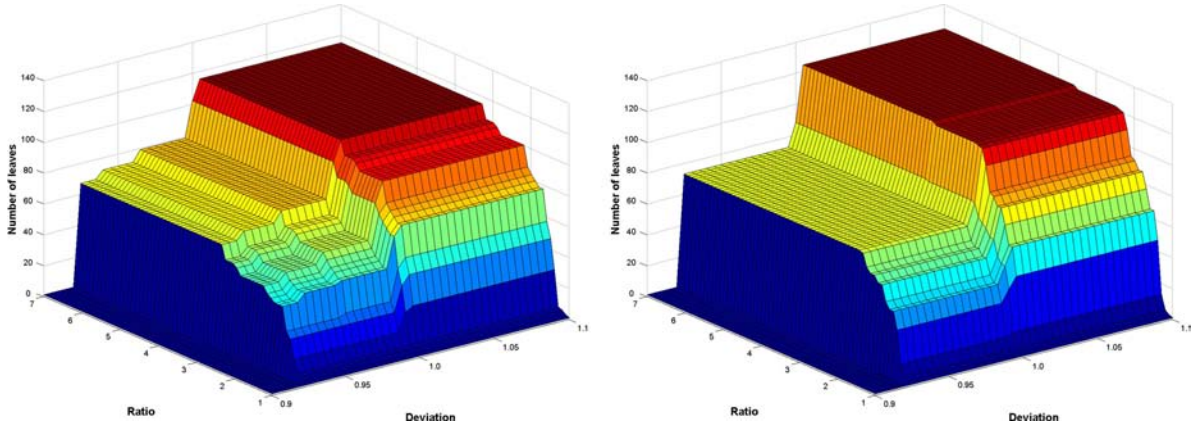


Figure 4.9: Pruning of decision trees using the approach developed in this study.
Note: Deviation and ratio are displayed on the bottom and the corresponding number of leaves after pruning on the z-axis. The goal of pruning is the reduction of the inherent complexity in the tree without decreasing and potentially even increasing the accuracy. The highest plateau indicates no pruning. The stable intermediate plateau for deviations slightly lower than 1 and ratios greater 4 is most desirable.

Both cut-off values, six for ratio and 0.975 for deviation, were derived by empirical tests and found suitable for several decision trees. Figure 4.9 depicts the number of leaves for selected decision trees after pruning for pairs of deviation and ratio. The surface plot always indicates an intermediate plateau, which is rather stable at the intersection of both cut-off values. The upper plateau does not yield a reduction of leaves and all other plateaus with ratio-values between one and two are not stable among all decision trees.

4.2.6 Classification and Merge

Following tree generation and pruning, each decision tree is applied to the entire dataset. For classification the decision tree is seen as a set of rules with binary conditions. Each split formulates a question like: if a metric is lower than a threshold value then choose the left branch, otherwise choose the right branch. The splitting process terminates at each leaf. Commonly, classification trees assign the class with the majority of training samples (Equation A.8), even if some samples indicate other class memberships. This study, however, makes a first step to a fuzzy class indicator by storing the class proportions indicated in each leaf.

Having applied all decision trees to the dataset, the separate fuzzy classifications are merged into an overall soft classification. Merging was performed using pixel-wise averaging of the class proportion of ten independent fuzzy classifications. Simple averaging was preferred in this study, because each individual decision tree is considered a weak learner (following the terminology of meta-strategies), whereas the merged classification yields an improved result. Weighting individual classifications is an option, but the criterion on which weighting should be based is rather unclear. The overall classification accuracy of each individual classification is clearly an option. For instance, Conrad (2006) employed the KHAT statistic as a measure for weighting individual classifications from classical classification trees with discrete classes. Another option for rating classes differently is the user's accuracy. Because of the lack of knowledge, however, the more conservative non-weighted approach of simple averaging seems more useful in this study.

The merged fuzzy classification can be transformed into discrete classes by applying the majority rule. In this case, the class with the highest membership is selected as "the best guess." The classification is transformed from three-dimensional fuzzy class estimate with considerable difficulties in interpretability to a simple map with defined classes. Besides the class with the highest proportion, a supplementary dataset stores the membership of the assigned class. This classification certainty is simply the fuzzy class estimate of the selected discrete class, and indicates the reliability of each pixel in the hard classification.

In addition to the proportional class membership of each decision tree, the identifier of each leaf is stored in a supplementary dataset. Although the identifier of each terminal node does not directly contain contextual information, this by-product indicates dominating leaves in the classification tree. It can be used for numerous explorative analyses, e.g. retrieval of the path which the classified pixel has taken through the set of conditions. In this study the map of leaf identifiers, in the following termed leaf map, is employed as an indicator of distinct class properties. For instance, a class such as grassland is often indicated by multi-modal frequency distribution, e.g. with class sub-types such as winter and summer peaks in the phenology. These different class properties are likely to be reflected by different terminal nodes. In many ways, the leaf map is similar to an unsupervised classification.

4.2.7 Accuracy Assessment

As already noted in 4.1.3 Background on Accuracy Assessment, this study considers hard and fuzzy accuracy assessment. In particular soft classification assessment imposes a challenge, because of the lack of established techniques and protocols. Therefore, a strategy is applied which borrows from other fields of continuous data analysis, e.g. the inter-comparison of remote-sensing data (Brown et al. 2006, Fensholt et al. 2006, Pisek and Chen 2007). Since this study aims at the development of an approach to time-series classification, detailed accuracy assessment is important for the evaluation of the classification approach.

First, the process of accuracy assessment will be outlined. Next, a section will discuss issues of the reference map, which requires a detailed understanding of the entire classification procedure. Finally, separate sections briefly explain the accuracies and error measures for hard and soft classification assessment.

Process of Accuracy Assessment

While Figure 4.4 provides the general workflow for time-series classification, Figure 4.10 specifically depicts the processes for accuracy assessment of this study. The basic inputs for accuracy assessment are outputs of the multi-scale analysis and the merge of fuzzy classifications as reference and classified data, respectively. Three outputs have been derived from both multi-scale analysis and fuzzy classification: the actual fuzzy map with memberships of each class, the class with the highest proportion as discrete map using the majority rule, and the membership value of this class as an indication of discrete classification certainty (see also 4.2.6 Classification and Merge). During sampling, eleven sample sets were generated, and one set was not used for decision-tree generation.

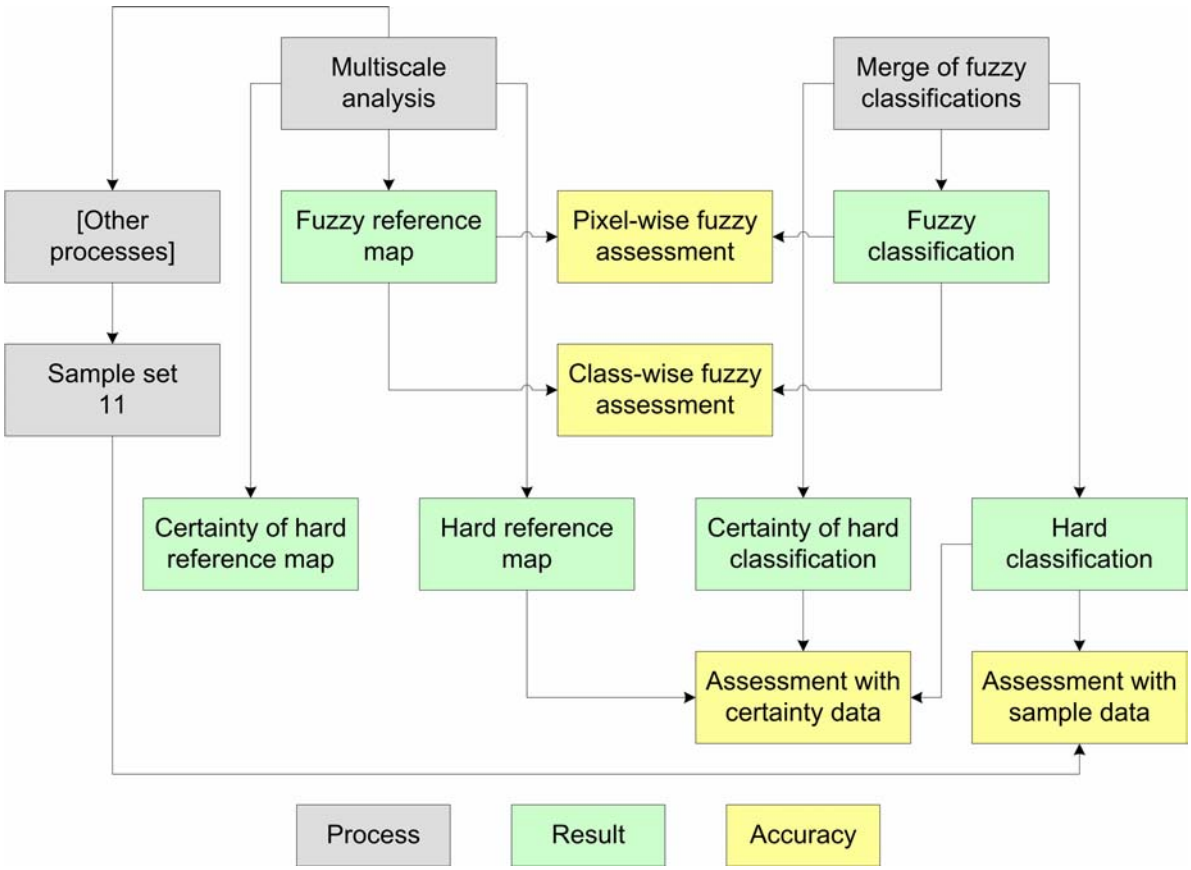


Figure 4.10: Detailed flow chart for the results of time series classification and accuracy assessment.
Note: The process indicates that each fuzzy map can be transformed to a discrete map (e.g. using the majority rule) and the corresponding membership indicates the certainty of class assignment. In total, four assessments of the classifications have been conducted. For details on the fuzzy assessment protocol see Figure 4.11. For the general classification procedure see Figure 4.4.

Two techniques evaluate the hard classification accuracy. A first assessment employs the independent eleventh sample set. This approach assesses the accuracy of selected homogeneous pixels, i.e. the estimated accuracy is biased to rather distinctive samples and tends to be too high. It is similar to the common approaches, for which point-based samples serve for map evaluation. A second assessment compares the hard reference map with hard classification. Because a simple 1:1 comparison would be an inadequate assessment of the uncertainty in discrete classes for mixed land surfaces, the evaluation is only performed for pixels above a defined classification certainty. The accuracy increases with higher certainty thresholds, but the dataset considerably decreases, on which the evaluation is based.

Soft accuracy-assessment techniques directly employ the fuzzy reference and classification. The class-wise fuzzy assessment estimates the membership of each class separately, e.g. by computing difference images between the reference and classification and the calculation of summary statistics. The other perspective on fuzzy assessment compares all class memberships of each individual pixel between the reference and classification. The respective statistics for both fuzzy assessment strategies are provided below.

It should be noted that the protocol for hard and soft classification assessment can be easily extended to change detection analysis of two classifications. In this case, the reference map is simply replaced by a similar dataset of another year. This strategy will be employed to indicate the capabilities of the time-series classification procedure for land-cover update.

Issues of the Reference Data

During multi-scale analysis a high spatial resolution land-cover dataset has been transformed to coarse spatial resolution with fuzzy-like class estimates (see also 4.2.2 Multi Scale Analysis of High Resolution Datasets). Although these fuzzy estimates are regarded as reference and are compared to fuzzy time-series classification, they are not a reference in the sense of ground truth.

A true reference dataset requires field-data sampling. While field studies are possible for small areas and for specific periods, it is impossible to provide sufficient ground-truth data for classification assessment of large areas and in temporal continuity. The mixed-pixel issue causes considerable difficulties when sampling field data for the evaluation of coarse-resolution maps. The incorporation of uncertainty measures in the reference data can mitigate this issue (Gopal and Woodcock 1994; see also 4.1.3 Background on Accuracy Assessment). Regional to global networks for product validation are being established, such as the CEOS-BELMANIP initiative (Baret et al. 2006; see also 3.1.3 Time Series Comparison and Validation).

Therefore, it is common practice in remote sensing to employ secondary datasets, such as maps or air photographs for either classifier training or accuracy assessment. Essentially, this approach has been employed for one hard classification assessment of this study. Similarly, the map evaluation comparing the discrete classification and reference map modified by class-certainty measure, mimics common accuracy protocols but is based on a much larger population.

Since it is practically impossible to sample fuzzy ground-truth data, another approach has to be employed for fuzzy reference generation. While discrete high spatial resolution maps indicate one class per unit, a scale transformation to a coarser grid can account for the heterogeneity of the land surface and the uncertainty in discrete class assignment. This fuzzy reference is generated by multi-scale analysis as described in 4.2.2 Multi Scale Analysis of High Resolution Datasets). Although the scale transformation mitigates some inaccuracies in high-resolution classification, the fuzzy reference is still biased by the analyst who generated the input data. An even more important issue is the MMU, which can dramatically decrease the surface heterogeneity of the reality. Visually interpreted maps in particular cannot account

for the full spatial variability beyond the mapping scale. Furthermore, the discrete class assignment at high spatial resolution does not reveal broad transition zones between land-cover classes. This issue is exceptionally clear for ecotones in South Africa. In this regard the reference map can bias the assessment considerably and should also be questioned during analysis. Obviously, the bias decreases with increasing the ratio between fine and coarse spatial resolution. Finally, the reference map has often been mapped from data of a different year than the data to be classified. Hence, there is a potential of uncertainty related to land-cover change.

Despite these limitations, the reference map has been employed for an assessment of pixel- and class-wise fuzzy classifications. The selected scale ratio of 1:10 for multi-scale analysis applied throughout this study seems sufficient. The scale-bias is decreased using a considerable neighborhood, accounting for the uncertainty in geometric alignment between both datasets (see 4.2.2 Multi Scale Analysis of High Resolution Datasets). The land-cover change is assumed to be small, and the likelihood of class-mismatch in training data is also mitigated by sample selection using multi-scale analysis. Furthermore, classification trees are known for their robustness, if a portion of the sample data is not correct. Although not a reason in the strict sense, the approach chosen in this study appears to be the only feasible way to generate fuzzy reference data.

It could be argued whether the accuracy assessment is merely a map comparison, where the truth is still unknown (Atkinson and Foody 2002). In this regard, the word error, as often used in this study, has to be softened, because the analysis merely indicates a difference. Another difference to a common accuracy assessment is the use of the entire population of data, i.e. no data sampling has been employed for fuzzy assessments. Due to the vast amount of pixels there is no significant bias when employing samples already used for training the classifier.

Assessment of Hard Classification

The assessment of hard classification accuracy comprised a sample-wise analysis using the eleventh set not employed for decision-tree generation and a 1:1 reference-classification map comparison modified by the classification certainty. Both approaches yield a confusion matrix, where conventional statistics such as overall accuracy, KHAT, and class-specific user's and producer's accuracy are calculated (Congalton and Green 1999; see Figure A.3 and Equation A.12 to Equation A.15).

The 1:1 map comparison has been computed for different classification-certainty thresholds to indicate the level of increase in accuracy with rising certainty, but decreasing population. Therefore, class-specific and overall accuracies as well as KHAT are plotted against certainty,

and an additional graph often shows the decrease in area-proportion. At some stage accuracy computation had to be terminated. A threshold for the minimum number of samples or sample proportion can be specified, but this approach highly depends on the study-area characteristics. Empirical tests have shown that the accuracy often dramatically decreases when the set of remaining pixels for analysis becomes too small. Therefore, the increase in certainty is stopped, if the decrease in accuracy is greater than 5 %. The ultimate termination occurs at a classification certainty of 100 %, which rarely occurs.

Assessment of Fuzzy Classification

It is clear that fuzzy accuracy assessment can only be carried out, if multi-scale analysis of the input data was feasible. In this regard, the level of detail of the vegetation map of Namibia did not allow for multi-scale analysis. Despite the minimum mapping unit for NLC1995 (South Africa) and CLC2000 (Germany) of 25 ha, multi-scale transformation was conducted. The NLCD2001 (Minnesota) is available without minimum mapping unit. The raster-cell sizes of the high-resolution data as employed for multi-scale analysis in this study can be obtained from the classification protocols (Table A.23 to Table A.26).

It has been noted that there are two distinct views on fuzzy classification assessment. Figure 4.11 depicts the process of fuzzy accuracy assessment as conducted in this study. Class-wise analysis compares the memberships between reference and classification for each class. On the other hand, the pixel-wise approach computes error estimates of class memberships for each cell separately. Several statistics are applicable to both class- and pixel-wise assessment.

For class-wise analysis, difference images of class membership between reference and classification are very informative. The so-called relative-difference image (considering the sign) indicates the over- or underestimation for each class. Computing the absolute of the difference image indicates the intensity, also called magnitude, of the error without direction.

Basic univariate statistics such as mean and standard deviation are derived from each class-wise difference map and indicate the error. The mean of the relative differences shows the overall trend of the class-wise error, and the mean of the absolute difference indicates the magnitude for each class. The standard deviations measure the variance in the respective difference images.

Two other indicators, the correlation coefficient, *r* and the RMSE, are derived for each class membership in class-wise assessment.

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{N}(y_i - \bar{y})^2}} \qquad \text{Equation 4.14}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}} \qquad \text{Equation 4.15}$$

N...*number of elements*
x...*reference image*
y...*classification image*

The correlation (Equation 4.14) is a measure of relative correspondence between the reference image and the classification result. It merely considers the normalized joint variance (covariance is contained in the numerator) excluding the bias between both datasets. In contrast, the RMSE (Equation 4.15) reveals the absolute differences between fuzzy class estimates of reference and classification including the bias. In many ways the RMSE is similar to the absolute mean statistic of the difference image, but pronounces higher discrepancies due to the square.

Essentially, the same basic statistics, relative and absolute mean and corresponding standard deviations, correlation coefficient, and RMSE can also be derived for pixel-wise analysis. The computation, however, is performed for all class memberships of each individual pixel, and therefore yields an image of each error statistic. Since the sum of all memberships is constrained to 1 (or 100 %), the relative mean is always zero. Removing the sign, the mean absolute difference is a good indicator of the general differences of class memberships between reference and classification. Similarly, the correlation indicates a relative correspondence and the RMSE reveals absolute differences. Furthermore, a set of basic statistics is derived from each error image. The minimum, maximum, mean, and standard deviation are good indicators of general agreement or disagreement between fuzzy reference and classification.

However, a statistical issue regarding the independence of class memberships needs to be discussed. The constraint that the sum of all memberships of each pixel must equal 1 (or 100 %) exposes some statistical concerns, especially for the pixel-wise fuzzy assessment. It becomes clear that fuzzy classification assessment, as carried out in this study, is

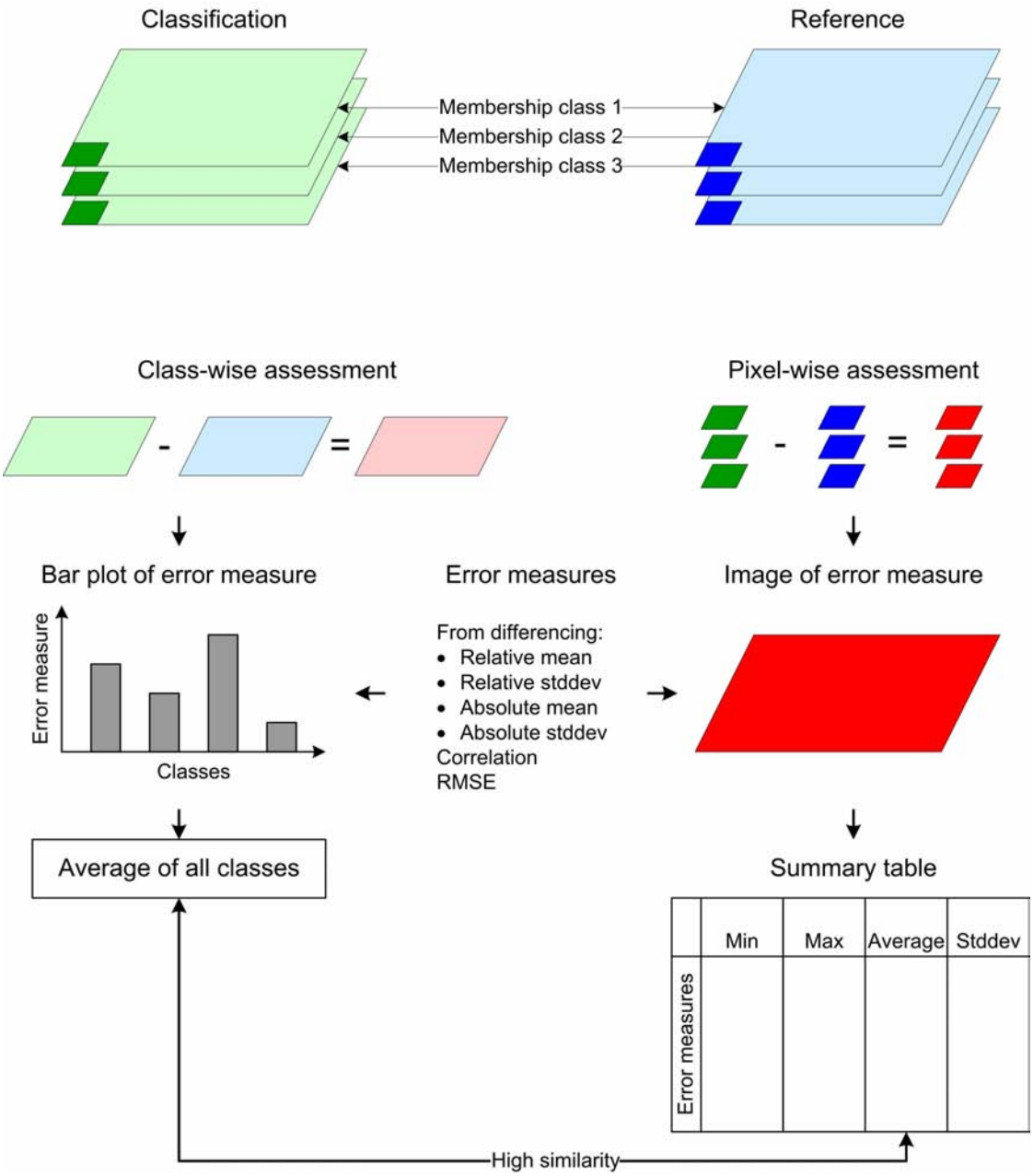


Figure 4.11: Accuracy assessment protocol of fuzzy classifications.
Note: The difference between classification and reference is computed for each class (layer; left-hand side) or each pixel (vector; right-hand side). Next, error measures are applied. The class-wise assessment yields error statistics for each class (layer), which are displayed e.g. as bar-plots. In contrast, the pixel-wise assessment computes the error measure for each pixel (vector of all classes) yielding an image. A summary table indicates simple univariate statistics of the error image. The comparison of the average of the summary table and the average of the class-wise bar-plot indicates negligible differences.

stochastically dependent among all other class proportions. For pixel-wise analysis, for instance, an overestimation of one class membership must lead to an underestimation of another class membership and *vice versa*. However, the difference can, and usually is, compensated by many other classes. Furthermore, the balancing is not equal among all other classes. A thorough accounting of the error dependencies among all classes is not trivial and requires the analysis of $class^{class}$ interdependencies, which is beyond the scope of this study.

Therefore, all error statistics have to be evaluated with caution in terms of the reference data and statistical issues. Nevertheless, the stochastically questionable pixel-wise assessment appears useful, because it provides a spatial view on the error distribution by considering the correspondence of all class memberships at once. Furthermore, there is usually a high similarity between the average of the pixel-wise error map (e.g. RMSE) and the average of all class-wise errors of the same error measure. This is, of course, not a statistical proof, but an empirical indicator for the suitability of the approaches.

4.3 Time Series Classification Analysis

The necessary basics for time-series classification and the rationale for choosing the decision-tree approach have been provided in the theoretical background of this chapter. The previous section described the methodology for time-series classification and measures for hard and fuzzy classification assessment partly developed or adapted in this study. The aim of this section is the analysis and especially the assessment of classifications. The analysis and accuracy of classifications will be demonstrated for selected study sites and datasets. The primary focus is on land cover and land-cover update, but the approach has also been extended to remapping vegetation types. On the other hand, this section applied a detailed accuracy-assessment protocol to hard and soft classifications. In particular, the impact of parameters of the developed classification procedure on map accuracy has been assessed, and similar statistics have also been used for evaluating the potential for land-cover update.

The next section will describe classification results using a general set of classification parameters for both main study sites, i.e. South Africa and Germany. It also contains a detailed accuracy assessment of hard and soft classifications. The second section provides a detailed sensitivity analysis of critical classification parameters. In a third section, the potential for classification update will be discussed. A last section will illustrate the use of the classification procedure beyond classical land-cover and land-use mapping. Here, an existing vegetation map of Namibia has been remapped using MODIS time-series data.

4.3.1 Land Cover Classification of Germany and South Africa

The aim of the first section is a detailed analysis of land-cover maps using a general set of classification parameters. The analysis is conducted for South Africa and Germany. Since this is the beginning of the assessment of land-cover classifications, both sections for South Africa will briefly recapitulate the accuracy measures and demonstrate their usefulness. Details of

the classification process are outlined in section 4.2 Classification Procedure. Protocols for parameters used for classifications are provided in Table A.23 and Table A.24 for Germany and South Africa, respectively.

Hard Land Cover Classification of South Africa

A first analysis focuses or a visual assessment of the discrete classification when compared to the reference map (Figure 4.12). Again, the reference presented on the right-hand side results from a transformation of high to coarse spatial resolution. The discrete classes were derived using the majority rule. The certainty depicts the membership for the assigned class.

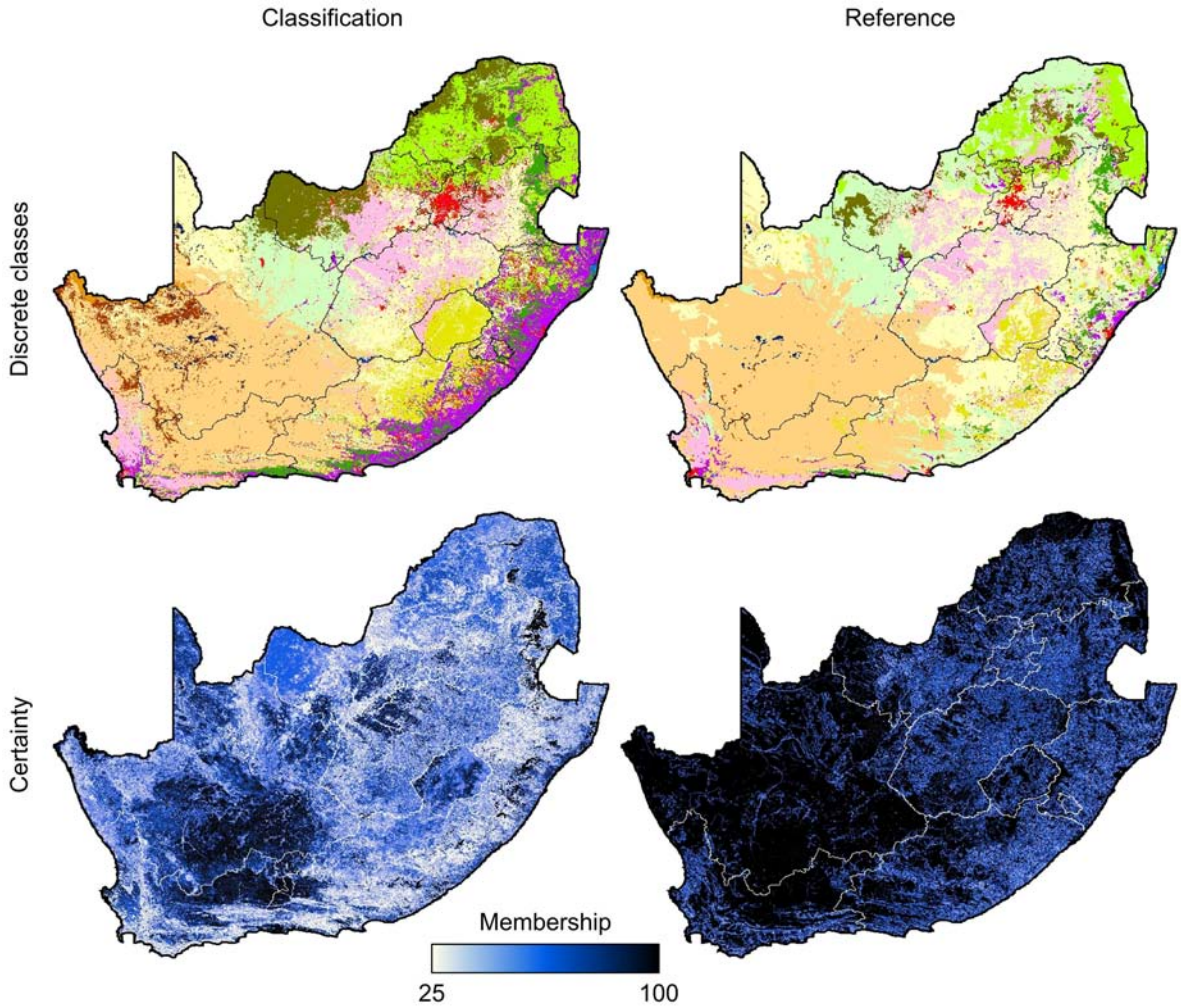


Figure 4.12: Discrete classification and reference map of South Africa as well as the corresponding certainty maps of correct class assignment.

Note: The certainty of class assignment of the reference has been derived by the multi-scale analysis. Since large portions are homogeneous in the discrete reference map, the certainty of class assignment is markedly higher. This has implications on the accuracy assessment (see also discussion in 4.2.7 Accuracy assessment). For the legend see Figure 2.3.

A visual analysis indicates that the discrete classes of the classification and reference correspond well for large portions of the study area (Figure 4.12, upper row, see also 2.3 Study Sites and Land Cover Datasets for legend and provinces). Small differences are shown for “urban” in the Gauteng area. Furthermore, class “degraded forest, woodland, thicket, and bushland” is clearly overestimated at the border to Botswana. While class “thicket and bushland” is well mapped in the center of the country, the northern areas have been assigned to contextually related classes such as “forest and woodland”. Class “irrigated agriculture”, however, is clearly overestimated, covering the entire coastal lowlands and eastern face of the escarpment. This misclassification results from the difficulty to map irrigated agriculture, which is spectrally and temporally similar to many high biomass units in well-vegetated regions.

The obvious differences in the certainty between classification and reference image (Figure 4.12, bottom row) result from the approach how both datasets were generated. While multiple decision-tree classifications were combined for the classification, multi-scale analysis was employed for the certainty of the reference data. Since large portions of the reference image comprise only a single land-cover type, the certainty of this class is 100 %. The issue is clearly revealed in the fuzziness of the reference map, which has been discussed in 4.2.7 Accuracy Assessment and will be analyzed below.

Table 4.2: Error matrix for land cover classification of South Africa for 2001.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	User's [%]
1	Forest	91	1	2	0	0	3	1	2	0	0	0	0	0	1	90
2	Forest and woodland	0	85	18	1	0	2	0	0	4	8	1	7	8	5	61
3	Thicket and bushland	0	3	40	3	0	5	0	0	2	0	0	0	1	0	74
4	Shrubland and low fynbos	0	0	2	87	0	7	1	0	5	0	0	0	0	0	85
5	Herbland	0	0	0	0	100	0	0	0	0	0	0	0	0	0	100
6	Dry grassland	1	0	3	2	0	60	1	3	3	0	9	1	6	2	66
7	Water	0	0	0	0	0	0	94	1	0	0	0	0	0	0	99
8	Wetland	0	0	1	0	0	0	1	81	7	0	0	0	0	0	90
9	Barren	0	0	3	4	0	3	0	8	70	0	0	0	0	5	75
10	Degraded forest…	0	4	28	0	0	1	0	0	0	90	1	3	9	4	64
11	Degraded grassland….	3	0	0	0	0	12	1	1	1	0	80	0	2	1	79
12	Irrigated agriculture	5	6	3	1	0	3	1	4	1	0	3	87	4	3	72
13	Dryland agriculture	0	0	0	2	0	3	0	0	1	0	2	0	69	3	86
14	Urban	0	1	0	0	0	1	0	0	6	2	4	2	1	76	82
	Producers [%]	91	85	40	87	100	60	94	81	70	90	80	87	69	76	

Note: **The (eleventh) independent sample set with 100 samples per class was used. User’s and producer’s accuracies are shown. “Degraded forest…”…”Degraded forest, woodland, thicket, and bushland”. “Degraded grassland…”…”Degraded grassland, shrubland, and low fynbos”. See Figure A.3 for a general outline of the error matrix.**

Employing the separate sample set for conventional accuracy assessment, the error matrix and the class-specific accuracies are shown in Table 4.2. The overall accuracy is 79 % and the KHAT yields 0.77. Two issues are noteworthy from both tables. The low producer's accuracy of class "ticket and bushland" results from a clear underestimation in area. Many test samples were mapped to class "degraded forest, woodland, thicket, and bushland", and a smaller portion to "forest and woodland" (see also above). On the other hand, the statistics of the error matrix do not indicate the visually obvious misclassification of class "irrigated agriculture". The user's accuracy of 72 % is still high. The effect is explained by finding areas which are "irrigated agriculture" in the reference (user's accuracy), and misclassification with classes, which are widely distributed and well mapped in other areas. Since samples of this analysis have only been selected for rather homogeneous areas, the sampling in this heterogeneous region is scarce. In this respect, the approach using the separated sample set drawn during multi-scale analysis only accounts for high-confidence pixels of the reference image.

The 1:1 comparison between classification and reference of Figure 4.12 was modified by the certainty estimate of the classification. Figure 4.13 plots classification accuracies against classification certainty. Generally, classification accuracy increases with increasing certainty. However, sample sets being too small may yield a reverse pattern, as indicated for high certainties of the overall accuracy and KHAT. Therefore, the dashed lines indicate the proportional decrease in samples with increasing certainty. Obviously, the set on which the assessment is based decreases in size with increasing certainty.

The parallel plots of overall accuracy and KHAT shown on the bottom of Figure 4.13 indicate a good agreement between reference and classification. Even at a confidence level of only 50 %, the overall accuracy already yields 69 % and rises up to 90 % for very high confidence samples. Both accuracy plots differ between 7 % and 10 %, a common range for many classifications. Often, the overall accuracy and KHAT are assumed to provide an upper respectively lower limit of true classification accuracy (Colditz 2003, Foody 1992).

The top row of Figure 4.13 shows user's and producer's accuracies and their respective proportional sample size (dashed lines). The selected classes are dominant land-cover types and their fuzzy classification will be analyzed in the next section in more detail. Because of significantly decreasing accuracies due to sample sets being too small, the graph of some classes terminates before reaching 100 % certainty. Except for class "forest", all other classes start at remarkably high user's accuracies and an area proportion of almost or above 50 %. The highest increase in user's accuracy is shown for forest (30 %). Interestingly, this graph shows distinctive jumps, which coincide with respective drops in sample proportion. In other words, the excluded pixels at specific certainty thresholds were always those which contributed to low user's accuracies. Although this is clearly the common pattern, it is not necessarily the case, as shown for the producer's accuracy of class "dry grassland".

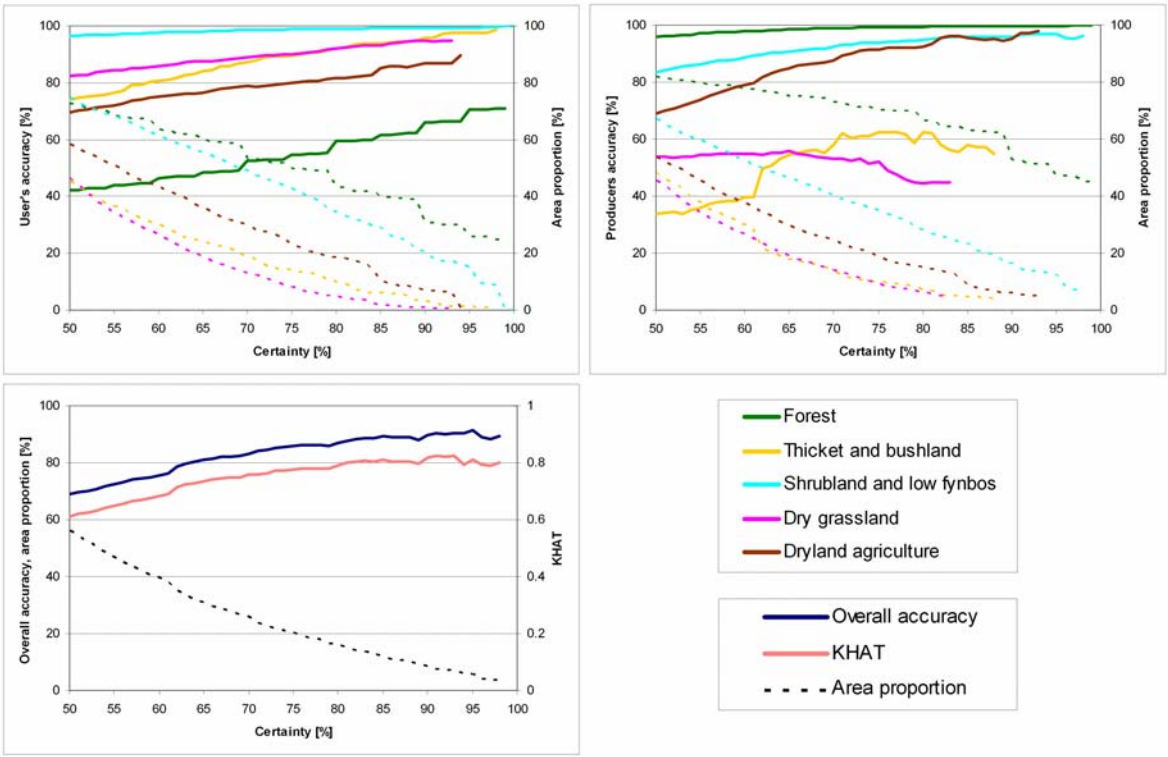


Figure 4.13: Accuracy assessment of the discrete map of South Africa using classification certainty.
Note: For selected classes, the user's accuracy (upper left) and producer's accuracy (upper right) as well as the overall accuracy and KHAT (lower left) are depicted. While the solid lines indicate the accuracy, dashed lines show the corresponding decrease in area proportion with increasing certainty in class assignment.

While the user's accuracy indicates the measure of commission, the omission error is shown in the producer's accuracy. The difference is well demonstrated for class "thicket and bushland" (see also Figure 4.12). The user's accuracy is remarkably high, because the area mapped as "thicket and bushland" in the classification coincides well with the reference. In this sense, the user's accuracy indicates the probability of a classified pixel being also mapped to this class in the reference. However, the area of "thicket and bushland" in the reference is clearly underestimated in the classification; hence the error in omission is high. Therefore, the producer's accuracy indicates the probability of a reference pixel being correctly classified.

Fuzzy Land Cover Classification of South Africa

This section analyzes and assesses the fuzzy estimates of the land-cover classification of South Africa in relation to the fuzzy reference. Figure 4.14 depicts memberships of the classification, reference, and relative difference for selected classes. The remaining classes are shown in Figure A.4 in the appendix. Displaying the directional or relative difference, i.e. with the sign, allows for a discussion of under- and over-estimation.

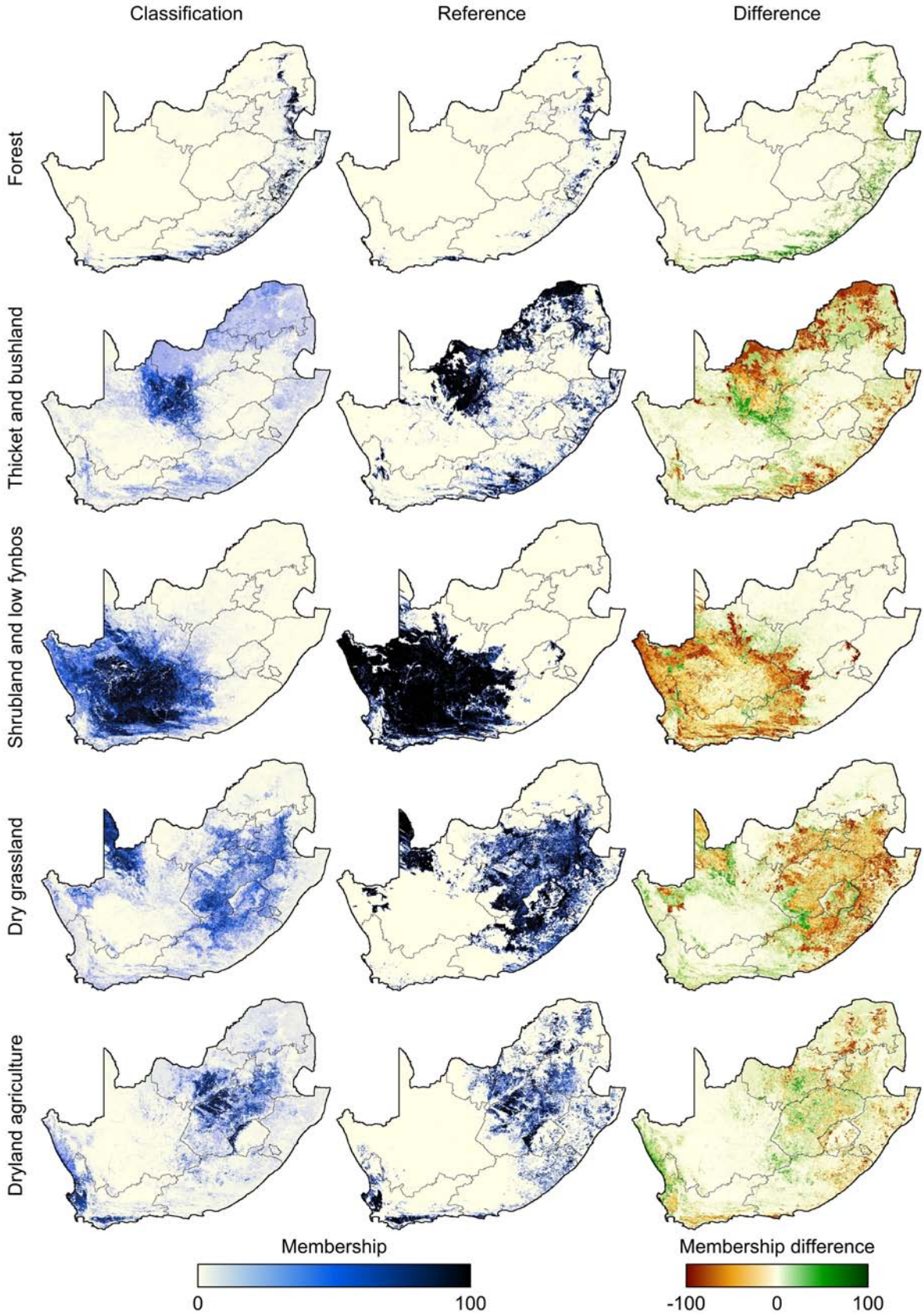


Figure 4.14: Selected class-wise comparison between fuzzy estimates of classification and reference and the corresponding difference for South Africa.

Note: Figure A.4 indicates the class-wise comparison for all other land cover classes of South Africa.

The membership of land-cover class “forest” is only slightly overestimated by area, in particular along the escarpment of the southern coast. However, the general match of memberships is remarkably good. As already indicated in the discussion of the hard classification, class “thicket and bushland” does not correspond well with the reference. There is a good agreement in the central portion of the study area, with highest memberships in the classification, and even an overestimation at the southern transition. However, the remaining areas of “thicket and bushland” show very low memberships. This may indicate a clear misclassification at first glance, but Figure A.4 indicates that contextually similar classes, such as “forest and woodland” and “degraded forest, woodland, thicket, and bushland”, are mapped instead. Obviously, “thicket and bushland” has a multi-modal frequency distribution, i.e. it is characterized by different cluster centers in feature space. If a cluster center is not well represented, then other, usually thematically related classes are mapped.

Land-cover classes’ “shrubland and low fynbos” and “dry grassland” (Figure 4.14) clearly have the highest area-proportion in the discrete reference map with 26 % and 15 %, respectively. While the extent between reference and classification matches well, both classes show remarkable discrepancies in memberships. The center portion of class “shrubland and low fynbos” is mapped accurately. But the edge shows increasing underestimations, because the reference also indicates memberships of 100 % for these portions. However, class “shubland and low fynbos” indicates a desert and semi-desert landscape in the Karroo, which has a wide transition to other natural and semi-natural landscapes. The ecotone has not been adequately represented in the high-resolution input dataset, and therefore cannot be revealed in multi-scale analysis (see also 4.2.7 Accuracy Assessment). It can be argued that the transition to other land-cover classes, including “barren”, both degradation classes (Figure A.4), and “dry grassland”, is better represented in the fuzzy classification than in the reference. Unfortunately, this argumentation cannot be proved with the quantitative assessment techniques applied in this study. However, the ecotone to “dry grassland”, bordering to the North and East, is plausibly indicated in the classification.

The low agreement of land-cover class “dry grassland”, suffers from an additional difficulty. Some areas in the Free State, North-West Province, and Mpumalanga as well as in the Eastern Cape and KwaZulu Natal are rather heterogeneous in the reference. This might have had some influence on automated training-data selection and mixed pixels may have caused difficulties during classification. The extensive areas in the northern Panhandle and the central part west of Lesotho are homogeneous in the reference and thus well mapped in the classification.

“Dryland agriculture” is also a class with several small patches (Figure 4.14, see also Figure 4.12). Furthermore, there are two distinct types, with summer rain in the central and eastern part of South Africa and winter rain in the Cape Region. The multi-modal frequency distribution exposes a significant challenge for automated classification. First, it can be noted

that reference and classification match up very well. In the Cape Region, “dryland agriculture covers the extensive croplands north of Cape Town. It extents further north along the Atlantic coast than in the reference, but with lower memberships. Furthermore, very small-patch agriculture in the Eastern Cape and KwaZulu Natal is not represented in the classification. Besides the remarkably good spatial agreement, there is also an excellent correspondence in membership values with only small under- and over-estimations. Furthermore, the distinct branch-like structure in the center of the study area and the intensive agriculture in the northwest of Lesotho are well represented. The reason for this good agreement of “dryland agriculture”, in contrast to similarly scattered “dry grassland”, is probably the effect of land-use with structures and boundaries *versus* semi-natural land with broad ecotones. The high-resolution map is capable of indicating structures, but inadequately accounts for transitions.

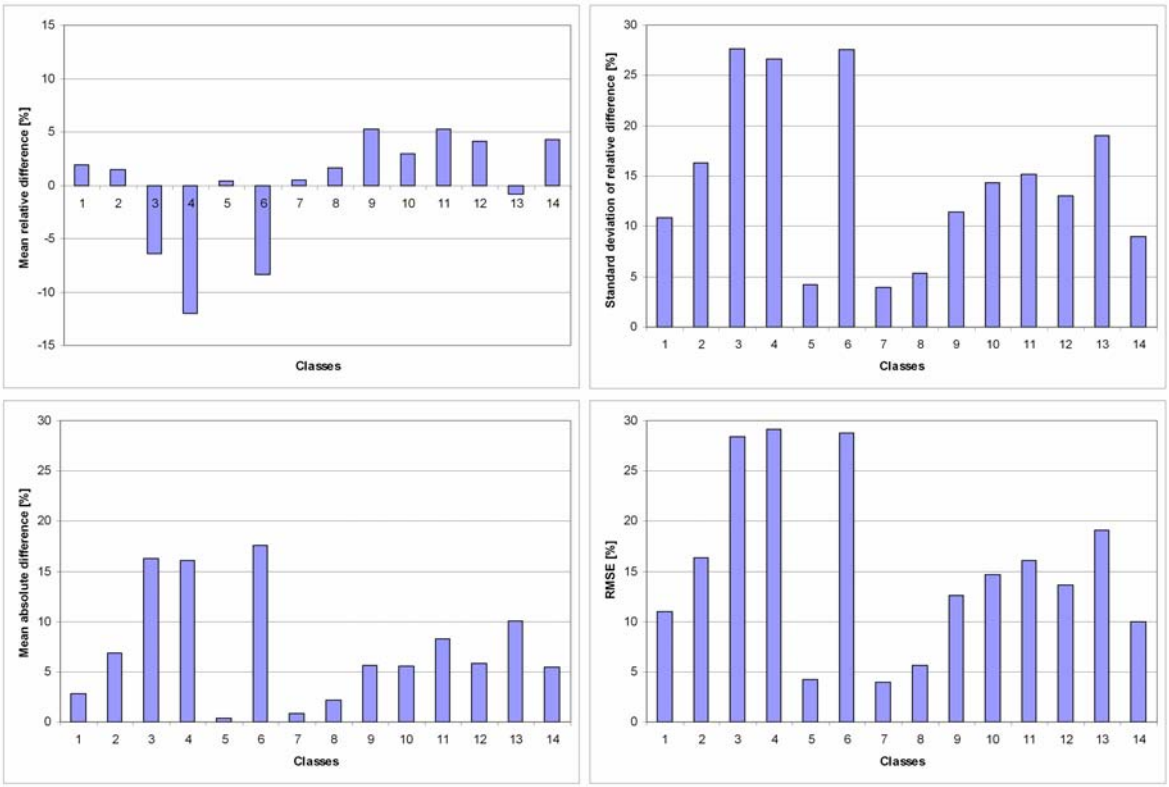


Figure 4.15: Selected error measures of class-wise fuzzy accuracy assessment for South Africa.
Note: The relative mean difference (upper left) depicts the general trend, i.e. under- or over-estimation compared to the reference. The corresponding standard deviation (upper right) indicates the variability in the relative differences. The absolute mean (lower left) and RMSE (lower right) show the magnitude in the difference, which can be markedly higher than the relative mean because of omitted balancing effects. For corresponding class names see Table A.17. For all error measures see Table A.27.

As illustrated in Figure 4.11, several error measures can be computed to summarize the spatial class-wise difference maps. Figure 4.15 shows bar-plots of four measures; the complete list is contained in Table A.27. The mean of the relative difference for each class in the upper left shows a directional agreement, or, in other words, the general trend. For instance, both forest classes (1 and 2) are slightly overrepresented in the classification, and class “13: dryland

agriculture", with small differences to both directions, yields a minor underestimation. Highest disagreements with clear underestimation are indicated for classes "3: thicket and bushland", "4: shrubland and low fynbos", and "6: dry grassland". It should be noted, though, that this analysis evaluates the entire area of South Africa, which increases the error for classes with higher area-proportions. A restriction to the pixels where the classes actually occur may seem useful at first glance. However, it would contradict the idea of fuzzy classification assessment, i.e. that each pixel has a membership for every class. In other words, a membership of zero, whether in the reference, classification, or both, is valid and has to be accounted for in the analysis.

The corresponding standard deviance of the relative difference in the upper right of Figure 4.15 depicts very high values for error-prone classes. However, even class "13: dryland agriculture" shows a remarkable variability of 19 %, caused by the small differences in either direction. However, the general pattern remains: classes dominant in area proportion have a higher variability.

The bottom row of Figure 4.15 depicts absolute error measures, namely the mean of the absolute difference and the RMSE. While the mean of the relative difference indicates the overall trend, these measures illustrate the magnitude. In other words, removing the sign, either by the mathematical absolute (mean) or the square (RMSE), prevents from the compensation in trends. This effect is particularly well indicated for classes "1: forest" and "13: dryland agriculture". While "1: forest" was mostly overestimated, the error only slightly increased for absolute measures. However, "13: dryland agriculture" with a small underestimation has an absolute error of 10 % and 19 % for absolute mean and RMSE, respectively. The ranging among classes using absolute mean or RMSE hardly differs in this example. Table A.27 also shows the standard deviations of absolute differences, which are always lower than the corresponding relative error variations. The class-wise correlation coefficient does not indicate a specific pattern.

The pixel-wise fuzzy classification assessment depicts the above-mentioned error measures in image fashion. Figure 4.16 shows two error maps, mean absolute difference and correlation coefficient; all other error measures reveal the same spatial pattern with different values. In contrast to difference images of individual classes (e.g. Figure 4.14), error images are a general indicator of membership agreement among all classes. The mean absolute difference, for instance, nicely shows areas with low discrepancies between reference and classification. Several parts of the study area have been mapped accurately, especially the central part of the Karroo ("shrubland and low fynbos"), Free State and Mpumalanga ("dry grassland" and "dryland agriculture"), areas in the Western Cape ("dryland agriculture"), and eastern Mpumalanga and Limpopo ("forest and woodland"). Other areas depict disagreements with the reference. The issue with an inadequate representation of transitions between natural

landscapes in high-resolution input data and therefore fuzzy reference has already been discussed, e.g. for the edge of the Karroo. Errors along the southern and eastern coastal lowlands and escarpment result from the overestimation of class “irrigated agriculture”. This class is trained on samples of multiple cover types of permanent and temporarily irrigated agriculture. Being partly a wetland, it can have multiple modes in feature space. Furthermore, it is located in a densely vegetated environment similar to the class properties itself, and is characterized by many small patches, well beyond the spatial resolution of 1 km. The pixel-wise analysis of the correlation on the right-hand side of Figure 4.16 indicates similar spatial patterns compared to the other error statistics. Most areas have a clearly positive correlation coefficient.

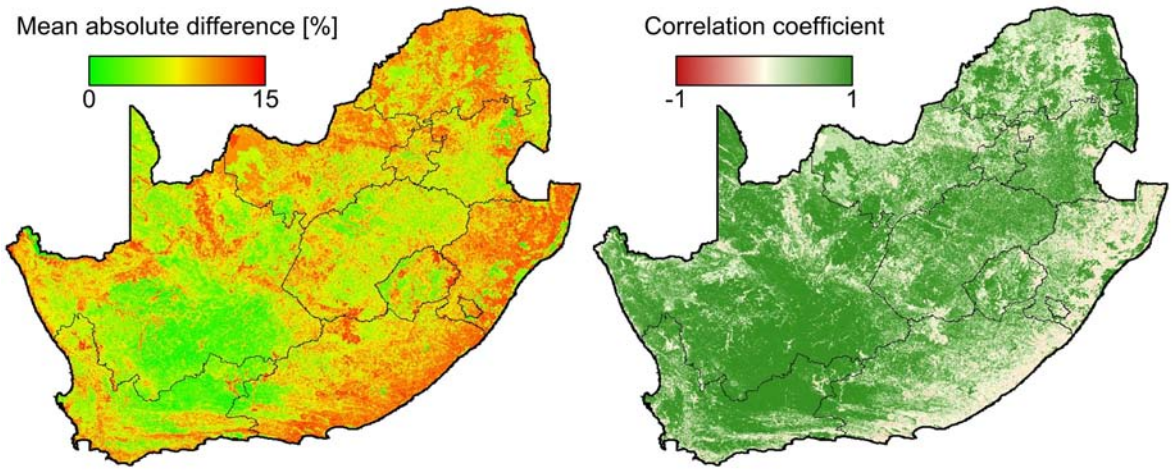


Figure 4.16: Selected error measures of pixel-wise fuzzy accuracy assessment for South Africa.
Note: The correlation coefficient is rarely below zero and green tones indicate values above 0.7. All other error measures indicate the same spatial pattern of disagreement. For summary statistics see Table A.28.

Finally, Table A.28 shows the summary statistics of the pixel-wise error measures. Basically, the average and standard deviation of each error measure can be regarded as an overall metric. While the average indicates the magnitude, the standard deviation shows the corresponding variation of the error between reference and classification. Clearly, the relative class mean must be close to zero, because errors level-out each other. All other statistics indicate the error according to their individual calculation. However, it is noteworthy that the average absolute mean is only 7 %. Furthermore, it had been noted that the average over all pixels in Table A.28 yields values similar to the average among all classes (Table A.27), which empirically confirms the chosen validation approach.

Hard Land Cover Classification of Germany

Figure 4.17 shows discrete classes and the certainty of class assignment of classification and reference for Germany. A visual interpretation of the hard classification indicates that class “arable” is often replaced by class “heterogeneous agriculture”, especially in southeastern and northwestern Germany, where the certainty of the classification is low. Furthermore, the area-proportion of “mixed forest” is clearly increased for upland areas in central Germany. The spatial pattern of class “artificial” matches well with the spatial pattern in the reference map, but is slightly larger in the classification. Altogether, there is a good agreement of all discrete classes in eastern Germany, but notable discrepancies are revealed for large portions of the remaining country. The cause of these differences will be explained in the next section on fuzzy classification analysis of Germany.

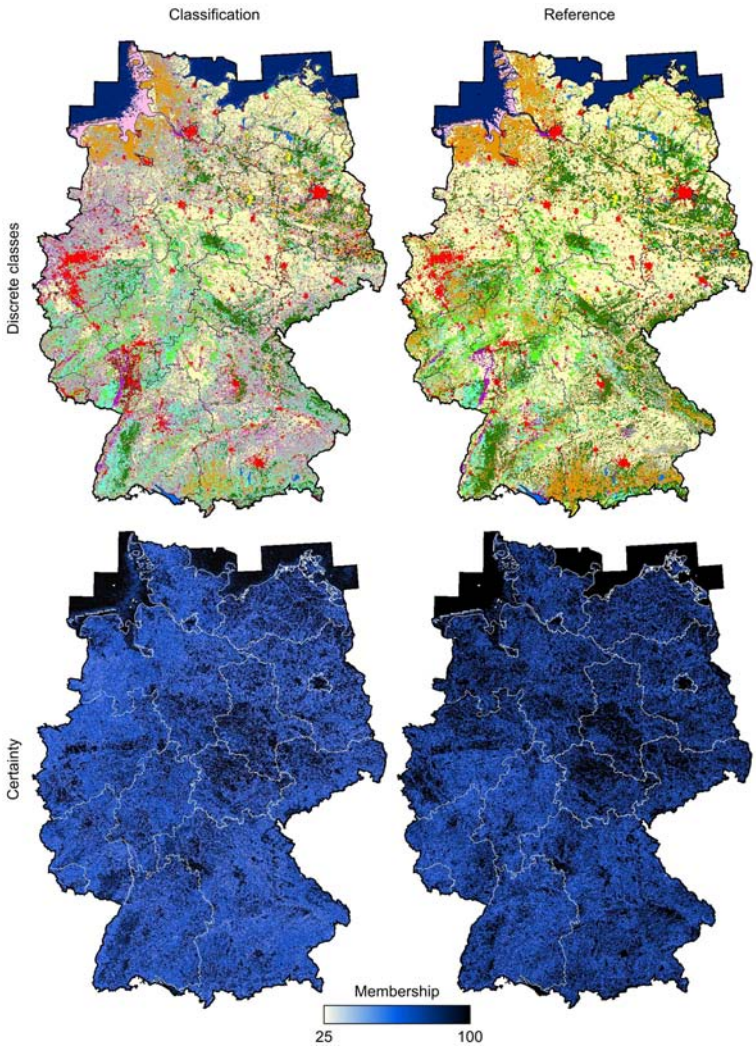


Figure 4.17: Discrete classification and reference map of Germany with the corresponding certainty of correct class assignment.

Note: The certainty of class assignment of the reference has been derived by the multi-scale analysis. Since large portions are homogeneous in the discrete reference map, also due to the MMU of 25 ha in CLC2000, the certainty of class assignment is markedly higher for the reference. For the legend see Figure 2.4.

The error matrix for the independent sample set of homogeneous pixels in the reference shows a generally good agreement (Table 4.3). The overall accuracy is 83 % and the Kappa coefficient yields 0.81. The user's accuracy is almost always higher than 80 %. Only class "heterogeneous agriculture" (58 %) shows a misclassification with "arable" and "permanent crops", but the other clearly overestimated class, "mixed forest", does not indicate high errors of commission. The producer's accuracy, however, reveals disagreements of "mixed forest" with "broadleaved forest" and "coniferous forest".

Table 4.3: Error matrix for land cover classification of Germany for 2001.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	User's [%]
1	Artificial	81	1	0	0	0	0	0	0	1	7	0	0	0	0	90
2	Arable	1	76	2	0	11	0	0	0	0	0	2	1	0	0	82
3	Permanent crops	2	1	72	1	4	0	0	0	0	0	3	0	0	0	87
4	Pasture	0	1	5	90	13	0	0	0	1	1	0	0	0	0	81
5	Heterogeneous agriculture	0	15	16	9	70	3	0	0	2	2	4	0	0	0	58
6	Broadleaved forest	0	0	0	0	1	92	0	15	0	0	1	1	0	0	84
7	Coniferous forest	0	0	0	0	0	0	98	18	2	0	1	0	1	0	82
8	Mixed forest	0	0	0	0	1	4	2	66	0	0	2	0	0	0	88
9	Herbaceous vegetation	2	1	4	0	0	0	0	0	78	3	9	0	0	0	80
10	Barren	12	4	0	0	0	0	0	0	6	85	2	0	0	0	78
11	Inland wetland	2	1	1	0	0	1	0	1	10	0	75	0	0	0	82
12	Coastal wetland	0	0	0	0	0	0	0	0	0	1	1	98	1	5	92
13	Inland water	0	0	0	0	0	0	0	0	0	1	0	0	95	6	93
14	Marine water	0	0	0	0	0	0	0	0	0	0	0	0	3	89	97
	Producers [%]	81	76	72	90	70	92	98	66	78	85	75	98	95	89	

Note: The (eleventh) independent sample set with 100 samples per class was used. User's and producer's accuracies are shown. See Figure A.3 for a general outline of the error matrix.

The upper row of Figure 4.18 depicts plots of the user's and producer's accuracy for the main land-cover classes of Germany. The plot of the user's accuracies shows generally high values for contextually uniform classes, while all mixed land-cover types, namely "heterogeneous agriculture" and "mixed forest", depict remarkable errors of commission. It is also notable that "heterogeneous agriculture" cannot be mapped at high certainty, which puts the high uncertainties in discrete class assignment in perspective. The related producer's accuracies indicate similar patterns. Clearly, class "heterogeneous agriculture" shows problematic values, even dropping with increasing certainty. Besides the inherent contextual heterogeneity, another reason is related to sampling and explained in the following section on fuzzy assessment. A remarkable gain in the producer's accuracy is indicated for "pasture" (47 %), which increasingly concentrates on high-confidence, large-patch pasture lands in northwestern Germany. The producer's accuracy of "mixed forest" is high, i.e. pixels of this class in the reference map are correctly assigned in the classification. At first glance this seems to contradict the findings of the confusion matrix analysis. However, samples of the

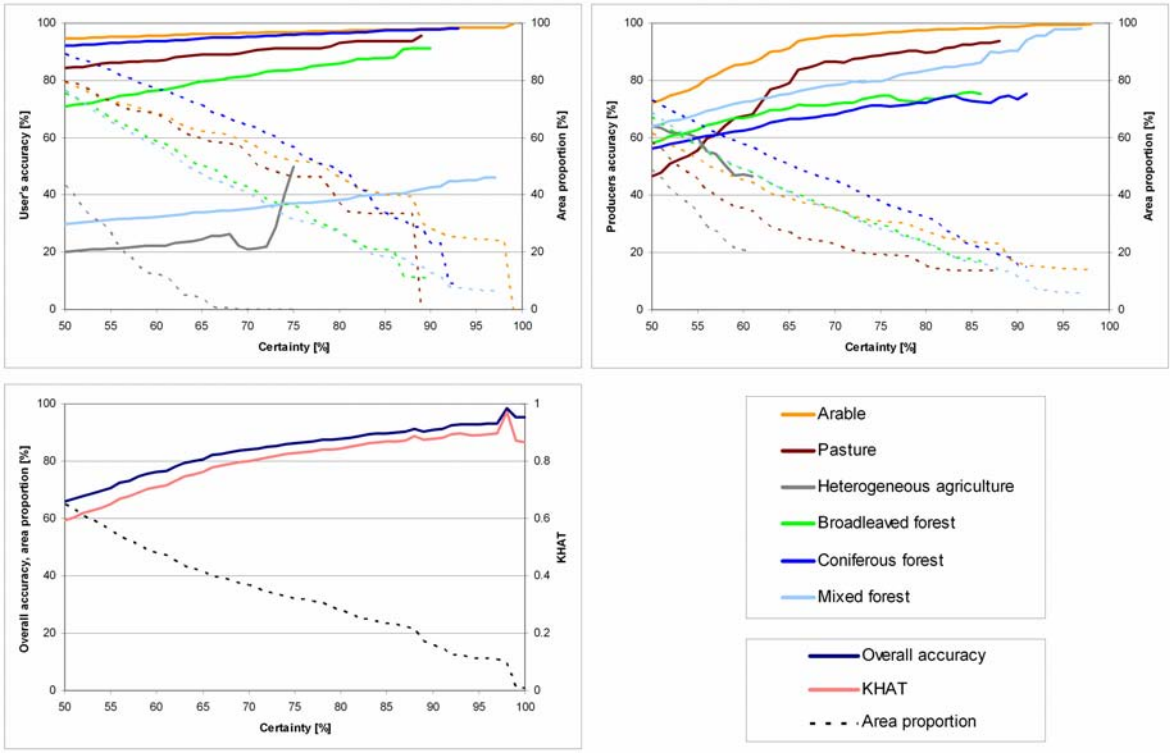


Figure 4.18: Accuracy assessment of the discrete map of Germany using classification certainty.
Note: For selected classes, the user's accuracy (upper left) and producer's accuracy (upper right) as well as the overall accuracy and KHAT (lower left) are depicted. While the solid lines indicate the accuracy, dashed lines show the corresponding decrease in area proportion with increasing certainty in class assignment.

error matrix are based on multi-scale analysis derived from reference data, while the certainty measure for the plots of Figure 4.18 is retrieved from the fuzzy classification.

The overall accuracy and KHAT steadily increase from 65 % (0.59) to 95 % (0.86). The small peak at very high confidence results from the drop in the available data. The difference between both accuracies ranges between 3 % and 6 %. Altogether, the general accuracy measures appear satisfactory.

Fuzzy Land Cover Classification of Germany

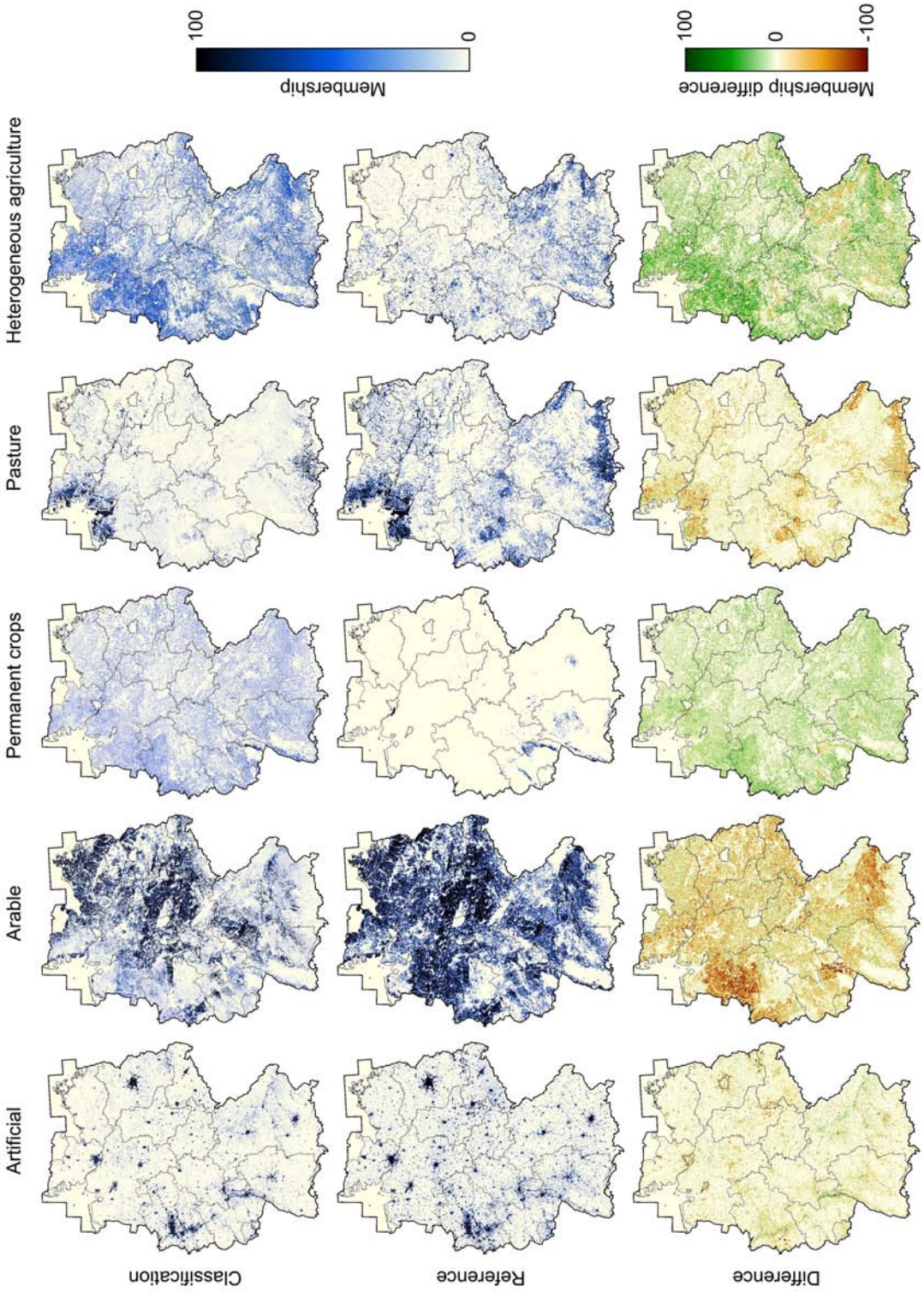
Including (excluding) marine water, agricultural lands cover an area of 57 % (62 %). Another broad category, that of forest, yields a proportion of 26 % (28 %). Together they comprise 84 % (91 %), hence the clear majority of the land. Therefore, it is interesting to analyze each category in more detail.

The first part of Figure 4.19 shows four agricultural classes: "arable", "permanent crops", "pasture", and "heterogeneous agriculture" (for water and wetlands see Figure A.5). The membership values of class "arable" are underestimated throughout Germany (see also Figure

4.20, top left, class 2). Furthermore, there are areas which are clearly underestimated, including northwestern Germany, the northern portion of the Rhine Graben, and southwestern Germany. While class “permanent crops” matches at the main sites, i.e. for vineyards in the Rhine Graben, hops north of Munich, and orchards in southwestern Germany and near Hamburg, it also shows some scatter throughout Germany. Clearly, this heterogeneous class of permanent agriculture, with temporally barren to densely covered areas, is difficult to map with automated approaches. Pasture lands in northwestern and central-southern Germany are revealed in the classification, but not for central Germany. Class “heterogeneous agriculture” shows a clearly differing pattern between reference and classification. Extensive portions of northwestern and southeastern Germany indicate high memberships, areas which have been “arable” in the reference.

The reason for this difference of memberships for class “heterogeneous agriculture” is related to the small-patch landscape, minimum mapping unit (MMU), and data sampling. The landscape in northwestern and south-central Germany is highly structured and the patch size is often beyond the resolution of 500 m. Secondly, CLC2000, the high-resolution input dataset, has a MMU of 25 ha. It has been noted in the discussion of the reference image generation (4.2.7 Accuracy Assessment) that the MMU can have a significant impact on the reference dataset generation, especially for highly structured small-patch regions. Despite the fact that the high-resolution reference map does not adequately indicate these small patches, image acquisition is still influenced by these fractional surface-cover types. Therefore, the high membership for class “heterogeneous agriculture” compensates for the underestimation of the uniformly defined class “arable land”. This may even seem more correct at a coarse spatial resolution, where a cell integrates several surface types.

A note has to be made on sampling of training and test data. The multi-scale analysis is the basic procedure for identifying potential sample pixels, starting at 100 % homogeneity (the importance of the homogeneity will be shown in 4.3.2 Sensitivity Analysis of the Classification Procedure). Among potential samples, areas with a high proportion of homogeneous cover will be preferred. This does not contradict the idea of adequate spatial distribution of samples. The samples are still equally spatially distributed among all candidates (see discussion in 4.1.2 Background on Decision Trees and for methodology in 4.2.4 Sample Generation and Subsetting). In the case of arable land, large fields in eastern Germany significantly contribute to sample candidates, more than in any other region. This becomes clear when analyzing the certainty image of the reference (Figure 4.17), where “arable” in northwestern and southeastern Germany does not yield 100 %. Eventually, a significant portion of samples for class “arable” is drawn from fields in eastern Germany. The sample set for “heterogeneous agriculture” is much smaller and candidates are mainly located in southeastern and northwestern Germany.


100
0
Membership
100
0
-100
Membership difference
Heterogeneous agriculture
Pasture
Permanent crops
Arable
Artificial
Classification
Reference
Difference

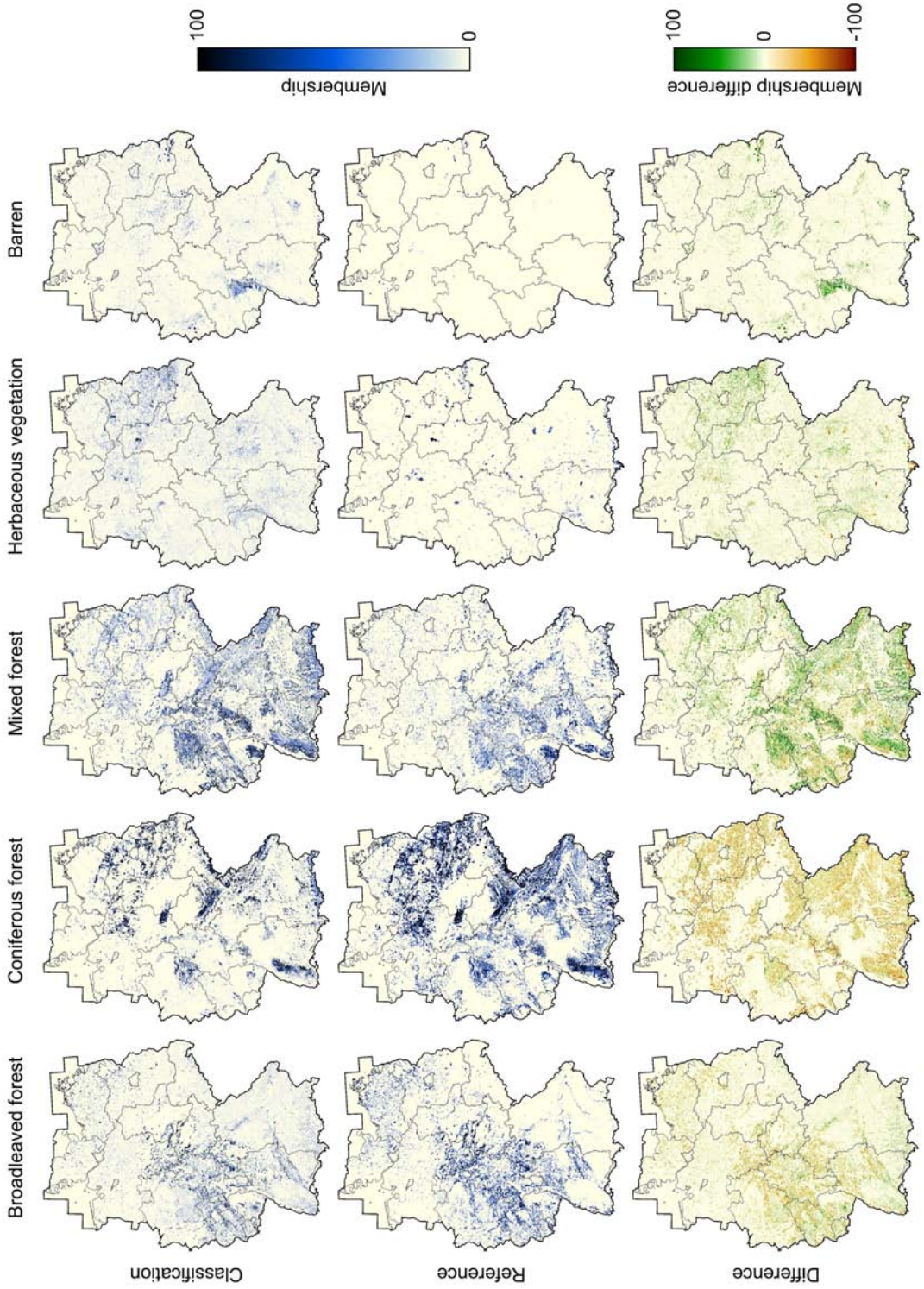


Figure 4.19: **Selected class-wise comparison between fuzzy estimates of classification and reference and the corresponding difference for Germany.**

Note: **Figure A.5 indicates the class-wise comparison for all other land cover classes of Germany.**

The other broad category, that of forests, comprises classes “broadleaved forest”, “coniferous forest”, and “mixed forest” (second part of Figure 4.19). There is a good spatial match between the reference and classification for “broadleaved forest”, and differences in membership values are remarkably low (Figure 4.20, top left, class 6). Needle-leaved forest areas, however, only coincide in the spatial arrangement for larger uplands such as the Black Forest, Thuringian Forest, Harz Mountains, and Westerwald as well as the Alpine Foothills and selected areas in the lake district of northeastern Germany. Still, there is a difference in the extent and membership values (Figure 4.20, top left, class 7). Interestingly, most of the difference is compensated for by “mixed forest”, which is overestimated in both area and membership values in the classification (Figure 4.20, top left, class 8). Connective mixed woodlands such as the Palatinate Forest and the west-facing slope of the Black Forest are matched well.

With regard to the general overestimation of forested land, there applies the same discussion as for classes “arable” and “heterogeneous agriculture”, i.e. an under-representation of the contextually uniform class “coniferous forest” is compensated for by a mixed class. The reason is less related to sampling, because both classes are well distributed in space with sufficiently homogeneous areas in all regions. It is rather an issue of the MMU limiting the variability on the ground and the contextual heterogeneity of class “mixed forest”. It is the wider distribution in feature space with multiple modes that led to the overestimation of class “mixed forest”. Nevertheless, at a coarser spatial resolution, class “mixed forest” may be an adequate characterization of many forests in Germany.

Class “herbaceous vegetation” mainly indicates military training areas in the reference, which are clearly matched in the classification. However, there is a notable scatter in eastern Germany, indicating the proportion of open land in the rather sparsely populated areas of southern Brandenburg. Furthermore, class “barren” reveals a clear spatial disagreement of “arable” in the northern Rhine Graben. This area obviously differs in its signature from other arable lands, e.g. in eastern Germany. Secondly, this area is intensively used, e.g. by vineyards or orchards; hence potentially being barren.

While settlements itself are mapped very well, matching all larger centers, there is an interesting effect among the classes “artificial” and “barren”. It compensates for an issue introduced by the generation of the legend and indicates the robustness of the approach and contextual correctness of the results. All artificial surfaces of the original CLC legend were merged into a single class (see Table A.19). However, this aggregation to CLC level 1 also contains a land-use class called “mines, dumps, and construction sites”, a unit which can also be regarded as barren from a land-cover point of view. The classifier recognizes these areas as “barren” with a high membership, although having been trained on different “barren” signatures such as beaches, sand, rock, or sparse vegetation (see Table A.19). Of course, the

effect results from the clear dominance of urban fabric in the class “artificial”, but the contextual correctness in class assignment is remarkable.

Larger mining areas are located west of Cologne, south of Leipzig, and in far eastern Germany. Although most strip mines in eastern Germany have been or are in the process of being re-naturalized, they still show clearly barren surface characteristics. Because the reference is based on the 14-class scheme, these areas are indicated as highly under- and overestimated spots for classes “artificial” and “barren”, respectively. In other words, although being referenced to “artificial” due to the definition in the sense of land use, they are classified to “barren” as the better land-lover descriptor for satellite data.

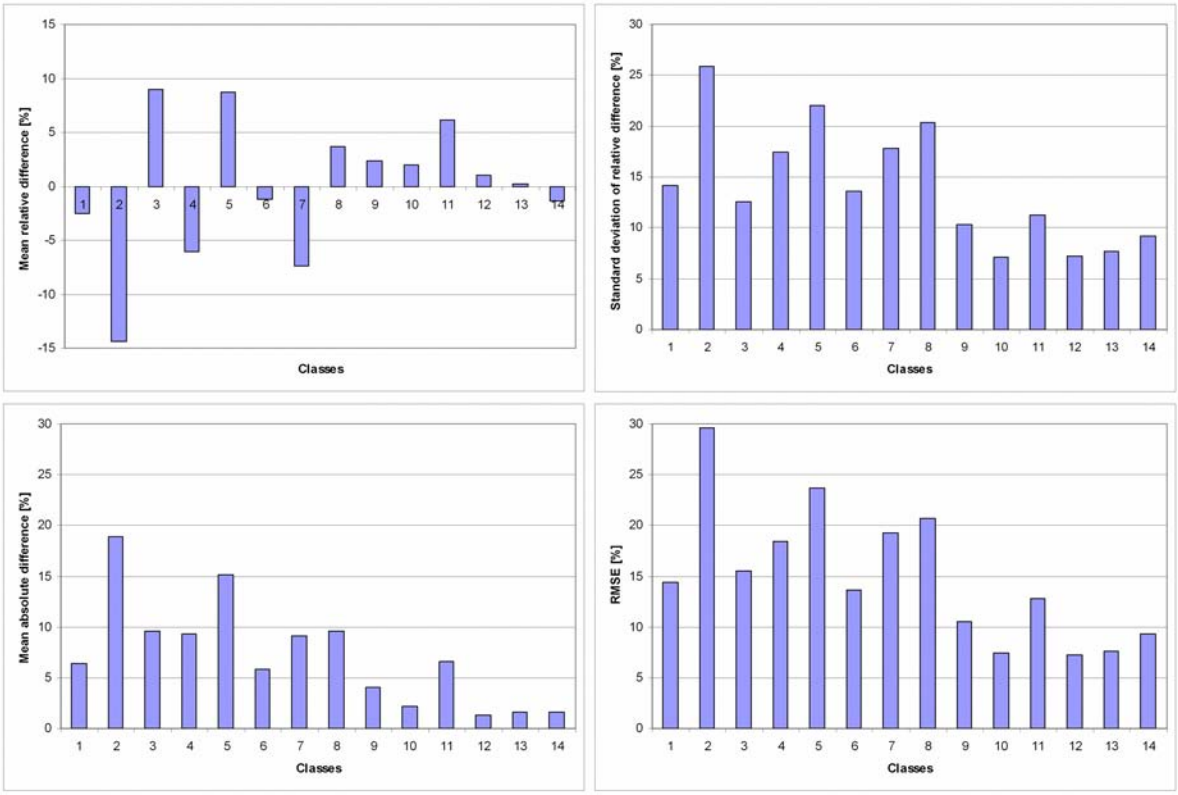


Figure 4.20: Selected error measures of class-wise fuzzy accuracy assessment for Germany.
Note: The relative mean difference (upper left) depicts the general trend, i.e. under- or over-estimation compared to the reference. The corresponding standard deviation (upper right) indicates the variability in the relative differences. The absolute mean (lower left) and RMSE (lower right) show the magnitude in the difference, which can be markedly higher than relative mean because of omitted balancing effects. For corresponding class names see Table A.19. For all error measures see Table A.29.

The error measures of the class-wise analysis (Figure 4.20, complete list in Table A.29) have been partly analyzed in previous paragraphs. The mean absolute difference and RMSE depict the expected pattern, a higher error for classes with either a large area proportion or contextually mixed surface type. Except for “2: arable” and “5: heterogeneous agriculture” all classes show an absolute mean error of less than 10 %. The clearer differences of memberships for “4: pasture” yield a higher RMSE compared to “3: permanent crops”, which

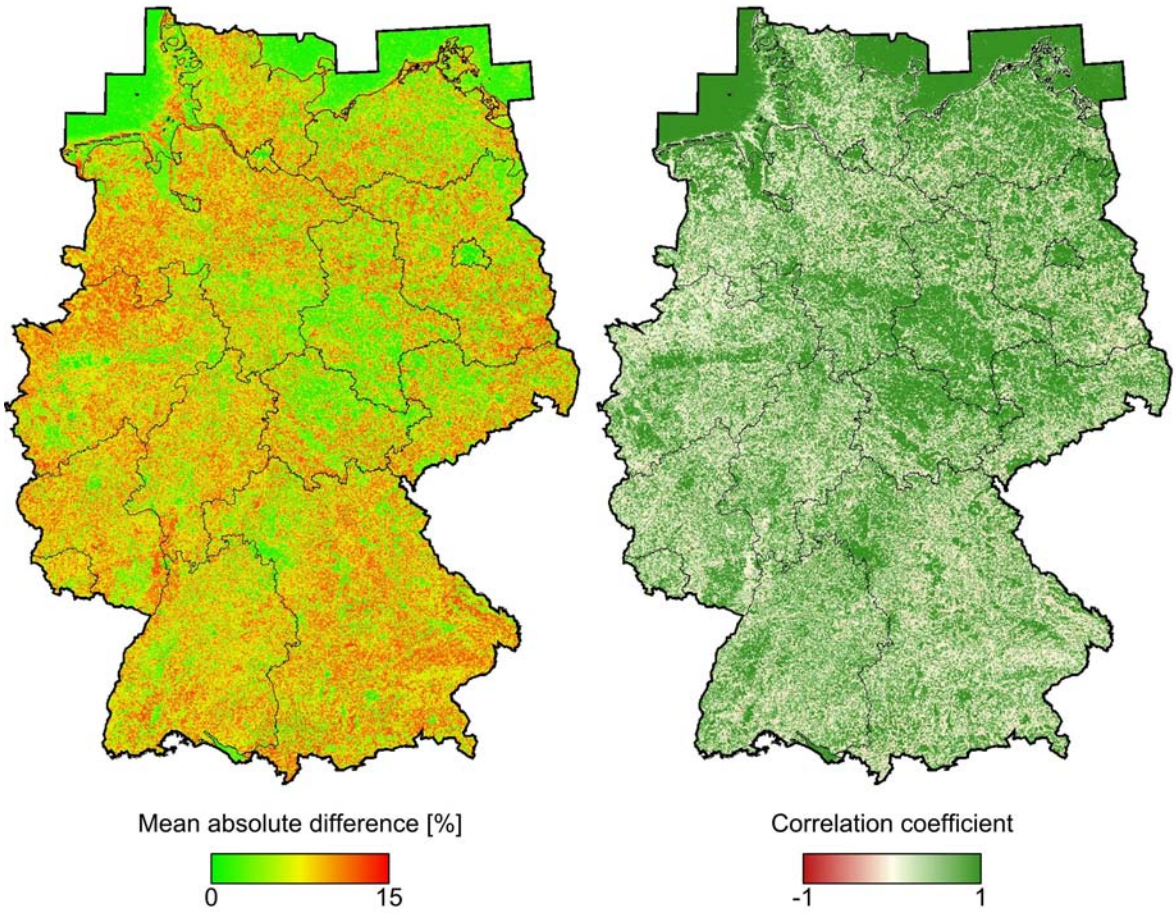


Figure 4.21: Selected error measures of pixel-wise fuzzy accuracy assessment for Germany.
Note: The correlation coefficient is rarely below zero and green tones indicate values above 0.7. All other error measures indicate the same spatial pattern of disagreement at different magnitudes. For summary statistics see Table A.30.

resulted in a reversal with the mean absolute error. Error compensation is also shown for the relative mean difference of classes “1: artificial”, “6: broadleaved forest”, and “8: mixed forest”.

The pixel-wise analysis, i.e. the comprehensive spatial display of error differences, is depicted in Figure 4.21. The mean absolute difference indicates low errors for urban centers, large agricultural areas in central and northeastern Germany, parts of grazing land in northwestern Germany, and selected coniferous and mixed forests of uplands. It should be noted that the agricultural district with large fields south of Würzburg has been particularly well mapped, which is also indicated by a correlation coefficient close to one. The majority of the image indicates small absolute mean differences of up to 10 %. Higher discrepancies are depicted for the underestimated arable region in the Northeast and Southwest. Further disagreements are shown for mismatches in forested land. Table A.30 indicates that the average of the absolute mean difference is only 7 % with a 4 % standard variation. Also, other error measures show acceptable differences.

Summary and Implications

This section has demonstrated a detailed analysis and assessment of both hard and fuzzy classification results for South Africa and Germany. It has applied the accuracy assessment strategy described in section 4.2.7 Accuracy Assessment for an optimal parameter setting of the time-series classification procedure developed in this study. Especially the hard and fuzzy analysis of the land-cover classification of Germany has discussed the importance of each individual assessment strategy.

Metrics-based decision-tree classification using MODIS time series and advanced classification strategies yield satisfactory results, in particular for the fuzzy estimation. Statistical summary values such as average absolute mean differences of 7 % or an average RMSE of 14 % to 15 % are indicators for a good agreement.

It can be argued whether the comparison to the reference, originating from the multi-scale comparison, is a valid approach. For some examples it had been shown that the fuzzy classification might result in better estimates of fractional surface cover than the reference, although this cannot be proved quantitatively. The inadequate representation of broad ecotones in the reference has been addressed for South Africa. This transition is well represented in the class-wise memberships, though, and therefore the assessment approach may introduce false errors.

An underestimation of class-memberships for one class is often compensated for by an overestimation of a contextually related land-cover type. This issue has been demonstrated for “thicket and bushland” in South Africa, which have been estimated to be “forest and woodland” or “degraded forest, woodland, thicket, or bushland”. In this respect, all classes would be recognized well at a thematically aggregated level such as “woody cover”. On the other hand, the considerable scatter of “irrigated agriculture” along the eastern escarpment is problematic. Land uses such as “irrigated agriculture” can be mapped well in semi-arid regions, but may be difficult to distinguish from other densely vegetated land-cover types. The clear overrepresentation of irrigated agriculture in the discrete map can be mitigated by introducing ancillary data or contextual knowledge, either using regional stratification or weights. However, this study has deliberately excluded additional and contextual knowledge in the classification process, because it aims at a rapid, automated, and fully transferable land-cover mapping procedure.

In the case of Germany, the assessment has been hindered by the small-patch landscape, which is insufficiently represented in the reference data due to a minimum mapping unit of 25 ha. Furthermore, each broad category of Germany, either agricultural land or forests, contains several thematically uniform and one broadly defined land-cover class. While the

tight definitions of uniform classes are applicable at a mapping scale of 1:100,000 with Landsat imagery (although some information is lost by the MMU of 25 ha), many of those classes mix at the scale of 1:1,000,000. Therefore, the clear increase in area-proportion of both mixed classes - heterogeneous agriculture (reference: 7 %, MODIS classification: 24 %) and mixed forest (reference: 5 %, MODIS classification: 11 %) - in the discrete classification can be still an accurate estimation employing 500 m MODIS data. A further validation of this statement could be conducted using GIS techniques or employing a dataset without a MMU.

Using contextual knowledge for the study region of Germany, the transition from uniform to mixed classes can be correct considering the scale of the analysis. The detailed class-wise membership analysis of the spatial distribution indicates that areas with small patches are preferably assigned to mixed classes. The rural small-patch landscapes of northwestern and southern Germany in particular are estimated to be “heterogeneous agriculture”. Both areas also may serve as potential sources for training mixed classes. On the other hand, large agricultural fields, especially in eastern Germany, remain as uniform classes with a high certainty. It can therefore be stated that the proportion of underestimated memberships of uniform classes is compensated for by the higher proportion of the corresponding mixed class.

Generally, the legend applied in this analysis has been rather detailed considering the variability and heterogeneity of Germany. In this regard, the accuracies and the low errors of class-memberships for many classes are very satisfactory. It could be argued, if the consideration of thematically mixed classes (“hererogeneous agriculture” and “mixed forest”) into this classification was appropriate. However, a significant portion of Germany is mixed, even at a high-resolution mapping scale of Landsat (with an MMU). This portion obviously increases with a coarser resolution.

However, sampling and MMU can only partly explain some difficulties during mapping. In particular agricultural land is a very heterogeneous class. Numerous crops are grown with different spectral and temporal characteristics. Furthermore, harvests in winter and summer are included. Although being uniform at first glance, in this regard agricultural classes are internally diverse, i.e. have a multi-modal frequency distribution. The characteristics depend on the spectral characteristics of the crops, the growing period (time), and eventual on functional use of the land.

Still, there is a good agreement in discrete classes among all broad categories of Germany, such as “urban”, “agricultural lands”, “forest and semi-natural land”, “wetland”, and “water”. In other words, even if there is some disagreement between reference and classification within each category, partly due to the reference map and the coarse mapping scale, the general class pattern matches well at a broad thematic definition (see 4.3.3 Capabilities for Classification Update and Change Detection).

4.3.2 Sensitivity Analysis of the Classification Procedure

This section aims at the analysis of the classification results, in particular their accuracy, and analyzes the influence of parameters which can be adjusted in the classification procedure. Obviously, this sensitivity analysis cannot present the entire protocol of classification assessment as provided in the previous section. It therefore selects suitable accuracies and error measures to illustrate the impact of the classification parameters on classification results and explains possible reasons for changes in accuracy. Some parameters will also show no significant impact on time-series classification. The section will evaluate results of both main study regions, except for the impact of the MMU, utilizing the NLCD2001 of Minnesota. The next sections will follow the sequence of modules of the classification process (partly shown in Figure 4.4), i.e. time-series generation, metrics derivation, multi-scale analysis, MMU issues, sampling schemes, and decision-tree parameters. A comprehensive summary and discussion will be provided in the last section.

Time Series Quality

Many regional to global mapping efforts employ time series. Since the only dataset of early studies originated from AVHRR data, the quality of time series ingested into the classification process may have an impact on classification accuracy. Of course, this issue has not been accounted for in earlier studies, because there was a limit of approaches to improve time-series generation.

In this study, time series of different quality levels were generated for assessing the influence on classification accuracy. For South Africa, four time series were processed: layer stacking without considering the data quality, and three levels of data-quality settings (see Table A.31 for NBAR and Table A.32 for LST). It should be noted, though, that original data (layer stack) already contain a remarkably high data quality. Hence, for South Africa, the improvement of time series by quality analysis is rather low. The remaining classification parameters are the same as in the general classification protocol (Table A.24).

The impact on the overall accuracy and KHAT of the error matrix derived from the independent sample set indicated a difference of less than 1 %. Furthermore, the assessment of hard classification and reference using classification certainty showed low discrepancies and no clear pattern among time-series quality settings.

Figure 4.22 illustrates the mean of the relative and absolute difference of each class membership, showing the general direction of the error and the overall magnitude, respectively. However, this class-wise comparison does not indicate differences between

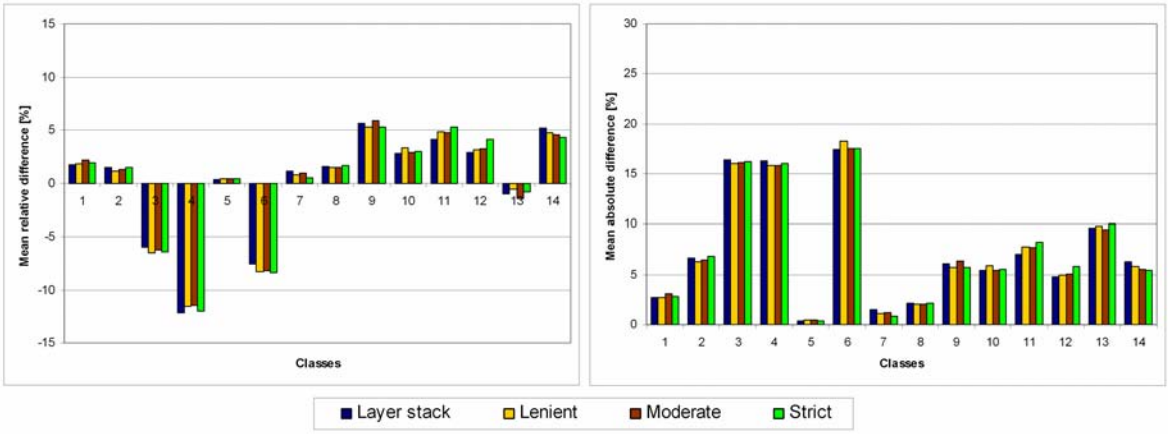


Figure 4.22: Differences between time series quality settings for South Africa using class-wise fuzzy error measures.

Note: The class-wise relative (left) and absolute mean (right) of the fuzzy assessment protocol indicate the negligible differences in trend and magnitude. For specifics of the quality settings for time series generation see Table A31 and Table A.32. For corresponding class names see Table A.17.

time-series quality settings. Highest discrepancies are less than 2 %, while, on average, there is less than 1 % difference. The pixel-wise analysis of the fuzzy classification indicates no clear differences in time-series quality for the spatial pattern. The minor discrepancies are solely due to some variation in the datasets, and have no significant impact on the discrete maps, especially for problematic classes such as “thicket and bushland”, “forest and woodland”, or “irrigated agriculture”. The summary table of the pixel-wise comparison showed a maximum difference in RMSE of much less than 1 %, and even lower average mean absolute differences.

The same analysis was conducted for Germany. The classification protocol of Table A.23 has only been modified by the time-series quality settings of (Table 3.2) and complemented by a layer stack. The remarkable impact of quality analysis on time-series quality of Germany has been shown in 3.3.2 Comparison between Improved and Original Time Series.

The overall accuracy varied between 82 % for lenient setting H7, and 84 % for intermediate snow setting H7-CS. Neither class-specific user’s and producer’s accuracy nor the 1:1 assessment of hard classification and reference with classification certainty indicated a clear ranking among time-series qualities.

The fuzzy class-wise error measures of the mean relative and absolute difference are depicted in Figure 4.23. The mean relative difference of class “6: broadleaved forest” shows a small overestimation for the layer stack and lenient time series H7, but all stricter quality settings resulted in an underestimation of class memberships. Despite this interesting pattern, the differences among time-series qualities are low, with 3 % at the most and 1 % on average. The class-wise difference images indicated minor discrepancies in the spatial pattern. For

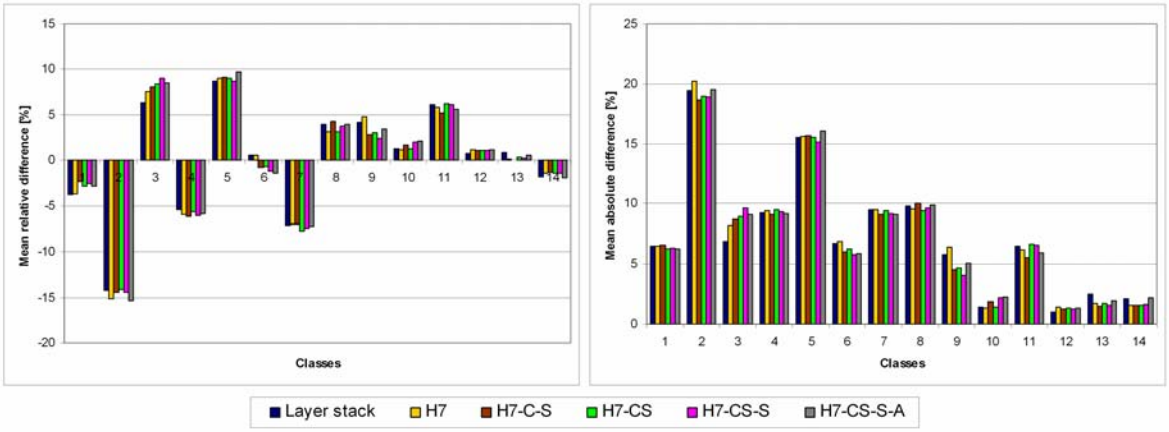


Figure 4.23: Differences between time series quality settings for Germany using class-wise fuzzy error measures.

Note: The class-wise relative (left) and absolute mean (right) of the fuzzy assessment protocol indicate the negligible differences in trend and magnitude. For specifics of the quality settings for time series generation see Table 3.2. For corresponding class names see Table A.19.

instance, layer stack and lenient quality setting H7 showed a clear underestimation of coniferous forest at the highest elevations of the Harz Mountains and Thuringian Forest. This effect can be attributed to clouds, which considerably influenced the signature / metric for classification. Pixel-wise fuzzy assessment does not depict clearly differing spatial patterns, and summary tables reveal a RMSE difference of clearly less than 1 %.

Both study regions, covering different environmental settings, indicated that the quality of the time series has a minor to no influence on the classification result. Although this was expected for South Africa with a low impact of quality analysis on time series, it is surprising that considerable changes in the temporal dynamics in Germany do not yield any notable discrepancies in classifications. On the contrary, it is imaginable that some noise patterns in the data may even have contributed to an accurate classification. However, the main reason for the low impact, the metrics, will be analyzed in the next section.

Metrics

All time-series classifications which have been found in the literature are based on sets of metrics, often rather simple statistics such as mean or standard deviation derived from time-series records. In many ways, metrics are regarded as stabilizers of the input data. Because a time series such as of NDVI data does not show a smooth plot, but several small variations related to noise in the data, it is very difficult to account for these uncertainties. Smoothing time series by filtering, e.g. harmonic analysis, removes the noise but also reduces the information content, which is disadvantageous for class separation. In other words, signatures of different thematic classes may become too similar in feature space.

Although the computation of metrics also reduces the information content compared to time series, it mitigates another difficulty, especially prevalent in smoothed data. Several algorithms, including decision trees and ANN, have difficulties if highly correlated data are used as input. Time series are commonly correlated in time, in particular for steadily increasing and decreasing variables such as surface-reflectance, vegetation indices, or biophysical measures. During tree generation, each parameter, i.e. observation of the time series, would be analyzed for the best separation of training data. If a decision is made, the respective observation together with its threshold would be used as a split criterion for class discrimination. However, if there is a small temporal shift for other regions in a large study area, the parameter may not be optimal for class separation. From this point of view, robust metrics appear more suitable. DeFries et al. (1995) provided a detailed study of suitable metrics sets and concluded that simple univariate statistics are most suitable.

Since metrics generation, in particular statistics employed in this study, is seen as a stabilizing process of noisy time series, the minor impact of time-series quality on classification accuracy is put in perspective. The metrics calculation reduced the higher variability and data noise to a high degree. The resulting data, input to subsequent decision-tree classification, do not show significant differences and therefore yield very similar classifications.

The following paragraphs will discuss two basic issues of metric generation, while all other parameters of image classification are retained (Table A.23 for Germany and Table A.24 for South Africa). On the one hand, the number of time-series parameters, e.g. surface-reflectance bands, can affect the classification. Tests using only NDVI metrics were compared to metrics sets of all available time series. For all classifications the basic set of statistics, i.e. mean, standard deviation, minimum, maximum, and range, was ingested into the classification.

On the other hand, temporal segmentation is believed to have a considerable impact on classification accuracy (see also Figure 4.5 and 4.2.1 Metrics). Annual time series were partitioned into three, two, or no temporal segment before calculating metrics, using the basic set of statistics (see above). However, if a time series was divided in three segments, this classification also considered metrics of two segments and no segmentation. Therefore, having five basic statistics, three segments resulted in 30, two in 15, and no partition in 5 metrics per time series. Each metrics set per time-series parameter was complemented with the number of harmonics and the respective explained variance. Altogether, the time-series classification of Germany using eight time series (seven surface-reflectance bands and NDVI) yielded 256 metrics, while for South Africa day and night LST were added, resulting in a set of 320 metrics. In a second step, temporal segments were generated from a thematic point of view rather than a static temporal partition. Here, a vegetation dynamics product (MOD12Q2) discriminating four phases of phenological development has been employed (Figure 4.5; see also 2.2.2 MODIS Land Datasets).

First, the analysis of metrics focuses on the study region of South Africa and employed the common approach, i.e. hard and soft classification assessment. Both, the number of segments and the metrics of time-series parameters (NDVI only or the full set of time series) were analyzed simultaneously. Assessing hard classification with the independent sample set did not indicate clear differences for a reduced number of segments for the full set of metrics (approximately 79 % overall accuracy). However, the NDVI metrics only showed a decreasing overall accuracy, from 61 % for three segments to 53 % for a single segment. Results for the phenology-based segmentation were intermediate with 56 %.

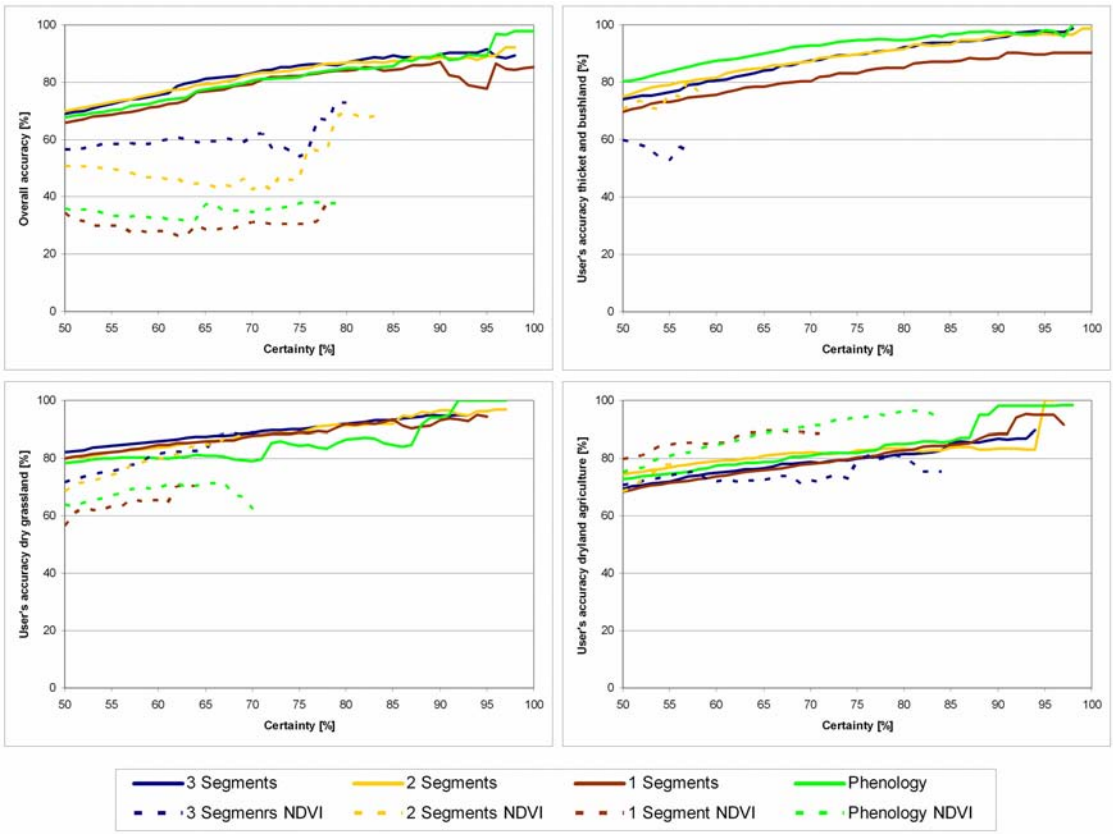


Figure 4.24: Comparison of temporal segmentation approaches and different sets of metrics for South Africa using discrete map comparison and classification certainty estimates.

Note: Overall accuracies (upper left) and user's accuracies for "thicket and bushland" (upper right), "dry grassland" (bottom left), and "dryland agriculture" (bottom right) are depicted. Solid lines indicate the accuracy using the full set of metrics (seven surface reflectance bands, NDVI, and day/night LST). Dashed lines show the accuracy using metrics of the NDVI only. Colors indicate the respective method of temporal partition.

The 1:1 comparison of discrete maps using the classification certainty is depicted in Figure 4.24. The overall accuracy in the upper left differs by not more than 5 % among segments of the entire metrics set. Only highest certainties show an increasing difference. The dashed lines illustrate the corresponding temporal partitions only using NDVI. Not only the accuracies are remarkably lower, but also high certainties are not reached. The ranking among segmentations using NDVI is very clear, with differences of up to 30 %, and it corresponds with the findings of the other hard classification assessment (see above).

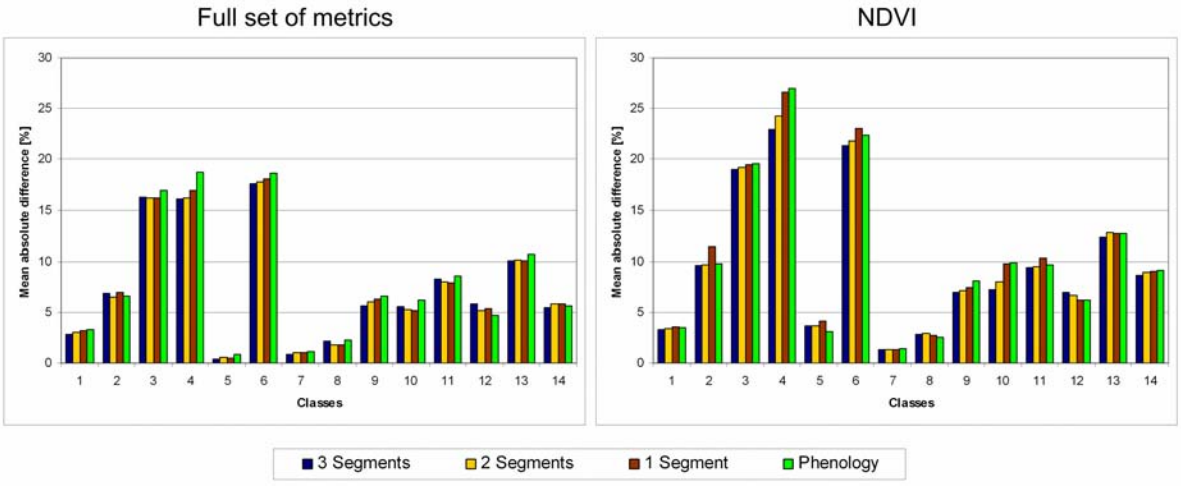


Figure 4.25: Differences between temporal segmentation approaches and different sets of metrics for South Africa using the mean absolute difference of the class-wise fuzzy accuracy assessment.

Note: Clear differences are shown with lower errors for the full set of metrics (left) compared to NDVI only (right). Temporal segmentation does not markedly improve the accuracy for the full set of metrics, but indicated some improvements for the NDVI. For corresponding class names see Table A.17.

In addition, Figure 4.24 depicts user's accuracies for selected classes. Class "ticket and bushland" exhibits a ranking among segmentations of the full metrics set. Phenological segmentation is indicated as being most useful for this vegetation class, and no segmentation yielded lowest user's accuracy. This pattern is completely reversed for "dry grassland" with differences of up to 10 %. It can be speculated whether the seasonality in the temporal signal of "dry grassland" is insufficient for a clear phenological characterization. For class "dryland agriculture", phenological segmentation is again a reasonably good strategy. Interestingly, however, class "dryland agriculture" can be mapped well by exclusively using NDVI data.

Fuzzy class-wise assessment is depicted in bar-plots using the absolute mean for magnitude derivation from the fuzzy reference set (Figure 4.25). Solely using NDVI metrics leads to higher errors compared to the entire set. For selected classes such as "4: shrubland and low fynbos" or "6: dry grassland", the difference to the full set of metrics on the left-hand side is more than 10 % respective 5 %. However, the segmentation strategies show only minor differences, in particular for the full set of metrics. Some classes, e.g. "4: shrubland and low fynbos", depict higher discrepancies of up to 4 % for the NDVI. Interestingly, the temporal partition using vegetation dynamics data yielded higher errors for several classes. However, discrepancies are not very distinct. The pattern indicated that merely sparsely- and non-vegetated classes, such as "4: shrubland and fynbos" or "9: barren" yield higher errors with phenology-based temporal partition, in particular for metrics based on NDVI. In case of the full set of metrics these differences are also attributed to the information of individual spectral bands, which does not necessarily relate to timing of vegetation phenology.

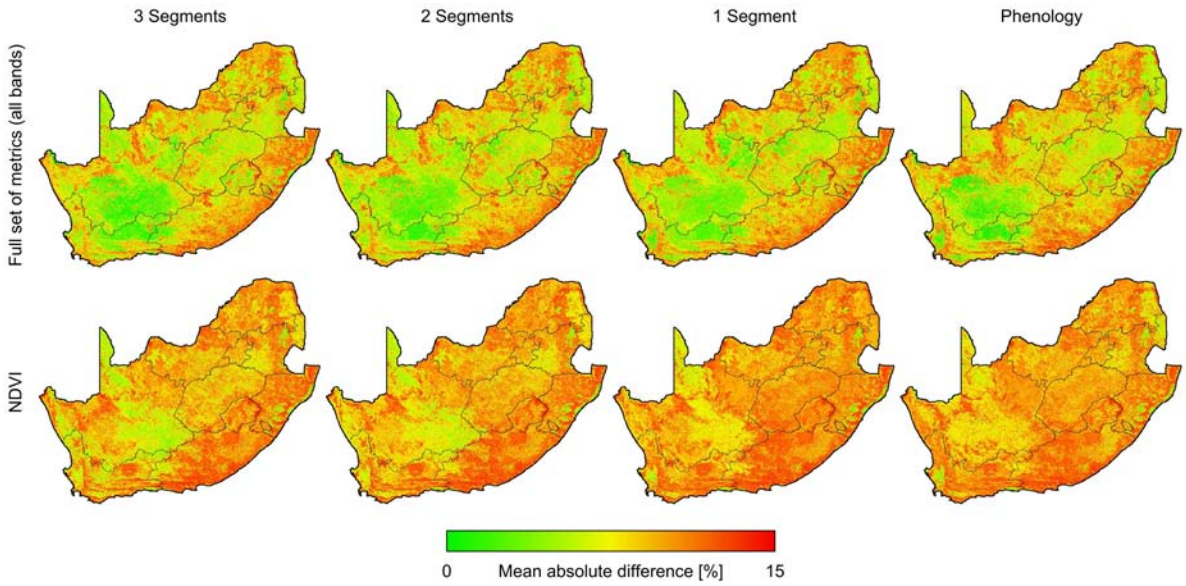


Figure 4.26: Comparison of temporal segmentation approaches and different sets of metrics for South Africa using the mean absolute difference of the pixel-wise fuzzy accuracy assessment.
Note: The spatial display reveals lower errors for the full set of metrics compared to the NDVI. No differences in the spatial pattern are shown for temporal segmentation of the full set of metrics (upper row). The lower row illustrating the use of metrics from the NDVI only shows an increase in error with no segmentation. The use of the temporal partition using phenological dates is not as good as using three segments strictly.

A last assessment to account for the spatial distribution of error in the fuzzy classification is shown in Figure 4.26, using the mean absolute difference measure. It becomes immediately clear that there is no significant spatial variability in error occurrence, if all time series were used for metric generation. Statistically, the average of the mean absolute difference (or RMSE) among partitions is low and ranges between 7.4 % (15.4 %) and 7.9 % (16.3 %) for the full set of metrics. In contrast, using only metrics of NDVI yields markedly higher errors. Even for three segments, the eastern and central part depicts higher errors. Furthermore, this effect increases if only two segments or no segmentation are generated. The result of the phenology-based approach is visually similar to using two segments. The summary statistics such as average mean absolute difference error (or RMSE) for NDVI-based metrics indicate this increase by values between 9.7 % (19.3 %) and 10.6 % (20.9 %). Thus, the difference between full metrics sets and exclusively using NDVI is revealed, but the discrepancy among segmentation strategies is not clearly indicated by fuzzy summary statistics.

The study region of Germany has been evaluated with the same statistics. The overall accuracy of hard classification is presented in Figure 4.27, either employing the independent sample set (left-hand side) or the certainty of classification (right-hand side). Using the independent sample set, overall accuracies indicate a 20 % difference between the full set of metrics and NDVI. While there is no difference among segmentations for the full set of metrics, overall accuracies using exclusively NDVI ranges between 56 % and 67 %. Although the sample set did not indicate a ranking for the full set of metrics, the 1:1 assessment using

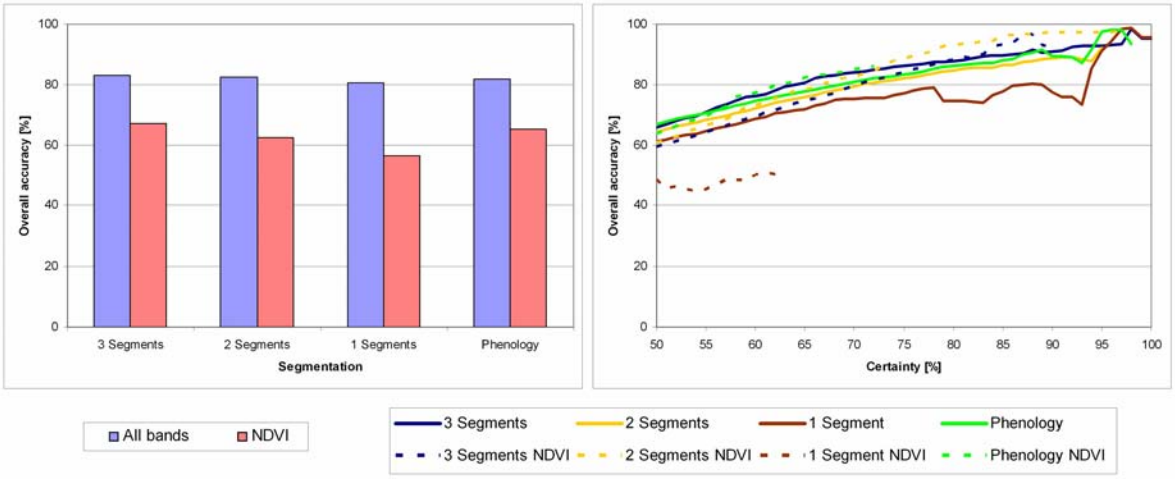


Figure 4.27: Comparison of temporal segmentation approaches and different sets of metrics for Germany using discrete map assessment techniques.

Note: The independent eleventh sample set has been used for the bar plot of overall accuracies on the right-hand side. It reveals dramatic differences between the sets of metrics. The left-hand side depicts the discrete map comparison using classification certainty estimates. Here differences among metric sets and segmentation procedures are not clearly indicated.

classification certainty indicates differences of up to 20 %, in particular for higher confidence assignments (right-hand side of Figure 4.27). Even more interesting, the employment of NDVI metrics only is partly indicated to perform better than the entire set of metrics, especially for two segments. This result contradicts the clear findings of the independent sample set (left-hand side of Figure 4.27) and requires further analysis using fuzzy measures.

Figure 4.28 depicts the class-wise accuracy assessment. The expected pattern, i.e. NDVI metrics performing worse than employing the full set of metrics, is clearly shown. The discrepancy among segmentation strategies is again low for the entire set of metrics. For the NDVI-based approach (right-hand side), class “arable” and “permanent crops” show clear differences. Hence, segmentation has an impact of up to 12 % of the mean absolute difference, if only NDVI metrics are employed. A spatial pattern in the pixel-wise accuracy assessment was not shown. The pixel-based summary statistics indicate an average mean absolute difference (or RMSE) of 7.3 % (14.5 %) for the full set of metrics. No differences are shown for segmentation strategies. Using NDVI only, the average errors range between 8.1 % (16.4 %) for three segments and 9.2 % (17.8 %) for no segmentation. Similar to South Africa, the difference between using only NDVI or the full set of metrics, the summary statistics of the fuzzy classification indicate a less clear picture among different segmentations.

Clearly, metrics computation has a high impact on the classification. This is particularly well illustrated with hard-classification assessment, albeit both approaches only consider either homogeneous samples of the reference data or rather certain data of the classification. The result appears less clear for fuzzy assessments, especially with regard to the temporal

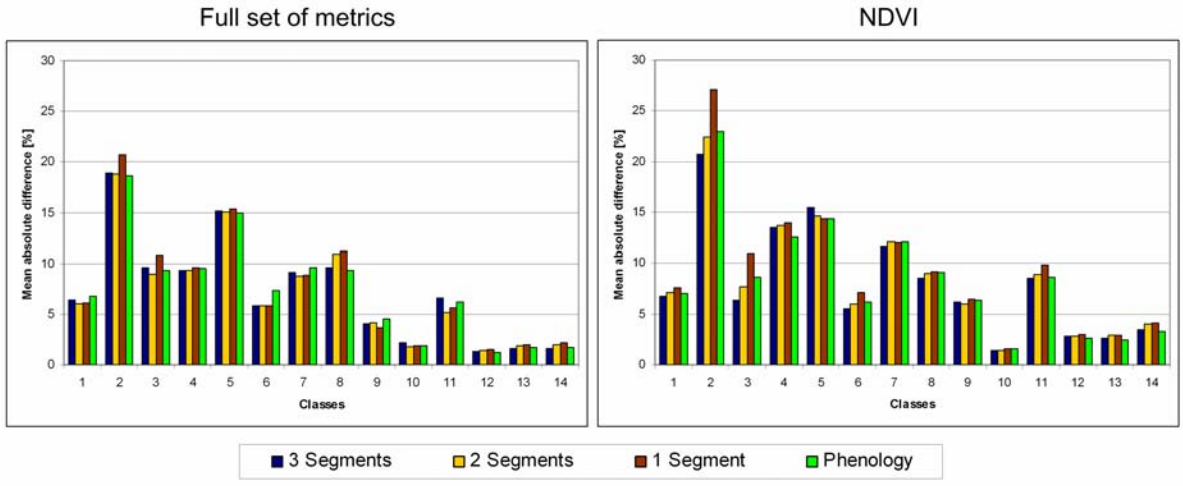


Figure 4.28: Differences between temporal segmentation approaches and different sets of metrics for Germany using the mean absolute difference of the class-wise fuzzy accuracy assessment.
Note: Clear differences are shown with lower errors for the full set of metrics (left) compared to NDVI only (right). Temporal segmentation does not markedly improve the accuracy for the full set of metrics. Improvements are indicated for the NDVI. In contrast to South Africa the use of partition with phenological dates seems to perform better for vegetated classes. For corresponding class names see Table A.19.

segmentation. Besides issues related to the reference map (for a detailed discussion of both study sites see also 4.3.1 Land cover Classification of Germany and South Africa) the reason is two-fold. First, fuzzy assessment examines all pixels and also accounts for low memberships. Still, class- as well as pixel-wise assessment reveals moderate differences. Secondly, overall statistics, such as average mean absolute difference or average RMSE from pixel-wise assessment, provide highly reduced information of the error measure due to averaging over a large region and memberships of 14 classes. Therefore, even small differences of 1 % may indicate a significant impact, if they are caused by a small subset, i.e. the error image has to be considered.

However, all temporal segmentation strategies reveal a ranking of NDVI-based classifications. For both study sites, three segments of 4-month periods seem most appropriate. For Germany, phenology-based temporal partition is ranked next, while half-year segments appear second-best for South Africa. No segmentation was always worst. This ranking, however, is only a general view and can be different for distinct classes. Vegetation classes seem to profit from temporal segmentation, but the impact of non-vegetated units is less clear. Especially phenology-based segmentation performs well for vegetated areas with a clear temporal difference in both study sites. In contrast, results for sparsely to non-vegetated regions indicate higher errors than employing rigid segments. In this respect, dynamic temporal segmentation can be useful for well-vegetated regions and if the number of time series is limited to only NDVI data.

It could be speculated whether a further temporal division could have improved the result of NDVI-based classifications. However, the partition in three segments reduced the number of composites to 8 for South Africa (23 composites per year) and 15 for Germany (46 composites per year). A further partition would even further reduce the number of composites employed for statistics utilized for metrics generation.

On the other hand, the full set of metrics does not indicate markedly decreasing accuracies with less than three segments. Even the use of annual metrics does not yield increasing errors. Obviously, employing a suite of annual metrics based on spectral data sufficiently discriminates classes. Or expressed differently, the information added by temporal segmentation does not clearly improve the accuracy of the classification.

For South Africa it can be speculated whether an exclusion of the LST from classification would have led to a decrease in the accuracies. LST can be supportive for image classification in semi-arid regions, because subtle differences might not be represented in surface-reflectance bands. Day and night LST, however, add rather different information, and are sometimes good indicators of land-cover differences in sparsely vegetated regions (Lambin and Ehrlich 1996, Lambin and Ehrlich 1997). Furthermore, LST is often used as a split parameter in the decision trees at higher hierarchies, which indicates the importance of this time-series parameter. A third reason for this speculation is the slightly higher difference between NDVI-based results and full sets of metrics for South Africa compared to Germany.

Multi Scale Analysis

The multi-scale analysis is a fundamental module of the classification procedure. It aims at scaling-up high-resolution thematic data. Simple and commonly applied operations to spatial aggregation for continuous and thematic data are the majority rule (retrieving the most frequent value), maximum or mean value assignment, or resampling with nearest neighbor or cubic convolution functions. Besides the majority rule, none of the above techniques is deemed useful for categorical data.

As the spatial scale ratio between coarse MODIS and high land-cover resolution is remarkable, always set to 1:10 in this study, it appears suitable to compute the proportion of each class for every coarse-resolution cell. This operation has several advantages. First, it provides the possibility to assess the homogeneity of each coarse-resolution pixel, which is, for instance, employed for sampling of training and test data. The accurate sampling of training data is a crucial but often tedious process if carried out manually, because classifiers heavily rely on the correctness of the sample sets. Essentially, training data are the source on which the classifier is built. In turn, the classifier serves for decision making if applied to all

other data. As this study aims at automated classification, an approach to automate robust training-data selection had to be developed. Selecting homogeneous pixels from training data seems to be an appropriate possibility. Secondly, the fractional estimates of each class can be regarded as a fuzzy reference map. As such they are employed for fuzzy accuracy-assessment strategies (for a discussion see 4.2.7 Accuracy Assessment). Thirdly, multi-scale analysis has also helped deriving the classification scheme, because it analyzes the potential of accurate classification from a data perspective.

In order to mitigate inaccuracies due to slight geometrical displacements between high and coarse spatial resolution, high scan angles, or the adjacency effect of the point spread function of a sensor, the multi-scale analysis also accounts for adjacent high-resolution cells (see also 4.2.2 Multi Scale Analysis of High Resolution Datasets). It is assumed that the results will be more accurate if the multi-scale analysis also accounts for neighboring areas. This section of the sensitivity analysis will focus on the importance and limitations of this neighborhood consideration using data of South Africa and Germany.

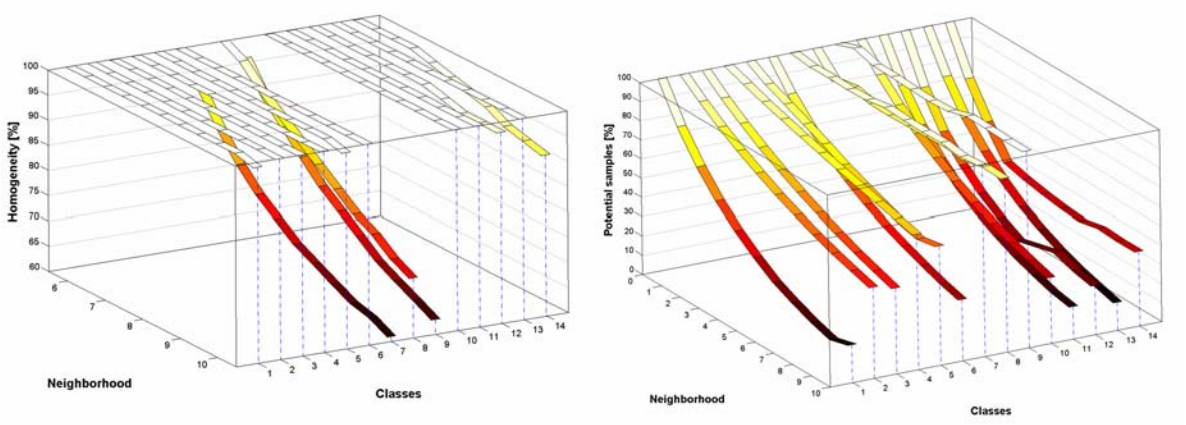


Figure 4.29: Decrease in homogeneity and proportion of potential samples with increasing neighborhood of multi-scale analysis for South Africa.
Note: The left-hand side depicts the homogeneity of each class which has to be accepted to provide at least 1,100 samples. Some classes show a decrease in homogeneity with increasing neighborhood. This usually corresponds with a leveling in progression of the proportional potential samples (right), e.g. for class "14: urban". For class names see Table A.17.

Figure 4.29 depicts the class-wise homogeneity and the proportion of potential samples as a function of neighborhood for South Africa. Phrased in simple terms, by increasing the neighborhood the number of samples candidates decreases. However, this study employs equal proportions of samples for each class and the size is set to 1,100 samples per class. For each neighborhood and class the process starts at homogeneity of 100 %. If there are more than 1,100 candidates a subset will be selected ensuring equal distribution in space. However, if the set of candidates is lower, the homogeneity is reduced by one percent, until at least 1,100 samples are derived. The left hand side of Figure 4.29 depicts the decrease in

homogeneity with increasing neighborhood, and the right-hand side indicates the corresponding proportional decrease of samples, scaled to the sample size at no neighborhood. For instance, several classes show no decrease in homogeneity in order to achieve the minimum of 1,100 samples. However, the set of candidates proportionally decreases due to an enlarging area for multi-scale analysis. Other classes such as “8: wetland” or “14: urban” decrease in homogeneity. At the same time the graph of the potential sample set levels off at the respective proportion for approximately 1,100 samples. Class “7: water” and “9: barren” are leveled in the proportional sample set from the beginning. In both cases, even if no neighborhood was employed, a reduction in homogeneity was necessary to achieve 1,100 samples.

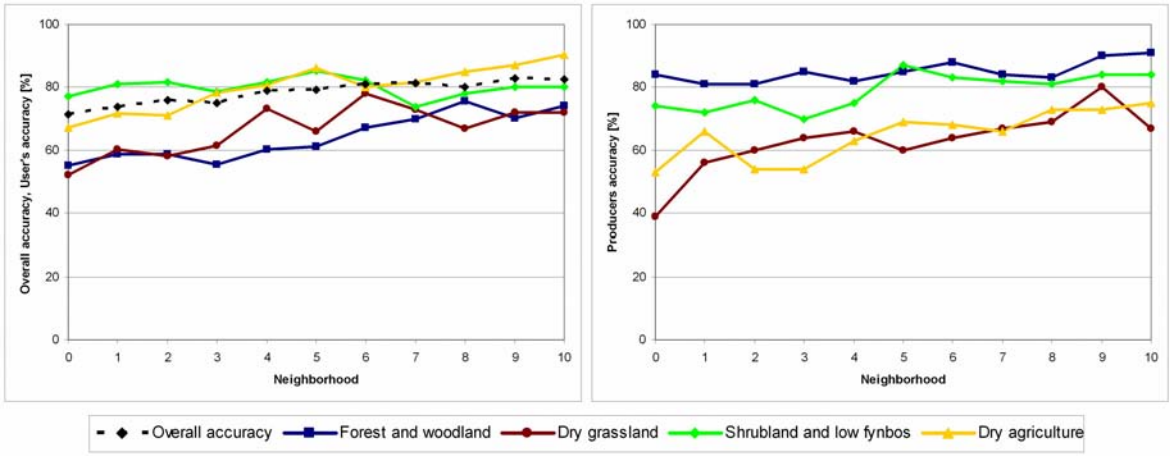


Figure 4.30: Course of classification accuracies with increasing neighborhood for South Africa.
Note: The assessment was conducted with the independent (eleventh) sample set. The overall accuracy is depicted by the dashed line in the left diagram. Selected user’s accuracies for most important classes are shown on the left-hand side and corresponding producer’s accuracies on the right-hand side. Small neighborhoods have resulted in an increase in accuracy, caused by higher certainty in training data. The plots level off for larger neighborhoods due to smaller or even heterogeneous sample sets.

Figure 4.30 depicts overall, user’s (left), and producer’s accuracies (right) using the independent sample set for hard classification. The overall accuracy indicates a steady increase with enlarging neighborhoods, ranging between 71 % and 82 %. Some user’s and producer’s accuracies show increases in accuracy of up to 20 %, but the average increase is approximately 10 %. The graph, however, does not illustrate a steady increase, but variations of up to 5 %. Still, the trend of increasing accuracy with a growing area is clear.

The analysis using 1:1 hard-map comparison between reference and classification plotted against certainty reveals a slightly different pattern (Figure 4.31). For instance, the overall accuracy shows highest values for a small neighborhood of three pixels. No enlargement and a moderate size of five pixels always depict 2 % to 3 % less accuracy. Large neighborhood settings indicate another 5 % to 8 % decrease in accuracy. The pattern of lower accuracies for large neighborhood sizes is also indicated for the user’s accuracy of class “dryland

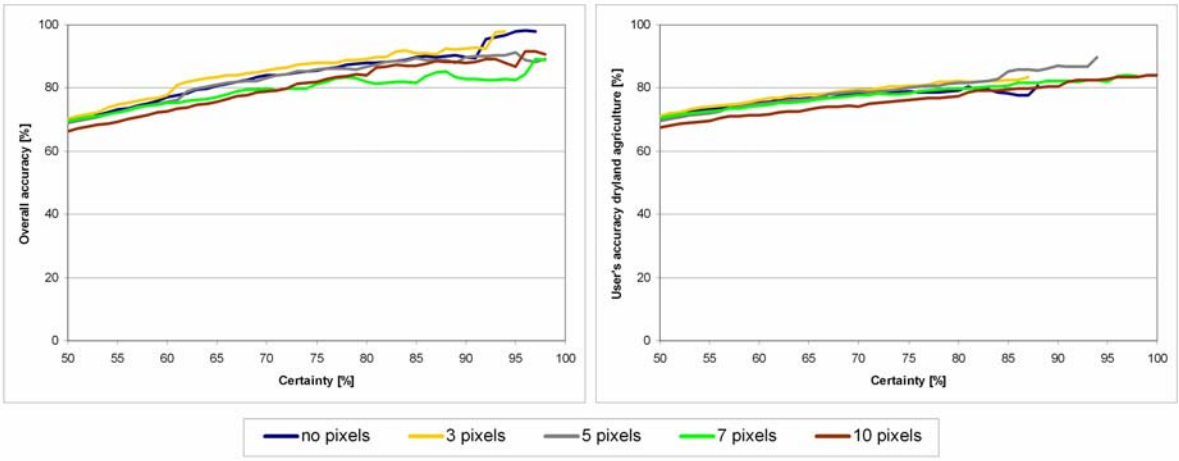


Figure 4.31: Difference between selected neighborhoods using discrete map assessment with classification certainty for South Africa.

Note: The overall accuracy on the left shows best values for small neighborhoods. The user's accuracy of class "dryland agriculture" (right) indicates low accuracies for large neighborhoods.

agriculture" at the right-hand side of Figure 4.31. The difference among neighborhoods is 4 % to 6 %. Again, a neighborhood of ten pixels indicates lower accuracies. Although it is still possible to draw 1,100 samples from class "dryland agriculture" even at large neighborhoods, the set of candidates decreased markedly from 10,740 for no enlargement to 1,833 (17 % compared to no enlargement) at a neighborhood of ten. Potentially, this reduction in samples introduced regional bias, i.e. it is not possible to derive samples indicating the full variability of class "dryland agriculture" in feature space.

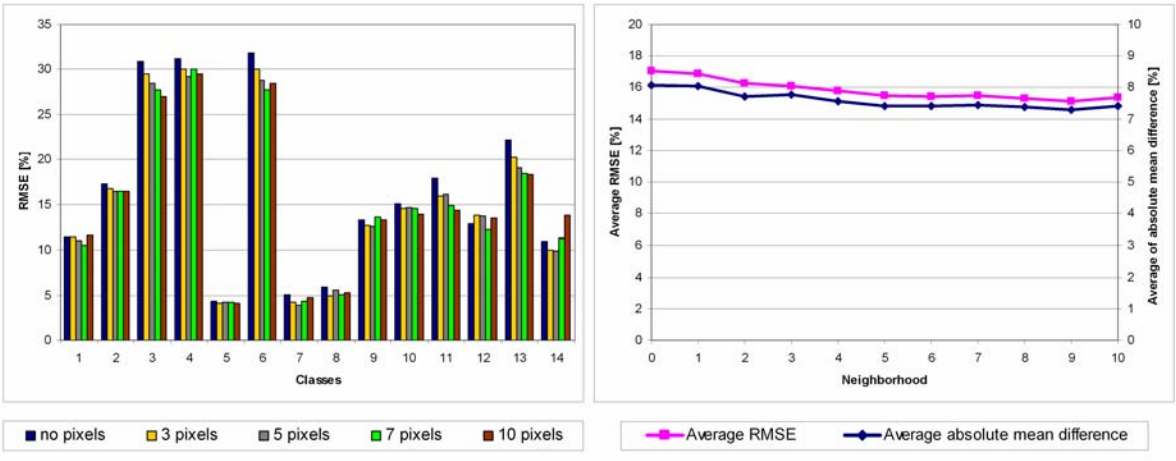


Figure 4.32: Difference and course of neighborhoods using fuzzy class-wise and pixel-wise assessment for South Africa.

Note: The class-wise RMSE measure (left) indicates a trend towards higher errors for no neighborhood and very large neighborhoods. For class names see Table A.17. The course of summary statistics such as average RMSE and average of absolute mean difference (right) show the decrease in error for small neighborhoods, which levels off for larger areas.

Soft accuracy assessments for South Africa are shown in Figure 4.32 using the RMSE for class-wise assessment (left-hand side) or average RMSE and mean of absolute differences from the summary table of pixel-wise analysis (right-hand side). The class-wise assessment indicates decreasing errors with increasing neighborhoods, especially for dominating land-cover classes such as “3: thicket and bushland”, “4: shrubland and low fynbos”, “6 dry grassland”, and “13: dryland agriculture”. The difference between neighborhood sizes is of up to 5 %. It appears that the decrease in RMSE is highest for smaller neighborhoods and levels off for larger adjacent areas. This pattern of first decreasing and then constant error for higher neighborhoods is confirmed by the summary statistics of the pixel-wise analysis.

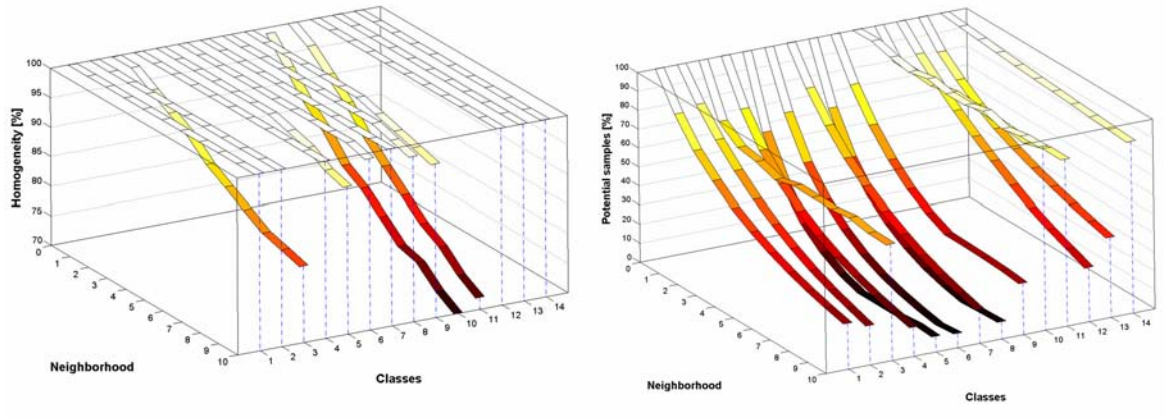


Figure 4.33: Decrease in homogeneity and proportion of potential samples with increasing neighborhood of multi-scale analysis for Germany.
Note: The left-hand side depicts the homogeneity of each class which has to be accepted to provide at least 1,100 samples. Some classes show a decrease in homogeneity with increasing neighborhood. This usually corresponds with a leveling in progression of the proportional potential samples (right), e.g. for class “3: permanent crops”. For class names see Table A.19.

The class-wise homogeneity with regard to growing neighborhoods of Germany is shown in Figure 4.33. It may seem surprising that the required 1,100 samples can be drawn for many classes, even for high neighborhoods. Only scattered classes “10: barren” and “11: inland wetlands” require a significant reduction to 70 % homogeneity. Due to the higher heterogeneity of land cover in Germany, it was expected that a decrease in homogeneity would be necessary. In contrast to 1 km-resolution data for South Africa, MODIS data of Germany have 500 m spatial resolution. Besides the impact of the MMU of 25 ha, it becomes clear that the homogeneity is also a function of the absolute size of the coarse-resolution cells as well as the extent of the study area to gather potential samples.

Figure 4.34 illustrates the overall user’s, and producer’s accuracy using the independent sample set. First of all, the increasing graphs seem to be less varying compared to class-wise accuracies of South Africa (Figure 4.30). The overall accuracy rises with growing neighborhoods from 76 % to 85 %. Similar increases are also shown for selected classes in

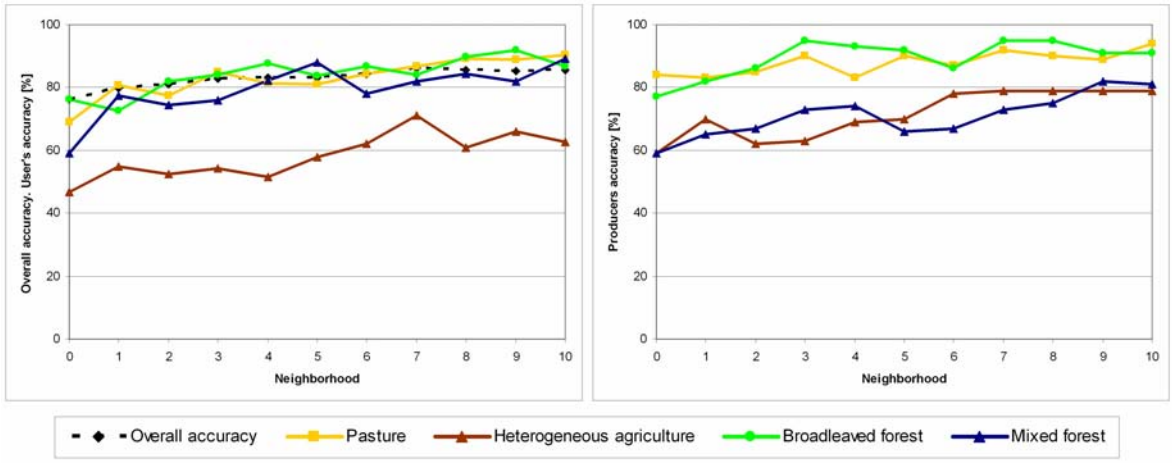


Figure 4.34: Course of classification accuracies with increasing neighborhood for Germany.
Note: The assessment was conducted with the independent (eleventh) sample set. The overall accuracy is depicted by the dashed line in the left diagram. Selected user's accuracies for most important classes are shown on the left and corresponding producer's accuracies on the right-hand side. Small neighborhoods have resulted in an increase in accuracy, caused by higher certainty in training data. The plots level off for larger neighborhoods due to smaller or even heterogeneous sample sets.

user's and producer's accuracy. Mixed forest indicates the highest gain with 30 % in user's accuracy. Some plots indicate a leveled accuracy for large neighborhoods. This pattern, e.g. for class "heterogeneous agriculture", can be explained in comparison with homogeneity depicted in Figure 4.33. The plot levels off at a neighborhood of seven pixels, exactly when the minimum threshold of 1,100 samples can only be reached by a reduction in homogeneity. Of course, a slight reduction in homogeneity for sample selection does not necessarily coincide with leveled or decreasing accuracies but is a possible reason.

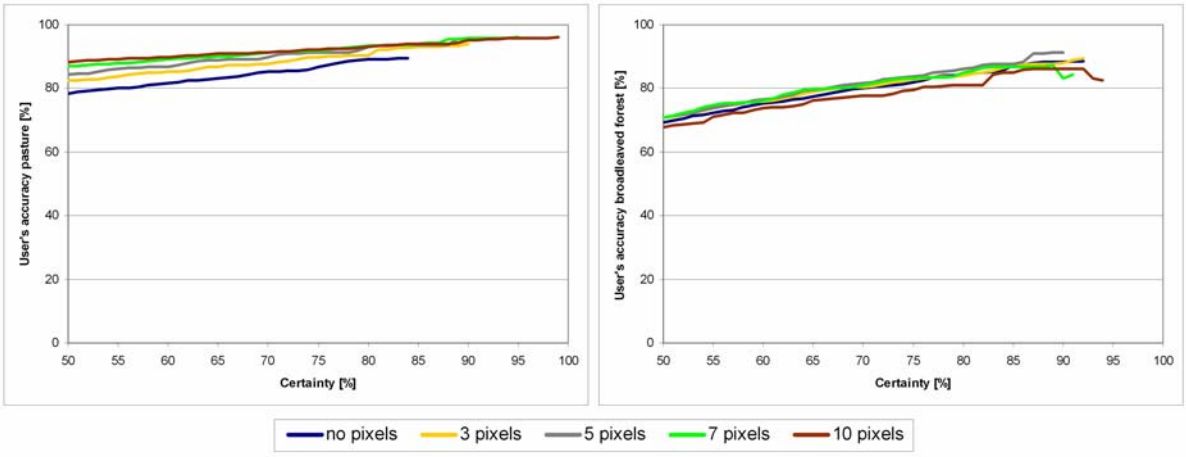


Figure 4.35: Difference between selected neighborhoods using discrete map assessment with classification certainty for Germany.
Note: The user's accuracy oft class "pasture" (left) shows best results for large neighborhoods. In contrast, the user's accuracy of class "broadleaved forest" (right) indicates best accuracies for moderate neighborhoods.

The evaluation of hard-classification accuracy using the classification certainty is shown for the user's accuracy of classes "pasture" and "broadleaved forest" (Figure 4.35). The difference in accuracy between neighborhood extents is 10 % and 5 % for pasture and broadleaved forest, respectively. However, a different ranking is exhibited. Class "pasture" (left-hand side) shows a clear increase from no to large neighborhoods. The pattern for "broadleaved forest" (right-hand side) indicates lowest accuracies for highest neighborhoods, followed by no neighborhood. Intermediate adjacency extents are best, and only differ slightly among each other. The interpretation of this pattern leads to the conclusion that there could be a threshold in neighborhood increase, on which the accuracy begins to decrease. Possible explanations such as insufficient representation of class characteristics in feature space due to spatially restricted sampling have already been discussed for South Africa (see above).

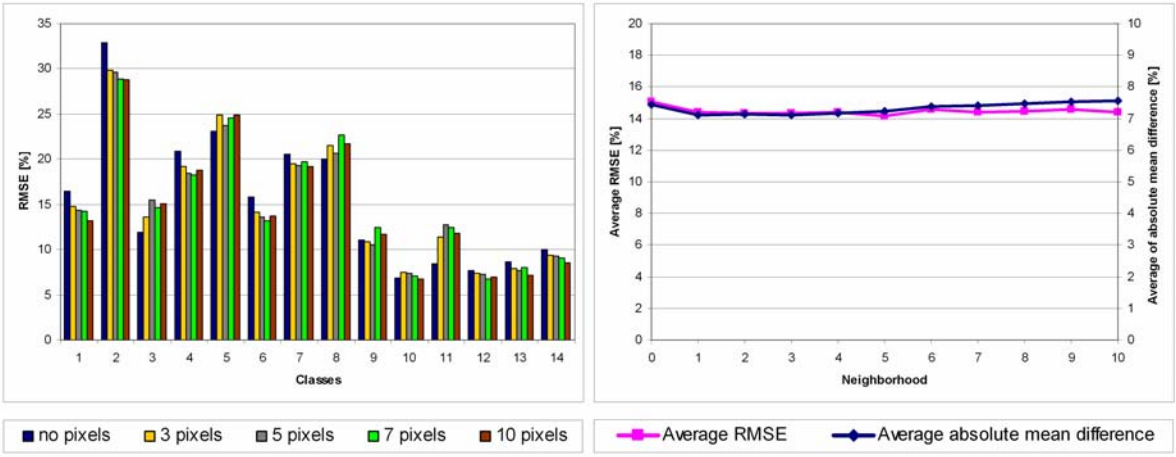


Figure 4.36: Difference and course of neighborhoods using fuzzy class-wise and pixel-wise assessment for Germany.

Note: The class-wise RMSE measure (left) indicates a trend of higher errors for no neighborhood and very large neighborhoods. For class names see Table A.19. The course of summary statistics such as average RMSE and average of absolute mean difference (right) show the decrease in error for smallest neighborhoods, but a reverse for neighborhoods larger than 5.

Fuzzy assessments using class- and pixel-wise statistics are depicted in Figure 4.36. Both indicate less clear results compared to South Africa. The class-wise assessment on the left-hand side indicates a decrease of the RMSE with growing neighborhood of up to 4 % for dominating and contextually uniform land-cover classes, including "2: arable", "4: pasture", 6: broadleaved forest", and "7: coniferous forest". Also, class "1: urban" shows a marked decrease of error with increasing neighborhood, indicating the importance of sampling very homogeneous areas. Mixed classes, such as "5: heterogeneous agriculture" and "8: mixed forest", however, indicate an unclear pattern and partly a slight increase in RMSE. This is attributed to the inherent mixed content of these classes, and the spatial heterogeneity with decreasing homogeneity for large neighborhoods. The summary plot on the right-hand side of Figure 4.36 shows a range of only 1 %. Both graphs rapidly drop in error for small

neighborhoods, but in particular the average of the absolute mean difference begins to increase for more than five adjacent pixels.

The presentation of the results for both study sites showed the importance of increasing the neighborhood around each coarse-resolution cell for more robust results. This has a high impact on samples derived by multi-scale analysis. A part of this sample set remains independent and is used for conventional accuracy assessment. In particular the assessment using this sample set indicates a clear gain in accuracy of, on average, 10 %, which can be markedly higher for specific classes. Hard assessment using the certainty indicates an increase of approximately 5 %, if neighborhoods are employed.

The picture is not so prominent for fuzzy assessment techniques. Despite generally smaller differences of fuzzy assessment techniques, the selection of the reference image is also a critical issue. Essentially, the enlargement in area by multi-scale analysis results in an individual reference dataset for each neighborhood. In order to introduce no bias, every fuzzy classification has been compared to its corresponding reference data. In other words, the reference data differ, which has an impact on the difference in error among neighborhoods. Still, the importance of a moderately increased neighborhood for more accurate results is shown for both study sites.

However, hard and fuzzy accuracy assessments also indicate that highest neighborhoods do not necessarily yield better results. This is mainly attributed to a restricted set of potential samples and, even more important, a decrease in homogeneity to draw 1,100 samples. The first fact is due to the robust, but solely statistical design of this approach. Spatially distributed sampling is assured at all levels, considering the set of candidates. However, multi-scale analysis with increasing neighborhoods draws smaller sample sets, potentially with a regional bias, by excluding more heterogeneous areas. In other words, if a class is broadly defined or contextually heterogeneous, a multi-modal distribution in feature space is quite likely. If these multiple cluster centers cannot be represented due to biased training data, the class will be inaccurately mapped. In this sense, a small-patch class is difficult to map with increasing neighborhoods. The effect becomes even more serious if the minimum threshold of samples is reached. In this case, the homogeneity criterion is reduced until a sufficiently large sample set can be drawn. This obviously introduces heterogeneous pixels into the sampling, which can have rather unpredictable effects on the classification process. For selected classes and large neighborhoods, homogeneity had been reduced to 70 % for Germany and 61 % for South Africa.

In this respect, a note should be made on the definition of the classification schemes employed in this study. As noted in 2.3.2 Land Cover Datasets, classes of the initial land-cover classifications have been partly aggregated. The recodes were conducted by means of

contextual knowledge and by considering the homogeneity of potential samples for multi-scale analysis. The first process employed some knowledge of the study site in conjunction with the hierarchical structures of the initial classification. Classes, in particular land-use units not detectable with remote-sensing data, have to be merged or excluded. Furthermore, the coarse spatial resolution limited mapping of classes with small patches. On the other hand, the analysis of the homogeneity usually requires class-wise testing, at which homogeneity the minimum number of samples (1,100 in this study) can be achieved. If the homogeneity is too low, the class has to be merged with other, thematically related classes. Both procedures yielded the classification schemes employed in this study. Therefore, it can be stated that multi-scale analysis is a central module in image classification of this study, which requires some interaction with the analyst to achieve the most suitable classification scheme. From a data point of view, it weights the quality of sample data against its abundance for training and partly testing the classifier.

Minimum Mapping Unit

Many high-resolution land-cover classifications are spatially aggregated using a defined minimum mapping unit. This operation reduces the noise, also known as classification speckle, while the map itself remains at the same high spatial resolution. In this sense, defining a minimum mapping unit (MMU) is not a multi-scale operation, but merely a contextually-based spatial aggregation. The reduction is carried out by identifying each individual patch using a clumping algorithm with a four- or eight-neighbor rule. If the area of a patch is below the MMU threshold, these pixels are assigned to neighboring patches. For both main input land-cover maps, NLC1995 and CLC2000, the MMU was set to 25 ha by the data provider.

The reduction of thematic accuracy seems suitable for many applications at a finer scale, but clearly has implications on the multi-scale analysis described above. This section therefore mimics four levels of MMU and illustrates their effect on multi-scale analysis and classification accuracy. It is clear that both common study regions cannot be employed. Therefore the NLCD2001 of Minnesota was used for analysis, where no MMU had been defined initially. First, the reference map was reprojected to 100 m spatial resolution. Next, datasets with a MMU of 9 ha, 25 ha, and 100 ha were generated using the eight-neighbor rule for clumping. The classification process is similar to all other study areas, and details can be obtained from Table A.25. Similar to the multi-scale analysis, the reference image had to be generated for each MMU separately. This influences the assessment, in particular for the fuzzy accuracies (for discussion see above).

The number of patches reduced from almost 2 Mio for no MMU to 11,358 for a MMU of 100 ha. The proportion and partly the spatial distribution vary considerably according to the dominating patch sizes and its relation to all other classes. The impact of the MMU on homogeneity and available samples is depicted in Figure 4.37 for each class. While, for instance, class "1: water" is homogeneously distributed, even an MMU of 100 ha had only a minor effect on the sample set which is depicted proportionally in relation to the maximum sample set (right-hand side of Figure 4.37). For all other classes, an increase of the MMU generated more homogeneous patches. In order to draw 1,100 samples, the homogeneity did not need to be severely reduced. For instance, for class "5: grassland and pasture" the homogeneity was reduced to 79 % for no MMU to draw a set of 1,100 samples (left-hand side of Figure 4.37). The homogeneity increased with enlarging minimum patch size. For 100 ha, much more than 1,100 sample candidates were available, which causes the dramatic difference in the set of potentially available data (right-hand side of Figure 4.37). Therefore, two effects are possible. For a moderate MMU, classes with patches above the MMU threshold will gain in size, because they replace the adjacent areas of patches below the MMU. Due to this reduction in heterogeneity of the map, multi-scale analysis will result in more homogeneous pixels. For an extreme MMU, on the other hand, highly fractional classes might disappear or do not yield sufficient samples. However, the latter case was not reached in this example.

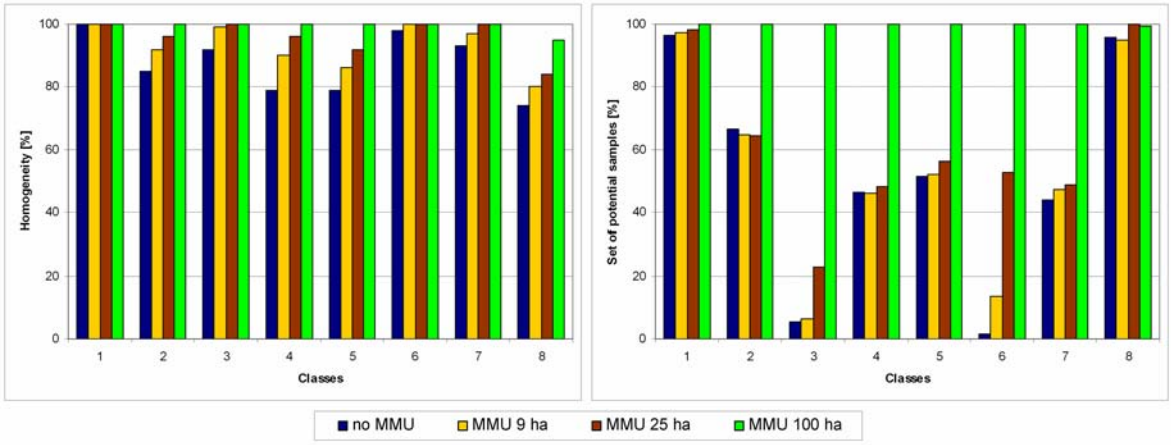


Figure 4.37: Decrease in homogeneity and proportion of potential samples with increasing minimum mapping unit for Minnesota.
Note: The left-hand side depicts the homogeneity of each class. Some classes show a decrease in homogeneity for no or a small MMU to draw at least 1,100 samples. The corresponding bar-plot on the right-hand side depicts the proportion of potential samples. For class names see Table A.20.

The overall accuracy of the hard classification with the independent sample set yielded decreasing accuracies with an increasing MMU, ranging from 96 % for no restrictions to 89 % for a MMU of 100 ha. Class-specific user's and producer's accuracy generally confirm this ranking. On the other hand, the hard classification assessment using the classification

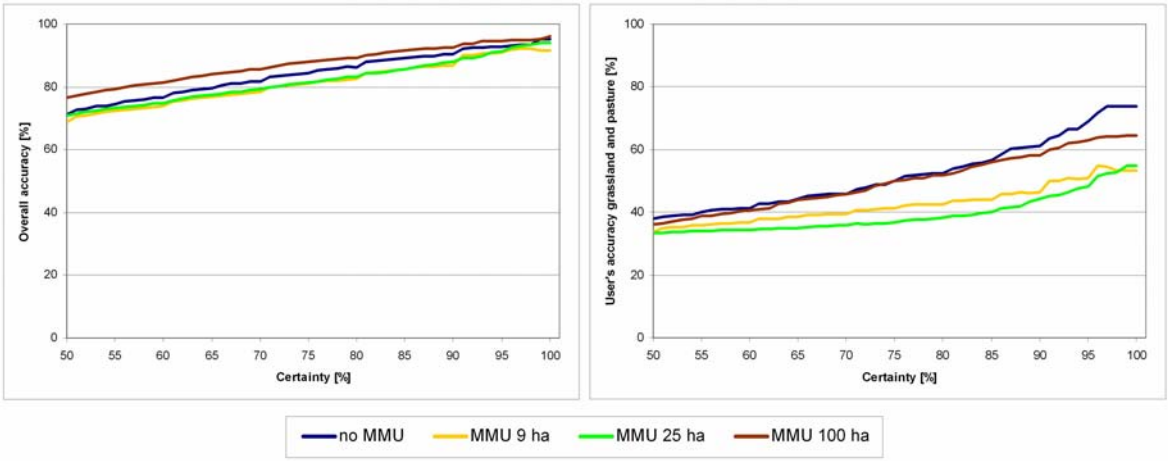


Figure 4.38: Differences in discrete classification accuracy using different minimum mapping units for Minnesota.
Note: The overall accuracy (left) indicates best results for largest MMU followed by no MMU. The user's accuracy of class "grassland and pasture" (right) shows similar patterns with best results for no MMU. The reason for good results with very high MMU is related to the homogeneous area with high certainty in the reference map.

certainty reveals a different pattern (Figure 4.38). The plot of the overall accuracy of the left-hand side indicates highest values for a MMU of 100 ha. Second best is no MMU, and low and moderate MMU thresholds did not show clear differences among each other. This pattern was confirmed by several user's accuracies, as exemplarily shown for class "grassland and pasture". Here, however, no MMU and a 100 ha minimum patch size is similar at lower certainties, but the accuracy of no patch restrictions increases more rapidly. Again, both intermediate MMU settings are markedly lower.

Figure 4.39 provides a spatial view of the effect of the MMU on reference data for class "grassland and pasture". This, in turn, has an impact on sample-set definition and therefore affects the classifications. The difference is depicted in the bottom row. With regard to the reference, no MMU depicts a wide spatial distribution of "grassland and pasture". Because also small patches contributed to the memberships of the reference image derived by multi-scale analysis, this reference map appears rather smooth. Small and moderate MMU of 9 ha and 25 ha already reduced class proportions in area and resulted in a rougher appearance. This effect becomes markedly clear for a minimum patch size of 100 ha. Areas with low memberships in the reference without MMU disappear in the reference of a large MMU which, in turn, affects multi-scale analysis. On the other hand, areas with already larger patches of class "grassland and pasture", e.g. in the center of the study area, increased in membership value. Here, small patches of other classes were assigned to surrounding "grassland and pasture" patches. Therefore it can be stated that an increase in MMU significantly affects the heterogeneity of the input data and therefore will influence the reference map as well as the classification.

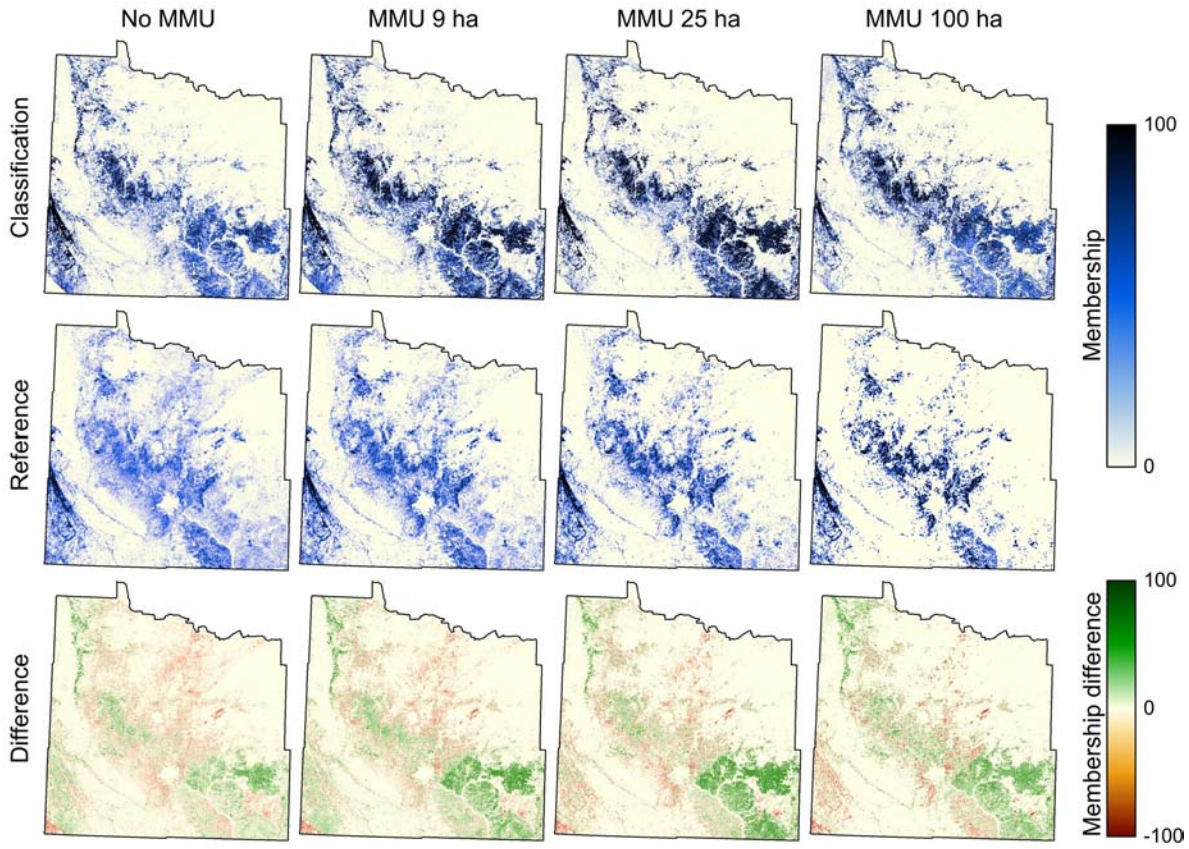


Figure 4.39: Comparison of class-memberships for class "grassland and pasture" in Minnesota using different minimum mapping units.
Note: The reference and classification reveal marked differences between MMUs. This proves the high influence of the MMU on the multi-scale analysis and the classification result.

This effect of different MMU thresholds on multi-scale analysis has a direct impact on potential samples for classification (top row of Figure 4.39, see also Figure 4.37 class 5). The fuzzy membership of the classification without MMU appears rather smooth. Intermediate MMUs of 9 ha and 25 ha reveal spatially similar results, but with higher memberships in the southeastern portion of the study area than in the reference. The effect is reversed for an even higher aggregation to 100 m, but the spatial extent is also reduced. The difference maps indicate that "grassland and pasture" is well mapped for most areas, indicated by small under- and over-estimations. The southeastern portion, which is dominated by class "agriculture", appears to be problematic, because no reference indicates this area as rangeland.

Fuzzy assessment using class-wise bar-plots for the mean absolute difference and RMSE are depicted in Figure 4.40. The ranking of both error measures among the MMUs does not agree for some classes such as "3: deciduous forest", "5: grassland and pasture", or "6: agriculture", owing to significant differences in some regions. Both error measures indicate the magnitude of differences, but the RMSE pronounces higher disagreements in class membership. However, both statistics indicate an increase in error with increasing MMU. On the other hand, for some classes the error seems to decrease for a large MMU such as 100 ha. This

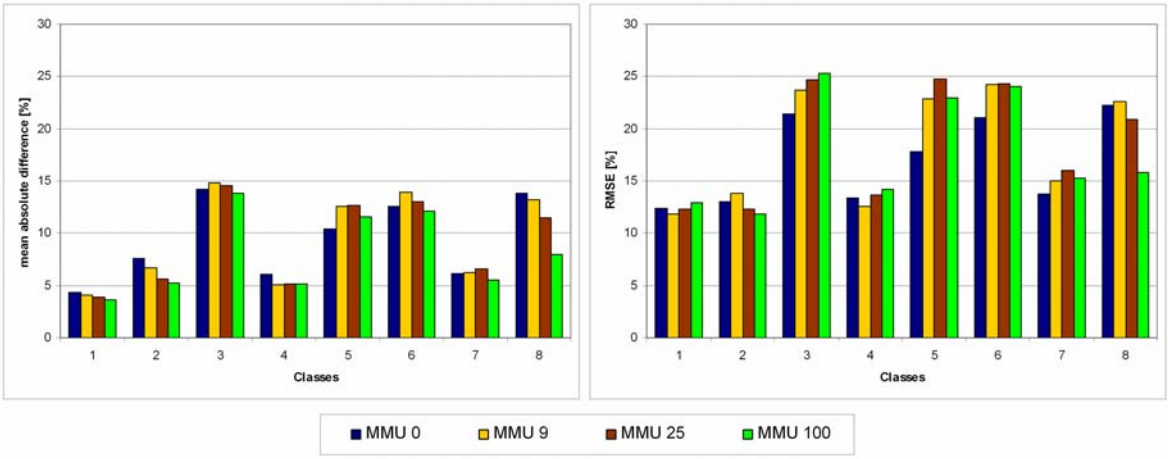


Figure 4.40: Differences in class-wise error measures for classifications using different minimum mapping units for Minnesota.
Note: The mean absolute difference (left) and RMSE (right) generally show lower errors for no MMU and largest MMU. For class names see Table A.20.

pattern is also revealed in summary statistics of the pixel-wise fuzzy evaluation. For instance, with increasing MMU the average of the absolute means (or RMSE) yields, 9.3 (14.5), 9.5 (15.4), 9.1 (15.0), and 8.1 (13.7) percent.

The analysis of the influence of the MMU on classification accuracy indicates an increase in error with increasing MMU. However, the chosen assessment procedure also implies that there seems to be a maximum MMU, at which the error begins to drop. The first issue appears rather reasonable, since the MMU reduces the heterogeneity of the surface. The finding that moderate MMU thresholds yield decreasing accuracies has been supported by all assessment strategies including the fuzzy approaches. The discrepancies between no and moderate MMU are not too severe regarding the spatial extent, but merely influences the difference in membership of fuzzy accuracy assessment. This, again, led to differences in hard-classification accuracy of approximately 5 %.

The second conclusion that large MMUs reduce the error in classification requires an in-depth discussion, also reconsidering the assessment process. Since there is no definitive ground truth data, choosing a reference image is a critical issue, in particular for the fuzzy classification assessment. This issue has already been addressed above. In order to introduce no bias, each classification has been compared to the corresponding reference image, derived by multi-scale analysis using data of the same MMU. This excludes small patches from the assessment and therefore might be biased. However, initial tests of the classification with a large MMU against the reference with no MMU did not result in clearly different relative and absolute fuzzy errors. Although the fragmented patches are widely distributed, the low contribution in membership does not have a high impact compared to clear classification errors. Therefore, it rather depends on the scale of the analysis and the characteristics of the study area, whether there is such a critical threshold being approached or not.

Sampling Schemes

It is often noted that a potential disadvantage of classification trees is their bias towards classes with a higher proportion in area, which are commonly represented by more samples (Hansen et al. 2000). This simply results from the approach as to how trees partition data into smaller, more homogeneous units. In contrast to distance-based approaches, the generation of a decision tree retrieves a threshold in feature space by computing the probability of each class according its number of samples and in relation to the samples of all other classes. According to a particular splitting rule (for its impact see next section), decision trees eventually select the best partition and continue recursively. However, classes with larger sample sets than others can become overemphasized in tree construction, especially if a smaller set is mixed with a larger set in feature space. Of course, a distinctive small sample set will be still separated during the partitioning process.

This section evaluates the influence of sampling strategies on classification accuracy for both, Germany and South Africa (for other classification parameters see Table A.23 and Table A.24). In order to mitigate the potential difficulty described above, equalized sampling by drawing 1,100 samples per class is applied as standard procedure to all classifications. It is compared to a linear-scaled and a logarithmic-scaled sample design. It should be noted that the bias among classes, introduced by these modifications, has no effect on the sample distribution in space. Both non-equal sample sets were constrained to a lower bound of 1,100 samples, and an upper bound of not more than 5,500 samples. In other words, 5,500 samples were drawn for the class with most potential samples, and the class with the least candidates had to yield 1,100 samples. In case of linear sampling, the required sample set of intermediate classes is located on a straight line between the highest and lowest sample set. The logarithmically-scaled approach transformed the potential set of samples prior to scaling, thus mitigates the effect of overrepresented samples of a clearly dominating class (see also 4.2.4 Sample Generation and Subsetting).

The overall accuracy for classifications of South Africa does not show differences among sample sets. Although user's and producer's accuracies vary up to 20 % (5 % on average) for different sampling schemes, there is no clear indication for a preferable approach. The hard-classification assessment using class certainty yields higher accuracies for linear and logarithmic sampling with values of up to 10 % for low and 4 % for higher certainties. While the assessment with samples drawn for homogeneous areas may be biased, this evaluation considers a larger dataset and partly heterogeneous surfaces. A further judgment with user's and producer's accuracies, however, did not indicate a clear pattern.

Fuzzy assessment, however, indicates clearer patterns. Figure 4.41 depicts plots of class-wise error measures. The relative mean of the difference image shows a reduction in the directional

error when using non-equalized sampling approaches. This is especially evident for spatially dominating classes, which had been mapped with partly moderate satisfaction using the general setting, e.g. “3: thicket and bushland”, “4: shrubland and low fynbos”, “6: dry grassland”, and both degraded classes (10, 11). Although the absolute mean of the differencing between classification and reference indicates that the directional error was partly compensated for by estimations into the other direction, there is still a remarkable difference among classes. Proportionally biased sampling has not significantly increased the error for any class. In contrast, errors in magnitude decreased of up to 3 % for spatially dominating classes.

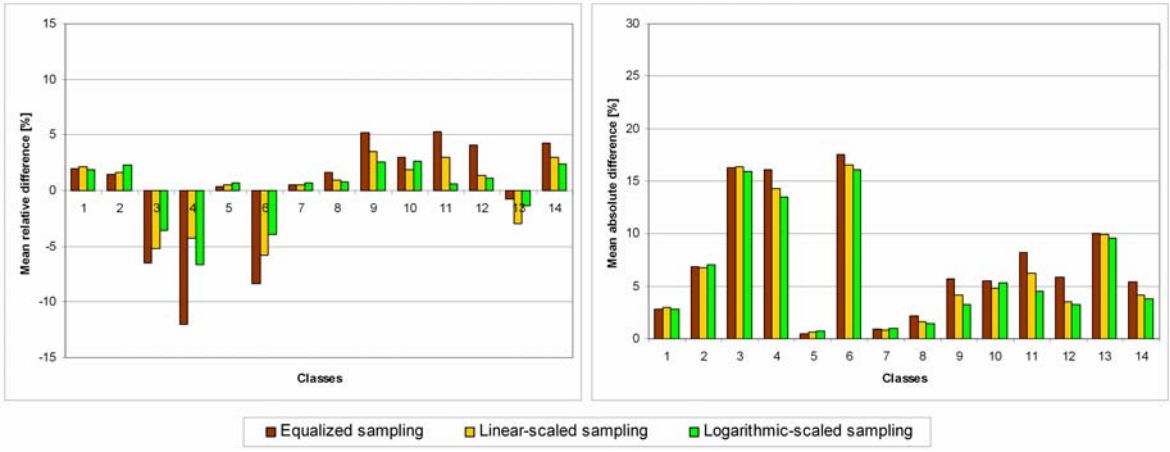


Figure 4.41: Differences between sampling schemes using class-wise error measures for South Africa.
Note: The mean relative difference (left) and mean absolute difference (right) indicate lower errors using proportional sampling schemes. For class names see Table A.17.

This picture is also revealed spatially using the mean absolute difference (Figure 4.42). High errors shown for equalized sampling decreased by both biased sampling schemes, e.g. for the transitional portions of the Karroo, for the woodlands at the border to Botswana and Zimbabwe, and for the southeastern coastal lowlands and escarpment area. Even the agricultural area of the central part of South Africa seems to have a better agreement. The improvement is also revealed in the summary statistics. Equalized sampling has a RMSE of 15.5 %, which drops to 13.6 % and 13.1 % for the linear- and logarithmically-scaled sampling scheme, respectively.

The assessment of sampling schemes for Germany already indicated a better performance using the independent sample set. While the overall accuracy of the equalized sampling is 83 %, it rises to 86 % and 85 % for the linear- and logarithmic-scaled sampling, respectively. However, for Germany the classification evaluation using the certainty is not as indicative as it was for South Africa. While the overall accuracy shows an increase of 6 % for lower certainties for the proportionally biased sample schemes, equalized sampling soon reaches the same level and traverses. Class-specific assessments did not provide further insights.

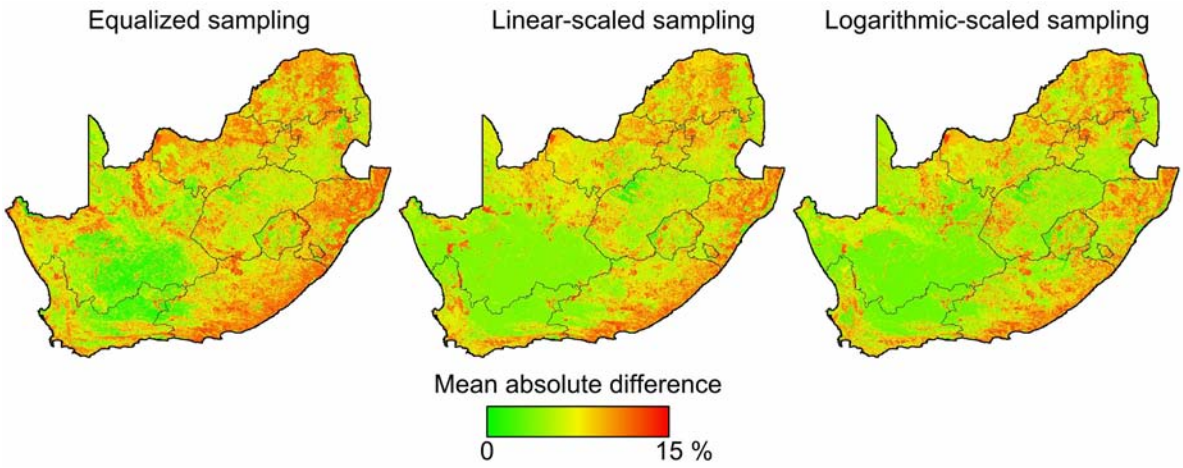


Figure 4.42: Spatial distribution of the fuzzy pixel-wise mean absolute difference for South Africa using different sampling schemes.
Note: Equalized sampling generally yields higher errors, especially in transitional zones, vegetated areas in the eastern part, and heterogeneous regions in the southeast.

Again, the fuzzy classification illustrated the gain of biased sampling. In particular class "2: arable", "3: permanent crops", "5: heterogeneous agriculture", and "11: inland wetlands" showed remarkable differences in relative class-mean differences (Figure 4.43). The magnitude in error reduction is between 2 % to 4 % for the classes mentioned above. A detailed analysis indicates that class-biased sampling better accounted for uniform classes. It reduced the membership of mixed classes, in particular for the agricultural units, which mitigated the classification noise in discrete maps. For instance, the arable land in the northeast of Germany is best represented with linear-scaled sampling, albeit the respective certainty is still low. The proportional reduction of mixed classes also affected class "inland wetland", scattered throughout Germany with low memberships. However, sampling schemes

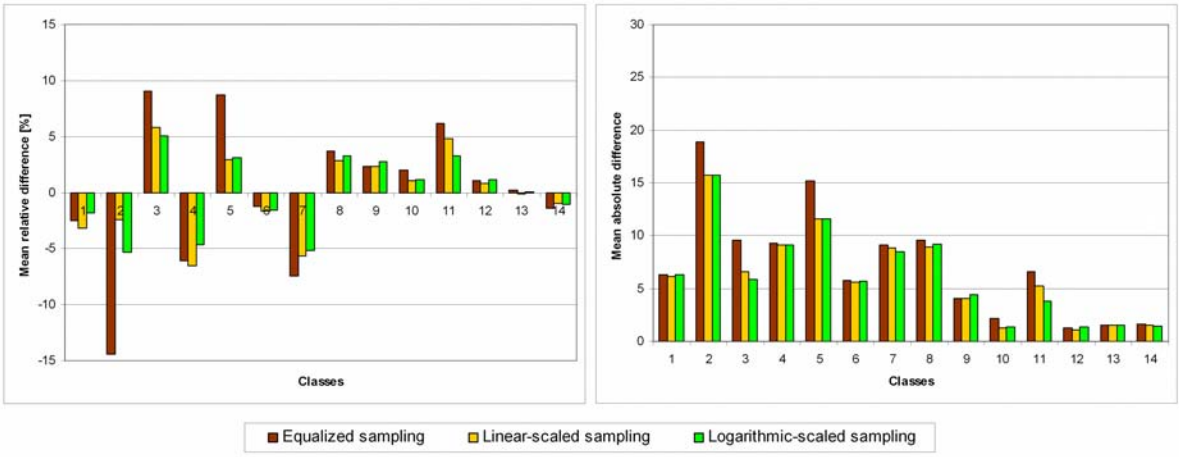


Figure 4.43: Differences between sampling schemes using class-wise error measures for Germany.
Note: The mean relative difference (left) and mean absolute difference (right) indicate lower errors with proportional sampling schemes. Highest differences are indicated for "2: arable" and "5: heterogeneous agriculture", which were better mapped with proportional sampling. For class names see Table A.19.

did not affect forested land with regard to the magnitude of the error. Finally, pixel-wise summary statistics such as average RMSE reveal 14.2 %, 12.4 %, and 12.1 % for equalized, linear, and logarithmic sampling, respectively.

This analysis of sampling schemes indicated that drawing an equalized sampling set for all classes does not necessarily improve image classification. Although decision trees have a potential to overestimate classes with larger sample sets from a theoretical point of view, moderate differences in sample size can improve the results. In this study, problematic areas with low classification certainties, e.g. along transitional zones in South Africa or in very heterogeneous settings in Germany, were better represented by proportional sample designs. The reason for this improvement is certainly related to an increase in sampling of training data, also in problematic areas. Issues related to large fields of arable land in eastern Germany, in contrast to more scattered agriculture in other regions, had been discussed in 4.3.1 Land Cover Classification of Germany and South Africa. However, the change in the classification affected areas with low certainty in class assignment. High-certainty regions remained or were possibly enlarged, such as the Karroo in South Africa.

Parameters of Decision Tree Generation

Decision-tree classifiers provide the possibility to adjust some parameters, including the split criterion, stopping rules, and pruning options. Although these parameters significantly alter the internal structure of the tree, none of these options led to clearly different accuracies of the classifications.

In this study, three splitting rules were tested, Entropy (used as standard in this study), Gini, and Towing. The hard assessment revealed discrepancies of less than 1 %, and the soft error-measures showed a difference of less than 0.2 %. It is interesting, though, that the internal tree structure is rather different. First, there is a difference between the number of nodes and leaves for trees of South Africa and Germany. A tree of South Africa has approximately 240 nodes, of which 120 are leaves. For Germany the mean number of nodes is 180 with 120 leaves.

It is often noted that Gini tends to separate homogeneous areas close to the root, which yields an unbalanced tree. Both other measures split the dataset in rather equal partitions and extract homogeneous sample sets at lower hierarchies. These structural differences, however, do not significantly affect the number of nodes and leaves in a tree. Again, the internal structural differences have no impact on the classification results and their accuracies.

The termination of tree generation, either by stopping or pruning, is the other possibility for adjusting the individual classifiers. Stopping has not been quantitatively assessed by means of classification accuracies and fuzzy-error measures. However, moderate stopping rules such as a threshold for minimum samples per node as applied in this study have no or a very marginal impact. Its effect is merely an earlier termination of tree growth before yielding meaningless leaves. These insignificant leaves are likely to be removed by pruning. Pruning had been applied to the decision trees of this study employing a new algorithm. Using this approach, tree size, i.e. the number of inner nodes and leaves, was cut by between 25 % and 50 %. The accuracies of pruned and un-pruned classifications showed no distinctive patterns. Sometimes classification accuracies were 1 % to 2 % better for pruned trees or *vice versa*. The same unclear pattern has been revealed by fuzzy measures, but the difference was always lower than 0.5 %.

Although classification parameters for decision-tree generation clearly influence the structure of the individual classifier, they have a very limited impact on classification accuracy. In contrast to Gini, splitting rules such as Entropy and Twoing yield balanced trees. Pruning cuts a remarkable number of nodes, which are merely insignificant for decision making. As the classification accuracy of these simplified trees remains at the same level, it can be concluded that those decisions are unlikely to be employed in the classification of the dataset.

Summary and Implications

This sensitivity analysis has provided a detailed analysis of the classification process and indicated more and less important parameters of the procedure. While a general classification protocol has been employed, several parameters were varied to analyze their impact on the classification accuracy. Interestingly, the findings do not clearly differ for both study regions, albeit the environmental settings are clearly different. Furthermore, it can be stated that meta-strategies and the selection of suitable approaches prior to the actual classification have the highest influence on the classification accuracy. The following paragraphs will place the results described in the previous sections in a broader perspective.

To the knowledge of the author, so far no other study has investigated the impact of the quality of time series on classification accuracy. It has been found that the time-series quality has a negligible impact on the accuracy. Although this appears to be surprising at first glance, especially for regions with considerable cloud cover such as Germany, it becomes clear that the transformation of time series to sets of metrics reduces the noise of lower-quality data. This conclusion is significant, because it has been unknown whether former classifications, especially those based on AVHRR data, could have suffered from low data-quality issues.

The generation of metrics is of central importance for a high classification accuracy. Simple statistics such as mean or standard deviation have been found to be most suitable by other studies (DeFries et al. 1995) and are commonly employed for regional and global mapping (Hansen et al. 2000). These statistics sufficiently have reduced the error contained in lower-quality time series. Furthermore, metrics computation diminishes the dimensionality and the correlation in feature space. The selection of all surface-reflectance bands for metrics computation has resulted in clearly higher accuracies than exclusively employing NDVI, with differences of up to 25 %. Secondly, temporal segmentation, employed by many studies (Conrad et al. 2007, Hansen et al. 2000), has been evaluated. The effect on NDVI time-series metrics is remarkable, with increases of 10 % in overall accuracy, and also clearly notable in the fuzzy evaluation. For NDVI metrics, the temporal segmentation has not only increased the number of features but also extracted new information, which resulted in a better classification. Phenology-based temporal partition can be recommended for vegetated regions. However, no difference in the temporal segmentation has been shown if all surface-reflectance bands were used. In this case, the extracted information is either redundant or did not significantly contribute to an improved partition of the data. This result puts finding of Conrad et al. (2007) in a boarder perspective, who concluded that temporal segmentation yields better results. This has been simply due to solely selecting the NDVI as time series for metrics definition. Employing surface-reflectance data, however, does not require further temporal segmentation, which can significantly reduce the amount of data to be processed and markedly increases the accuracy of the classification compared to NDVI.

Another key aspect of the classification process developed in this study has been the multi-scale analysis. It provides the possibility to generate a soft reference map for fuzzy classification and allows the automated extraction of sample data, also important for defining an appropriate classification scheme. The initial idea has been expanded to a more robust selection process by enlarging the area of scale transformation. This has decreased the uncertainty, e.g. due to geometrical displacement between the high-resolution map and the MODIS data, high scan angles, or marginal errors in the high-resolution classification. The increase in neighborhood has yielded higher overall classification accuracies of up to 10 %. However, there is a limitation for large neighborhoods where classification accuracy begins to decrease, as the number of sample candidates drops significantly. This potentially has introduced a regional bias in sampling, i.e. it is not possible to sample for all class-modes in feature space. At some stage, homogeneity has to be decreased, which introduced mixed pixels into the sample set. A neighborhood size of half of a coarse-resolution cell - a neighborhood of five high-resolution pixels in this study with a scale ratio 1:10 - has seemed to be most suitable. MODIS has a generally good geometric accuracy, with offsets of approximately 50 m at nadir (Wolfe et al. 2002). Previous comparisons with Landsat 7ETM$^+$

have shown a generally good spatial agreement with a displacement of clearly less than one MODIS pixel.

The effect of the minimum mapping unit has been analyzed using the NLCD2001 of Minnesota, a high-resolution land-cover dataset without aggregation to larger patches. The tests have indicated that the overall accuracy decreased of up to 7 % and errors in fuzzy image classification increase even for a small MMU. In other words, the reduction of the true thematic variability on the ground has an impact on time-series classification. Of course, the MMU does not directly affect the classification accuracy. However, the automated extraction of homogeneous sample data employs the results of the multi-scale analysis. If the high-resolution input map does not correctly represent the heterogeneity on the ground, there is a potential to select pixels of mixed surface types. Although it is impossible to correct for effects caused by the MMU, it is important to know and quantify its impact.

The sampling scheme is an important consideration for almost all classifiers. It relates to biases in both, space and among classes. The partitioning process of decision trees is often considered error-prone, if sampling is biased towards classes with a high area proportion. Therefore, the standard sampling scheme in this study has been defined as equalized sampling, i.e. each class is sampled with the same number (1,100). The results of this analysis indicate, however, that larger sample sets of prominent classes have yielded lower errors in the fuzzy assessment. The conclusion that sampling based on the class-proportion is better than equalized sampling should be taken with care. Reverse cases are clearly possible; in particular if a class with a very small sample set cannot be sufficiently represented and would be replaced by a class with a high area proportion. The ratio among sample size with 1:5 for proportionally small to high classes is rather moderate. For instance, random sampling in Germany would hardly detect classes such as “heterogeneous agriculture”, “herbaceous vegetation”, or “barren”. Possibly random stratified sampling is the most suitable approach, but the question of the minimum set of samples still remains.

The adjustments of parameters of the actual decision-tree classifier have no impact in the classification accuracy. The termination criteria, such as stopping during tree generation or *a posteriori* pruning, have yielded a reduction in complexity of the tree structure. However, an impact on the classification cannot be shown in this study. In particular for pruning, this may be partly related to the routine itself, and advanced techniques will be tested in future studies.

4.3.3 Capabilities for Classification Update and Change Detection

The previous chapters of time-series classification have provided a detailed view on results of the year 2001 for South Africa and Germany and analyzed the impact of classification parameters on map accuracy. This section will focus on a third aspect and a goal of this thesis: the potential of the time-series classification procedure for land-cover updating. Although it is clear that the classification procedure outlined in the methods section can be applied to different years, there remains a question as to how well the agreement between two or more years might be. In other words, this section aims at change detection between classifications of annual time series. In order to simplify the analysis, biannual change detection between consecutive years will be conducted.

Change detection often suffers from estimating too much change between two observations, usually related to the uncertainty in cover-type estimation. Using contextual knowledge for both study sites, however, it can be assumed that the very clear majority of land-cover types will not change. On the other hand, larger conversions or even dramatic temporally changes in land cover should be indicated in the bi-annual comparison. Therefore, this analysis is a considerable challenge and an indirect assessment of the robustness of the classification procedure for its transferability to other years.

Change detection is a well-explored field in remote sensing with a wide range of techniques, especially for comparing two datasets (Coppin et al. 2002, Jensen 2005). This study partly relates to techniques of post-classification change detection, i.e. the comparison of two maps with discrete classes. Similar to fuzzy classification assessment, there is a considerable lack of methods for the comparison of fuzzy maps. However, the flexible framework for accuracy assessment only requires a simple adjustment in order to be employed for change-detection studies (see also 4.2.7 Accuracy Assessment and Figure 4.10). The modification is simply the replacement of the reference image by another hard or fuzzy classification. Furthermore, the discrete map assessment using the classification certainty considers the confidence value of correct class assignment of both classifications. Although further statistics are deemed suitable for analyzing the agreement of discrete classes in cross-tables, e.g. the contingency coefficient (Hudson et al. 2006), this study only employs the overall agreement, similar to overall accuracy (Equation A.14) in the classification assessment protocol. Furthermore, the "error measures" of the fuzzy-assessment protocol are employed as indicators for differences in class membership (Figure 4.11). In this respect, they can be appropriate for detecting subtle land-cover changes.

The following sections on the capabilities of land-cover update and the detection of change will follow the same principle as the previous sections. For each main study region, first South Africa and then Germany, hard and fuzzy maps of several years will be analyzed.

Obviously, some approaches require a selection of specific years and classes, which illustrate the potential of the techniques. The last section will provide a summary of findings

Classification Update and Change Detection in South Africa

The land-cover classification of South Africa has followed the classification protocol of Table A.24, employing time series of the respective year. First, three discrete maps of annual time series will be analyzed visually (Figure 4.44). At first glance, there is a high agreement among the maps. All major land-cover types are located in the same region with approximately the same spatial extent. Of course, there are discrepancies, e.g. the location and extent of class "barren" in the Karroo or the extent of "urban" in Gauteng. However, these changes are mostly mapped with low certainty, i.e. another class is almost as likely as the assigned one. It is noteworthy, though, that transition zones do not vary significantly in space. Despite the low classification certainty, the boundary between "shrubland and low fynbos" and adjacent classes such as "thicket and bushland" or "dry grassland" remains. The high stability of classes, even for lower-confidence areas and transition zones, indicates the robustness of the classification procedure, since no dramatic landscape change can be expected for a short period.

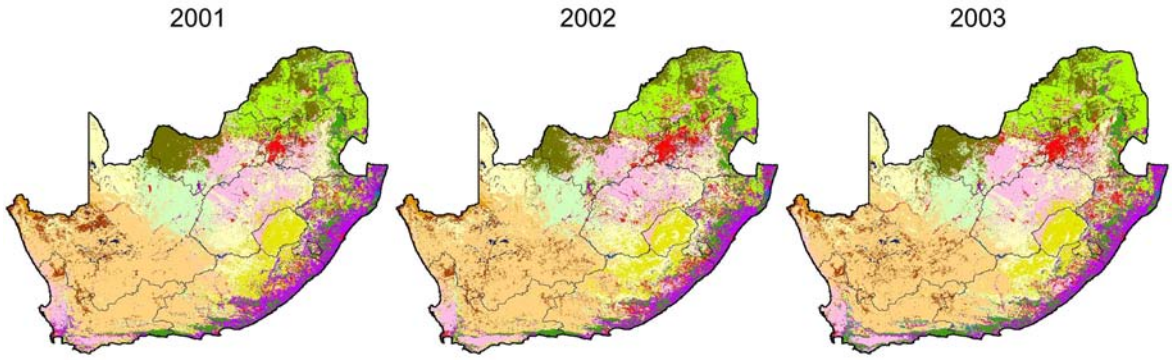


Figure 4.44: Discrete classifications of South Africa for the years 2001, 2002, and 2003.
Note: Significant changes are small and mainly occur in the same general land cover category, e.g. woodland. For the legend see Figure 2.3.

The plots shown in Figure 4.45 indicate the low amount of change using hard and soft bi-annual change detection of consecutive years. The change detection using 1:1 map comparison and classification certainty is depicted in the upper left. The graphs indicate that there is a considerable agreement, i.e. no change, between 68 % and 70 %, even if no classification confidence is exploited. Utilizing the classification confidence, the agreement between pair-wise land-cover updates increases, while the proportional areas of map comparison decreases (dashed line). It is notable that a classification certainty of above 50 % in both maps yields an agreement of more than 90 %, still accounting for 40 % of the map.

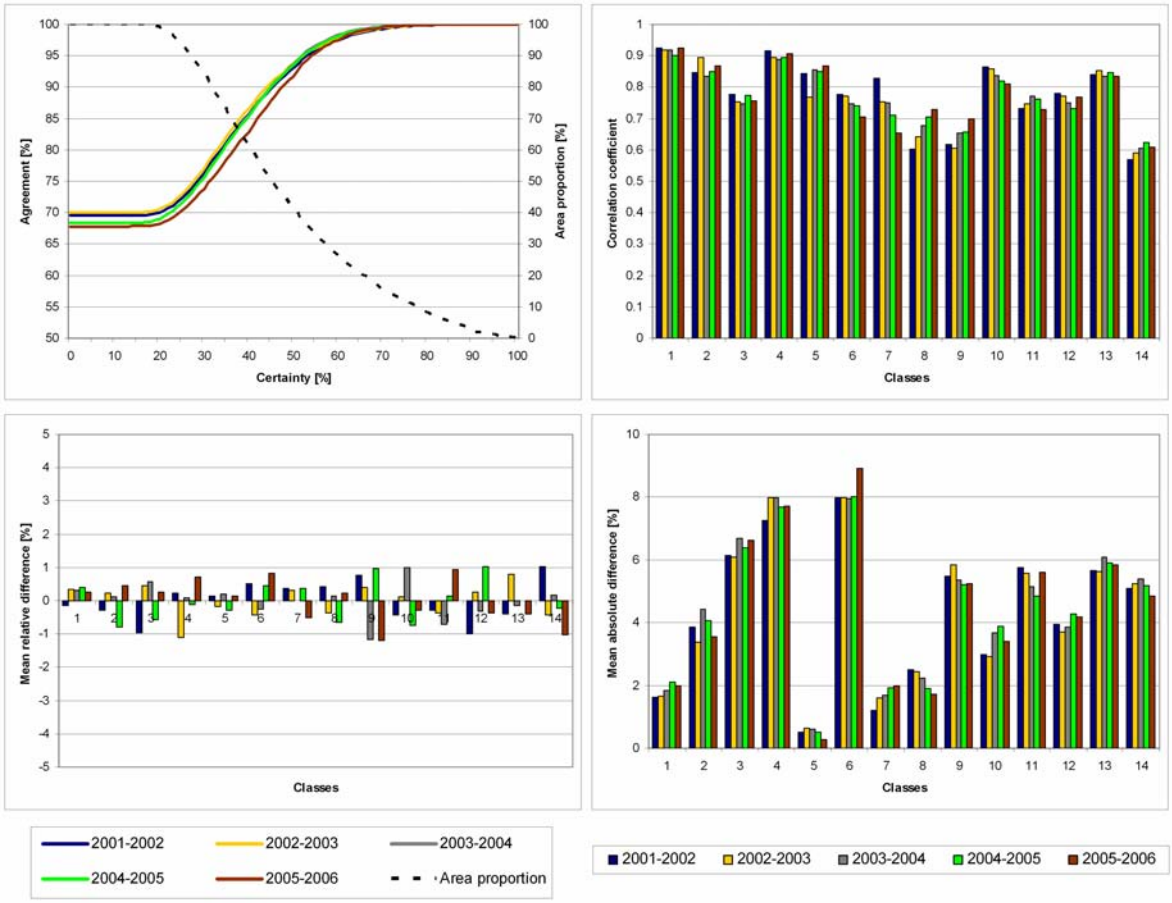


Figure 4.45: Change detection for South Africa with post-classification comparisons between consecutive years.

Note: The upper left diagram illustrates the comparison of discrete maps using the classification certainty with a good agreement for high certainties above 70 %. The dashed line indicates the area proportion considered in the analysis, which was very similar for all comparisons. All other plots are based on the class-wise fuzzy assessment protocol. Mean relative differences (lower left) are very low, and changes of one comparison were usually compensated by the next. Mean absolute differences (lower right) depict low values and no trends. The correlation coefficient (upper right) shows high values indicating the good correspondence. For class names see Table A.17.

Discrete class assignments with high certainty, e.g. above 70 %, result in an above 99 % agreement (15 % of the area proportion). Another measure for the good agreement is the low variability between the graphs of different bi-annual comparisons. The difference is never higher than 3 %. Both, the high agreement of bi-annual comparisons and the similarity of the graph for multiple comparisons indicate the high stability of the land cover for a major portion of the study region, especially in the core of each class. This is also another indirect indicator for the accuracy and robustness of the classification procedure.

The bar-plots of Figure 4.45 show statistics of fuzzy class-wise land-cover update using error measures (see 4.2.7 Accuracy Assessment). The relative mean-difference plots in the lower left indicate that there is a very low proportion in directional change of memberships, on average clearly less than 1 %. Furthermore, directional changes are usually compensated for

in the next or next-but-one bi-annual fuzzy comparison. The corresponding absolute mean indicates the magnitude of bi-annual changes. This measure is important, because the directional analysis only depicts an overall trend which can be significantly compensated for by reverse patterns in the same class. Although the proportion is clearly higher for some classes, the magnitude of differences is still very low, indicating no significant change in the landscape. Finally, the high correlation coefficients itself and their low variability between class-wise bi-annual comparisons indicate the good agreement between class memberships.

Classification Update and Change Detection in Germany

The land-cover classification of Germany was updated using the general classification protocol of Table A.23. Figure 4.46 depicts discrete land-cover classifications of three years. A detailed view on the classification shows some changes, especially for the map of 2003. However, the changes are merely between uniform and mixed classes, e.g. classes “arable” and “heterogeneous agriculture”, which belong to the same general category “agricultural land”. This stability of major structures of CLC level 1 cover types, including “urban”, “agricultural land”, “forest and semi-natural land”, “wetland”, and “water”, is shown in the recoded close-ups of Figure 4.46. Not only large physical structures such as the Harz Mountains in the center of the focus, but also small woodlands in Thuringia, Hesse, or Lower Saxony are accurately detected in each classification. Similarly, urban structures are indicated correctly. Only “wetland”, a highly heterogeneous class, has some scattered pixels with low classification certainty (see also 4.3.1 Land Cover Classification of Germany and South Africa).

Again using the full legend of 14 classes, Figure 4.47 illustrates the classification accuracies for classifications from 2001 to 2006. The upper-left plot depicts the overall agreement of the annual classifications. Using no certainty of class assignment, the similarity of bi-annual comparison indicates 60 % to almost 70 %. The agreement is higher than 80 % for certainties above 50 % and still accounts for 46 % (55 %) of the data. The overall agreement is 97 % above the 70 % certainty threshold.

The most prominent feature in the graphs, however, is the difference in bi-annual change analysis for the year 2003. This also necessitated two graphs of data availability, either with or without the comparisons to 2003. The difference becomes particularly clear for lower confidence assignments. It has to be noted that the summer in 2003 was particularly hot and dry in Central Europe. This striking difference is revealed in the detailed map of 2003 (Figure 4.46). Even visually, there are higher proportions of uniform classes in 2003. However, the alteration in land cover is mainly within broad categories, because there is no significant

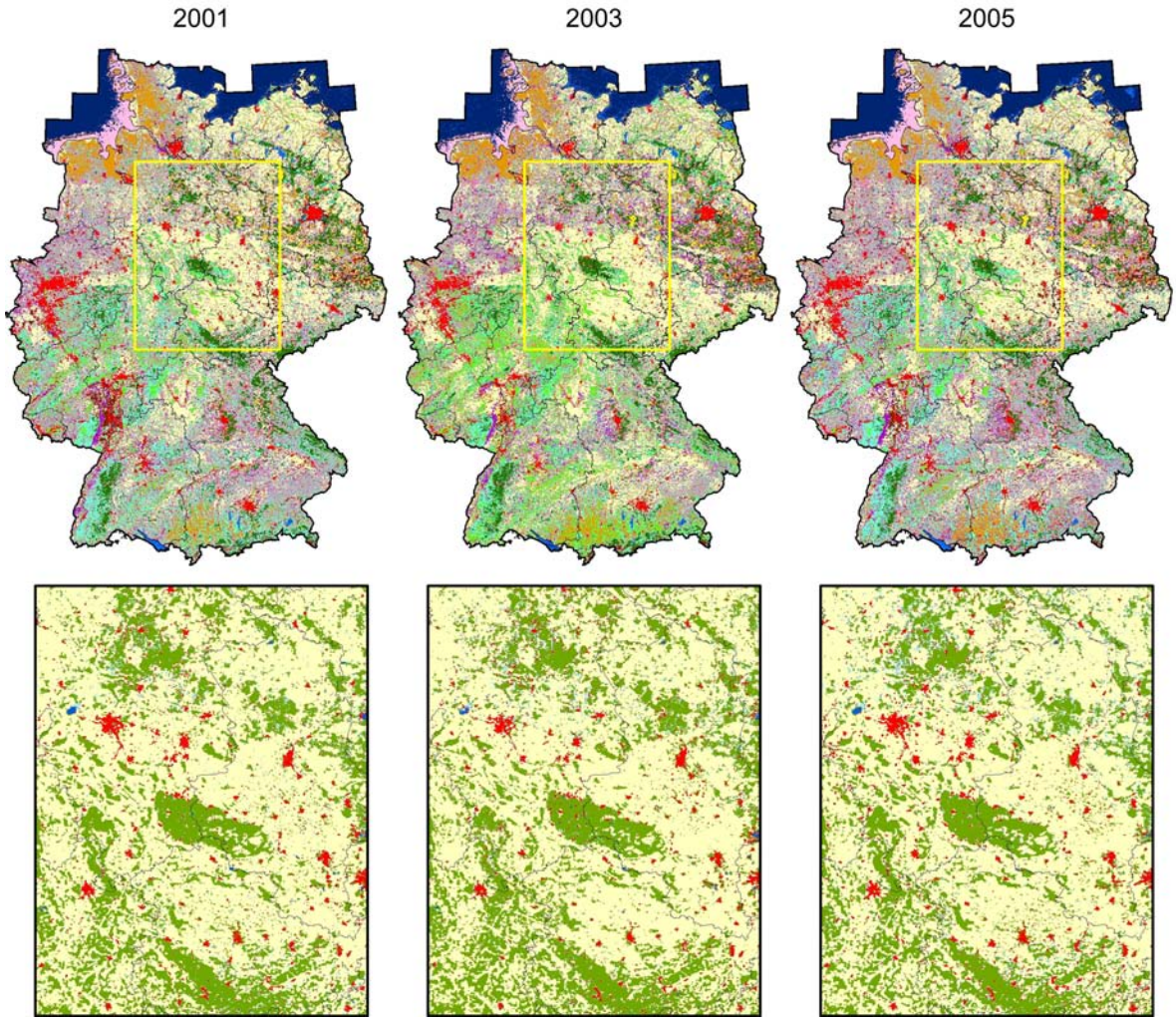


Figure 4.46: Discrete classifications of Germany for the years 2001, 2003, and 2005.
Note: Changes occur between uniform and mixed classes of the same broad category, i.e. forest or agricultural land. In particular 2003 depicts a different assemblage, caused by differences in the annual cycle of vegetation due to the exceptionally hot summer. The close-ups at the bottom row indicate the regionally good correspondence of general categories, even for 2003 (red: urban, yellow: agricultural land, green: forest and natural land, cyan: wetland, blue: water). For the legend of the upper row see Figure 2.4.

difference in close-ups of Figure 4.46. Several studies have addressed the impact on the landscape and ecosystem of the exceptional summer of 2003 (Gobron et al. 2005, Reichstein et al. 2007). This has altered the land-surface characteristics and slightly changed memberships of land-cover classes, which influenced the discrete classification for low confidence assignments. Former studies also stated that the anomaly had no impact on subsequent years. This statement is indicated by this analysis as well, because all comparisons not taking into account the year 2003 show a highly similar curve.

Fuzzy land-cover update indicates similar trends (Figure 4.47). While the directional analysis is generally low, some remarkable discrepancies are indicated for 2002-2003, which almost always show a reverse direction in the next comparison. Class “6: broadleaved forest” indicates the highest discrepancy, owing to the higher proportion of uniform classes in 2003.

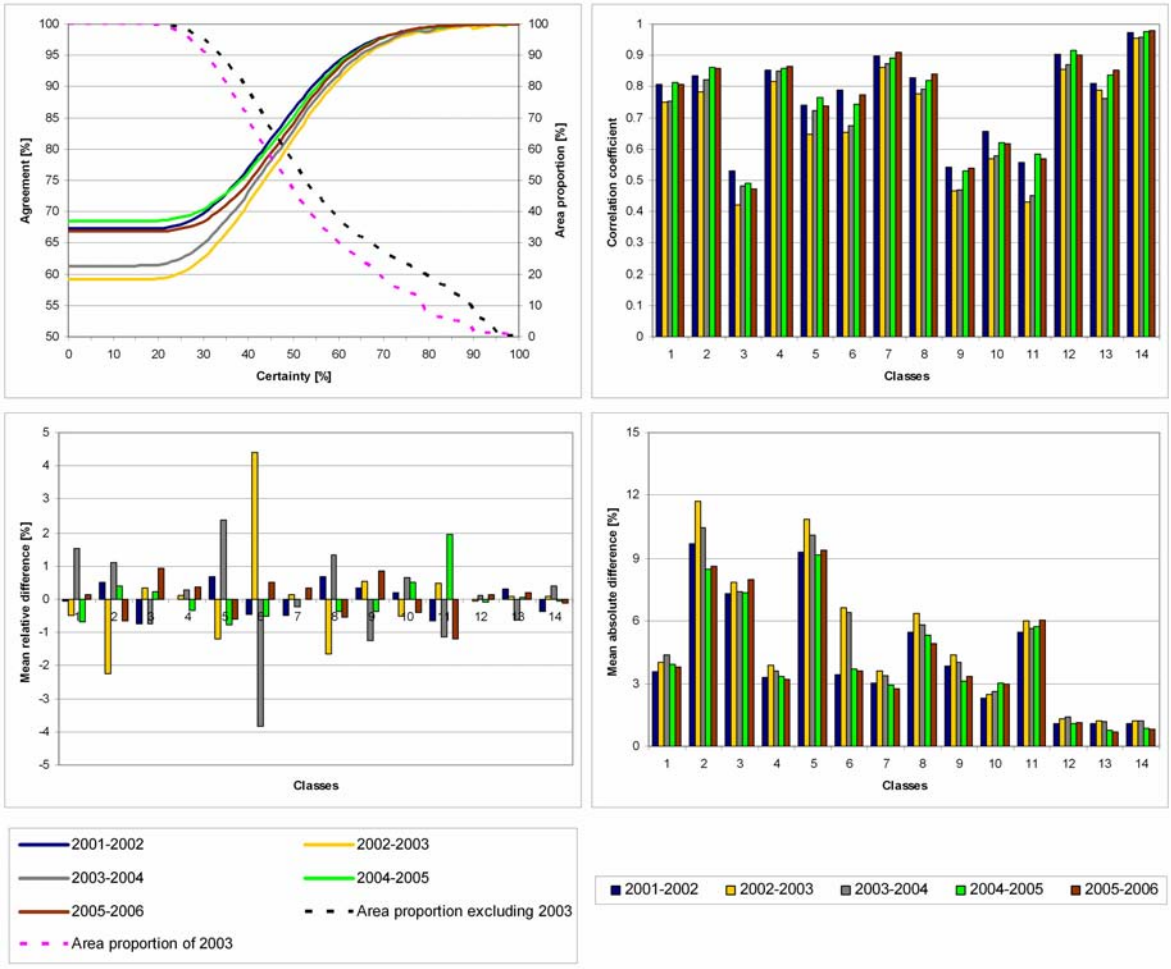


Figure 4.47: Change detection for Germany with post-classification comparisons between consecutive years.

Note: The upper left diagram illustrates the comparison of discrete maps using the classification certainty with a good agreement for high certainties above 70 %. Only comparisons with 2003 yielded lower agreements. All other plots are based on the class-wise fuzzy assessment protocol. Mean relative differences (lower left) are very low except for comparisons with 2003. Clearly, changes are compensated by the next year. Mean absolute differences (lower right) also depict higher disagreements for 2003. Similar patterns are indicated by the correlation coefficient (upper right), but values still indicate a good correspondence. For class names see Table A.19.

The magnitude of the difference in class membership for bi-annual change detection is acceptable. It indicates expected and already explained patterns with higher discrepancies for comparisons with the year 2003. Furthermore, the correlation of class-wise memberships is sufficient.

The pixel-wise analysis yields images of changes between two years (Figure 4.48). All bi-annual comparisons, except for 2003, depict low differences in all statistics. For 2003 there are some differences in both, the image and the summary tables, with an increase from, on average, 8.6 % to 10.1 % in RMSE. However, there are no spatial patterns of change in Germany, which simply indicates the general difference of the summer period for the entire country (compare center of Figure 4.48 to right-hand side).

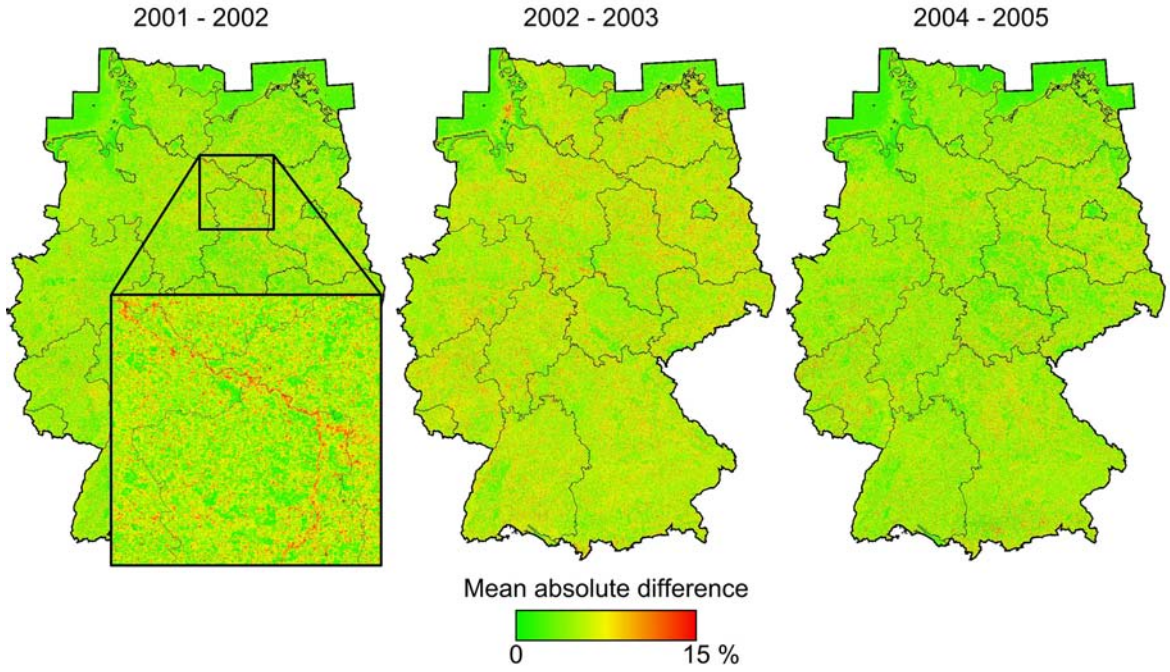


Figure 4.48: Change detection of fuzzy classification for Germany.
Note: The mean absolute difference measure has been employed similar to the fuzzy pixel-wise accuracy assessment protocol. The comparison of 2001 - 2002 generally depicts a good agreement but clearly indicates the flooding of the Elbe River in the summer of 2002. At the right-hand side of the inset, the damming of inflow from the Havel River is clearly shown. The change detection of 2002 - 2003 compared to 2004 - 2005 indicates a general higher disagreement without clear spatial patterns due to the very hot summer of 2003. This indicates some differences in class composition, also shown by the change detection statistics. Still, the general picture is represented.

However, another indicator of the reliability of the classification procedure for accurate land-cover update and change indication is shown in the inset for 2001-2002 (right-hand side of Figure 4.48). The severe flood of the Elbe River in the summer of 2002 has clearly had an impact on class-memberships. This effect is revealed in the absolute mean difference with a normal year, indicating that there is a higher potential for a change in land cover. It not only shows the flooding along the Elbe River, but also the damming of water coming from tributaries such as the Havel River at the right-hand side of the inset. Clearly, this episodic change in land-cover composition has had no long-term effect on the land cover. Still, the example indicates the potential of time-series classification for large-scale change detection or subtle change indication.

Summary and Implications

The classification procedure of time series has been applied to six years of MODIS data for South Africa and Germany. All classifications indicate a high correspondence in visual and statistical results for discrete classes and soft estimates. This pattern has been expected, because the period of six years does not suggest clear land-cover alterations in the study

regions. However, changes in land-cover and class-membership composition due to significant impacts on the environment are revealed for Germany by the exceptionally hot summer of 2003 and the severe flooding of the River Elbe in 2002.

Post-classification bi-annual change detection seemed most appropriate for quantifying the agreement between consecutive years. While the change detection of the discrete classifications has been straightforward, using a variant of post-classification change detection modified by classification certainties, the protocol for accuracy assessment has been slightly modified. The error measures for accuracy assessment have become indicators of dissimilarity of fuzzy maps. Since there is a lack of techniques for analyzing continuous results of image classification, the methods adopted from other fields in this study may be regarded as a step towards a fuzzy change-detection protocol.

It is important to note that the sample data, mainly employed for training the classifier, have always been derived from the same location. This study does not employ any random selection process. Sampling has been solely based on the multi-scale analysis and the rigid sampling scheme. Although the same location is sampled, determent by the multi-scale analysis, the extraction of signatures from time-series metrics has been carried out for each year. Due to the variability of the data in time and intensity, a simple transfer of a class signature of one year to a dataset of another year is not possible. This indicates the need for stable training data, an issue addressed in the scientific literature for several years (Baret et al. 2006, Muchoney et al. 1999). The previous sensitivity analysis has indicated the importance of accurate training data for high classification accuracies. Even robust classifiers such as decision trees can only partly compensate for errors in training samples. Homogeneous training sites can be employed for multi-year analysis, but should be checked for internal land-cover modifications or conversions if applied to several years before or after site definition. It should be noted that the NLC1995 has been applied to MODIS data of 2001 and onwards, which is not optimal and could introduce unpredictable uncertainties. Initiatives for global land-cover databases and validation or inter-comparison sites, including the STEP database or CEOS-BELMANIP, are highly necessary for future regional to global mapping.

4.3.4 Re-Mapping Vegetation Types of Namibia

The goal of this final section of the analysis of automated time-series classification is the application to a contextually different type of data. Besides land-cover and land-use mapping, biologically defined vegetation types can be partly mapped with remotely sensed data. Although remote sensing cannot substitute for detailed botanical inventories by field campaigns, it has the potential to provide supportive data and continuously map large regions

in space and time (Nagendra 2001, Turner et al. 2003). Medium to coarse spatial resolution time series in particular can be useful, due to mapping phenological differences in the vegetation assemblage. Furthermore, repeated mapping with time series can indicate variations to subtle modifications and, of course, intensive changes in vegetation composition, sometimes hard to detect by local and temporary fieldwork.

This section aims at re-mapping an existing vegetation map of Namibia using MODIS time series. The following sections will briefly outline data processing issues and demonstrate mapping results including capabilities for improved vegetation monitoring. The results of this section have been published in Colditz et al. (2007d).

Data Processing

The basic classification procedure follows the workflow of this study as outlined in Figure 4.4. The remote-sensing baseline data have been annual time series from 2001 to 2003. Input data have been seven bands of NBAR, NDVI, and day / night LST. However, one adoption had to be made. Since the vegetation baseline data (see also Figure 2.6 and 2.3.2 Land Cover Datasets; Mendelsohn et al. 2002) is mapped at a rather general level of detail, multi-scale analysis does not seem suitable for this dataset. In order to draw meaningful sample data for signature extraction and classification assessment, a 10 km buffer was generated along each class boundary. Furthermore, missing multi-scale analysis only allows for the assessment of the hard classification accuracy using the independent sample set and classification certainty. The remaining sampling and classification procedure follows the process applied to all other datasets and the respective classification protocol is provided in Table A.26.

Analysis of the Remotely Sensed Vegetation Map of Namibia

Figure 4.49 depicts a color composite of a wet and dry-stage image of Namibia. The differences clearly reveal the seasonality of the year, in particular for the wetter, northern part. The superimposed boundaries of the vegetation types indicate the often only subtle differences between classes. Other vegetation types depict clearly distinctive core areas, but broad transitions to other units. This imposes a challenge for classification and makes a fuzzy representation of the data more meaningful. Some other classes, for instance larger polygons the southern part, contain a mixture of surface types.

The classification accuracy using the independent sample set reveals very similar results for all years of analysis. The overall accuracies and KHATs are 85 % and 0.84, respectively.

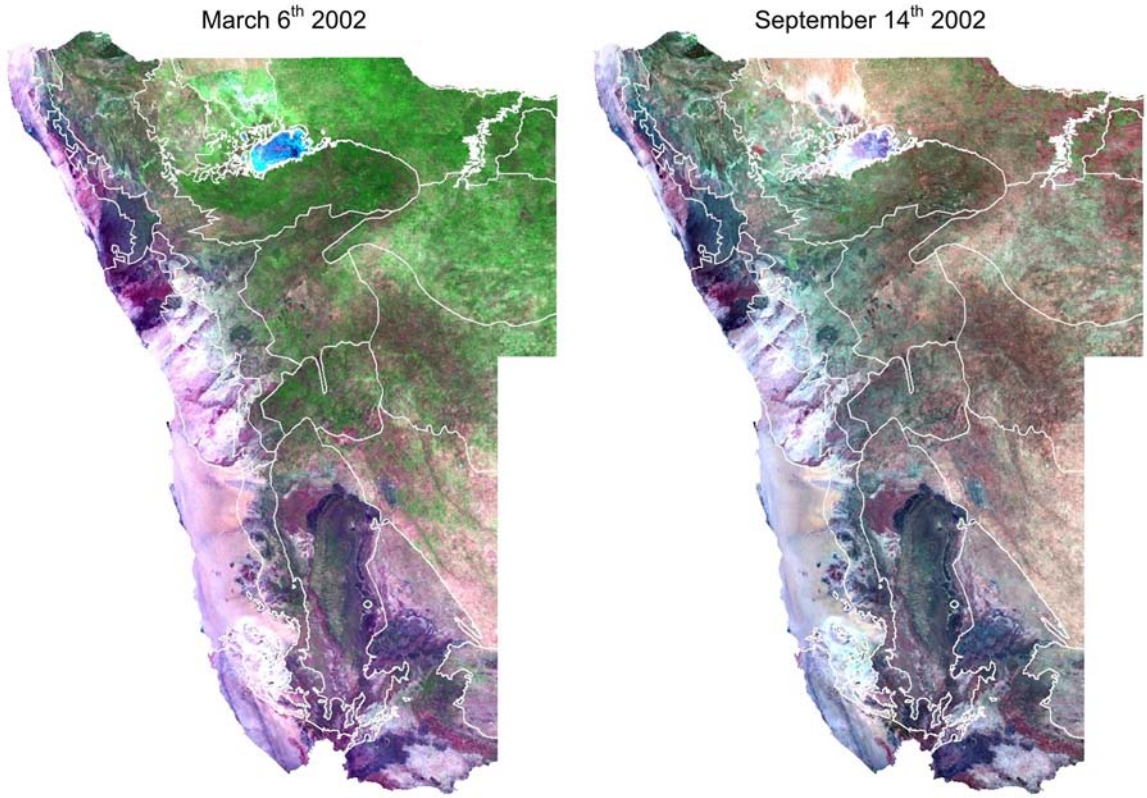


Figure 4.49: Multispectral MODIS images of Namibia.
Note: The image at the right-hand side is a 16-day composite starting on March 6th 2002 (day 65), and on the left-hand side on September 14th 2002 (day 257). They correspond with peak and minimum in phenology, respectively. The band combination of surface reflectance is (RGB): 7 (SWIR3), 2 (NIR), 1 (VIS red). Superimposed white lines indicate the vegetation types of Namibia (Mendelsohn et al. 2002).

Except for very few classes user's and producer's accuracies are well above 80 % and often higher than 90 %. Class-specific accuracies are never lower than 70 %. The assessment using classification certainty of 50 % reveals an overall accuracy of approximately 83 %. This rises to 90 % for a certainty of 70 %.

The classification and the respective certainty of class assignment are shown in Figure 4.50 for 2002. Issues addressed in the previous paragraph are well revealed in this discrete map and the corresponding confidence estimate. First, there is a remarkable agreement between the reference (superimposed lines or Figure 2.6) and the classification. The agreement is similar in the classifications of 2001 and 2003. Even vegetation types with only marginal differences in the input data (see Figure 4.49) in the central, eastern, and southern part aligned well with the reference. For instance, the prominent "Karstveld" in the north-central portion south of the Etosha Pan is accurately delineated from it surrounding vegetation types. On the other hand, a clear spatial shift towards the east is depicted for the "North-western escarpments and inselbergs". Furthermore, there is a clearly distinctive spot enclosed by vegetation types of the "Southern Kalahari", which is also different in the surface-reflectance image.

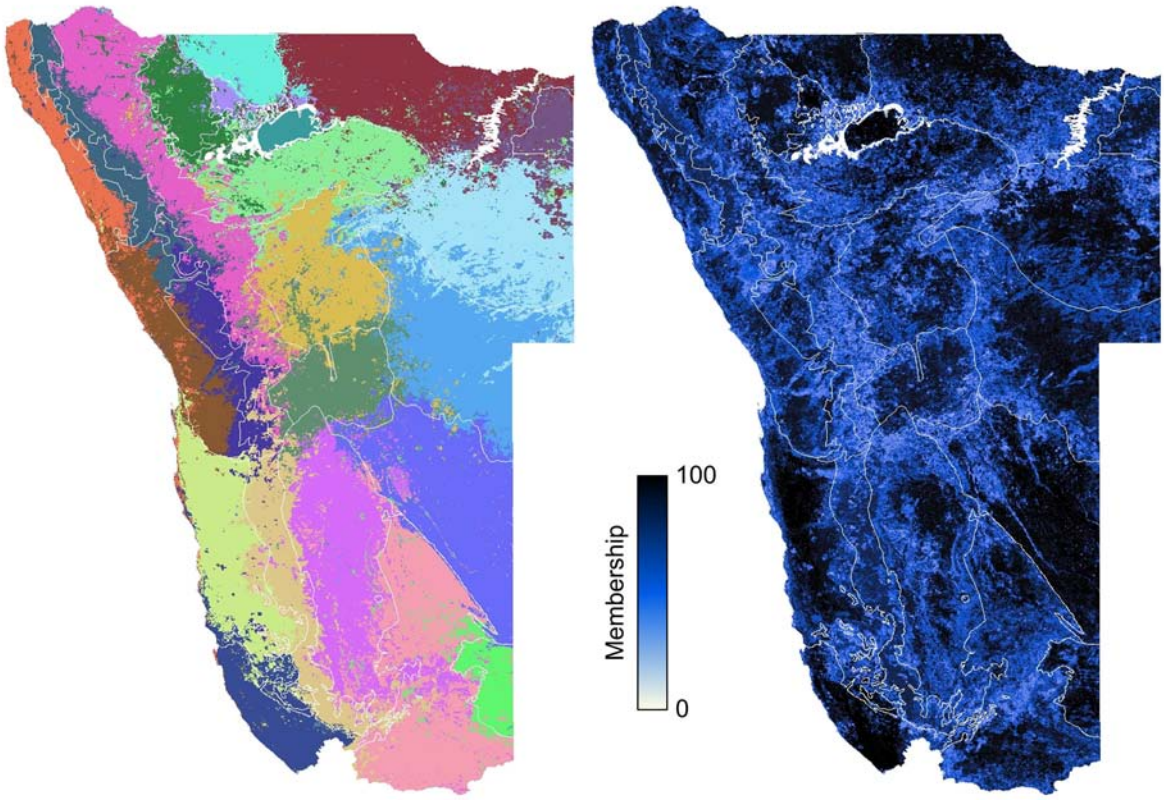


Figure 4.50: Discrete classification and class certainty of vegetation types of Namibia for 2002.
Note: The classification (left) generally shows a good correspondence with the superimposed boundaries of the reference image. For legend see Figure 2.6. The classification certainty (right) indicates broad transition zones between core areas, especially for the Kalahari vegetation types in the eastern part of Namibia.

Fuzzy classification provides the possibility for mapping class transitions, a very useful property for this semi-arid region with broad ecotones. Interpreting the classification of Figure 4.50 together with its certainty estimate exhibits a comprehensive view of the vegetation composition. For instance, all vegetation types of the Kalahari in the eastern portion and many classes in the central part of Namibia indicate clear core regions with a high certainty in the classification. This confidence decreases towards the edge, which reveals the broad transition to another unit.

An exploratory by-product of decision-tree classification is the leaf-map of each individual decision tree (see also Figure 4.4). Figure 4.51 depicts two leaf-maps of different trees for the classification of 2002. The coloring, based on the identifier of each leaf, is not important, unlike the spatial arrangement of different patches. Interestingly, many leaves map rather contiguous patches, concentrating in one area. In other words, there is a clear spatial pattern in both maps. With regard to the content, the explorative leaf maps are similar to an unsupervised classification, i.e. there is a pattern without a contextual link. Of course, each leaf is connected to a thematic class, which would be the common output of a classification tree. However, this display merely explores different sub-units of the classification tree.

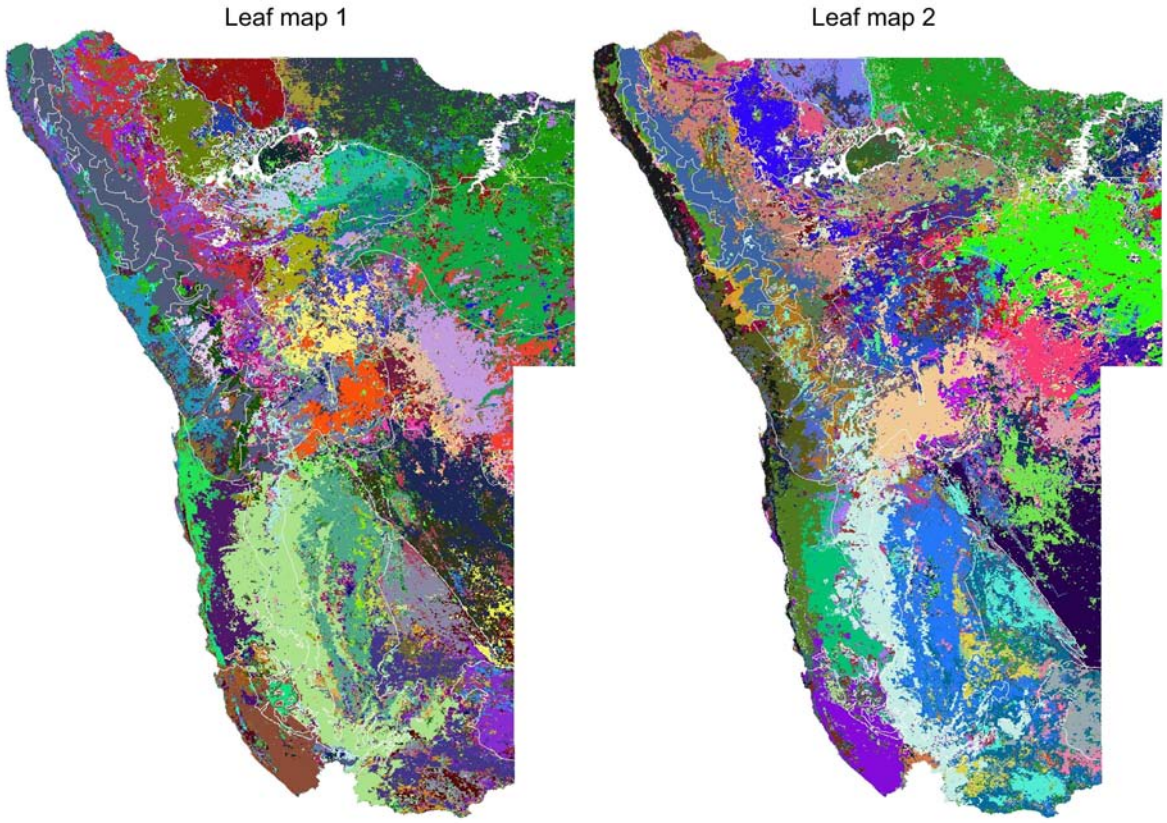


Figure 4.51: Examples of leaf maps of the classification of the vegetation types of Namibia for 2002.
Note: The leaf maps result from individual classification trees (see Figure 4.4). Leaf maps depict that one class can be represented by several leaves in the classification tree, i.e. that one class is made up of several clusters in feature space. This can be considered an explorative method and the result is related to an unsupervised classification. The boundaries of the vegetation types of Namibia are superimposed.

There are some marked similarities for some large units in the southern part. This indicates a good correspondence and can be an indicator of a higher confidence sub-unit. Similar patterns are also shown for the northern regions and the escarpment area. Other areas might only be contiguous in one individual map. This may still indicate a distinctive sub-unit. In this sense, each individual tree-map can support vegetation mapping if some field knowledge is available. Therefore, local experts may be able to identify distinctive vegetation types hardly detectable in its spatial composition in the field. However, approaches have to be developed to combine different tree-maps to a more comprehensive product.

Bi-annual change detection between years revealed very small discrepancies in both, hard and fuzzy classification. A 1:1 map comparison yields an overall agreement of 75 %. At a classification certainty of 50 % and 70 %, the agreements rise to 86 % and 98 %, respectively. Fuzzy class-wise changes of the relative mean indicating trends are lower than 0.5 %. Even the magnitude using the absolute mean is negligible, with values below 4 % at the most and 3 % on average. Furthermore there are no trends between the bi-annual change detections. The pixel-wise variant does not reveal a spatial pattern. Only minor changes are indicated

north of the Etosha Pan, an area which has been and is in the process of being transformed due to the comparatively high population density.

Summary and Implications

This case study has demonstrated the usefulness of the classification procedure also for other thematic applications than land-cover mapping. The vegetation map of Namibia has been successfully re-mapped, yielding very similar results compared to the reference. This good agreement in the spatial pattern may indirectly indicate the correctness of both, reference and re-classification. Furthermore, the fuzzy result provides the possibility to identify and monitor transition zones, which are an integral part of the vegetation composition in Namibia.

Areas with disagreement could indicate a change in vegetation type composition. For instance, the densely populated and intensively used northern portion of Namibia is undergoing some considerable environmental changes, which clearly affect the vegetation composition of the land. In this regard, the annual re-mapping and subsequent change detection with hard and soft classification could provide change indicators.

Chapter 5

Conclusions and Future Research

This study has aimed at the development and application of approaches to time-series generation and classification. The main challenge of this research has been to ensure high accuracy, robustness, and the full transferability to other datasets and other regions. Although solely MODIS data have been employed for analysis, the methods can be, will be, and partially have already been adopted or applied to data of other sensors or data assemblages. The main study regions, South Africa and Germany, are located in different environmental settings and provide a challenge for automated mapping. Two minor study sites, Minnesota and Namibia, have been added for specific purposes or to illustrate the flexibility for vegetation mapping.

This final chapter will put the results of previous chapters 3 (Time Series Compilation) and 4 (Automated Time Series Classification) in a broader perspective, and will illustrate potentials and challenges for future research. The conclusions for time-series generation and automated time-series classification will be addressed in separate sections. Some final remarks will be noted at the end.

5.1 Time Series Generation and Analysis

Remote-sensing time series are often a prerequisite for all studies involving a detailed, spatial analysis of processes at the Earth's surface. In this regard, time series of satellite data directly contribute to the aims of global-change research, identifying the current state of the environment and its transformation. It has been outlined that one, two, or even several image acquisitions cannot be considered a time series, but merely a sequence of snap-shots. In this respect, there are specific considerations with regard to sensors suitable for time series. The most important characteristic is the frequent and consistent observation of the same spot on Earth. Global, long-term acquisition strategies are most appropriate, but limit the sensors to either geostationary satellites with a very coarse spatial resolution or near polar-orbiting systems with a global coverage within one to two days. In this field of remote sensing the AVHRR instruments have been the first, and for a long time the only systems for terrestrial applications. Since the late 90ies of the last century, other systems such as SeaWiFS, SPOT, MODIS, and MERIS have complemented this important field in remote sensing, and future missions such as NPP and NPOESS are planned.

Besides the requirement of suitable sensors, there is a need for appropriate methods for time-series generation. Compositing partly mitigates some difficulties, in particular for cloudy observations. However, simple stacking of datasets to produce a time series is inappropriate for many regions and several applications. Often, filtering functions have been applied to smooth the temporal course, which can be helpful for processing meaningful time series. However, making use of additional information, specifically indicating the quality of each record, has been found to be a suitable approach to compiling time series.

The Time Series Generator (TiSeG), a software tool which has been developed in this study, aims at interpreting data-quality indicators provided with each MODLand dataset. The interactive software evaluates these pixel-level quality indicators and derives two important data-availability indices, describing the feasibility of temporal interpolation: the number of invalid pixels and the maximum gap length. The quality settings can be modified spatially and for different periods. Eventually, the pixels regarded as invalid can be either masked or interpolated by temporal and spatial techniques.

The current implementation of TiSeG has been focused on the specifics of MODIS data. However, other sensors such as MERIS also provide extensive data-quality indicators. Since the scientific community increasingly employs data-quality issues into data analysis, future missions will also provide pixel-level quality estimations. Although there have been other studies on data-quality evaluation and subsequent processing, focusing on particular MODIS products (Erasmi et al. 2006, Lunetta et al. 2006, Neteler 2005), TiSeG is clearly the most

comprehensive software and enables processing of all kinds of MODLand data. Since its first release in 2006, it has been employed by several other institutions and organizations world-wide.

It is planned to provide an automated time series generator for several sensors along with the current interactive environment of TiSeG. A spin-off processor has been successfully implemented for a daily 8 km AVHRR NDVI (PAL) time series of Africa from 1981 to 2000. This automated approach has utilized a ranking of supplementary data such as cloud flags or angular information and applied stepwise interpolation (Colditz et al. 2006b). In this respect it is stated that the idea of identifying low data quality and subsequent interpolation with temporal techniques can and partly has been adapted to other sensors.

The impact of different quality specifications on time series has been demonstrated using the NDVI for the study area of Germany. Several levels ranging from no quality analysis to lenient, moderate, strict, and very strict settings have been analyzed using visual and statistical methods. It has become clear that no quality analysis and the selection of lenient settings insufficiently represented the phenological development of vegetation. On the other hand, very strict settings have excluded too many records for a meaningful temporal interpolation. Therefore, it is concluded that weighting data quantity against data quality is a critical step for producing meaningful time series for many applications in land-change research, including long-term phenological analysis, crop monitoring, and environmental modeling. With regard to updated MODIS processing algorithms, it is important to consider changes in product generation in order to obtain comparable results. The currently released collection 5 data have seemed to yield improved time series with regard to temporal stability. In comparison with the NDVI, the EVI has shown a smoother curve, which is useful for high-quality phenological analysis.

The appropriate quality setting can vary significantly between different study regions and during the year. For instance, the impact of the quality analysis on time series of large portions of South Africa is rather low, because the atmospheric conditions are comparatively good and cloud cover is largely eliminated during compositing. On the other hand, clouds are more persistent in Germany, even during the summer season. In this regard, the importance to analyze quality indicators has been clearly demonstrated. Other studies have shown even more severe impacts of the quality analysis on time series, for instance in tropical western Africa, with persistent cloudy periods of up to eight months (Colditz et al. 2006a, Wagner et al. 2007).

Statistical methods have been utilized for evaluating and comparing time series. Typically, these statistical techniques evaluate the noise component and often conclude that time series with long-term interpolations are best. As noted above, the strictest setting often excludes too

many observations for a meaningful representation of the temporal course. Therefore, these comparisons have only a limited potential for a contextually correct assessment.

Since statistical functions applied to time series have been insufficient for examination, other approaches to assessment have to be developed. The currently on-going efforts of product inter-comparison seem to be the most promising alternative (Brown et al. 2006, Fensholt et al. 2006). However, no remote-sensing product is clearly independent from the dataset on which it is based. Therefore, the inherent differences due to the specifics of each sensor need to be considered and corrected (Gallo et al. 2005). An inter-comparison study of NDVI between AVHRR and MODIS, for instance, has yielded a good relative agreement but considerable absolute differences and clearly different dynamic ranges (Huete et al. 2002). The reasons for remarkable differences of the same product have been attributed to differences in spectral bands, susceptibility to water vapor, sensitivity to chlorophyll, and processing issues. In this respect, the product inter-comparison has only the capacity to indicate the level of agreement between two or more datasets, but it is not an approach for validation with reality. Phrased in drastic words, if both time series inaccurately represent the temporal development on the ground, but exhibit a high similarity, the comparison will nevertheless yield good results.

Large-scale and temporally continuous time-series validation is obviously not possible. Other potential validation techniques such as using model curves might adopt techniques from other fields of remote sensing, such as hyperspectral data analysis with field-measured endmember spectra. However, in contrast to physically-based spectral measurements in the electromagnetic spectrum, time series contain an inherent variability in time and intensity. Even temporally stable regions in the mid-latitudes, mainly driven by the seasonal increase and decrease in temperature, have a considerable variability in time. For instance, the onset of green-up can vary by up to two weeks. Similarly, the intensity can be considerably different, as has been indicated and assessed with remote-sensing products for the exceptionally hot summer of 2003 (Gobron et al. 2005, Reichstein et al. 2007). The thorough validation for highly variable ecosystems obviously becomes even more difficult. Therefore, it seems impossible to validate time series with model curves, because there are numerous other influences, including precipitation, water storage in the soil, land cover and land use, or fire.

Due to the limitations of quantitative assessment strategies, the visual assessment of temporal plots paired with contextual knowledge of the study region has been most important. TiSeG does not only provide capabilities for visualization of data-availability indicators. It also presents time-series plots of the original curve in comparison to an exemplary interpolation of the individual pixel, a small surrounding neighborhood, or a larger region. This visual assessment has been extensively employed during time-series generation and led to the conclusion that moderate quality specifications for time-series generation usually perform

best. This finding has been confirmed by other studies, for instance in tropical western Africa (Colditz et al. 2006a, Wagner et al. 2007).

The approach of time-series generation as developed in this study can be employed in conjunction with several other techniques for time-series processing and analysis. Because there is an evolving need for consistent long-term product generation and reliable long-term analysis, the TiSeG framework can be utilized as a first step in processing. Next, smoothing techniques, for instance, TIMESAT (Jönsson and Eklundh 2004) or HANTS (Roerink et al. 2000) can be employed for fine tuning, outlier reduction, and processing towards particular applications. There are also software packages for remote-sensing time-series analysis. Besides general statistical software packages, TIMESTATS (Udelhoven and Stellmes 2007), for instance, provide a multitude of functions for information extraction specifically devoted to time-series records of remote-sensing data. The conjunctive use of all tools and techniques mentioned above would make a powerful processing sequence from the raw data to a value-added product based on high-quality time series.

5.2 Time Series Classification and Analysis

The second main achievement of this study is the development of a robust and transferable classification procedure, also appropriate for time-series records. Fractional cover estimates are required due to the coarse spatial resolution of most suitable sensors. Since the small-patch landscape in developed regions and the transitional characteristics of natural ecosystems cause several difficulties with discrete classes, the class membership of each pixel is a much better estimator of the true surface characteristics.

Automated time-series classification directly serves the need of global change, and especially land-change research. Since time series can be produced in a highly automated fashion, it can be readily employed for land mapping and change detection. Several other studies have employed time series for regional and global classifications using a wide range of classifiers (DeFries and Townshend 1994, DeFries et al. 2000, Loveland et al. 2000). Most of the current, partly even operational classification approaches utilize decision-tree methodologies (Friedl et al. 2002, Hansen et al. 2003). Classification trees following the CART principle (Breiman et al. 1984) have been employed in this study. The satisfactory results indicate their general suitability. However, the good accuracy of automated time-series classification has been more an effect of the appropriate use of meta-strategies and the entire classification framework developed in this study rather than the individual decision tree.

State-of-the-art classification procedures do not simply focus on a satisfactory single classifier, but employ meta-strategies. These strategies combine individual classifiers, generally using the same technique (e.g. decision trees), by either varying the samples (bootstrapping) or the features (sub-spaces). The selection process is often random. In this study, a strategic variant of bagging (an acronym for bootstrap aggregation) has been employed. With respect to the framework of random forests (Breiman 2001), the meta-strategy of this study could be called “strategic forest”, since neither randomization nor the replacement of samples have been used.

In this study, individual classification trees, commonly yielding discrete results, have been modified to account for the uncertainty in leaves. Therefore, the proportion of each class has been stored for each individual tree, making a first step towards fuzzy classification. Several other parameters could also be modified in the decision-tree generation process. For instance, the employed splitting using thresholds could be adapted to fuzzy splits, which account for more uncertainty in the partitioning process (Quinlan 1993). Although pruning has not been found to have a significant impact on the results of this study, other pruning procedures may yield different results.

Another step of the fuzzy classification process has been the combination of all fuzzy classifications from individual decision trees using averaging. Alternatives such as the use of the kappa coefficient of the individual classification may introduce pseudo-fuzziness at a later stage of processing (Conrad et al. 2007). Other options for weighting might utilize the user’s and producer’s accuracy, but the consideration of error measures for softening discrete classifications only mimics a true fractional cover estimate. Boosting, for instance, would be a different technique for obtaining an improved fuzzy classification and will be tested in future studies.

In order to thoroughly assess the results of the classification and the impact of selected modules of the classification procedure, a detailed evaluation protocol has been developed for discrete maps and fuzzy class estimates. The simple use of high-confidence sample data, as often applied to evaluate classifications, is insufficient for state-of-the-art accuracy assessment, and in particular for a detailed sensitivity analysis. A hard and fuzzy reference map has been developed by multi-scale analysis. Altogether, four assessment strategies have been employed for accuracy assessment. Besides the use of standard high-confidence samples with error matrix statistics, a 1:1 comparison of discrete reference map and classification has been conducted, modified by the classification-certainty estimate. Because there is a lack of assessment approaches for fuzzy classifications, the evaluation protocol of this study has adopted error measures from other fields of data inter-comparison, including difference maps, relative and absolute mean, or RMSE. The assessments have been conducted for each class

and each pixel separately. The protocol developed in this study is considered useful to other studies concerned with a detailed accuracy assessment of hard and soft classifications.

A central part of the classification procedure has been the multi-scale analysis, transforming high spatial resolution data to a coarse resolution by storing the proportions of each class. This transformation has allowed a robust and reliable extraction of sample data for either training the classifier or assessing its results. Furthermore, the transformed dataset has been used as a fuzzy reference set for classification assessment (see above). Thirdly, the multi-scale analysis also has helped to derive appropriate classification schemes from a data point of view. The initial idea has been expanded to also account for adjacent high-resolution cells, making a more conservative estimate. The sensitivity study has found a marked increase in accuracy for moderate neighborhoods of up to 10 %. However, the enlargement of the neighborhood is inversely related to the availability of sample candidates. In this respect, adjacent areas being too large may yield a decreasing accuracy, because the remaining samples insufficiently represent all possible modes in feature space. An enlargement can also reduce the set of potential samples below the required threshold of 1,100 samples per class. In this case the homogeneity is reduced in order to achieve the required minimum sample set, also incorporating partly heterogeneous samples. However, a high number of heterogeneous samples can negatively affect the accuracy of the tree classifier. Also, other issues such as the minimum mapping unit have been investigated using the dataset of Minnesota. The minimum mapping unit has been found to have an influence on the multi-scale analysis and the classification results (7 %). Unfortunately it is not possible to correct for the effects inherent in the high-resolution data.

Another important consideration has been the selection of appropriate features and sample data, a conclusion which is valid for every classification process irrespective of the classifier. With regard to the first, features, it is uncommon to use time series directly. The high dimensionality and, yet more important, the high temporal correlation cause several difficulties, even if robust and advanced, non-parametric classifiers have been used. Therefore, time series are transformed to a set of metrics, usually simple univariate statistics such as mean and standard deviation (DeFries et al. 1998). This also restricts the impact of the quality of time series. It was found that the quality of time series has no effect on the accuracy of the classification. This result is important, because it proves that many pervious studies based on noisy AVHRR data have yielded reliable results, if simple metrics were employed.

However, the appropriate selection of metrics, representing the dimensions in feature space, has been very important. It has been shown that temporal segmentation for metrics computation yields higher accuracies (10 %), if NDVI time series have been utilized exclusively. Temporal partitioning of NDVI with phenological parameters has yielded slightly better results for vegetated classes. However, non-vegetated units have indicated

lower accuracies, because the temporal course of these surfaces does not correspond with phenological patterns. Applying temporal segmentation to the entire set of metrics including surface-reflectance bands has revealed no improvements in accuracy. This indicates that no significant information has been extracted to support the partitioning process of decision trees. However even more important, the classification accuracies have been up to 25 % higher with a full set of surface-reflectance data compared to using NDVI only. This puts results of Conrad (2006) in a broader perspective, which have shown the improvement of temporal segmentation with NDVI time series. Since the information content of surface-reflectance data is much higher, it is recommended to employ this full set without temporal segmentation instead of using NDVI and generating numerous new features with rather redundant information.

Sampling has been the other critical consideration in classification, in particular for decision trees, which have the potential to overestimate classes with larger sample sets (Hansen et al. 2000). Therefore, equalized sampling with the same number of samples for each class has been regarded to be most appropriate. However, it has been found that proportional sampling schemes perform better than strictly equalized sets. This effect may be due to thematically heterogeneous classes in the classification scheme, which become overrepresented in the tree due to their broader distribution in feature space. Stratified sampling schemes, considering all important sub-distributions of each class in feature space, are most desirable.

The classification procedure has been applied for land-cover mapping to both main study sites, South Africa and Germany. Furthermore, a vegetation map of Namibia has been successfully re-classified to illustrate the use for other purposes than land-cover mapping. The classification approach is useful for mapping broad transition and ecotones, especially important for natural and semi-natural landscapes. For instance, the wide transition of shubland and fynbos to grassland at the eastern part of the Karroo in South Africa has illustrated the merit of fuzzy classifications. Similar patterns with clear core areas and broad ecotones have been depicted for the vegetation map of Namibia; a continuous pattern, which better describes the reality than a discrete class assignment.

In Germany, the fuzzy class estimate has been useful to account for the heterogeneity of the small-patch landscape. The classification of contextually mixed classes has been difficult, because sample data are rather scattered in feature space. This has caused a general overestimation in discrete classifications, but the certainty of class assignment has been fairly low. Fuzzy results have indicated that the expected uniform class usually has yielded the second-highest membership. Hence there is a potential for regionally tuned results, for instance using weighting functions. This, however, would limit the transferability and therefore has been beyond the scope of this thesis. Furthermore, misclassifications of discrete

results usually have occurred between contextually related classes, for instance between different types of forest, but discrepancies between general categories have been rare.

It is concluded that time-series classification seems most appropriate to discriminate distinct vegetation types, for instance permanent and temporary grassland, woodland, and dense forests. It has also been possible to discriminate more detailed vegetation classes, as demonstrated for Namibia, but the result highly depends on specific information, i.e. appropriate features and training samples supporting class discrimination. In contrast to simple snapshots in time, time series often provide the additional phenological information needed for advanced vegetation discrimination. This has been well demonstrated by using the NDVI and temporal segmentation considering the phenological information. Nevertheless, the information content of surface-reflectance data is higher compared to using NDVI exclusively. Non-vegetated and sparsely vegetated classes in particular have been better characterized for large parts in South Africa, if surface reflectance was incorporated. Still, there is a unexplored potential for using phenological information directly, e.g. for improved discrimination of detailed classes. For instance, different types of savanna are a highly demanded product in biodiversity research. However, this requires specific information on the study area, which would limit the transferability and is therefore beyond the scope of this study.

Besides the goal of providing an approach for time-series classification independent from additional data, the capabilities for automated classification update have been explored and demonstrated for Germany and South Africa. Post-classification change-detection methods of the hard and fuzzy classification have been employed for consecutive years. It is assumed that there is negligible change in the land-cover composition of both regions. Indicating no change is a particular challenge, because most change-detection techniques tend to indicate more conversions than there are in reality, due to considerable uncertainty in the class assignment.

However, the robustness of the classification approach as well as the supplementary information of classification certainty supplied, together with the hard classification, provides the possibility to mitigate false changes. This has been successfully demonstrated for both study regions, indicating no change for higher confidence assignments. However, known large-scale alterations of the land surface, even if they were temporary, have been indicated for Germany using the fuzzy result as a change indicator. For instance, a distinct indication of change has been shown for the severe Elbe flood in 2002. Furthermore, a different land-cover composition - not necessarily a change in the land-cover class, but merely an indicator of differences in class components – has been depicted for the exceptionally hot year of 2003. In coincidence with other studies (Gobron et al. 2005, Reichstein et al. 2007), there have been no long-term effects of the hot summer on subsequent years.

This analysis of accurate quantitative change detection and subtle change indication of large areas is particularly interesting for several national and international organizations and programs concerned with global-change research. A plus of the methodology developed in this study is the highly automated process of time-series classification and change detection. This serves particular needs of land-change analysis, which requires rapid mapping results with an accurate indication of change processes. Knowing land-change hot spots, national and international organizations can focus their attention and analysis to these regions and utilize higher spatial-resolution data and regionally tuned methodologies for detailed studies. In this respect, it might be tested whether the automated classification process supports crisis information services by rapid mapping.

5.3 Final Remarks

Both, the semi-automatic time-series generation and the automatic fractional classification process have been extensively tested on MODIS data. The basic idea of TiSeG, the use of the quality information for time-series generation, has been fully automated and successfully applied to a long-term AVHRR record (Colditz et al. 2006b). This has demonstrated the transferability of the approach to other instruments. The transferability to advanced sensors with extensive quality indicators (e.g. MERIS) is rather clear and planned for the near future. Knowing that future sensors suitable for time series (e.g. VIIRS) will also contain ancillary data and data-quality indicators is a promising perspective for this software tool. In addition, the high spatial resolution time series of the upcoming Rapid Eye sensor fleet can be explored.

The automated classification process developed does not have particular prerequisites, except for a set of training data, because it is a supervised approach. Certainly a high spatial resolution dataset is most desirable for training and assessing the fuzzy classification. However, the study using the vegetation map of Namibia has indicated that other approaches such as buffering can also be used instead. Similarly, a point-based sample set derived by fieldwork can be employed using another sample-extraction module. The data to be classified is neither constrained to MODIS time series nor to time series at all. The classification process is clearly data-driven, and the partitioning process of decision trees as employed in this study extracts the most useful feature. Basically, all modules of the procedure, including the decision-tree classifier itself, can be substituted by other, more suitable modules for a specific application. In this respect, the classification procedure is a framework for automated classification, where modules most suitable for time-series classification have been assembled in this study.

The regional transferability has been extensively demonstrated for two study sites of environmentally different settings, South Africa and Germany. The TiSeG software does not have any constraints with regard to the study site; it is a completely data-driven process. The growing user community has employed the TiSeG software for processing several MODIS datasets in many distinctive environmental settings world-wide. For instance, TiSeG has been utilized for the phenological analysis of vegetation in Germany (Colditz et al. 2007a, Colditz et al. 2007b), long-term change-detection studies and relation to other environmental variables such as precipitation in South Africa (Colditz et al. 2007c), and crop-monitoring in Uzbekistan (Colditz et al. 2005, Conrad 2006, Conrad et al. 2007). Furthermore, TiSeG has been effectively applied to burn-severity mapping and the analysis of vegetation re-growth in a Mediterranean ecosystem in southwestern Australia (Walz et al. 2007) and the simulation of the water balance in western Africa (Wagner et al. 2007).

The automated classification process has been applied to both main study sites using the same assemblage of modules. Since the process is also data-driven, the classification scheme and the appropriate selection of features and samples determine the accuracy of the classification, which has been demonstrated for different metrics sets and sampling schemes. The high accuracies of the rather detailed legends (14 classes for both main study regions) considering the spatial resolution of MODIS data have indicated the good performance of the classification due to the fuzzy class estimation. The extension to a different thematic content, in this case vegetation types of Namibia, finally has demonstrated the flexibility of the modular design and transferability to another context (Colditz et al. 2007d).

In this respect, the methodological developments of this study provide the means for processing raw data to high-quality time series suitable for a multitude of applications in the framework of global-change studies. On the one hand, TiSeG provides the means for accurate time-series generation, a prerequisite for consistent datasets. On the other hand, the developed classification procedure employs time series for land-surface mapping. The robust and flexible approach of fractional cover estimation can serve regional to large-scale mapping studies of the land surface. The highly automated process allows for rapid product generation and map updating by providing classification-certainty estimates. These automatically and possibly even contemporary generated change indicators point to modification or conversion processes on the land surface. These hot-spots of land change may require further attention using high resolution data and regionally tuned methodologies.

In this respect the advanced time-series generation procedure and the robust but change-sensitive classification methodology serve as tools for state-of-the-art data processing. These tools can be and already have been employed by national and international organizations concerned with climate change, deforestation, land conversion, agricultural monitoring, biodiversity, and food security.

References

Ackerman, S.A., Strabala, K.I., Menzel, W.P., Frey, R.A., Moeller, C.C., Gumley, L.E., Baum, B.A., Wetzel Seeman, S. and Zhang, H., 2006. Discriminating clear-sky from cloud with MODIS, Algorithm Theoretical Basis Document (ATBD MOD35) Version 5.0, pp. 129.

Acocks, J.P.H., 1988. Veld types of South Africa. Botanical Research Institute, Department of Agriculture and Water Supply, Pretoria, 146 pp.

Anderson, J.R., Hardy, E.E., Roach J.T. and Witmer R.E., 1976. A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964, 28 p.

Anyamba, A., Tucker, C.J. and Mahoney, R., 2002. From El Niño to La Niña: vegetation response patterns over east and southern Africa during the 1997-2000 period. *Journal of Climate*, 15(21), pp. 3096-3103.

Asner, G.P., Townsend, A.R. and Braswell, B.H., 2000. Satellite observation of El Niño effects on Amazon forest phenology and productivity. *Geophysical Research Letters*, 27(7), pp. 981-984.

Asner, G.P. and Heidebrecht, K.B., 2003. Imaging spectroscopy for desertification studies: comparing AVIRIS and EO-1 Hyperion in Argentina drylands. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), pp. 1283-1296.

Atkinson, P.M. and Tatnall, A.R.L., 1997. Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), pp. 699-709.

Atkinson, P.M., Cutler, M.E.J. and Lewis, H., 1997. Mapping sub-pixel proportional land cover with AVHRR imagery. *International Journal of Remote Sensing*, 18(4), pp. 917-935.

Atkinson, P. M. and Foody, G. M., 2002. Uncertainty in Remote Sensing and GIS: Fundamentals. In Foody, G. M. and Atkinson, P. M. *Uncertainty in Remote Sensing and GIS.* John Wiley and Sons. New York, 326 pp.

Bailey, R.G., 2002. Ecoregions. In: A.R. Orme (Editor), The physical geography of North America. Oxford University Press, Oxford, pp. 235-245.

Ballantine, J.-A.C., Okin, G.S., Prentiss, D.E. and Roberts, D.A., 2005. Mapping North African landforms using continental scale unmixing of MODIS imagery. *Remote Sensing of Environment*, 97(4), pp. 470-483.

Barbieri, R., Montgomery, H., Qiu, S., Barnes, B., Knowles, D., jr, Che, N. and Goldberg, I.L., 1997. MODIS calibration plan: level 1B calibrated, geolocated radiances (MOD02), Draft of the MODIS Algorithm Theoretical Basis Document (ATBD MOD01) Version 2.0, pp. 70.

Baret, F., Morisette, J.T., Fernandes, R.A., Champeaux, J.L., Myneni, R.B., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S. and Nickeson, J.E., 2006. Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: Proposition of the CEOS-BELMANIP. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), pp. 1794-1803.

Baret, F., Weiss, M., Garrigue, S., Allard, D., Leroy, M., Jeanjean, H., Fernandes, R., Myneni, R.B., Morisette, J.T., Privette, J., Bohbot, H., Bosseno, R., Dedieu, G., Di Bella, C., Espana, M., Gond, V., Gu, X.F., Guyon, D., Lelong, C., Maisongrande, P., Mougin, E., Nilson, T., Veroustraete, F. and Vintilla, R., 2007. VALERI: A network of sites and a methodology for the validation of medium spatial resolution satellite products. *Remote Sensing of Environment*, pp. submitted.

Barnes, W.L., Pagano, T.S. and Salomonson, V.V., 1998. Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), pp. 1088-1100.

Barnes, W.L., Xiong, X. and Salomonson, V.V., 2003. Status of Terra MODIS and Aqua MODIS. *Advances in Space Research*, 32(11), pp. 2099-2106.

Bartholomé, E. and Belward, A.S., 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, 26(9), pp. 1959-1977.

Bauer, E. and Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning*, 36(1-2), pp. 105-139.

Becker, F. and Li, Z.-L., 1990. Toward a local split window method over land surface. *International Journal of Remote Sensing*, 11(3), 369-393.

Behrenfeld, M.J., Randerson, J.T., McClain, C.R., Feldman, G.C., Los, S.O., Tucker, C.J., Falkowski, P.G., Field, C.B., Frouin, R., Esaias, W.E., Kolber, D.D. and Pollack, N.H., 2001. Biospheric Primary Production During an ENSO Transition. *Science*, 291, pp. 2594-2597.

Belward, A.S., Estes, J.E. and Kline, K.D., 1999. The IGBP-DIS global 1km land cover data set DISCover: a project overview. *Photogrammetric Engineering and Remote Sensing*, 65(9), pp. 1013-1020.

Bezdek, J.C., 1981. Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York, 256 pp.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, Berlin, 738 pp.

Bittner, M., Offermann, D., Bugaeva, I.V., Kokin, G.A., Koshelkov, J.P., Krivolutsky, A., Tarasenko, D.A., Gil-Ojeda, M., Hauchecorne, A., Luebken, F.-J., de la Morena, B.A., Mourier, A., Nakane, H., Oyama, K.I., Schmidlin, F.J., Soule, I., Thomas, L. and Tsuda, T., 1994. Long period/large scale oscillations of temperature during the DYANA campaign. *Journal of Atmospheric and Terrestrial Physics*, 56(13-14), pp. 1675-1700.

Bonan, G.B., 2002. Ecological Climatology: Concepts and Applications. Cambridge University Press, 678 pp.

Borak, J.S. and Strahler, A.H., 1999. Feature selection and land cover classification of a MODIS-like data set for a semiarid environment. *International Journal of Remote Sensing*, 20(5), pp. 919-938.

Borak, J.S., Lambin, E.F. and Strahler, A.H., 2000. The use of temporal metrics for land cover change detection at coarse spatial scales. *International Journal of Remote Sensing*, 21(6-7), pp. 1415-1432.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., 1994. Time series analysis: forecasting and control. Prentice Hall, Englewood Cliffs, NJ, 587 pp.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984. Classification and regression trees. Wadsworth, Belmont, CA, 358 pp.

Breiman, L., 1996. Bagging predictors. *Machine Learing*, 24(2), pp. 123-140.

Breiman, L., 2001. Random forests. *Machine Learing*, 45, pp. 5-32.

Brockmann, C., 2004. Demonstration of the BEAM software . A tutorial for making best use of VISAT. In: H. Lacoste (Editor), Proceedings of MERIS User Workshop. ESA, ESA-ESRIN, Frascati, Italy.

Brown, M.E., Pinzon, J.E., Didan, K., Morisette, J.T. and Tucker, C.J., 2006. Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-Vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), pp. 1787-1793.

Camberlin, P., Martiny, N., Philippon, N. and Richard, Y., 2007. Determinants of the interannual relationships between remote sensed photosynthetic activity and rainfall in tropical Africa. *Remote Sensing of Environment*, 106, pp. 199-216.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H. and Rosen, D.B., 1992. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3(5), pp. 698-713.

Carpenter, G.A., Gjaja, M.N., Gopal, S. and Woodcock, C.E., 1997. ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(2), pp. 308-325.

Carpenter, G.A., Gopal, S., Macomber, S.A., Martens, S., Woodcock, C.E. and Franklin, J., 1999. A neural network method for efficient vegetation mapping. *Remote Sensing of Environment*, 70, pp. 326-338.

Chatfield, C., 2004. The analysis of time series: an introduction. Chapman & Hall, London, 352 pp.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B. and Eklundh, L., 2004a. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, 91(3-4), pp. 332-344.

Chen, Z.M., Babiker, I.S., Chen, Z.X., Komaki, K., Mohamed, M.A.A. and Kato, K., 2004b. Estimation of interannual variation in productivity of global vegetation using NDVI data. *International Journal of Remote Sensing*, 25(16), pp. 3139-3159.

Cihlar, J., Manak, D. and D'Iorio, M., 1994. Evaluation of compositing algorithms for AVHRR data over land. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2), pp. 427-437.

Cihlar, J., Ly, H., Li, Z. and Huang, F., 1997. Multitemporal, multichannel AVHRR data sets for land biosphere studies - artifacts and corrections. *Remote Sensing of Environment*, 60(1), pp. 35-57.

Clark, L.A. and Pergibon, D., 1992. Tree-based models. In: J.M. Chambers and T.J. Hastie (Editors), Statistical Models in S. Kluwer Academic Press, pp. 608.

Cohen, W.B. and Justice, C.O., 1999. Validating MODIS terrestrial ecology products: linking in situ and satellite measurements. *Remote Sensing of Environment*, 70(1), pp. 1-3.

Cohen, W.B., Maiersperger, T.K., Yang, Z., Gower, S.T., Turner, D.P., Ritts, W.D., Berterretche, M. and Running, S.W., 2003. Comparisons of land cover and LAI estimates derived from ETM+ and MODIS for four sites in North America: a quality assessment of 2000/2001 provisional MODIS products. *Remote Sensing of Environment*, 88.

Colditz, R.R., 2003. Land cover and geomorphological floodplain mapping of the lower Pánuco basin, Mexico, utilizing remote sensing and GIS-methods. Diplomarbeit Thesis, University of Wuerzburg, Wuerzburg, 198 pp.

Colditz, R.R., Conrad, C., Rücker, G.R., Schweitzer, C., Fistric, S., Schmidt, M. and Dech, S.W., 2005. Ableitung von phänologischen Verlaufsmustern aus MODIS-Zeitserien und Möglichkeiten der Anwendung. In: J. Strobl, T. Blaschke and G. Griesebner (Editors), AGIT. Beiträge zum 17. AGIT-Symposium Salzburg, Salzburg, pp. 94-99.

Colditz, R.R., Conrad, C., Wehrmann, T., Schmidt, M., Dech, S., 2006a Generation and assessment of MODIS time series using quality information. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2006, July 31st - August 4th 2006, Denver, CO, 6 pages.

Colditz, R.R., Conrad, C., Schmidt, M., Schramm, M., Schmidt, M. and Dech, S.W., 2006b. Mapping regions of high temporal variability in Africa, ISPRS mid-term symposium 2006 - remote sensing: from pixels to processes, Enschede, the Netherlands, pp. 6 pages.

Colditz, R.R., Wehrmann, T., Bachmann, M., Steinnocher, K., Strunz, G., Schmidt, M. and Dech, S.W., 2006c. Influence of image fusion approaches on classification accuracy - A case study. *International Journal of Remote Sensing*, 27(15), pp. 3311-3335.

Colditz, R.R., Conrad, C., Wehrmann, T., Schmidt, M., Dech, S., 2007a. TiSeG – A flexible software tool for time series generation of MODIS data utilizing the quality assessment science data set. *IEEE Transactions on Geoscience and Remote Sensing*, (submitted).

Colditz, R.R., Conrad, C., Wehrmann, T., Schmidt, M. and Dech, S.W., 2007b. Analysis of the quality of collection 4 and 5 vegetation index time series from MODIS. ISPRS Spatial Data Quality Symposium, Enschede, The Netherlands, accepted and CRC press.

Colditz, R.R., Gessner, U., Conrad, C., van Zyl, D., Malherbe, J., Newby, T., Landmann, T., Schmidt, M. and Dech, S.W., 2007c. Dynamics of MODIS time series for ecological applications in southern Africa, Fourth International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multitemp 2007), Leuven, Belgium, pp. CD-ROM.

Colditz, R.R., Keil, M., Strohbach, B., Gessner, U., Schmidt, M. and Dech, S.W., 2007d. Vegetation structure mapping with remote sensing time series: Capabilities and

improvements, 32nd International Symposium on Remote Sensing of Environment, San Jose, Costa Rica, pp. CD-ROM.

Congalton, R.G. and Green, K., 1999. Assessing the accuracy of remotely sensinf data: principles and prcatices. Lewis Publishers, 160 pp.

Conrad, C., Rücker, G.R., Colditz, R.R., Strunz, G. and Dech, S.W., 2004. Crop monitoring using multi-temporal MODIS remote sensing data in Khorezm, Uzbekistan, Proceedings of the Humboldt Kolleg 2004, Uzbekistan, July 6-10, 2004 Tashkent, Uzbekistan. The use of geographic information systems and simulation models for research and decision support in central asian river basins. Humboldt Kolleg, Tashkent, Uzbekistan, pp. 118-125.

Conrad, C., Colditz, R.R., Petrocchi, A., Rücker, G.R., Dech, S.W. and Schmidt, M., 2005. Time Series Generator – Ein flexibles Softwaremodul zur Generierung und Bewertung von Zeitserien aus NASA MODIS Datenprodukten. In: J. Strobl, T. Blaschke and G. Griesebner (Editors), AGIT. Beiträge zum 17. AGIT-Symposium Salzburg, Salzburg, pp. 100-105.

Conrad, C., 2006. Fernerkundungsbasierte Modellierung und hydrologische Messungen zur Analyse und Bewertung der landwirtschaftlichen Wassernutzung in der Region Khorezm (Usbekistan). Dissertation Thesis, University of Wuerzburg, Wuerzburg, 210 pp.

Conrad, C., Dech, S.W., Hafeez, M., Lamers, J., Martius, C. and Strunz, G., 2007. Mapping and assessing water use in a Central Asian irrigation system by utilizing MODIS remote sensing products. *Irrigation Drainage Systems*, DOI 10.1007/s10795-007-9029-z.

Coppin, P., Lambin, E.F., Jonckheere, I. and Muys, B., 2002. Digital change detection methods in natural ecosystem monitoring: a review. In: L. Bruzzone and P. Smits (Editors), Analysis of multi-temporal remote sensing images. World Scientific, New Jersey, pp. 440.

Crews-Meyer, K.A., Hudson, P.F. and Colditz, R.R., 2004. Landscape complexity and remote classification in eastern coastal Mexico: application of Landsat 7ETM data. *Geocarto International*, 19(1), pp. 45-57.

Datcu, M., Seidel, K. and Walessa, M., 1998. Spatial information retrieval from remote-sensing images—Part I: information theoretical perspective. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5), pp. 1431-1445.

Dean, W.R.J. and Milton, S.J., 1999. The Karroo: ecological patterns and processes. Cambridge University Press, Cambridge, 374 pp.

de Beurs, K.M. and Henebry, G.M., 2004a. Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. *Remote Sensing of Environment*, 89(4), pp. 423-433.

de Beurs, K.M. and Henebry, G.M., 2004b. Trend Analysis of the Pathfinder AVHRR Land (PAL) NDVI Data for the Deserts of Central Asia. *IEEE Transactions on Geoscience and Remote Sensing*, 1(4), pp. 282-286.

de Beurs, K.M. and Henebry, G.M., 2005a. Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high-latitude transects. *Global Change Biology*, 11, pp. 779-790.

de Beurs, K.M. and Henebry, G.M., 2005b. A statistical framework for the analysis of long image time series. *International Journal of Remote Sensing*, 26(8), pp. 1551-1573.

Dech, S.W., Tungalagsaikhan, P., Preusser, C. and Meissner, R.E., 1998. Operational value-adding to AVHRR data over Europe: methods, results, and prospects. *Aerospace Science and Technology*, 2(5), pp. 335-346.

Defourny, P., Vancutsem, C., Bicheron, P., F., N., Brockmann, C., Schouten, L. and Leroy, M., 2006. Towards a 300 m global land cover product - the GLOBCOVER initiative. In: M. Braun (Editor), Proceeding of the 2nd workshop of the EARSeL SIG on land use and land cover. EARSeL, Bonn, pp. 1.

DeFries, R.S. and Townshend, J.R.G., 1994. NDVI derived land cover classifications at a global scale. *International Journal of Remote Sensing*, 15(17), pp. 3567-3586.

DeFries, R.S., Hansen, M.C. and Townshend, J.R.G., 1995. Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data. *Remote Sensing of Environment*, 54(3), pp. 209-222.

DeFries, R.S., Hansen, M.C., Steininger, M., Dubayah, R.C., Sohlberg, R.A. and Townshend, J.R.G., 1997. Subpixel forest cover in central Africa from multisensor, multitemporal data. *Remote Sensing of Environment*, 60(3), pp. 228-246.

DeFries, R.S., Hansen, M.C., Townshend, J.R.G. and Sohlberg, R.A., 1998. Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, 19(16), pp. 3141-3168.

DeFries, R.S. and Los, S.O., 1999. Implications of land-cover misclassification for parameter estimates in global land-surface models: an example from the simple biosphere model (SiB2). *Photogrammetric Engineering and Remote Sensing*, 65(9), pp. 1083-1088.

DeFries, R.S., Townshend, J.R.G. and Hansen, M.C., 1999. Continuous fields of vegetation characteristics at the global scale at 1km resolution. *Journal of Geophysical Research*, 104(D14), pp. 16911-16923.

DeFries, R.S., Hansen, M.C. and Townshend, J.R.G., 2000. Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year 8 km AVHRR data. *International Journal of Remote Sensing*, 21(6-7), pp. 1389–1414.

De Lima, M.V.N., 2005. IMAGE2000 and CLC2000 - products and methods, JRC and EEA, Ispra.

Dennison, P.E. and Roberts, D.A., 2003a. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. *Remote Sensing of Environment*, 87, pp. 123-135.

Dennison, P.E. and Roberts, D.A., 2003b. The effects of vegetation phenology on endmember selection and species mapping in southern California chaparral. *Remote Sensing of Environment*, 87(2-3), pp. 295-309.

Dickinson, R.E., Henderson-Sellers, A., Kennedy, P.J. and Wilson, M.F., 1986. Biosphere-atmosphere transfer scheme (BATS) for the NCAR community climate model: NCAR Technical Note NCAR/TN275+STR, Boulder, CO. 69 p.

Didan, K. and Huete, A.R., 2006. MODIS vegetation index product series collection 5 change summary, Collection 5 changes, pp. 17.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learning*, 40(2), pp. 139-157.

Di Gregorio, A., 2005. Land Cover Classification System - Classification concepts and user manual for Software version 2, FAO Environment and Natural Resources Service Series, No. 8, Rome.

Dilley, M., 2003. Regional responses to climate variability in southern Africa. In: K. O'Brien and C. Vogel (Editors), Coping with climate variablity. Ashgate Publishing Company, Burlington, VT, pp. 35-47.

Duchemin, B. and Maisongrande, P., 2002. Normalisation of directional effects in 10-day global syntheses derived from VEGETATION/SPOT: I. Investigation of concepts based on simulation. *Remote Sensing of Environment*, 81(1), pp. 90-100.

Duda, R.O., Hart, P.E. and Stork, D.G., 2001. Pattern Classification. John Wiley & Sons, New York, 654 pp.

Eidenshink, J.C. and Faundeen, J.L., 1994. The 1km AVHRR global land data set: first stages in implementation. *International Journal of Remote Sensing*, 15(17), pp. 3443-3462.

el Saleous, N.Z., Vermote, E.F., Justice, C.O., Townshend, J.R.G., Tucker, C.J. and Goward, S.N., 2000. Improvements in the global biospheric record from the Advanced Very High Resolution Radiometer (AVHRR). *International Journal of Remote Sensing*, 21(6-7), pp. 1251-1277.

Erasmi, S., Bothe, M. and Petta, R.A., 2006. Enhanced filtering of MODIS time series data for the analysis of desertification in northern Brazil, ISPRS midterm symposium Remote sensing: from pixels to processes, Enschede,The Netherlands, pp. 6 pages.

Fensholt, R., Theis Nielsen, T. and Stisen, W., 2006. Evaluation of AVHRR PAL and GIMMS 10-day composite NDVI time series products using SPOT-4 vegetation data for the African continent. *International Journal of Remote Sensing*, 27(13), pp. 2719-2733.

Ferencz, C., Bognar, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timar, G., Molnar, G., Pasztor, S., Steinbach, P., Szekely, B., Ferencz, O.E. and Ferencz-Arkos, I., 2004. Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, 25(20), pp. 4113-4149.

Ferreira, L.G., Yoshioka, H., Huete, A.R. and Sano, E.E., 2003. Seasonal landscape and spectral vegetation index dynamics in the Brazilian Cerrado: an analysis within the large-scale biosphere–atmosphere experiment in Amazônia (LBA). *Remote Sensing of Environment*, 87(4), pp. 534-550.

Fomferra, N. and Brockmann, C., 2005. Beam - the ENVISAT MERIS and AATSR toolbox. In: H. Lacoste (Editor), Proceedings of the MERIS (A)ATSR Workshop 2005. ESA, Frascati, Italy.

Foody, G.M., 1992. On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, 58(1459-1460).

Foody, G.M., McCulloch, M.B. and Yates, W.B., 1995. The effect of training set size and composition on artificial neural network classification. *International Journal of Remote Sensing*, 16(9), pp. 1707-1723.

Foody, G.M., 1996. Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17(7), pp. 1317-1340.

Foody, G.M. and Arora, M.K., 1996. Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications. *Pattern Recognition Letters*, 17(13), pp. 1389-1398.

Foody, G.M. and Arora, M.K., 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing*, 18(4), pp. 799-810.

Foody, G.M., 2001. GIS: the accuracy of spatial data revisited. *Progress in Physical Geography*, 25, pp. 389-398.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp. 185-201.

Freund, Y. and Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), pp. 119-139.

Freund, Y. and Schapire, R.E., 1999. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5), pp. 771-780.

Friedl, M.A. and Brodley, C.E., 1997. Decision tree classification of landcover from remotely sensed data. *Remote Sensing of Environment*, 61(3), pp. 399-409.

Friedl, M.A., Brodley, C.E. and Strahler, A.H., 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), pp. 969-977.

Friedl, M.A., Muchoney, D., McIver, D.K., Gao, F., Hodges, J.C.F. and Strahler, A.H., 2000. Characterization of North American land cover from NOAA-AVHRR data using the EOS MODIS land cover classification algorithm. *Geophysical Research Letters*, 27(7), pp. 977-980.

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C.B.,

2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83(1-2), pp. 287-302.

Fu, R. and Li, W., 2004. The influence of the land surface on the transition from dry to wet season in Amazonia. *Theoretical and Applied Climatology*, 78, pp. 97-110.

Fuller, D.O., 1998. Trends in NDVI time series and their relation to rangeland and crop production in Senegal, 1987-1993. *International Journal of Remote Sensing*, 19(10), pp. 2013-2018.

Fuller, D.O., Jessup, T.C. and Salim, A., 2004. Loss of forest cover in Kalimatan, Indonesia, since the 1997-1998 El Niño. *Conservation Biology*, 18(1), pp. 249-254.

Gallo, K., Ji, L., Reed, B., Eidenshink, J.C. and Dwyer, J., 2005. Multi-platform comparisons of MODIS and AVHRR normalized difference vegetation index data. *Remote Sensing of Environment*, 99(3), pp. 221-231.

Gao, B.-C. and Kaufman, Y.J., 1998. The MODIS near-IR water vapor algorithm - total precipitable water, Algorithm Theoretical Basis Document (ATBD MOD05), pp. 25.

GCOS, 2004. http://www.wmo.ch/web/gcos/gcoshome.html (11. June 2007)

Geerken, R., Batikha, N., Celis, D. and Depauw, E., 2005. Differentiation of rangeland vegetation and assessment of its status: field investigations and MODIS and SPOT VEGETATION data analyses. *International Journal of Remote Sensing*, 26(20), pp. 4499-4526.

Geßner, U., Günther, K.P. and Maier, S.W., 2005. Land cover / land use map of Germany based on MERIS full-resolution data. In: ESA Proceedings of the 2004 Envisat & ERS Symposium, ESA-SP, 572, Envisat & ERS Symposium 2004, Salzburg, Austria, 6 - 10 September 2004.

Giess, W., 1971. A preliminary vegetation map of South West Africa. *Dinteria*, 4, pp. 5-114.

Glaser, R., Gebhardt, H. and Schenk, W., 2007. Geographie Deutschlands. Wissenschaftliche Buchgesellschaft, Darmstadt, 280 pp.

Gobron, N., Pinty, B., Melin, F., Taberner, M., Verstaete, M.M., Belward, A.S., Lavergne, T. and Widlowski, J.-L., 2005. The state of vegetation in Europe following the 2003 drought. *International Journal of Remote Sensing*, 26(9), pp. 2013–2020.

Goel, P.K., Prasher, S.O., Patel, R.M., Landry, J.A., Bonnell, R.B. and Viau, A.A., 2003. Classification of hyperspectral data by decision trees and artificial neural networks to identify

weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture*, 39, pp. 67-93.

GOFC-GOLD, 2007. http://www.gofc-gold.uni-jena.de/ (7. June 2007)

Gopal, S. and Woodcock, C.E., 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, 60(2), pp. 181-188.

Gopal, S., Woodcock, C.E. and Strahler, A.H., 1999. Fuzzy neural network classification of global land cover from a 1° AVHRR data set. *Remote Sensing of Environment*, 67(2), pp. 230-243.

Goward, S.N., Markham, B.L., Dye, D.G., Dulaney, W. and Yang, J., 1991. Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote Sensing of Environment*, 35(2-3), pp. 257-277.

Goward, S.N., Turner, S., Dye, D.G. and Liang, S., 1994. The University of Maryland improved global vegetation index product. *International Journal of Remote Sensing*, 15(17), pp. 3365-3395.

Guenther, B., Xiong, X., Salomonson, V.V., Barnes, W.L. and Young, J., 2002. On-orbit performance of the Earth Observing System Moderate Resolution Imaging Spectroradiometer; first year of data. *Remote Sensing of Environment*, 83(1-2), pp. 16-30.

Gutman, G. and Ignatov, A., 1996. The relative merit of cloud/clear identification in the NOAA/NASA Pathfinder AVHRR land 10-day composites. *International Journal of Remote Sensing*, 17(16), pp. 3295-3304.

Hall, D.K., Riggs, G.A. and Salomonson, V.V., 2001. MODIS Snow and Sea Ice-Mapping Algorithms (MOD10), Algorithm Theoretical Basis Document (ATBD) Version 5.0, pp. 45.

Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E. and Bayr, K.J., 2002. MODIS snow-cover products. *Remote Sensing of Environment*, 83(1-2), pp. 181-194.

Hall, F.G., Collatz, G., Los, S., Brown de Colstoun, E. and Landis, D., 2005. ISLSCP Initiative II. NASA. DVD/CD-ROM. NASA.

Han, K.-S., Champeaux, J.-L. and Roujean, J.-L., 2004. A land cover classification product over France at 1 km resolution using SPOT4/VEGETATION data. *Remote Sensing of Environment*, 92(1), pp. 52-66.

Hansen, M.C., Dubayah, R.C. and DeFries, R.S., 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, 17(5), pp. 1075-1081.

Hansen, M.C. and Reed, B.C., 2000. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *International Journal of Remote Sensing*, 21(6-7), pp. 1365-1373.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G. and Sohlberg, R.A., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21(6-7), pp. 1331-1364.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Marufu, L. and Sohlberg, R.A., 2002a. Development of a MODIS tree cover validation data set for Western Province, Zambia. *Remote Sensing of Environment*, 83(1-2), pp. 320-335.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R.A., DiMiceli, C. and Carroll, M.L., 2002b. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, 83(1-2), pp. 303-319.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C. and Sohlberg, R.A., 2003. Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, 7(10), pp. 1-15.

Hansen, M. and DeFries, R.S., 2004. Detecting long-term global forest change using continuous fields of tree-cover maps from 8km Advanced Very High Resolution Radiometer (AVHRR) data for the years 1982-99. *Ecosystems*, 7, pp. 695-716.

Herold, M., Woodcock, C.E., di Gregorio, A., Mayaux, P., Belward, A.S., Latham, J. and Schmullius, C., 2006a. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), pp. 1719-1727.

Herold, M., Latham, J.S., Di Gregorio, A. and Schmullius, C.C., 2006b. Evolving standards on land cover characterization. *Journal of Land Use Science*, 1(2-4), pp. 157-168.

Heuvelink G.B.M., 2002. Analyzing uncertainty propagation in GIS: Why is it not that simple. In Foody, G. M. and Atkinson, P. M. *Uncertainty in Remote Sensing and GIS*. John Wiley and Sons. New York, 326 pp.

Ho, T.K., 1995. Random decision forest, Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, Canada, pp. 278-282.

Ho, T.K., 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), pp. 832-844.

Ho, T.K., 2002. A data complexity analysis of comparative advantages of decision forest constructors. *Pattern Analysis and Applications*, 5, pp. 102-112.

Hoffman, T. and Ashwell, A., 2001. Land degradation in South Africa. University of Cape Town Press, Cape Town, 168 pp.

Holben, B.N., 1986. Characterization of maximum value composites from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11), pp. 1417-1434.

Holben, B.N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J.P., Setzer, A., Vermote, E.F., Reagan, J.A., Kaufman, Y.J., Nakajima, T., Lavenu, F., Jankowiak, I. and Smirnov, A., 1999. AERONET—a federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, 66(1), pp. 1-16.

Homer, C., Huang, C., Yang, L., Wylie, B. and Coan, M., 2004. Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, 70(7), pp. 829-840.

Hudson, P.F., Colditz, R.R. and Aguilar-Robledo, M., 2006. Spatial relations between floodplain environments and land use / land cover in a large humid tropical river valley: Pánuco basin Mexico. *Environmental Management*, 38(3), pp. 487-503.

Huemmrich, K.F., Privette, J.L., Mukelabai, M., Myneni, R.B. and Knyazikhin, Y., 2005. Time-series validation of MODIS land biophysical products in a Kalahari woodland, Africa. *International Journal of Remote Sensing*, 26(19), pp. 4381-4398.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), pp. 295-309.

Huete, A.R., Justice, C.O. and Liu, H., 1994. Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, 49(3), pp. 224-234.

Huete, A.R., Liu, H.Q., Batchily, K. and van Leeuwen, W.J.D., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59(3), pp. 440-451.

Huete, A., Justice, C.O. and van Leeuwen, W.J.D., 1999. MODIS Vegetation Index (MOD 13), Algorithm Theoretical Basis Document (ATBD) Version 3.0, pp. 129.

Huete, A.R., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1-2), pp. 195-213.

Huete, A.R., Didan, K., Shimabukuro, Y.E., Ratana, P., Saleska, S. R., Hutyra, L.R., Yang, W., Nemani, R.R., and Myneni, R., 2006. Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, 33.

IDL, 2007. http://www.ittvis.com/idl/ (May 30 2007).

Jakubauskas, M.E., Legates, D.R. and Kastens, J.H., 2001. Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing*, 67(4), pp. 461-470.

Jakubauskas, M.E., Legates, D.R. and Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, 37(1-3), pp. 127-139.

James, M.E. and Kalluri, S.N.V., 1994. The pathfinder AVHRR land data set: an improved coarse resolution data set for terrestrial monitoring. *International Journal of Remote Sensing*, 15(17), pp. 3347-3363.

Jensen, R.R., 2005. Introductory digital image processing. Prentice Hall, Upper Saddle River, NJ, 318 pp.

Jensen, R.R., 2007. Remote sensing of the environment. Prentice Hall, Upper Saddle River, NJ, 544 pp.

Ji, L. and Peters, A.J., 2003. Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices. *Remote Sensing of Environment*, 87(1), pp. 85-98.

Jönsson, P. and Eklundh, L., 2002. Seasonality extraction by function fitting to times-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8), pp. 1824-1832.

Jönsson, P. and Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, 30, pp. 833-845.

Justice, C.O., Townshend, J.R.G., Holben, B.N. and Tucker, C.J., 1985. Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, 6(8), pp. 1271-1318.

Justice, C.O., Vermote, E.F., Townshend, J.R.G., DeFries, R.S., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A.H., Lucht, W., Myneni, R.B., Knyazikhin, Y., Running, S.W., Nemani, R.R., Wan, Z., Huete, A.R., van Leeuwen, W.J.D., Wolfe, R.E., Giglio, L., Muller, J.-P., Lewis, P. and Barnsley, M.J., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), pp. 1228-1249.

Justice, C.O. and Townshend, J.R.G., 2002. Preface: Special issue on the moderate resolution imaging spectroradiometer (MODIS): a new generation of land surface monitoring. *Remote Sensing of Environment*, 83(1-2), pp. 1-2.

Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., el Saleous, N.Z., Roy, D.P. and Morisette, J.T., 2002a. An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, 83(1-2), pp. 3-15.

Justice, C.O., Giglio, L., Korontzi, S., Owens, J., Morisette, J.T., Roy, D.P., Descloitres, J., Alleaume, S., Petitcolin, F. and Kaufman, Y.J., 2002b. The MODIS fire products. *Remote Sensing of Environment*, 83(1-2), pp. 244-262.

Justice, C.O., Giglio, L., Boschetti, L., Roy, D.P., Csiszar, I., Morisette, J.T. and Kaufman, Y.J., 2006. MODIS fire products, Algorithm Theoretical Basis Document (ATBD MOD14) Version 2.3, pp. 34.

Kasischke, E.S. and French, N.H.F., 1997. Constraints on using AVHRR composite index imagery to study patterns of vegetation cover in boreal forests. *International Journal of Remote Sensing*, 18(11), pp. 2403-2426.

Kass, G.V., 1980. An explanatory technique for investigating large quantities of categorial data. *Applied Statistics*, 29(2), pp. 119-127.

Kaufman, Y.J. and Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), pp. 261-270.

Kaufman, Y.J., Herring, D.D., Ranson, K.J. and Collatz, G.J., 1998. Earth observing system AM1 mission to earth. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), pp. 1045-1055.

Kaufman, Y.J. and Justice, C.O., 1998. MODIS Fire Products (MOD14), Algorithm Technical Background Document (ATBD) Version 2.2, pp. 77.

Keil, M., Kiefl, R. and Strunz, G., 2005. CORINE land cover 2000 - Germany, German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen.

Kidwell, K., 1998. NOAA polar orbiter data user's guide. National Oceanic and Atmospheric Administration (NOAA) http://www2.ncdc.noaa.gov/docs/podug/index.htm (May 30 2007).

King, M.D., Closs, J., Wharton, S., Myers, M. and Parkinson, C.L. (Editors), 2004. EOS data products handbook, 1, Greenbelt, MD.

Klein, U., Sester, M. and Strunz, G., 1998. Segmentation of remotely sensed images based on the uncertainty of multispectral classification, GIS Between Visions and Applications. The International Archives of Photogrammetry and Remote Sensing, Stuttgart, Germany, pp. 299-305.

Klir, G.J. and Folger, T.A., 1988. Fuzzy sets, uncertainty and information. Prentice Hall, London.

Knyazikhin, Y., Martonchik, J.V., Diner, D.J., Myneni, R.B., Verstraete, M.M., Pinty, B. and Gobron, N., 1998a. Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data. *Journal of Geophysical Research*, 103(D24), pp. 32239-32256.

Knyazikhin, Y., Martonchik, J.V., Myneni, R.B., Diner, D.J. and Running, S.W., 1998b. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research*, 103(D24), pp. 32257-32275.

Knyazikhin, Y., Glassy, J., Privette, J.L., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J.T., Votava, P., Myneni, R.B., Nemani, R.R. and Running, S.W., 1999. MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15), Algorithm Theoretical Basis Document (ATBD) Version 4.0, pp. 130.

Kressler, F.P. and Steinnocher, K.T., 1999. Detecting land cover changes from NOAA-AVHRR data by using spectral mixture analysis. *International Journal of Applied Earth Observation and Geoinformation*, 1(1), pp. 21-26.

Lambin, E.F. and Strahler, A.H., 1994a. Change vector analysis in multitemporal space: a tool to detect and categorize land-cover change processes using high temporal-resolution satellite data. *Remote Sensing of Environment*, 48(2), pp. 231-244.

Lambin, E.F. and Strahler, A.H., 1994b. Indicators of land-cover change for change-vector analysis in multitemporal space at coarse spatial scales. *International Journal of Remote Sensing*, 15(10), pp. 2099-2119.

Lambin, E.F. and Ehrlich, D., 1996. The surface temperature- vegetation index space for land cover and land-cover change analysis. *International Journal of Remote Sensing*, 17(3), pp. 463-487.

Lambin, E.F. and Ehrlich, D., 1997. Land-cover changes in sub-saharan Africa (1982–1991): Application of a change index based on remotely sensed surface temperature and vegetation indices at a continental scale. *Remote Sensing of Environment*, 61(2), pp. 181-200.

Lambin, E.F., Geist, H.J. and Rindfuss, R.R., 2006. Introduction: local processes with global impacts. In: E.F. Lambin and H.J. Geist (Editors), Land-Use and Land-Cover Change. Springer, Berlin, pp. 1-8.

Landmann, T. Breda, F., Di Gregorio, A., Latham, J., Sarfatti P. and Giacomo, D., 2005. Looking Towards a new African Land Cover Dynamics Data Set: The medium resolution data-base for Africa (MEDA). presented at 3rd Proceedings of AFRICA GIS, Tshwane, South Africa.

LDOPE, 2007. http://landweb.nascom.nasa.gov/cgi-bin/QA_WWW/newPage.cgi (May 30 2007).

Leptoukh, G., Berrick, S., Rui, H., Liu, Z., Zhu, T. and Shen, S., 2005. NASA GES DISC On-line Visualization and Analysis System for Gridded Remote Sensing Data, Proc. of the 31th International Symposium of Remote Sensing of the Environment, St. Petersburg, Russia.

Li, W., and Fu, R., 2004. Transition of the Large-Scale Atmospheric and Land Surface Conditions from the Dry to the Wet Season over Amazonia as Diagnosed by the ECMWF Re-Analysis. *Journal of Climate*, 17 (13), pp. 2637-2651.

Li, X. and Strahler, A.H., 1992. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), pp. 276-292.

Li, Z. and Kafatos, M., 2000. Interannual variability of vegetation in the United States and its relation to El Niño/Southern Oscillation. *Remote Sensing of Environment*, 71(3), pp. 239-247.

Liang, S., Fang, H., Chen, M., Shuey, C.J., Walthall, C.L., Daughtry, C.S.T., Morisette, J.T., Schaaf, C.B. and Strahler, A.H., 2002. Validating MODIS land surface reflectance and albedo products: methods and preliminary results. *Remote Sensing of Environment*, 83(1-2), pp. 149-162.

Lillesand, T.M., Kiefer, R.W. and Chipman, J.W., 2003. Remote sensing and image interpretation. John Wiley and Sons, New York, 784 pp.

Liu, C., Frazier, P. and Kumar, L., 2007. Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment*, 107, pp. 606-616.

Lobell, D.B. and Asner, G.P., 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, 93, pp. 412-422.

Loh, W.-Y. and Vanichsetakul, N., 1988. Tree-structured classification via generalized discriminant analysis. *Journal of the American Statistical Association*, 83(403), pp. 715-728.

Loh, W.-Y. and Shih, Y.-S., 1997. Split selection methods for classification trees. *Statistica Sinica*, 7, pp. 815-840.

Los, S.O., Justice, C.O. and Tucker, C.J., 1994. A global 1° by 1° NDVI data set for climate studies derived from the GIMMS continental NDVI data. *International Journal of Remote Sensing*, 15(17), pp. 3493-3518.

Los, S.O., Weedon, G.P., North, P.R.J., Kaduk, J.D., Taylor, C.M. and Cox, P.M., 2006. An observation-based estimate of the strength of rainfall-vegetation interactions in the Sahel. *Geophysical Research Letters*, 33, pp. 5p.

Lotsch, A., Tian, Y., Friedl, M.A. and Myneni, R.B., 2003. Land cover mapping in support of LAI and FPAR retrievals from EOS-MODIS and MISR: classification methods and sensitivities to errors. *International Journal of Remote Sensing*, 24(10), pp. 1997-2016.

Loveland, T.R., Zhu, Z., Ohlen, D.O., Brown, J.F., Reed, B.C. and Yang, L., 1999. An analysis of the IGBP global land cover characterization process. *Photogrammetric Engineering and Remote Sensing*, 65(9), pp. 1021-1032.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L. and Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, 21(6), pp. 1303-1330.

Lovell, J.L. and Graetz, R.D., 2001. Filtering Pathfinder AVHRR Land NDVI data for Australia. *International Journal of Remote Sensing*, 22(13), pp. 2649–2654.

Lu, D., Mausel, P., Brondizio, E. and Moran, E., 2004. Change detection techniques. *International Journal of Remote Sensing*, 25(12), pp. 2365-2407.

Lucht, W., Schaaf, C.B. and Strahler, A.H., 2000. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2), pp. 977-998.

Lunetta, R.S., Knight, J.F., Ediriwickrema, J., Lyon, J. and Worthy, L.D., 2006. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*, 105, pp. 142-154.

Maiden, M.E. and Greco, S., 1994. NASA's pathfinder data set programme: land surface parameters. *International Journal of Remote Sensing*, 15(17), pp. 3333-3345.

Masuoka, E., Fleig, A.J., Wolfe, R.E. and Patt, F.S., 1998. Key characteristics of MODIS data products. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), pp. 1313-1323.

Mather, P.M., 2003. Computer processing of remotely-sensed images. John Wiley & Sons, Chichester, 292 pp.

McIver, D.K. and Friedl, M.A., 2001. Estimating pixel-scale land cover classification confidence using nonparametric machine learning methods. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9), pp. 1959-1968.

McIver, D.K. and Friedl, M.A., 2002. Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment*, 81(2-3), pp. 253-261.

McKnight, T.L., 1992. Regional Geography of the United States and Canada. Prentice Hall, Englewood Cliffs, NJ, 554 pp.

McVicar, T.R. and Bierwirth, P.N., 2001. Rapidly assessing the 1997 drought in Papua New Guinea using composite AVHRR imagery. *International Journal of Remote Sensing*, 22(11), pp. 2109-2128.

Mendelsohn, J.M. and Roberts, C.S., 1997. An environmental profile and Atlas of Caprivi. Directorate of Environmental Affairs, Namibia.

Mendelsohn, J.M., el Obeid S. and Roberts, C.S., 2000. A profile of north-central Namibia. Gamsberg Macmillan, Windhoek, 2000.

Mendelsohn, J.M., Jarvis, A., Roberts, C.S. and Robertson, T., 2002. Atlas of Namibia. David Philip Pubishers, Cape Town, 200 pp.

Mendelsohn, J.M. and el Obeid, S., 2003. Sand and Water. A profile of the Kavango Region. Struik Publishers & RAISON, Cape Town & Windhoek, 2003.

Menenti, M., Azzali, S., Verhoef, W. and van Swol, R., 1993. Mapping agroecological zones and time lag in vegetation growth by means of fourier analysis of time series of NDVI images. *Advances in Space Research*, 13(5), pp. 233-237.

Meynen, E. and Schmithüsen, J., 1953. Handbuch der naturräumlichen Gliederung Deutschlands. Remagen, Bundesanstalt für Landeskunde u. Raumforschung.

Middelkoop, H., 1990. Uncertainty in a GIS: a test for quanifying interpretation output. *ITC Journal*, 3, pp. 225-232.

Mika, J., Kerényi, J., Rimóczi-Paál, A., Merza, Á., Szinell, C. and Csiszár, I., 2002. On correlation of maize and wheat yield with NDVI: example of Hungary (1985-1998). *Advances in Space Research*, 30(11), pp. 2399-2404.

Mitchell, T.M., 1997. Machine Learning. McGraw-Hill, Singapore, 414 pp.

Miura, T., Huete, A.R. and Yoshioka, H., 2000. Evaluation of sensor calibration uncertainties on vegetation indices for MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3), pp. 1399-1409.

MODIS 2007a http://modis-land.gsfc.nasa.gov/ (March 30 2007).

MODIS 2007b http://edcdaac.usgs.gov/modis/dataproducts.asp (May 30 2007).

Moody, A. and Strahler, A.H., 1994. Characteristics of composited AVHRR data and problems in their classification. *International Journal of Remote Sensing*, 15(17), pp. 3473-3491.

Moody, A. and Woodcock, C.E., 1994. Scale-dependent errors in the estimation of land-cover proportions: implications for global land-cover datasets. *Photogrammetric Engineering and Remote Sensing*, 60(5), pp. 585-594.

Moody, A. and Johnson, D.M., 2001. Land-surface phenologies from AVHRR using the discrete fourier transform. *Remote Sensing of Environment*, 75(3), pp. 305-323.

Morisette, J.T., Privette, J.L. and Justice, C.O., 2002. A framework for the validation of MODIS Land products. *Remote Sensing of Environment*, 83(1-2), pp. 77-96.

Muchoney, D., Strahler, A.H., Hodges, J.C.F. and LoCastro, J., 1999. The IGBP DISCover confidence sites and the system for terrestrial ecosystem parameterization: tools for validating global land cover data. *Photogrammetric Engineering and Remote Sensing*, 65(9), pp. 1061-1067.

Muchoney, D. and Strahler, A.H., 2002. Regional vegetation mapping and direct land surface parameterization from remotely sensed and site data. *International Journal of Remote Sensing*, 23(6), pp. 1125-1142.

Mücher, C.A., Steinnocher, K.T., Kressler, F.P. and Heunks, C., 2000. Land cover characterization and change detection for environmental monitoring of pan-Europe. *International Journal of Remote Sensing*, 21(6-7), pp. 1159-1181.

Murthy, C.S., Raju, P.V. and Badrinath, K.V.S., 2003. Classification of wheat crop with multi-temporal images: performance of maximum likelihood and artificial neural networks. *International Journal of Remote Sensing*, 24(23), pp. 4871-4890.

Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrar, G. and Nemani, R.R., 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature*, 386(698-702).

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M.A., Morisette, J.T., Votava, P., Nemani, R.R. and Running, S.W., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83(1-2), pp. 214-231.

Myneni, R.B., Yang, W., Nemani, R.R., Huete, A.R., Dickinson, R.E., Knyazikhin, Y., Didan, K., Fu, R., Negron Juarez, R.I., Saatchi, S.S., Hashimoto, H., Ichii, K., Shabanov, N.V., Tan, B., Ratana, P., Privette, J.L., Morisette, J.T., Vermote, E.F., Roy, D.P., Wolfe, R.E., Friedl, M.A., Running, S.W., Votava, P., el Saleous, N.Z., Devadiga, S., Su, Y. and Salomonson, V.V., 2007. Large seasonal swings in leaf area of Amazon rainforests. *Proceedings of the National Academy of Sciences of the United States of America*, 104, pp. 4820-4823.

Nagendra, H., 2001. Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, 22(12), pp. 2377-2400.

NASA, 2003a. National aeronautics and space administraton strategic plan 2003. NASA, Washington, 64 pp.

NASA, 2003b. National aeronautics and space administraton Earth science enterprise strategy 2003. NASA, Washington, 94 pp.

NASA, 2007. National aeronautics and space administraton science plan 2007. NASA, Washington, 173 pp.

Nemani, R.R., Keeling, C.D., Hashimoto, H., Jolly, W.M., Piper, S.C., Tucker, C.J., Myneni, R.B. and Running, S.W., 2003. Climate-driver increases in global terrestrial net primary production from 1982 to 1999. *Science*, 300, pp. 1560-1563.

Neteler, M., 2005. Time series processing of MODIS satellite data for landscape epidemiological applications. *International Journal of Geoinformatics*, 1(1), pp. 133-138.

Nicholson, S.E., Davenport, M.L. and Malo, A.R., 1990. A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. *Climatic Change*, 17, pp. 209-241.

Nishihama, M., Wolfe, R.E., Solomon, D., Patt, F.S., Blanchette, J., Fleig, A.J. and Masuoka, E., 1997. MODIS level 1A: earth location (MOD03), Algorithm Theoretical Basis Document (ATBD) Version 3.0, pp. 147.

Olson, J.S., 1994a. Global ecosystem framework-definitions: USGS EROS Data Center Internal Report, Sioux Falls, SD, 37 p.

Olson, J.S., 1994b. Global ecosystem framework-translation strategy: USGS EROS Data Center Internal Report, Sioux Falls, SD, 39 p.

Pal, M. and Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86, pp. 554-565.

Pal, M., 2006. Error-corrected output coding-based class decomposition approach for remote sensing classification. *International Journal of Remote Sensing*, 27(14), pp. 2863-2876.

Parkinson, C.L. and Greenstone, R. (Editors), 2000. EOS data products handbook, 2, Greenbelt, MD.

Parkinson, C.L., 2003. Aqua: an earth-observing satellite mission to examine water and other climate variables. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), pp. 173-183.

Parkinson, C.L., Chahine, M.T., Kummerow, C.D. and Alomonson, V.V.S., 2003. Foreword to the EOS aqua special issue. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), pp. 172.

Patel, N.R., Bhattacharjee, B., Mohammed, A.J., Tanupriya, B. and Saha, S.K., 2006. Remote sensing of regional yield assessment of wheat in Haryana, India. *International Journal of Remote Sensing*, 27(19), pp. 4071-4090.

Paterson, J.H., 1994. North America. Oxford University Press, Oxford, 529 pp.

Pax-Lenney, M., Woodcock, C.E., Macomber, S.A., Gopal, S. and Song, C., 2001. Forest mapping with a generalized classifier and Landsat TM data. *Remote Sensing of Environment*, 77, pp. 241-250.

Petitcolin, F. and Vermote, E.F., 2002. Land surface reflectance, emissivity and temperature from MODIS middle and thermal infrared data. *Remote Sensing of Environment*, 83(1-2), pp. 112-134.

Pisek, J. and Chen, J.M., 2007. Comparison and validation of MODIS and VEGETATION global LAI products over four BigFoot sites in North America. *Remote Sensing of Environment*, pp. in press.

Piwowar, J.M. and LeDrew, E.F., 2002. ARMA time series modelling of remote sensing imagery: a new approach for climate change studies. *International Journal of Remote Sensing*, 23(24), pp. 5225-5248.

Potter, C.S. and Brooks, V., 1998. Global analysis of empirical relations between annual climate and seasonality of NDVI. *International Journal of Remote Sensing*, 19(15), pp. 2921-2948.

Privette, J.L., Myneni, R.B., Knyazikhin, Y., Mukelabai, M., Roberts, G., Tian, Y., Wang, Y. and Leblanc, S.G., 2002. Early spatial and temporal validation of MODIS LAI product in the Southern Africa Kalahari. *Remote Sensing of Environment*, 83(1-2), pp. 232-243.

Privette, J.L. and Roy, D.P., 2005. Southern Africa as a remote sensing test bed: the SAFARI 2000 Special Issue overview. *International Journal of Remote Sensing*, 26(19), pp. 4141-4158.

Quinlan, J.R., 1993. C4.5 programs for machine learning. Morgan Kaufmann Publishers, San Mateo, CA, 302 pp.

Quinlan, J.R., 1996. Bagging, boosting, and C4.5, Proceedings of the 14th National Conference on Artificial Intelligence, Menlo Park, CA, pp. 725-730.

Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W. and Ohlen, D.O., 1994. Measuring phenological variability from satellite data. *Journal of Vegetation Science*, 5, pp. 703-714.

Reeves, M.C., Zhao, M. and Running, S.W., 2005. Usefulness and limits of MODIS GPP for estimating wheat yield. *International Journal of Remote Sensing*, 26(7), pp. 1403-1421.

Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S.W., Viovy, N., Cramer, W., Granier, A., Ogée, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., Grünwald, T., Heimann, M., Heinesch, B., Knohl, A., Kutsch, W., Loustau, D., Manca, G., Matteuc, G., Miglietta, F., Ourcival, J.M., Pilegaard, K., Pumpanen, J., Rambal, S., Schaphoff, S., Seufert, G., Soussana, J.-F., Sanz, M.-J., Vesala, T. and Zhao, M., 2007.

Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Global Change Biology*, 13, pp. 634-651.

Remer, L.A., Tanre, D. and Kaufman, Y.J., 2006. Algorithm for remote sensing of tropospheric aerosol from MODIS, Algorithm Theoretical Basis Documents (ATBD MOD04), pp. 87.

Richards, J.A. and Jia, X., 2006. Remote sensing digital image analysis. Springer, Berlin, 340 pp.

Rizzi, R., Rudorff, B.F.T., Shimabukuro, Y.E. and Doraiswami, P.C., 2006. Assessment of MODIS LAI retrievals over soybean crop in southern Brazil. *International Journal of Remote Sensing*, 27(19), pp. 4091-4100.

Roberts, D.A., Smith, M.O. and Adams, J.B., 1993. Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment*, 44, pp. 255-269.

Roberts, D.A., Gardner, M.E., Church, R., Ustin, S.L., Scheer, G. and Green, R.O., 1998. Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, 65, pp. 267-279.

Roberts, D.A., Dennison, P.E., Gardner, M.E., Hetzel, Y., Ustin, S.L. and Lee, C.T., 2003. Evaluation of the potential of Hyperion for fire danger assessment by comparison to the airborne visible/infrared imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), pp. 1997-1310.

Roebeling, R.A., van Putten, E., Genovese, G. and Rosema, A., 2004. Application of Meteosat derived meteorological information for crop yield predictions in Europe. *International Journal of Remote Sensing*, 25(23), pp. 5389-5401.

Roerink, G.J., Menenti, M. and Verhoef, W., 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, 21(9), pp. 1911-1917.

Roerink, G.J., Menenti, M., Soepboer, W. and Su, Z., 2003. Assessment of climate impact on vegetation dynamics by using remote sensing. *Physics and Chemistry of the Earth*, 28(1-3), pp. 103-109.

Rogan, J., Franklin, J. and Roberts, D.A., 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment*, 80, pp. 143-156.

Rosenfield, G.H. and Fitzpatrick-Lins, K., 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52, pp. 223-227.

Ross, J.K., 1981. The radiation regime and architecture of plant stands. Norwell, MA: Dr. W. Junk, 392 pp.

Roy, D.P., 1997. Investigation of the maximum Normalized Difference Vegetation Index (NDVI) and the maximum surface temperature (Ts) AVHRR compositing procedure for the extraction of NDVI and Ts over forest. *International Journal of Remote Sensing*, 18(11), pp. 2383-2401.

Roy, D.P., 2000. The impact of misregistration upon composited wide field of view satellite data and implications for change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4), pp. 2017-2032.

Roy, D.P., Borak, J.S., Devadiga, S., Wolfe, R.E., Zheng, M. and Descloitres, J., 2002a. The MODIS Land product quality assessment approach. *Remote Sensing of Environment*, 83(1-2), pp. 62-76.

Roy, D.P., Lewis, P.E. and Justice, C.O., 2002b. Burned area mapping using multi-temporal moderate spatial resolution data—a bi-directional reflectance model-based expectation approach. *Remote Sensing of Environment*, 83(1-2), pp. 263-286.

Running, S.W., Loveland, T.R., Pierce, L.L., Nemani, R.R. and Hunt, E.R., jr., 1995. A remote sensing based vegetation classification logic for global land cover analysis. *Remote Sensing of Environment*, 51(1), pp. 39-48.

Running, S.W., Nemani, R.R., Glassy, J. and Thornton, P.E., 1999a. MODIS Daily Photosynthesis (PSN) and Annual Net Primary Production (NPP) product (MOD17), Algorithm Theoretical Basis Document (ATBD) Version 4.0, pp. 59.

Running, S.W., Baldocchi, D.D., Turner, D.P., Gower, S.T., Bakwin, P.S. and Hibbard, K.A., 1999b. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of Environment*, 70(1), pp. 108-127.

Russell, S.J. and Norvig, P., 2003. Artificial Intelligence: A Modern Approach. Prentice Hall, 1132 pp.

Saleska, S.R., Miller, S.D., Matross, D.M., Goulden, M.L., Wofsy, S.C., da Rocha, H.R., de Camargo, P.B., Crill, P., Daube, B.C., de Freitas, H.C., Hutyra, L., Keller, M., Kirchhoff, V.,

Menton, M., Munger, J.W., Hammond Pyle, E., Rice, A.H. and Silva, H., 2003. Carbon in Amazon forests: unexpected seasonal fluxes and disturbance-induced losses. *Science*, 302, pp. 1554-1557.

Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X., Tsang, T., Strugnell, N.C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M.J., Hobson, P.D., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R.P., Hug, B., Liang, S., Privette, J.L. and Roy, D.P., 2002. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83(1-2), pp. 135-148.

Schultz, P.A. and Halpert, M.S., 1993. Global correlation of temperature, NDVI and precipitation. *Advances in Space Research*, 13(5), pp. 277-280.

Seemann, S.W., Borbas, E.E., Li, J., Menzel, W.P. and Gumley, L.E., 2006. MODIS atmospheric profile retrieval. *Algorithm Theoretical Basis Document (ATBD MOD07) Version 6.0*, pp. 40.

Sellers, P.J., Mintz, Y., Sud, Y.C. and Dalcher A., 1986. A simple biosphere model (SiB) for use within general circulation models. *Journal of Atmospheric Science*, 43, pp. 505-531.

Sellers, P.J., Tucker, C.J., Collatz, G.J., Los, S.O., Justice, C.O., Dazlich, D.A. and Randall, D.A., 1994. A global 1° by 1° NDVI data set for climate studies. Part 2: the generation of global fields of terrestrial biophysical parameters from the NDVI. *International Journal of Remote Sensing*, 15(17), pp. 3519-3545.

Sellers, P.J., Randall, D.A., Collatz, G.J., Berry, J.A., Field, C.B., Dazlich, D.A., Zhang, C., Collelo, G.D. and Bounoua, L., 1996a. A revised land surface parameterization(SiB2) for atmospheric GCMs. Part I: Model formulation. *Journal of Climate*, 9(4), pp. 676-705.

Sellers, P.J., Los, S.O., Tucker, C.J., Justice, C.O., Dazlich, D.A., Collatz, G.J. and Randall, D.A., 1996b. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data. *Journal of Climate*, 9(4), pp. 706-737.

Shi, Z., Ruecker, G. R., Mueller, M., Conrad, C., Ibragimov, N., Lamers, J. P. A., Martius, C., Strunz, G., Dech, S. and Vlek, P. L. G., 2007. Modeling of cotton yields in the Amu Darya river floodplains of Uzbekistan integrating multi-temporal remote sensing and minimum field data, Agronomy Journal, accepted.

Skidmore, A.K., Turner, B.J., Brinkhof, W. and Knowles, E., 1997. Performance of a neural network: mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 63(5), pp. 501-514.

Snyder, W.C. and Wan, Z., 1998. BRDF models to predict spectral reflectance and emissivity in the thermal infrared. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1), pp. 214-225.

Snyder, W.C., Wan, Z., Zhang, Y. and Feng, Y.-Z., 1998. Classification-based emissivity for land surface temperature measurement from space. *International Journal of Remote Sensing*, 19(14), pp. 2753-2774.

Stehman, S.V. and Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, 64(3), pp. 331-334.

Stowe, L.L., McClain, E.P., Carey, R., Pellegrino, P., Gutman, G.G., Davis, P., Long, C. and Hart, S., 1991. Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Advances in Space Research*, 11(3), pp. 51-54.

Strahler, A.H., Lucht, W., Schaaf, C.B., Tsang, T., Gao, F., Li, X., Muller, J.-P., Lewis, P. and Barnsley, M.J., 1999a. MODIS BRDF/Albedo Product, MODIS Algorithm Theoretical Basis Document (ATBD MOD43) Version 5.0, pp. 53.

Strahler, A.H., Muchoney, D., Borak, J.S., Friedl, M.A., Gopal, S., Lambin, E.F. and Moody, A., 1999b. MODIS Land Cover and Land-Cover Change (MOD12), Algorithm Theoretical Basis Document (ATBD) Version 5.0, pp. 72.

Strohbach, B.J., 2001. Vegetation survey of Namibia. *Journal of the Namibia Scientific Society*, 49, pp. 93-124.

Stroppiana, D., Grégoire, J.-M., Pereira, J.M.C., 2003. The use of SPOT VEGETATION data in a classification tree approach for burnt area mapping in Australian savanna. *International Journal of Remote Sensing*, 24(10), pp. 2131-2151.

Strugnell, N.C. and Lucht, W., 2001. An algorithm to infer continental-scale albedo from AVHRR data, land cover class, and field observations of typical BRDFs. *Journal of Climate*, 14(7), pp. 1360-1376.

Thompson, M., 1996. A standard land-cover classification scheme for remote-sensing applications in South Africa. *South African Journal of Science*, 92, pp. 34-42.

Tottrup, C. and Rasmussen, M.S., 2004. Mapping long-term changes in savannah crop productivity in Senegal through trend analysis of time series of remote sensing data. *Agriculture, Ecosystems and Environment*, 103(3), pp. 545-560.

Townshend, J.R.G. and Justice, C.O., 1986. Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing*, 7(11), pp. 1435-1445.

Townshend, J.R.G., 1994. Global data sets for land applications from the advanced very high resolution radiometer: an introduction. *International Journal of Remote Sensing*, 15(17), pp. 3319-3332.

Townshend, J.R.G., Zhan, X., DeFries, R., Hansen, M., Dimiceli, C., Sohlberg, R. and Huang, C., 1999. MODIS Enhanced Land Cover and Land Cover Change Product (MOD44), Algorithm Theoretical Basis Documents (ATBD) Version 2.0, pp. 93.

Townshend, J.R.G. and Justice, C.O., 2002. Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. *Remote Sensing of Environment*, 83(1-2), pp. 351-359.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), pp. 127-150.

Tucker, C.J., Newcomb, W.W. and Dregne, H.E., 1994. AVHRR data sets for determination of desert spatial extent. *International Journal of Remote Sensing*, 15(17), pp. 3547-3565.

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F. and el Saleous, N.Z., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26(20), pp. 4485–4498.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, 18(6), pp. 306-314.

Tyson, P.D. and Preston-Whyte, R.A., 2000. The weather and climate of southern Africa. Oxford University Press, Oxford, 396 pp.

Udelhoven, T. and Stellmes, M., 2007. Changes in land surface conditions on the Iberian peninsula (1989 to 2004) detected by means of time series analysis from hypertemporal remote sensing data. In: G.J.M. De Lannoy, R. Hoeben, N.E.C. Verhoest, W.W. Verstraeten, S. Bruneel and P. Coppin (Editors), Multitemp 2007. IEEE, Leuven, Belgium, pp. CD-ROM.

UNESCO, 1973. International Classification and mapping of vegetation. *Ecology and Conservation*, 6, pp. 15-37.

van der Merwe, H., 1994. Illustrated Atlas of southern Africa. Reader's Digest, Cape Town, 232 pp.

van Leeuwen, W.J.D., Laing, T.W. and Huete, A.R., 1997. Quality assurance of global vegetation index compositing algorithms using AVHRR data, Proceedings of IGARSS '97, Singapore, pp. 341-343.

van Leeuwen, W.J.D., Huete, A.R. and Laing, T.W., 1999. MODIS vegetation index compositing approach: a prototype with AVHRR data. *Remote Sensing of Environment*, 69(3), pp. 264-280.

Venables, W.N. and Ripley, B.D., 2002. Modern applied statistics with S. Springer, Berlin, 495 pp.

Vermote, E.F., el Saleous, N.Z., Justice, C.O., Kaufman, Y.J., Privette, J.L., Remer, L., Roger, J.C. and Tanre, D., 1997a. Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces: background, operational algorithm and validation. *Journal of Geophysical Research*, 102(D14), pp. 17131-17141.

Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M. and Morcrette, J.-J., 1997b. Second simulation of the satellite signal in the solar spectrum, 6S: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3), pp. 675-686.

Vermote, E.F. and Vermeulen, A., 1999. Atmospheric correction algorithm: spectral reflectances (MOD09), MODIS Algorithm Technical Basis Document (ATBD) Version 4.0, pp. 107.

Vermote, E.F., el Saleous, N.Z. and Justice, C.O., 2002. Atmospheric correction of MODIS data in the visible to middle infrared: first results. *Remote Sensing of Environment*, 83(1-2), pp. 97-111.

Vikhamar, D. and Solberg, R., 2003. Snow-cover mapping in forests by constrained linear spectral unmixing of MODIS data. *Remote Sensing of Environment*, 88, pp. 309-323.

Viovy, N., Arino, O. and Belward, A., 1992. The Best Index Slope Extraction (BISE): a method for reducing noise in NDVI time-series. *International Journal of Remote Sensing*, 13, pp. 1585–1590.

Viovy, N., 2000. Automatic Classification of Time Series (ACTS): a new clustering method for remote sensing time series. *International Journal of Remote Sensing*, 21(6-7), pp. 1537-1560.

Vogel, C., 2003. Climate an climatic change: causes and consequences. In: R. Fox and K. Rowntree (Editors), The geography of South Africa in a changing world. Oxford University Press, Oxford, pp. 284-303.

Vogel, C. and O'Brien, K., 2003. Climate forecast in southern Africa. In: K. O'Brien and C. Vogel (Editors), Coping with climate variability. Ashgate Publishing Company, Burlington, VT, pp. 3-34.

Wagner, S. Kunstmann, H., Bárdossy, A., Conrad, C., Colditz, R.R., 2007. Water balance simulations in a poorly gauged basin using different meteorological and land surface data sources. *Physics and Chemistry of the Earth*, submitted.

Walz, Y., Maier, S.W., Dech, S.W., Conrad, C. and Colditz, R.R., 2007. Classification of burn severity using Moderate Resolution Imaging Spectroradiometer (MODIS): A case study in the jarrah-marri forest of southwest Western Australia. *Journal of Geophysical Research*, 112(G02002), pp. 14 pp.

Wan, Z. and Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4), pp. 892-905.

Wan, Z. and Li, Z.-L., 1997. A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4), pp. 980-996.

Wan, Z., 1999. MODIS Land-Surface Temperature (MOD11), Algorithm Theoretical Basis Document (ATBD) Version 3.3, pp. 77.

Wan, Z., Zhang, Y., Zhang, Q. and Li, Z.-L., 2002. Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sensing of Environment*, 83(1-2), pp. 163-180.

Wan, Z., Zhang, Y., Zhang, Q. and Li, Z.-L., 2004. Quality assessment and validation of the MODIS global land surface temperature. *International Journal of Remote Sensing*, 25(1), pp. 261-274.

Wang, F., 1990. Fuzzy supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 28(2), pp. 194-201.

Wannebo, A. and Rosenzweig, C., 2003. Remote sensing of US cornbelt areas sensitive to the El Niño-Southern Oscillation. *International Journal of Remote Sensing*, 24(10), pp. 2055-2067.

Wanner, W., Strahler, A.H., Hu, B., Lewis, P., Muller, J.-P., Li, X., Barker Schaaf, C.L. and Barnsley, M.J., 1997. Global retrieval of bidirectional reflectance and albedo over land from EOS MODIS and MISR data: theory and algorithm. *Journal of Geophysical Research*, 102(D14), pp. 17143-17161.

Wehrmann, T., 2007. Automatisierte Klassifikation von Landnutzung durch Objekterkennung am Beispiel von CORINE Land Cover, University of Wuerzburg, Wuerzburg, 152 pp.

Weiss, E., Marsh, S.E. and Pfirman, E.S., 2001. Application of NOAA-AVHRR NDVI time-series data to assess changes in Saudi Arabia's rangelands. *International Journal of Remote Sensing*, 22(6), pp. 1005-1027.

Weiss, M., Baret, F., Garrigues, S. and Lacaze, R., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. *Remote Sensing of Environment*, pp. in press.

Wolfe, R.E., Roy, D.P. and Vermote, E.F., 1998. MODIS land data storage, gridding, and compositing methodology: level 2 grid. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), pp. 1324-1338.

Wolfe, R.E., Nishihama, M., Fleig, A.J., Kuyper, J.A., Roy, D.P., Storey, J.C. and Patt, F.S., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sensing of Environment*, 83(1-2), pp. 31-49.

Woodcock, C.E. and Gopal, S., 2000. Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science*, 14(2), pp. 153-172.

Wuest, S. and Bittner, M., 2006. Non-linear resonant wave–wave interaction (triad): Case studies based on rocket data and first application to satellite data. *Journal of Atmospheric and Solar-Terrestrial Physics*, 68, pp. 959-976.

Xavier, A.C., Rudorff, B.F.T., Shimabukuro, Y.E., Berka, L.M. and Moreira, M.A., 2006. Multi-temporal analysis of MODIS data to classify sugarcane crop. *International Journal of Remote Sensing*, 27(4), pp. 755-768.

Xiong, X. and Barnes, W.L., 2006. An overview of MODIS radiometric calibration and characterization. *Advances in Atmospheric Sciences*, 23(1), pp. 69-79.

Yang, C.-C., Prasher, S.O., Enright, P., Madramootoo, C., Burgess, M., Goel, P.K. and Callum, I., 2003. Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems*, 76, pp. 1101-1117.

Yang, L., Homer, C., Hegge, K., Huang, C., Wylie, B. and Reed, B., 2001. A Landsat 7 scene selection strategy for a national land cover database, Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium IGARSS2001, Sydney, Australia, pp. CD-ROM.

Yu, F., Price, K.P., Ellis, J. and Shi, P., 2003. Response of seasonal vegetation development to climatic variations in eastern central Asia. *Remote Sensing of Environment*, 87(1), pp. 42-54.

Yu, F., Price, K.P., Ellis, J., Feddema, J.J. and Shi, P., 2004. Interannual variations of the grassland boundaries bordering the eastern edges of the Gobi Desert in central Asia. *International Journal of Remote Sensing*, 25(2), pp. 327-346.

Zhan, X., DeFries, R.S., Townshend, J.R.G., Dimiceli, C., Hansen, M.C., Huang, C. and Sohlberg, R.A., 2000. The 250m global land cover change product from the Moderate Resolution Imaging Spectroradiometer of NASA's Earth Observing System. *International Journal of Remote Sensing*, 21(6-7), pp. 1433-1460.

Zhan, X., Sohlberg, R.A., Townshend, J.R.G., DiMiceli, C., Carroll, M.L., Eastman, J.C., Hansen, M.C. and DeFries, R.S., 2002. Detection of land cover changes using MODIS 250 m data. *Remote Sensing of Environment*, 83(1-2), pp. 336-350.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C. and Huete, A.R., 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3), pp. 471-475.

Appendix of Figures

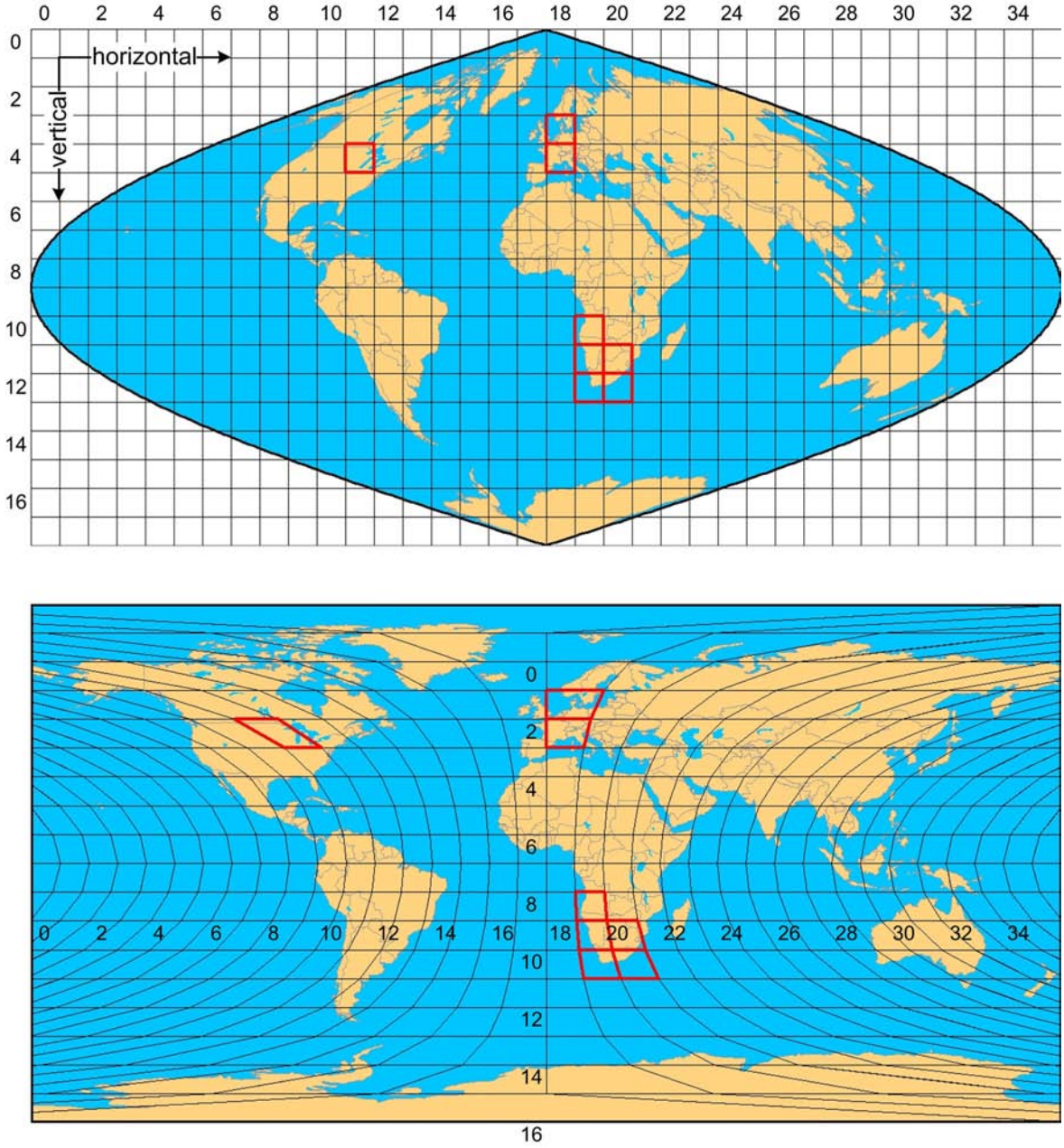


Figure A.1: Global distribution of MODIS tiles.
Note: All tiles used in this study are highlighted in red color. The upper figure illustrated the Earth in sinusoidal projection with the respective location of MODIS tiles. The lower part depicts the MODIS tiles projected to geographic latitude and longitude. Note that in this representation 90˚ North and South are located at the top and bottom line; hence the 0˚ meridian could intersect anywhere on this line.

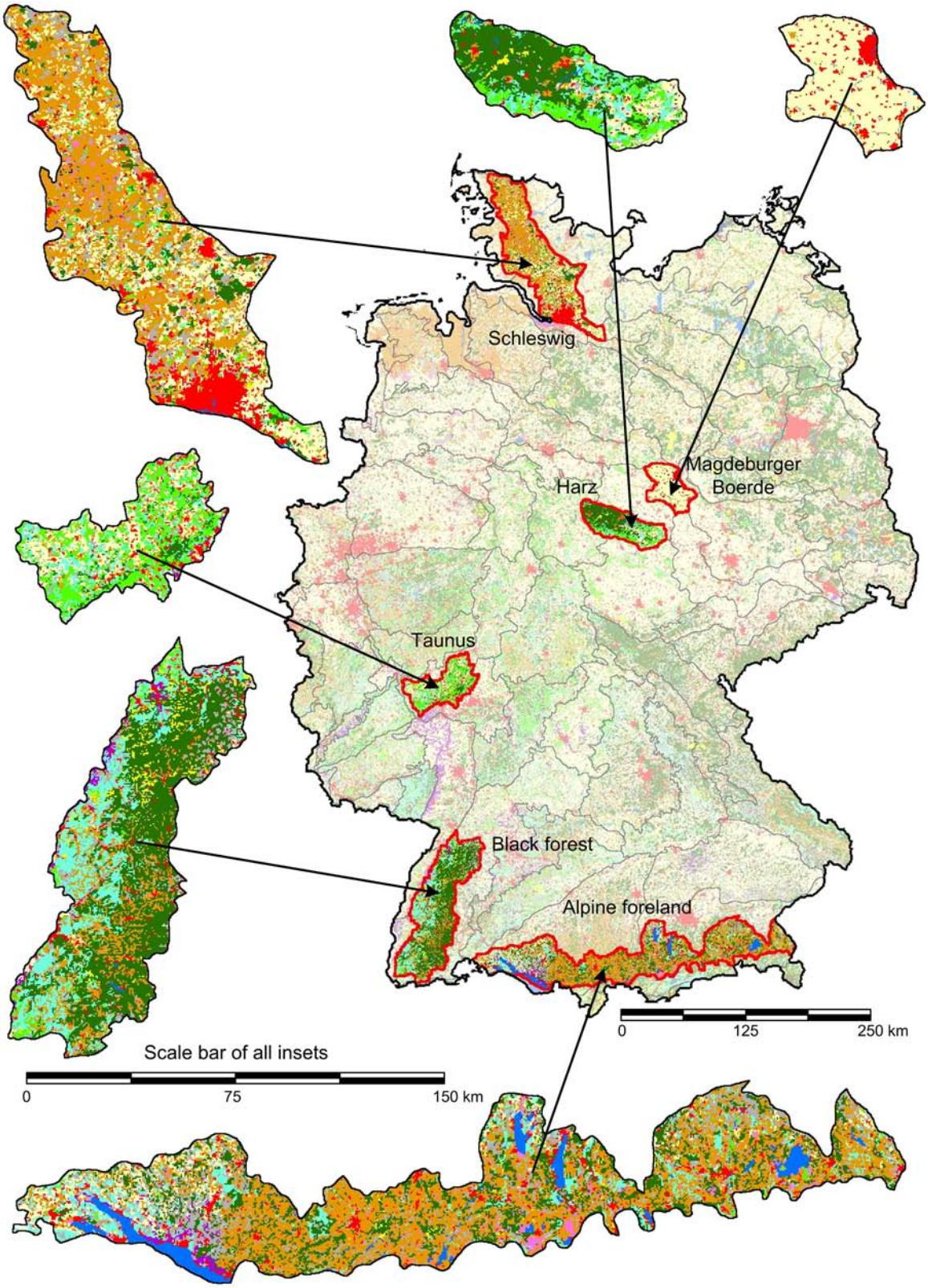
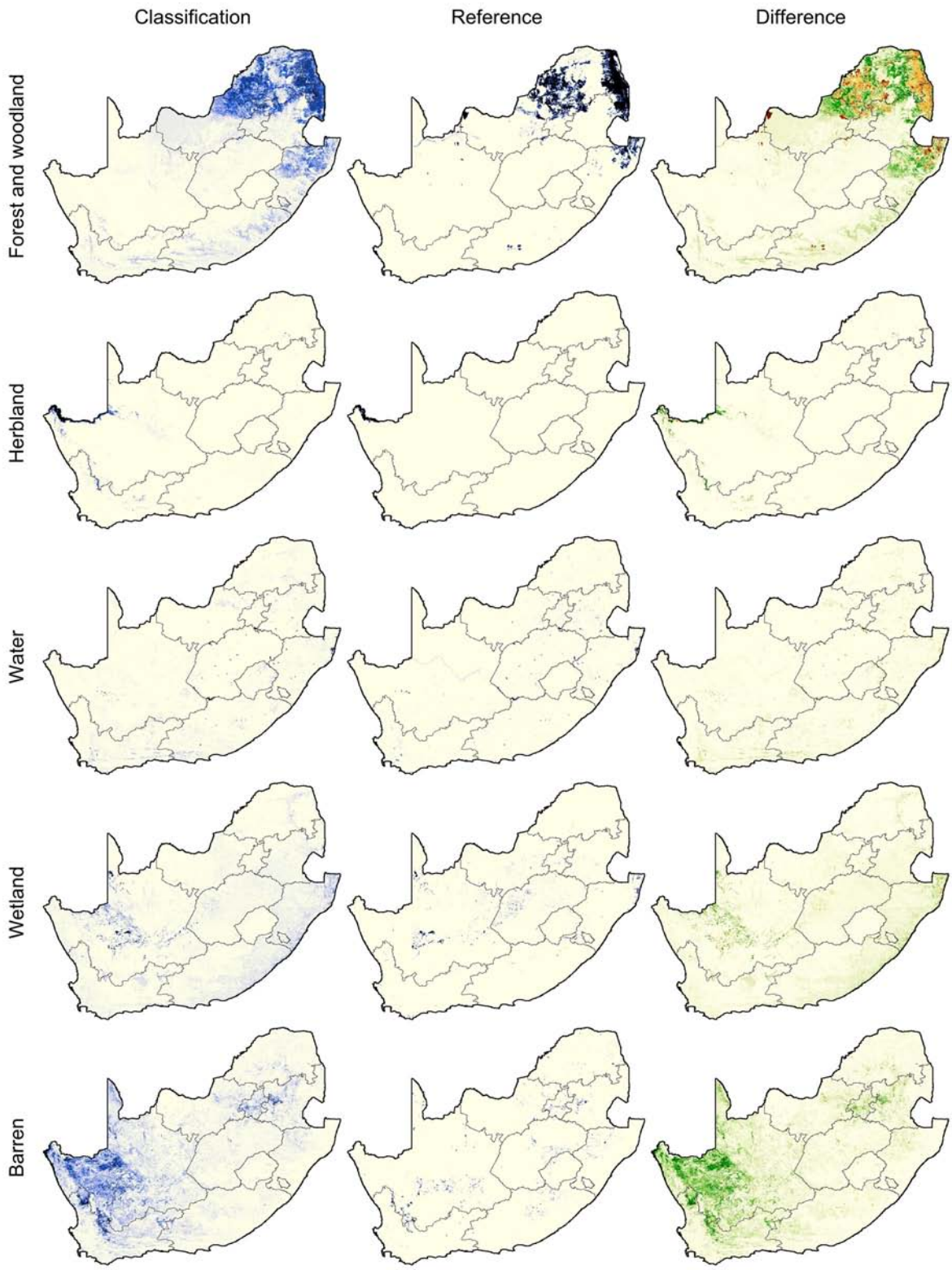


Figure A.2: Close-up of selected natural regions and CORINE land cover classes of Germany.
Note: Grey lines indicate approximately 90 natural regions according to Meynen and Schmithüsen (1953).

		reference				
		A	B	C	D	Σ
classification	A	n_{AA}	n_{AB}	n_{AC}	n_{AD}	n_{A+}
	B	n_{BA}	n_{BB}	n_{BC}	n_{BD}	n_{B+}
	C	n_{CA}	n_{CB}	n_{CC}	n_{CD}	n_{C+}
	D	n_{DA}	n_{DB}	n_{DC}	n_{DD}	n_{D+}
	Σ	n_{+A}	n_{+B}	n_{+C}	n_{+D}	n

Figure A.3: Principle of the error matrix.



Continued on next page

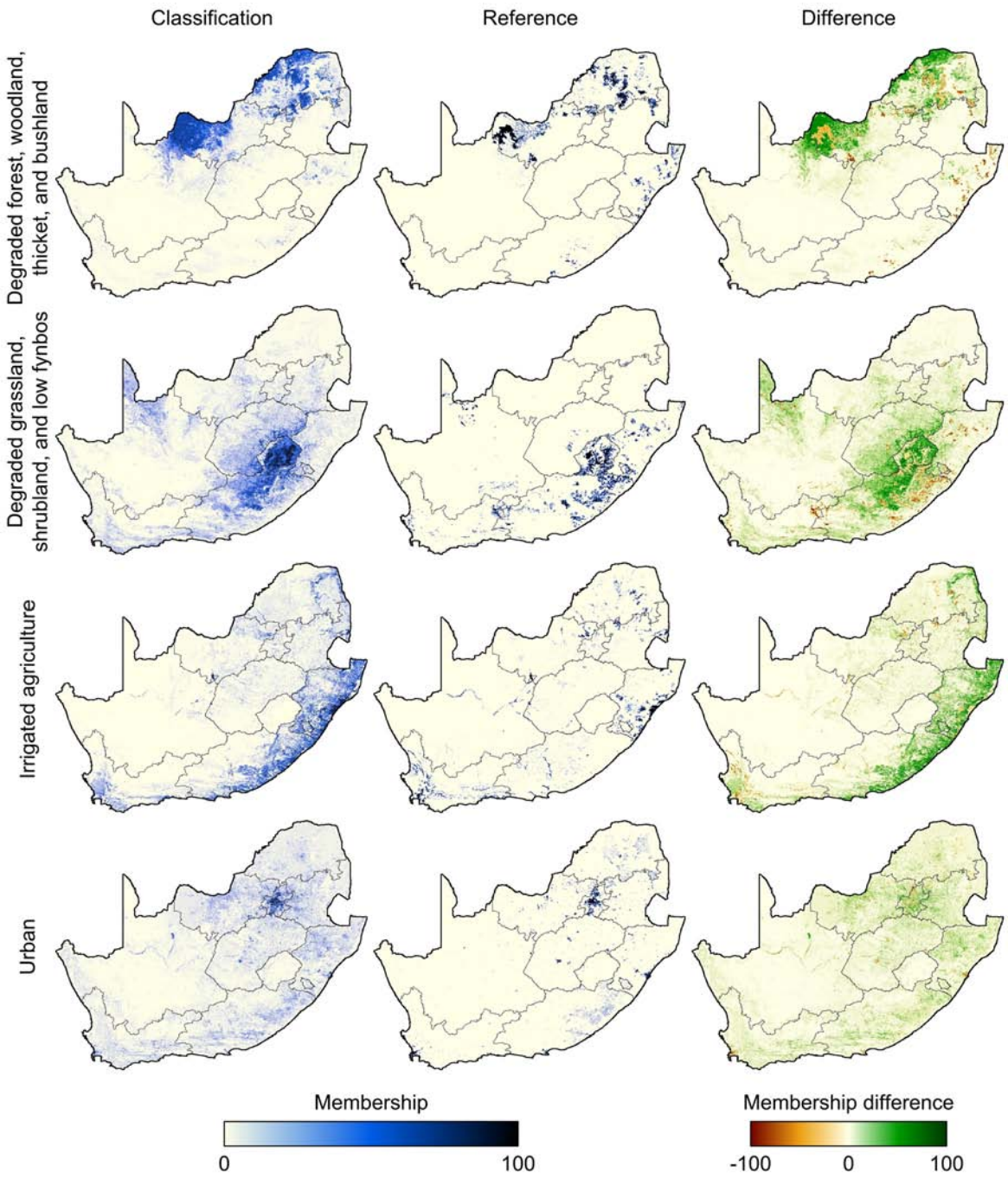


Figure A.4: **Class-wise comparison between fuzzy estimates of classification and reference and the corresponding difference for South Africa.**

Note: **Figure 4.14 indicates the remaining class-wise comparison for all main land cover classes of South Africa.**

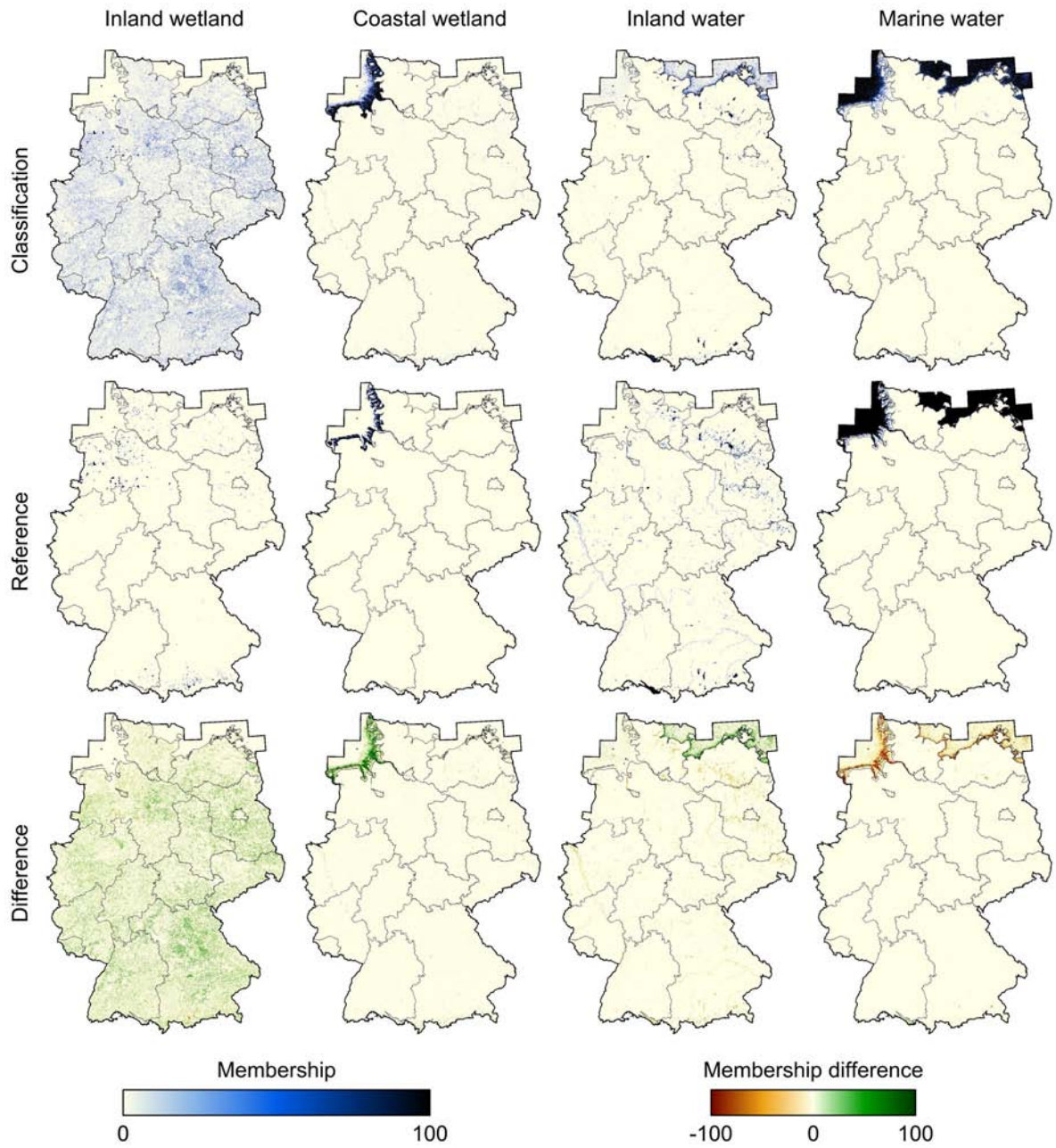


Figure A.5 **Class-wise comparison between fuzzy estimates of classification and reference and the corresponding difference for Germany.**

Note: **Figure 4.19 indicates the remaining class-wise comparison for all main land cover classes of Germany. "Inland wetland" depicts a scatter with low memberships throughout Germany, which introduces significant classification speckle at the edge of patches in discrete mapping. "Coastal wetland", indicating the large marshland along the North Sea Coast, shows clear disagreement with "marine water" due to strong tidal shifts. Because of their variability, both classes are difficult to describe in feature space.**

Appendix of Tables

Table A.1: Division of wavelengths.

Name	Abbreviation	Wavelength [µm]
Ultraviolet radiation	UV	0.2 - 0.4
Visible light	VIS	0.4 - 0.7
Near infrared	NIR	0.7 - 1.0
Short wave infrared	SWIR	1.0 - 2.5
Middle infrared	MIR	2.5 - 5.0
Thermal infrared	TIR	5.0 - 50
Far infrared	FIR	50 - 300
Microwaves	MW	1mm - 1m
Radiowaves	RW	1m - 10m

Note: **No definite distinction of wavelengths groups has been found in the literature. The used divisions have been assembled from multiple references in remote sensing and physics.**

Table A.2: List of all MODLand products.

Parameter	Product	Spatial res. [m or °]	Temporal res. [days]	Level	Collection	Platform	TiSeG
Surface reflectance	09GA	500, 1000	1	2G	5	MOD, MYD	
Surface reflectance	09GQ	250	1	2G	5	MOD, MYD	
Surface reflectance	09CMG	0.05	1	3	5	MOD, MYD	
Surface reflectance	09A1	500	8	3	4, 5	MOD, MYD	x
Surface reflectance	09Q1	250	8	3	4, 5	MOD, MYD	x
Surface reflectance	09GHK	500	1	2G	4	MOD, MYD	x
Surface reflectance	09GQK	250	1	2G	4	MOD, MYD	x
Surface reflectance quality	09GST	1000	1	2G	4	MOD, MYD	
LST / emissivity	11_L2	1000	5 min (swath)	2	4, 5	MOD, MYD	x
LST / emissivity	11A1	1000	1	3	4, 5	MOD, MYD	x
LST / emissivity	11A2	1000	8	3	4, 5	MOD, MYD	x
LST / emissivity	11B1	4600	1	3	4, 5	MOD, MYD	x
LST / emissivity	11C1	0.05	1	3	4, 5	MOD, MYD	
LST / emissivity	11C2	0.05	8	3	4, 5	MOD, MYD	
LST / emissivity	11C3	0.05	monthly	3	4, 5	MOD, MYD	
Land cover type	12C1	0.05	yearly	3	4, 5	MOD	
Land cover type	12Q1	1000	yearly	3	4, 5	MOD	
Land cover dynamics	12Q2	1000	bi-yearly	3	4, 5	MOD	
Vegetation indices	13Q1	250	16	3	4, 5	MOD, MYD	x
Vegetation indices	13A1	500	16	3	4, 5	MOD, MYD	x
Vegetation indices	13A2	1000	16	3	4, 5	MOD, MYD	x
Vegetation indices	13A3	1000	monthly	3	4, 5	MOD, MYD	x
Vegetation indices	13C1	0.05	16	3	4, 5	MOD, MYD	
Vegetation indices	13C2	0.05	monthly	3	4, 5	MOD, MYD	

Continued on next page

Thermal anomalies / fire	14	1000	5 min (swath)	2	4, 5	MOD, MYD	
Thermal anomalies / fire	14A1	1000	1	3	4, 5	MOD, MYD	x
Thermal anomalies / fire	14A2	1000	8	3	4, 5	MOD, MYD	x
Leaf area index / FPAR	15A2	1000	8	4	4, 5	MOD, MYD	x
Net photosynthesis	17A2	1000	8	4	4, 5	MOD, MYD	x
Net primary productivity	17A3	1000	yearly	4	5	MOD, MYD	
BRDF / albedo model parameters	43A1	500	16	3	5	MCD	
BRDF / albedo quality	43A2	500	16	3	5	MCD	
Albedo	43A3	500	16	3	5	MCD	
Nadir BRDF-adjusted reflectance	43A4	500	16	3	5	MCD	
BRDF / albedo model parameters	43B1	1000	16	3	4, 5*	MOD, MCD	x
BRDF / albedo quality	43B2	1000	16	3	5	MCD	
Albedo	43B3	1000	16	3	4, 5*	MOD, MCD	x
Nadir BRDF-adjusted reflectance	43B4	1000	16	3	4, 5*	MOD, MCD	x
BRDF / albedo model parameters	43C1	0.05	16	3	4, 5*	MOD, MCD	
BRDF / albedo snow-free quality	43C2	0.05	16	3	5	MCD	
Albedo	43C3	0.05	16	3	4, 5*	MOD, MCD	
Nadir BRDF-adjusted reflectance	43C4	0.05	16	3	4, 5*	MOD, MCD	
Vegetation cover conversion	44A	250	quarterly	4	4, 5	MOD	
Vegetation continuous fields	44B	500	yearly	4	4, 5	MOD	
Burned area	45A1	500	monthly	4	5	MCD	
Geolocation angles	MGGAD	1000	1	2G	4	MOD, MYD	
Observation pointers	PT1KD	1000	1	2G	4	MOD, MYD	
Observation pointers	PTHKM	500	1	2G	4	MOD, MYD	
Observation pointers	PTQKM	250	1	2G	4	MOD, MYD	

Note: **Cryosphere products MOD10 and MOD29 (see Figure 2.2) are excluded because they do not strictly belong the MODLand product group. *…version 4 only available for MOD (Terra) data. Column TiSeG indicates products, which can be processed with the Time Series Generator (see chapter 3). Not all products are currently available, in particular products only available for version 5. Most products are distributed through the LP DAAC but some are also available through the GLCF at the University of Maryland or other data servers. Highlighted products are used in this study. Information is based on multiple sources including Justice et al. (1998), Masuoka et al. (1998), Justice et al. (2002a), Roy et al. (2002a), King et al. (2004), MODIS (2007a), and MODIS (2007b).**

Table A.3: Science data sets of MOD09A1 (8-day 500 m surface reflectance) product (MODIS 2007b).

SDS	Unit	Valid range	Scale factor	Offset	Fill value	Data type
Band 1 (VIS red)	reflectance	-100 - 16000	0.0001	0	-28672	16-bit signed integer
Band 2 (NIR)	reflectance	-100 - 16001	0.0001	0	-28672	16-bit signed integer
Band 3 (VIS blue)	reflectance	-100 - 16002	0.0001	0	-28672	16-bit signed integer
Band 4 (VIS green)	reflectance	-100 - 16003	0.0001	0	-28672	16-bit signed integer
Band 5 (SWIR1)	reflectance	-100 - 16004	0.0001	0	-28672	16-bit signed integer
Band 6 (SWIR2)	reflectance	-100 - 16005	0.0001	0	-28672	16-bit signed integer
Band 7 (SWIR3)	reflectance	-100 - 16006	0.0001	0	-28672	16-bit signed integer
500m quality control flags	bit field	0 - 4294966275	1	0	4294967295	32-bit unsigned integer
Solar zenith	degree	0 - 18000	0.01	0	0	16-bit signed integer
View zenith	degree	0 - 18000	0.01	0	0	16-bit signed integer
Relative azimuth	degree	-18000 - 18000	0.01	0	0	16-bit signed integer
500m state flags	bit field	0 - 57343	1	0	65535	16-bit unsigned integer
Day of year	julian Day	0 - 366	1	0	65535	16-bit unsigned integer

Table A.4: Description of MOD09A1 QA-SDS with band specific quality flags (SDS 500m quality control of Table A.3; MODIS 2007b).

Bit range	Bit setting	Description
0 - 1		MODLand QA bits
	00	corrected product produced at ideal quality -- all bands
	01	corrected product produced, less than ideal quality -- some or all bands
	10	corrected product not produced due to cloud effects -- all bands
	11	corrected product not produced for other reasons -- some or all bands may be fill value
2 - 5		band 1 data quality
	0000	highest quality
	1000	dead detector; data copied from adjacent detector
	1001	solar zenith >= 86 degrees
	1010	solar zenith >= 85 and < 86 degrees
	1011	missing input
	1100	internal constant used in place of climatological data for at least one atmospheric constant
	1101	quality so low that it is not useful
	1110	L1B data faulty
	1111	not useful for any other reason / not processed (e.g. clouds, ocean)
6 - 9		band 2 data quality (see band 1)
10 - 13		band 3 data quality (see band 1)
14 - 17		band 4 data quality (see band 1)
18 - 21		band 5 data quality (see band 1)
22 - 25		band 6 data quality (see band 1)
26 - 29		band 7 data quality (see band 1)
30		atmospheric correction performed
	0	no
	1	yes
31		adjacency correction performed
	0	no
	1	yes

Table A.5: Description of MOD09A1 QA-SDS with surface and atmosphere state flags (SDS 500m state flags of Table A.3; MODIS 2007b).

Bit range	Bit setting	Description
0 - 1		cloud state
	00	clear
	01	cloudy
	10	mixed
	11	not set, assumed clear
2		cloud shadow
	0	no
	1	yes
3 - 5		land/water flag
	000	shallow ocean
	001	land
	010	ocean coastlines and land shorelines
	011	shallow inland water
	100	ephemeral water
	101	deep inland water
	110	ocean
	111	surface unknown (treated as land)
6 - 7		aerosol quantity
	00	climatology
	01	low
	10	average
	11	high
8 - 9		cirrus detected
	00	none
	01	small
	10	average
	11	high
10		internal cloud mask
	0	clear
	1	cloudy
11		internal fire mask
	0	no fire
	1	fire
12		snow/ice flag
	0	no
	1	yes
13 - 14		BRDF correction performed
	00	no
	01	Montana methodology
	10	Boston methodology
15		internal snow mask
	0	no snow
	1	snow

Table A.6: Science data sets of MOD11A2 (8-day 1 km LST and emissivity) product (MODIS 2007b).

SDS	Unit	Valid range	Scale factor	Offset	Fill value	Data type
LST day	kelvin	7500 - 65535	0.02	0	0	16-bit unsigned integer
Quality day	bit field	0 - 255	1	0	0	8-bit unsigned integer
View time day	hours	0 - 240	0.1	0	0	8-bit unsigned integer
View angle day	degree	0 - 130	1	-65	255	8-bit unsigned integer
LST night	kelvin	7500 - 65535	0.02	0	0	16-bit unsigned integer
Quality night	bit field	0 - 255	1	0	0	8-bit unsigned integer
View time night	hours	0 - 240	0.1	0	0	8-bit unsigned integer
View angle night	degree	0 - 130	1	-65	255	8-bit unsigned integer
Emissivity band 31		1 - 255	0.002	0.49	0	8-bit unsigned integer
Emissivity band 32		1 - 255	0.002	0.49	0	8-bit unsigned integer
Clear day coverage		0 - 255	1	0	0	8-bit unsigned integer
Clear night coverage		0 - 255	1	0	0	8-bit unsigned integer

Table A.7: Description of MOD11A2 QA-SDS (identical for SDS quality day and quality night of Table A.6; MODIS 2007b).

Bit range	Bit setting	Description
0 - 1		mandatory flags
	00	LST produced, good quality, not necessary to examine more detailed QA
	01	LST produced, other quality, recommend examination of more detailed QA
	10	LST not produced due to cloud effects
	11	LST not produced primarily due to reasons other than cloud
2 - 3		data quality flag
	00	good data quality
	01	LST affected by thin cirrus and/or sub-pixel clouds
	10	not processed due to missing pixels
	11	not processed due to poor quality
4 - 5		emissivity error flag
	00	average emissivity error <= 0.01
	01	average emissivity error <= 0.02
	10	average emissivity error <= 0.04
	11	average emissivity error > 0.04
6 - 7		LST error flag
	00	average LST error <= 1K
	01	average LST error <= 2K
	10	average LST error <= 3K
	11	average LST error > 3K

Table A.8: Science data sets of MOD12Q2 (bi-yearly 1 km land cover dynamics; MODIS 2007b).

SDS	Unit	Valid range	Scale factor	Offset	Fill value	Data type
Onset greenness increase	days	1 - 32766	1	0	32767	16-bit unsigned integer
Onset greenness maximum	days	1 - 32766	1	0	32767	16-bit unsigned integer
Onset greenness decrease	days	1 - 32766	1	0	32767	16-bit unsigned integer
Onset greenness minimum	days	1 - 32766	1	0	32767	16-bit unsigned integer
NBAR EVI greenness minimum	EVI	0 - 10000	0.0001	0	32767	16-bit unsigned integer
NBAR EVI greenness maximum	EVI	0 - 10000	0.0001	0	32767	16-bit unsigned integer
NBAR EVI area	EVI, cummulative	0 - 32766	0.01	0	32767	16-bit unsigned integer
Dynamics quality	flags	0 - 65535	1	0	65535	16-bit unsigned integer

Table A.9: Science data sets of MOD13A1 (16-day 500 m vegetation indices) product of collection 4 (MODIS 2007b).

SDS	Unit	Valid range	Scale factor	Offset	Fill value	Data type
NDVI	NDVI	-2000 - 10000	0.0001	0	-3000	16-bit signed integer
EVI	EVI	-2000 - 10000	0.0001	0	-3000	16-bit signed integer
NDVI quality	bit field	0 - 65535	1	0	65535	16-bit unsigned integer
EVI quality	bit field	0 - 65535	1	0	65535	16-bit unsigned integer
Red reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
NIR reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
Blue reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
SWIR3 reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
Average view zenith	degree	-9000 - 9000	0.01	0	-10000	16-bit signed integer
Average sun zenith	degree	-9000 - 9000	0.01	0	-10000	16-bit signed integer
Average relative azimuth	degree	-3600 - 3600	0.1	0	-4000	16-bit signed integer

Table A.10: Description of MOD13A1 QA-SDS of collection 4 (identical for SDS NDVI quality and EVI quality of Table A.9; MODIS 2007b).

Bit range	Bit setting	Description
0 - 1		mandatory flags
	00	VI produced with good quality
	01	VI produced but with unreliable quality
	10	VI produced but contaminated with clouds
	11	VI not produced due to bad quality

Continued on next page

2 - 5		VI usefulness index
	0000	perfect quality
	0001	high quality
	0010	good quality
	0011	acceptable quality
	0100	fair quality
	0101	intermediate quality
	0110	below intermediate quality
	0111	average quality
	1000	below average quality
	1001	questionable quality
	1010	above marginal quality
	1011	marginal quality
	1100	low quality
	1101	no atmospheric correction performed
	1110	quality too low to be useful
	1111	not useful for other reasons
6 - 7		aerosol quantity
	00	climatology used for atmospheric correction
	01	low
	10	intermediate
	11	high
8		atmosphere adjacency correction
	0	no adjacency correction performed
	1	adjacency correction performed
9		atmosphere BRDF correction
	0	no atmosphere-surface BRDF coupled correction performed
	1	atmosphere-surface BRDF coupled correction performed
10		mixed clouds
	0	no mixed clouds
	1	possible existence of mixed clouds
11 - 12		land/water mask
	00	ocean/inland water
	01	coastal region
	10	wetland
	11	land
13		snow/ice
	0	no snow/ice
	1	possible existence of snow/ice
14		shadow
	0	no shadow
	1	possible existence of shadow
15		compositing method
	0	BRDF composite method used for compositing
	1	constraint view angle MVC (CV-MVC) method used for compositing

Table A.11: Science data sets of MOD13A1 (16-day 500 m vegetation indices) product of collection 5 (MODIS 2007b).

SDS	Unit	Valid range	Scale factor	Offset	Fill value	Data type
NDVI	NDVI	-2000 - 10000	0.0001	0	-3000	16-bit signed integer
EVI	EVI	-2000 - 10000	0.0001	0	-3000	16-bit signed integer
Quality	bit field	0 - 65535	1	0	65535	16-bit unsigned integer
Red reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
NIR reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
Blue reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
SWIR3 reflectance	reflectance	0 - 10000	0.0001	0	-1000	16-bit signed integer
Average view zenith	degree	-9000 - 9000	0.01	0	-10000	16-bit signed integer
Average sun zenith	degree	-9000 - 9000	0.01	0	-10000	16-bit signed integer
Average relative azimuth	degree	-3600 - 3600	0.1	0	-4000	16-bit signed integer
Composite day of year	day	1 - 366	1	0	-1	16-bit signed integer
Pixel reliability	rank	0 - 4	1	0	-1	8-bit signed integer

Table A.12: Description of MOD13A1 QA-SDS of collection 5 (see Table A.11; MODIS 2007b).

Bit range	Bit setting	Description
0 - 1		mandatory flags
	00	VI produced with good quality
	01	VI produced but with unreliable quality
	10	VI produced but contaminated with clouds
	11	VI not produced due to bad quality
2 - 5		VI usefulness index
	0000	perfect quality
	0001	high quality
	0010	good quality
	0011	acceptable quality
	0100	fair quality
	0101	intermediate quality
	0110	below intermediate quality
	0111	average quality
	1000	below average quality
	1001	questionable quality
	1010	above marginal quality
	1011	marginal quality
	1100	low quality
	1101	no atmospheric correction performed
	1110	quality too low to be useful
	1111	not useful for other reasons

Continued on next page

6 - 7		aerosol quantity
	00	climatology used for atmospheric correction
	01	low
	10	intermediate
	11	high
8		atmosphere adjacency correction
	0	no adjacency correction performed
	1	adjacency correction performed
9		atmosphere BRDF correction
	0	no atmosphere-surface BRDF coupled correction performed
	1	atmosphere-surface BRDF coupled correction performed
10		mixed clouds
	0	no mixed clouds
	1	possible existence of mixed clouds
11 - 13		land/water mask
	000	shallow ocean
	001	land
	010	ocean coastlines and lake shorelines
	011	shallow inland water
	100	ephemeral water
	101	deep inland water
	110	moderate or continental ocean
	111	deep ocean
14		snow/ice
	0	no snow/ice
	1	possible existence of snow/ice
15		shadow
	0	no shadow
	1	possible existence of shadow

Table A.13: Description of MOD13A1 reliability SDS of collection 5 (see Table A.11; MODIS 2007b).

Rank	Summary	Description
-1	no data	not processed
0	good data	use with confidence
1	marginal data	useful, but look at other QA information
2	snow / ice	target covered with snow/ice
3	cloudy	target not visible, covered with cloud

Table A.14: Science data sets of MOD43B4 (16-day 1 km nadir BRDF-adjusted surface reflectance) product (MODIS 2007b).

SDS	Unit	Valid range	Scale factor	Offset	Fill value	Data type
Nadir reflectance	reflectance	0 - 32766	0.0001	0	32767	16-bit signed integer
Nadir reflectance quality	bit field	0 - 4294967294	1	0	4294967295	32-bit unsigned integer

Table A.15: Description of MOD43B4 QA-SDS with surface state flags (first part of three-dimensional SDS nadir reflectance quality of Table A.14; MODIS 2007b).

Bit range	Bit setting	Description
0 - 1		mandatory flags
	00	processed, good quality
	01	processed, see other QA
	10	not processed due to cloud effects
	11	not processed due to other effects
2 - 3		period used
	00	16 days
	01	32 days
4 - 7		land/water
	000	shallow ocean
	001	land
	010	ocean and lake shorelines
	011	shallow inland water
	100	ephemeral water
	101	deep inland water
	110	moderate or continental ocean
	111	deep ocean
8 - 10		AM/MISR/PM
	000	AM
	001	AM/PM
	010	AM/PM/MISR
	011	AM/MISR
	100	PM
	101	PM/MISR
	110	MISR
11 - 15		mean solar angle
	00000	0-5 degrees
	00001	5-10 degrees
	00010	10-15 degrees
	00011	15-20 degrees
	00100	20-25 degrees
	00101	25-30 degrees
	00110	30-35 degrees
	00111	35-40 degrees
	01000	40-45 degrees

Continued on next page

	01001	45-50 degrees
	01010	50-55 degrees
	01011	55-60 degrees
	01100	60-65 degrees
	01101	65-70 degrees
	01110	70-75 degrees
	01111	75-80 degrees
	10000	80-90 degrees
16 - 17		snow
	0	no snow
	1	snow present
31		quality fill
	0	not fill-value
	1	fill-value

Table A.16: Description of MOD43B4 QA-SDS with band-specific flags (second part of three-dimensional SDS nadir reflectance quality of Table A.14; MODIS 2007b).

Bit range	Bit setting	Description
0 - 3		band 1 quality
	0000	RMSE good, WoD(NBAR) good, WoD(WSA) good
	0001	RMSE good, WoD(NBAR) good, WoD(WSA) moderate
	0010	RMSE good, WoD(NBAR) moderate, WoD(WSA) good
	0011	RMSE good, WoD(NBAR) moderate, WoD(WSA) moderate
	0100	RMSE moderate, WoD(NBAR) good, WoD(WSA) good
	0101	RMSE moderate, WoD(NBAR) good, WoD(WSA) moderate
	0110	RMSE moderate, WoD(NBAR) moderate, WoD(WSA) good
	0111	RMSE moderate, WoD(NBAR) moderate, WoD(WSA) moderate
	1001	magnitude inversion (numobs >3&<7)
	1010	magnitude inversion (numobs <=3)
	1011	bus-in DB parameters
	1111	fill value
4 - 7		band 2 quality (see band 1)
8 - 11		band 3 quality (see band 1)
12 - 15		band 4 quality (see band 1)
16 - 19		band 5 quality (see band 1)
20 - 23		band 6 quality (see band 1)
24 - 27		band 7 quality (see band 1)
31		QA fill
	0	not fill value
	1	fill value

Table A.17: Recode of the original NLC1995 classification scheme of South Africa to 14 classes employed in this study.

NLC1995 classification scheme		Classification scheme of this study	
Code	Class name	Code	Class name
01-00-000	Forest and woodland	2	Forest and woodland
01-02-000	Forest	1	Forest
02-00-000	Thicket and bushland	3	Thicket and bushland
03-00-000	Shrubland and low fynbos	4	Shrubland and low fynbos
04-00-000	Herbland	5	Herbland
05-01-000	Unimproved grassland	6	Dry grassland
05-02-000	Improved grassland	12	Irrigated agriculture
06-00-000	Forest plantation	1	Forest
07-00-000	Waterbody	7	Water
08-00-000	Wetland	8	Wetland
09-01-000	Barren rock	9	Barren
09-02-001	Dongas and sheet erosion scars	9	Barren
09-02-002	Degraded: forest and woodland	10	Degraded forest, woodland, thicket, and bushland
09-02-003	Degraded: thicket and bushland	10	Degraded forest, woodland, thicket, and bushland
09-02-004	Degraded: unimproved grassland	11	Degraded grassland, shrubland, and low fynbos
09-02-005	Degraded: shrubland and low fynbos	11	Degraded grassland, shrubland, and low fynbos
09-02-006	Degraded: herbland	11	Degraded grassland, shrubland, and low fynbos
10-01-006	Cultivated: permanent -commercial irrigated	12	Irrigated agriculture
10-01-007	Cultivated: permanent – commercial dryland	13	Dryland agriculture
10-01-010	Cultivated: permanent – commercial sugarcane	12	Irrigated agriculture
10-02-006	Commercial: temporary – commercial irrigated	12	Irrigated agriculture
10-02-007	Cultivated: temporary – commercial dryland	13	Dryland agriculture
10-02-009	Cultivated: temporary – semi-commercial/subsistence dryland	13	Dryland agriculture
11-01-000	Urban/built-up land: residential	14	Urban
11-01-008	Urban/built-up land: residential (small holdings: woodland)	14	Urban
11-01-009	Urban/built-up land: residential (small holdings: bushland)	14	Urban
11-01-010	Urban/built-up land: residential (small holdings: shrubland)	14	Urban
11-01-011	Urban/built-up land: residential (small holdings: grassland)	14	Urban
11-02-000	Urban/built-up land: commercial	14	Urban
11-03-000	Urban/built-up land: industrial/transport	14	Urban
12-00-000	Mines and quarries	9	Barren

Note: **To some extent the recode follows the hierarchical approach used by NLC1995 class definition (indicated in the NLC1995 code column) and described by Thompson (1996), but modifications have been necessary to meet requirements of 1,000m MODIS data.**

Table A.18: Projection parameters of study sites used for all datasets in this study.

	South Africa	Germany	Minnesota	Namibia
Projection	Albers equal area	UTM	UTM	Albers equal area
Datum	WGS 84	WGS 84	WGS 84	WGS 84
Spheroid	WGS 84	WGS 84	WGS 84	WGS 84
Zone		32N	15N	
1. standard parallel	20° N			20° N
2. standard parallel	23° S			23° S
Central meridian	25° E	9° E	93° W	25° E
Origin of latitude	0	0	0	0
False easting	0	500,000	500,000	0
False northing	0	0	0	0
Upper left X	-921,000	278,000	157,000	-1,367,000
Upper left Y	-2,563,000	6,118,000	5,496,000	-1,995,000
Lower right X	818,000	947,000	798,000	-410,000
Lower right Y	-3,941,000	5,226,000	4,804,000	-3,322,000

Table A.19: Recode of the original CLC2000 classification scheme of Germany to 14 classes employed in this study.

CLC2000 classification scheme		Classification scheme of this study	
Code	Class name	Code	Class name
111	Continuous urban fabric	1	Artificial
112	Discontinuous urban fabric	1	Artificial
121	Industrial or commercial units	1	Artificial
122	Road and rail networks and associated land	1	Artificial
123	Port areas	1	Artificial
124	Airports	1	Artificial
131	Mineral extraction sites	1	Artificial
132	Dump sites	1	Artificial
133	Construction sites	1	Artificial
141	Green urban areas	1	Artificial
142	Sport and leisure facilities	1	Artificial
211	Non-irrigated arable land	2	Arable
212	*Permanently irrigated land*		
213	*Rice fields*		
221	Vineyards	3	Permanent crops
222	Fruit trees and berry plantations	3	Permanent crops
223	*Olive groves*		
231	Pastures	4	Pasture
241	*Annual crops associated with permanent crops*		
242	Complex cultivation patterns	5	Heterogeneous agriculture
243	Land principally occupied by agriculture, with significant areas of natural vegetation	5	Heterogeneous agriculture
344	*Agro-forestry areas*		
311	Broad-leaved forest	6	Broadleaved forest

Continued on next page

312	Coniferous forest	7	Coniferous forest
313	Mixed forest	8	Mixed forest
321	Natural grasslands	9	Herbaceous vegetation
322	Moors and heathland	9	Herbaceous vegetation
323	*Sclerophylous vegetation*		
324	Transitional woodland - shrub	9	Herbaceous vegetation
331	Beaches, dunes, sands	10	Barren
332	Bare rocks	10	Barren
333	Sparsely vegetated areas	10	Barren
334	Burnt areas	10	Barren
335	Glaciers and perpetual snow	10	Barren
411	Inland marshes	11	Inland wetland
412	Peat bogs	11	Inland wetland
421	Salt marshes	12	Coastal wetland
422	*Salines*		
423	Intertidal flats	12	Coastal wetland
511	Water courses	13	Inland water
512	Water bodies	13	Inland water
521	Coastal lagoons	14	Marine water
522	Estuaries	14	Marine water
523	Sea and ocean	14	Marine water

Note: **The recode follows the hierarchical approach used by the CLC2000 class definition (indicated in the CLC2000 code column), but varies throughout all levels. For instance, "urban" is in accordance to level 1 but "forests" have been mapped at level 3. Seven classes were not mapped in Germany and are indicate in italic. Class "burnt areas" may exist in Germany but have not been mapped in CLC2000.**

Table A.20: Recode of the original NLCD2001 classification scheme of Minnesota to 8 classes employed in this study.

NLCD2001 classification scheme		Classification scheme of this study	
Code	Class name	Code	Class name
11	Open water	1	water
12	*Perennial ice / snow*		
21	Developed, open space	2	urban
22	Developed, low intensity	2	urban
23	Developed, medium intensity	2	urban
24	Developed, high intensity	2	urban
31	Barren land (rock / sand / clay)	2	urban
32	*Unconsolidated shore*		
41	Deciduous forest	3	Decidious forest
42	Evergreen forest	4	Evergreen forest
43	Mixed forest	4	Evergreen forest
51	*Dwarf scrub*		
52	Shrub / scrub	5	Herbaceous
71	Grassland / herbaceous	5	Herbaceous
72	*Sedge / herbaceous*		
73	*Lichens*		
74	*Moss*		
81	Pasture / hay	5	Herbaceous
82	Cultivated crops	6	Agriculture
90	Woody wetlands	7	Woody wetlands
95	Emergent herbaceous wetlands	8	Herbaceous Wetlands

Note: **The recode follows the contextual ideas of the NLCD2001 classification scheme (Homer et al. 2004) and constraints of the coarse spatial resolution of MODIS data. Recoding varies throughout both levels. For instance, "urban" is in accordance to level 1 but "deciduous forest" is mapped at level 2. Six classes were not mapped in Minnesota and are indicate in italic.**

Table A.21: Recode of the original map of vegetation types of Namibia to 23 classes employed in this study.

Vegetation types of Namibia classification scheme		Classification scheme of this study	
Code	Class name	Code	Class name
ETO-1	Pans	1	Pans
CE-2	Western highlands	2	Western highlands
CE-1	Central-western escarpment and inselbergs	3	Central-western escarpment and inselbergs
CD-5	Central desert	4	Central desert
FP	Caprivi floodplains		
CUV-1	Cuvelai drainage	5	Cuvelai drainage
DKT	Dwarf shrub / southern Kalahari transition	6	Dwarf shrub / southern Kalahari transition
MW	Caprivi mopane woodland		
ED-1	Eastern drainage	7	Eastern drainage
HS	Highland shrubland	8	Highland shrubland
ND-1	Northern desert	9	Northern desert
GIE-11	Northern Kalahari	10	Northern Kalahari
RIV-3	Okavango valley		
RIV-4	Omatako drainage		
SN-2	Southern desert	11	Southern desert
TS	Thornbush shrubland	12	Thornbush shrubland
WK-1	Western Kalahari	13	Western Kalahari
RIV-11	Riverine woodlands and islands		
DST-2	Desert / dwarf shrub transition	14	Desert / dwarf shrub transition
GIE-9	Dwarf shrub savanna	15	Dwarf shrub savanna
KW-8	North-eastern Kalahari woodlands	16	North-eastern Kalahari woodlands
DS-1	Karas dwarf shrubland	17	Karas dwarf shrubland
KAR-3	Karstveld	18	Karstveld
DOL-2	Mopane shrubland	19	Mopane shrubland
NE-7	North-western escarpment and inselbergs	20	North-western escarpment and inselbergs
ETO-2	Etosha grass and dwarf shrubland		
SS-2	Succulent steppe	21	Succulent steppe
GIE-12	Central Kalahari	22	Central Kalahari
SK-2	Southern Kalahari	23	Southern Kalahari

Note: **Some classes have been excluded, because they are exclusively located in the Caprivi area ("Caprivi floodplains", "Caprivi mopane woodland", and "Riverine woodlands and islands"), which has not been mapped in this study. Other classes along rivers ("Okavango valley" and "Omatako drainage") and scattered classes ("Etosha grass and dwarf shrubland") have been excluded due to a too small spatial extent.**

Table A.22: Area proportion in km² and percent of selected CORINE land cover classes of Germany and specific natural regions.

Land cover classes	Germany		Schleswig		Magdeburger Börde		Harz		Taunus		Black forest		Alpine foreland	
	km²	%	km²	%	km²	%	km²	%	km²	%	km²	%	km²	%
Arable	137228	35	1614	25	1262	83	268	12	576	25	69	1	651	6
Pasture	45753	12	2780	42	24	2	86	4	175	8	861	15	3859	38
Broadlevead	24173	6	168	3	7	0	427	19	759	33	151	3	62	1
Coniferous	56799	15	370	6	2	0	942	43	163	7	2578	44	1814	18

Table A.23: Standard land cover classification protocol for Germany.

Inputs / operation	Parameters / settings / specifications
Input data	
Land cover	CLC2000, 50 m, recoded to 14 classes (see Table A.19 and Figure 2.4)
MODIS	8-day 500 m surface reflectance (MOD09A1, see Table A.3)
Time series preprocessing	
Tiles	h18v03, h18v04 (see Figure A.1)
Period	annual, January to December
Transformations	mosaicing, reprojection, spatial subset using MRT (see also Table A.18)
Time series generation	
Quality settings LST	H7-CS-S (see Table 3.2)
Interpolation	linear temporal interpolation of all invalid values
Number of time series	8 (seven surface reflectance bands and NDVI derived from red and NIR)
Harmonic analysis	Maximum of 3 harmonics or 80 % explained variance
Metrics	
Set of statistics	minimum, maximum, mean, standard deviation, range
Set of segments	1 × annual, 2 × six months, 3 × four months
Total per time series	32 (number of harmonics + explained variance of harmonics + five statistics × six segments)
Total for land cover mapping	256
Multi-scale analysis	
Ratio	1:10 (50 m land cover, 500 m time series metrics)
Neighborhood	5 high resolution pixels (250 m)
Sampling	
Spatial	area-equalized
Class	class-equalized
Total samples per class	1100
Subsets per class	11 with 100 samples per class
Classification tree	
Number of individual trees	10
Split criterion	Entropy with ln
Stopping	Termination with homogeneous leaf, minimum samples per node: 10, minimum samples per leaf: 1, root deviance: 0
Pruning	yes (ratio: 6, deviation 0.975)

Table A.24: Standard land cover classification protocol for South Africa.

Inputs / operation	Parameters / settings / specifications
Input data	
Land cover	NLC1995, 100 m, recoded to 14 classes (see Table A.17 and Figure 2.3)
MODIS	16-day 1 km NBAR (MOD43B4, see Table A.14) and 8-day 1 km LST (MOD11A1, see Table A.6)
Time series preprocessing	
Tiles	h19v11, h19v12, h20v11, h20v12 (see also Figure A.1)
Period	annual, January to December
Transformations	mosaicing, reprojection, spatial subset using MRT (see also Table A.18)
Time series generation	
Quality settings NBAR	Strict, see Table XX14
Quality settings LST	Strict, see Table XX13
Interpolation	linear temporal interpolation of all invalid values
Number of time series	10 (seven surface reflectance bands, NDVI derived from red and NIR, and 16-day averaged day / night LST)
Harmonic analysis	Maximum of 3 harmonics or 80 % explained variance
Metrics	
Set of statistics	minimum, maximum, mean, standard deviation, range
Set of segments	1 × annual, 2 × six months, 3 × four months
Total per time series	32 (number of harmonics + explained variance of harmonics + five statistics × six segments)
Total for land cover mapping	320
Multi-scale analysis	
Ratio	1:10 (100 m land cover, 1 km time series metrics)
Neighborhood	5 high resolution pixels (500 m)
Sampling	
Spatial	area-equalized
Class	class-equalized
Total samples per class	1100
Subsets per class	11 with 100 samples per class
Classification tree	
Number of individual trees	10
Split criterion	Entropy with ln
Stopping	Termination with homogeneous leaf, minimum samples per node: 10, minimum samples per leaf: 1, root deviance: 0
Pruning	yes (ratio: 6, deviation 0.975)

Table A.25: Standard land cover classification protocol for Minnesota.

Inputs / operation	Parameters / settings / specifications
Input data	
Land cover	NLCD2001, 100 m, recoded to 8 classes (see Table A.20 and Figure 2.5)
MODIS	16-day 1 km NBAR (MOD43B4, see Table A.14)
Time series preprocessing	
Tiles	h11v04 (see also Figure A.1)
Period	annual, January to December
Transformations	mosaicing, reprojection, spatial subset using MRT (see also Table A.18)
Time series generation	
Quality settings NBAR	Snow: no, Fill: 2 × no, Bands: Obs > 3 & < 7
Interpolation	linear temporal interpolation of all invalid values
Number of time series	8 (seven surface reflectance bands and NDVI derived from red and NIR)
Harmonic analysis	Maximum of 3 harmonics or 80 % explained variance
Metrics	
Set of statistics	minimum, maximum, mean, standard deviation, range
Set of segments	1 × annual, 2 × six months, 3 × four months
Total per time series	30 (five statistics × six segments)
Total for land cover mapping	240
Multi-scale analysis	
Ratio	1:10 (100 m land cover, 1000 m time series metrics)
Neighborhood	5 high resolution pixels (500 m)
Sampling	
Spatial	area-equalized
Class	class-equalized
Total samples per class	1100
Subsets per class	11 with 100 samples per class
Classification tree	
Number of individual trees	10
Split criterion	Entropy with ln
Stopping	Termination with homogeneous leaf, minimum samples per node: 10, minimum samples per leaf: 1, root deviance: 0
Pruning	no

Table A.26: Standard land cover classification protocol for Namibia.

Inputs / operation	Parameters / settings / specifications
Input data	
Land cover	Vegetation map of Namibia, 1000 m, recoded to 23 classes (see Table A.21 and Figure 2.6)
MODIS	16-day 1 km NBAR (MOD43B4, see Table A.14) and 8-day 1 km LST (MOD11A1, see Table A.6)
Time series preprocessing	
Tiles	h19v10, h19v11, h19v12 (see also Figure A.1)
Period	annual, January to December
Transformations	mosaicing, reprojection, spatial subset using MRT (see also Table A.18)
Time series generation	
Quality settings NBAR	see Table XX15
Quality settings LST	General: good, acceptable, LST: 1-2 K
Interpolation	linear temporal interpolation of all invalid values
Number of time series	10 (seven surface reflectance bands, NDVI derived from red and NIR, and 16-day averaged day / night LST)
Harmonic analysis	Maximum of 3 harmonics or 80 % explained variance
Metrics	
Set of statistics	minimum, maximum, mean, standard deviation, range
Set of segments	1 × annual, 2 × six months, 3 × four months
Total per time series	32 (number of harmonics + explained variance of harmonics + five statistics × six segments)
Total for land cover mapping	320
Buffering	
Buffer size	10 cells (10 km in total)
Sampling	
Spatial	area-equalized
Class	class-equalized
Total samples per class	1100
Subsets per class	11 with 100 samples per class
Classification tree	
Number of individual trees	10
Split criterion	Entropy with ln
Stopping	Termination with homogeneous leaf, minimum samples per node: 10, minimum samples per leaf: 1, root deviance: 0
Pruning	yes (ratio: 6, deviation 0.975)

Table A.27: Class-wise fuzzy assessment statistics of South Africa for 2001.

	Class	Rel. mean	Rel. stddev	Abs. mean	Abs. stddev	r	RMSE
1	Forest	1.94	10.87	2.82	10.68	0.68	11.05
2	Forest and woodland	1.48	16.34	6.85	14.91	0.63	16.41
3	Thicket and bushland	-6.44	27.70	16.23	23.35	0.52	28.44
4	Shrubland and low fynbos	-11.99	26.60	16.08	24.34	0.81	29.17
5	Herbland	0.37	4.20	0.41	4.20	0.69	4.22
6	Dry grassland	-8.38	27.57	17.55	22.86	0.62	28.82
7	Water	0.49	3.91	0.86	3.85	0.66	3.94
8	Wetland	1.61	5.36	2.14	5.18	0.47	5.60
9	Barren	5.22	11.43	5.69	11.21	0.23	12.57
10	Degraded forest…	3.00	14.37	5.55	13.59	0.44	14.67
11	Degraded grassland….	5.29	15.21	8.23	13.84	0.33	16.10
12	Irrigated agriculture	4.08	13.05	5.84	12.37	0.34	13.68
13	Dryland agriculture	-0.80	19.05	10.07	16.19	0.62	19.07
14	Urban	4.30	8.97	5.42	8.34	0.38	9.94
	Average	0.01	14.62	7.41	13.21	0.53	15.26

Note: **For error measure computation see Figure 4.11. rel…relative. abs…absolute. stddev…standard deviation. r…correlation coefficient. average…mean of one error measure for all classes. "Degraded forest…"…"Degraded forest, woodland, thicket, and bushland". "Degraded grassland…"…"Degraded grassland, shrubland, and low fynbos".**

Table A.28: Summary table with statistics of pixel-wise fuzzy assessment of South Africa for 2001.

	Minimum	Maximum	Average	Stddev
Rel. mean	0.00	0.0	0.00	0.00
Rel. stddev	0.00	39.22	16.04	8.24
Abs. mean	0.00	14.29	7.41	3.72
Abs. stddev	0.00	36.31	14.04	7.36
r	-0.42	1.00	0.61	0.37
RMSE	0.00	37.80	15.46	7.94

Note: **For error measure computation see Figure 4.11. rel…relative. abs…absolute. stddev…standard deviation. r…correlation coefficient.**

Table A.29: Class-wise fuzzy assessment statistics of Germany for 2001.

	Class	Rel. mean	Rel. stddev	Abs. mean	Abs. stddev	r	RMSE
1	Artificial	-2.50	14.15	6.35	12.88	0.67	14.36
2	Arable	-14.39	25.91	18.90	22.83	0.73	29.64
3	Permanent crops	9.02	12.57	9.61	12.13	0.20	15.47
4	Pasture	-6.09	17.43	9.35	15.93	0.65	18.47
5	Heterogeneous agriculture	8.69	22.06	15.15	18.24	0.32	23.71
6	Broadleaved forest	-1.24	13.58	5.80	12.35	0.64	13.64
7	Coniferous forest	-7.42	17.80	9.13	16.99	0.76	19.29
8	Mixed forest	3.70	20.38	9.60	18.35	0.45	20.71
9	Herbaceous vegetation	2.37	10.27	4.06	9.73	0.24	10.54
10	Barren	1.99	7.13	2.16	7.08	0.36	7.40
11	Inland wetland	6.15	11.20	6.56	10.97	0.15	12.78
12	Coastal wetland	1.06	7.19	1.27	7.15	0.74	7.27
13	Inland water	0.20	7.65	1.56	7.50	0.52	7.66
14	Marine water	-1.41	9.16	1.64	9.13	0.93	9.27
	Average	0.01	14.04	7.22	12.95	0.52	15.02

Note: For error measure computation see Figure 4.11. rel…relative. abs…absolute. stddev…standard deviation. r…correlation coefficient. average…mean of one error measure for all classes.

Table A.30: Summary table with statistics of pixel-wise fuzzy assessment of Germany for 2001.

	Minimum	Maximum	Average	Stddev
Rel. mean	0.00	0.00	0.00	0.00
Rel. stddev	0.00	39.22	14.70	8.46
Abs. mean	0.00	14.29	7.22	4.25
Abs. stddev	0.00	36.31	12.58	7.33
r	-0.52	1.00	0.54	0.42
RMSE	0.00	37.80	14.17	8.16

Note: For error measure computation see Figure 4.11. rel…relative. abs…absolute. stddev…standard deviation. r…correlation coefficient.

Table A.31: Quality settings for NBAR (MOD43B4) time series for South Africa.

Setting	Snow	Fill	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Fill
Lenient	no	no	Obs <= 3	Obs <= 3	Obs <= 3	Obs <= 3	Obs <= 3	Obs <= 3	Obs <= 3	no
Moderate	no	no	Obs > 3 & < 7	Obs > 3 & < 7	Obs > 3 & < 7	Obs > 3 & < 7	Obs > 3 & < 7	Obs > 3 & < 7	Obs > 3 & < 7	no
Strict	no	no	RMSE+ NBAR-WSA -	RMSE+ NBAR-WSA -	RMSE+ NBAR-WSA -	RMSE+ NBAR-WSA -	RMSE+ NBAR-WSA -	RMSE+ NBAR-WSA -	RMSE+ NBAR-WSA -	no

Note: For specifics of the QA-SDS see Table A.15 and Table A.16.

Table A.32: Quality settings for LST (MOD11A2) time series for South Africa.

Setting	General	LST
Lenient	Good, Acceptable	1, 2
Moderate	Good, Acceptable	1, 2
Strict	Good, Acceptable	1

Note: For specifics of the QA-SDS see Table A.7.

Appendix of Equations

Notations for Equations A.1 to Equation A.11

l...*left*
r...*right*
N...*number of samples*
Π...*prior probability*
p...*probability*
t...*node*
T...*tree*
$\tilde{t}$...*leaf*
$\tilde{T}$...*number of leaves*
j...*class*
$\hat{j}$...*assigned class*
J...*number of classes*
m...*misclassification*
$M(T)$...*misclassification error of tree*
i...*impurity*
$I(T)$...*impurity of tree*

$\Pi(j)$ estimates the prior probability of class j. The prior probability can also be provided by the analyst, which significantly influences the result. In this study prior probabilities are always estimated from the data. Due to equalized sampling, all classes have identical prior probabilities at the root.

$$\Pi(j) = N_j / N \qquad \text{Equation A.1}$$

$p(j,t)$ estimates the probability that a sample of class j belongs to node t. $n_j(t)$ denotes the number of samples of class j in node t.

$$p(j,t) = \Pi(j) * N_j(t) / N_j \qquad \text{Equation A.2}$$

p(t) estimates the probability that a sample belongs to node *t*.

$$p(t) = \sum_j p(j,t)$$ Equation A.3

p(j|t) estimates the probability that a sample which is in node *t* belongs to class *j*.

$$p(j \mid t) = p(j,t) / p(t)$$ Equation A.4

Therefore, the sum of all *p(j|t)* for all classes *j* equals 1.

$$\sum_j p(j \mid t) = 1$$ Equation A.5

$p_l(s,t)$ estimates the proportion of *left* sub-tree, i.e. with *p(t)* lower than potential threshold value of split *s*.

$$p_l(t) = p_l(t) / p(t)$$ Equation A.6

$p_r(s,t)$ estimates the proportion of *right* sub-tree, i.e. with *p(t)* lower than potential threshold value of split *s*.

$$p_r(t) = p_r(t) / p(t)$$ Equation A.7

$\hat{j}(\tilde{t})$ assigns class with most samples in leaf $\tilde{t}$

$$\hat{j}(\tilde{t}) = \arg\max_{j \in \{1 \ldots J\}} p(j \mid \tilde{t})$$ Equation A.8

$m(\tilde{t})$ estimates the probability of misclassification in leaf $\tilde{t}$

$$m(\tilde{t}) = 1 - \max_{j \in \{1 \ldots J\}} p(j \mid \tilde{t})$$ Equation A.9

M(T) estimates the resubstitution error of the decision tree *T*

$$M(T) = \sum_{t \in \tilde{T}} m(t) \cdot p(t) \qquad \text{Equation A.10}$$

I(T) estimates the impurity of the decision tree *T*

$$I(T) = \sum_{t \in \tilde{T}} i(t) \cdot p(t) \qquad \text{Equation A.11}$$

Notations for Equation A.12 to Equation A.15, for an illustration see Figure A.3

C...number of classes
n...samples

User's accuracy

$$UA_i = \frac{n_{ii}}{n_{i+}} \qquad \text{Equation A.12}$$

Producer's accuracy

$$PA_i = \frac{n_{ii}}{n_{+i}} \qquad \text{Equation A.13}$$

Overall accuracy

$$OA = \frac{\sum_{i=1}^{C} n_{ii}}{n} \qquad \text{Equation A.14}$$

Kappa coefficient

$$KHAT = \frac{n \cdot \sum_{i=1}^{C} n_{ii} - \sum_{i=1}^{C} n_{i+} \cdot n_{+i}}{n^2 \cdot \sum_{i=1}^{C} n_{i+} \cdot n_{+i}} \qquad \text{Equation A.15}$$

Eidesstattliche Erklärung

Hiermit erkläre ich, dass die von mir eingereichte Dissertation zum Thema „Time Series Generation and Classification of MODIS Data for Land Cover Mapping“ selbständig und nur unter Benutzung der in der Arbeit angegebenen Literatur und sonstigen Quellen angefertigt wurde. Indirekt übernommene Gedanken habe ich als solche kenntlich gemacht.

_____

Würzburg, den 04.September 2007

Curriculum vitae

Personal

Name:	René Roland Colditz
Address:	Am Altenberg 44 97078 Würzburg Germany
Phone:	+49-(0)931-282700
Email:	rene.colditz@mail.uni-wuerzburg.de

Date of birth:	February 15th 1979
Place of birth:	Jena
Nationality:	German
Martial status:	single

Education

09/1985 – 08/1991:	Primary and secondary school Polytechnische Oberschule Bertolt Brecht / Nordschule, Jena
09/1991 – 06/1997:	High school Gymnasium Johannes R. Becher / Angergymnasium, Jena
	school leaving exam (Abitur): overall grade: very good (1,4)
11/1997 – 08/1998:	Military service
11/1998 – 07/2001:	Study of Geography at the Julius-Maximilians-University Würzburg with minors Geology and Computer Science
07/2000:	intermediate exam with overall grade „good“
08/2001 – 05/2002:	Study at the University of Texas at Austin, Department of Geography, overall grade: very good (grade point average 4.0)

06/2002 – 03/2003:	Scientific research at the Department of Geography, University of Texas at Austin, writing of Diploma-thesis (German equivalent to Masters of Science)
04/2003 – 02/2004:	Study at the Julius-Maximilians-University Würzburg
01/2004:	Degree in Geography "Diplom-Geograph" with overall grade „very good"
08/2004 – 07/2007	Ph.D. student of the German Aerospace Center at the Remote Sensing Unit, Department of Geography, Julius-Maximilians-University Würzburg
since 08/2007:	Integrated expert at the Comisión nacional para el conocimiento y uso de la biodiversidad (CONABIO), Mexico City, supported by the Centrum für internationale Migration und Entwicklung (CIM)

Skills

In-depth skills:	Windows, Office, Erdas Imagine, ENVI, ArcView, ArcGIS Desktop, IDL, Pascal, Basic
Basic skills:	Linux, ArcInfo Workstation, Fortran, Java, C, C++
Languages:	English (fluent), Russian (basic), Spanish (basic)
Further skills:	Drivers license: BCE (car up to 7.5 tons and trailer)

Interests

Main interests:	Remote sensing, approaches of time series generation and analysis, image classification with emphasis on time series, validation of remote sensing products, image fusion
Further interests:	GIS, scale analysis, physical geography, North America, Africa with emphasis on southern Africa

Employments during study

07/2000 – 07/2001: Driver for Company Hensel Studiotechnik GmbH & Co. KG

01/2002 – 08/2002: Research assistant for remote sensing and GIS at the Department of Geography, University of Texas at Austin

11/2003 – 12/2003: Research assistant at the Remote Sensing Unit, Department of Geography, University of Würzburg

03/2004 – 07/2004: Research assistant and teaching assistant at the Remote Sensing Unit, Department of Geography, University of Würzburg

Internships during study

03/2000 – 04/2000: Assistant for geomorphological mapping and data processing, Dipl.-Geogr. Jürgen Jung, Department of Geography University of Würzburg

03/2001 – 04/2001: GIS-assistant at environmental planning office Zeeb, Dipl.-Geogr. Regina Zeeb, Zeeb Ökologie und Umweltplanung, Gernstr. 43, 89081 Ulm

09/2002 – 12/2002: Remote sensing and GIS-assistant as well as field assistant in the Pánuco project, Prof. Dr. Paul F. Hudson, Department of Geography, University of Texas at Austin.

10/2003 – 11/2003: Intern at the German Remote Sensing Data Center, German Aerospace Center (DFD-DLR) Oberpfaffenhofen, Unit: Environment and Security, project: Economic and Ecological Restructuring of Land and Water Use in the Khorezm Region (Uzbekistan)

Awards

11/2002: First place in poster-competition at the annual meeting of South Western Association of American Geographers (SWAAG), Laredo, TX; Colditz, R.R. Geomorphic mapping of lower Pánuco basin floodplain environments using satellite imagery.

Journal Articles

2007

Colditz, R.R., Conrad, C., Wehrmann, T., Schmidt, M., Dech, S., 2007, TiSeG – A flexible software tool for time series generation of MODIS data utilizing the quality assessment science data set. *IEEE Transactions on Geoscience and Remote Sensing*, (submitted).

Wagner, S. Kunstmann, H., Bárdossy, A., Conrad, C., Colditz, R.R., 2007. Water balance simulations in a poorly gauged basin using different meteorological and land surface data sources. *Physics and Chemistry of the Earth*, (submitted).

Walz, Y., Maier, S.W., Dech, S.W., Conrad, C., Colditz, R.R., 2007, Classification of burn severity using Moderate Resolution Imaging Spectroradiometer (MODIS): A case study in the jarrah-marri forest of southwest Western Australia. *Journal of Geophysical Research*, 112: 14 pp.

2006

Colditz, R.R., Wehrmann, T., Bachmann, M., Steinnocher, K., Schmidt, M., Strunz, G., Dech, S., 2006, Influence of Image Fusion Approaches on Classification Accuracy – A Case Study. *International Journal of Remote Sensing*, 27: 3311-3335.

Hudson, P.F., Colditz, R.R., Augilar-Robledo, M., 2006, Spatial relations between floodplain environments and land use – land cover of a large lowland tropical river valley: Pánuco basin, México. *Environmental Management*. 38: 487-503.

2004

Crews-Meyer, K.A., Hudson, P.F. and Colditz, R.R., 2004, Landscape complexity and remote classification in eastern coastal Mexico: application of Landsat 7ETM data. *Geocarto International*, 19: 45-56.

2003

Hudson, P.F. and Colditz, R.R., 2003, Flood delineation in a large and complex alluvial valley, lower Pánuco basin, Mexico. *Journal of Hydrology*. 280: 229-245. and *Geophysical Research Abstracts*, 5: 30-31.

Conferences

2007

Colditz, R.R, Geßner, U., Conrad, C., van Zyl, D., Malherbe, J., Newby, T., Landmann, T., Schmidt, M., Dech, S., 2007, Dynamics of MODIS time series for ecological applications in southern Africa. Fourth International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multitemp 2007), July 18th – 20th 2007, Leuven, Belgium, (CD-ROM).

Colditz, R.R., Conrad, C., Wehrmann, T., Schmidt, M. and Dech, S.W., 2007b. Analysis of the quality of collection 4 and 5 vegetation index time series from MODIS. ISPRS Spatial Data Quality Symposium, Enschede, The Netherlands, accepted and CRC press.

Colditz, R.R., Keil, M., Strohbach, B., Geßner, U., Schmidt, M., Dech, S., 2007, Vegetation structure mapping with remote sensing time series: Capabilities and improvements. 32nd International Symposium on Remote Sensing of Environment, June 25th – 29th 2007, San Jose, Costa Rica, (CD-ROM).

Landmann, T., Schmidt, M., Colditz, R.R., Dech, S., 2007, An object-conditional land cover classification system (LCCS) based wetland biodiversity characterization method for West African savannas using 250 meter MODIS observations. 32nd International Symposium on Remote Sensing of Environment, June 25th – 29th 2007, San Jose, Costa Rica, (CD-ROM).

2006

Landmann, T., Colditz, R.R., Schmidt, M., 2006, An object-conditional land cover classification system (LCCS) wetland probability detection method for west African savannas using 250-meter MODIS observations. In ESA-ESRIN GlobWetland symposium, October 19th -20th 2006, Frascati, Italy, 5 pages.

Colditz, R.R., Conrad, C., Wehrmann, T., Schmidt, M., Dech, S., 2006 Generation and assessment of MODIS time series using quality information. In: IEEE International Conference on Geoscience and Remote Sensing, 2006. IGARSS 2006. July 31st - August 4th 2006, Denver, CO, pp. 779-782.

Colditz, R.R., Conrad, C., Schmidt, M., Schramm, M., Schmidt, M., Dech, S., 2006, Mapping regions of high temporal variability in Africa. In ISPRS mid-term symposium Remote sensing: from pixels to processes, May 8th – 11th 2006, Enschede, the Netherlands, 6 pages.

2005

Colditz, R.R., Conrad, C., Ruecker, G.R., Schweitzer, C., Fistric, S., Schmidt, M., Dech, S, 2005, Ableitung von phänologischen Verlaufsmustern aus MODIS Zeitserien und Möglichkeiten der Anwendung. In: Strobl, J., Blaschke, T., and Griesebner G., (Eds), 17. Symposium und Fachmesse für Angewandte Geoinformatik (AGIT), July 6th - 8th 2005, Salzburg, Austria, pp. 94-99.

Conrad, C., Colditz, R.R., Petrocchi, A., Rücker, G.R., Dech, S.W. and Schmidt, M., 2005, Time Series Generator – Ein flexibles Softwaremodul zur Generierung und Bewertung von Zeitserien aus NASA MODIS Datenprodukten. In: Strobl, J., Blaschke, T., and Griesebner G., (Eds), 17. Symposium und Fachmesse für Angewandte Geoinformatik (AGIT), July 6th - 8th 2005, Salzburg, Austria, pp. 100-105.

Wehrmann, T., Colditz, R.R., Bachmann, M., Steinnocher, K. and Dech, S.W., 2005, Evaluation of image fusion techniques. In: Erasmi, S., Cyffka, B., Kappas, M. (Eds), Remote Sensing & GIS for Environmental Studies. Goettinger GIS and Remote Sensing Days (GGRS 2004), October 7th – 8th 2004, Goettingen, Germany, Goettinger Geographische Abhandlungen, 113: 296-302.

2004

Conrad, C., Rücker, G.R., Colditz, R.R., Strunz, G. and Dech, S.W., 2004, Crop monitoring using multi-temporal MODIS remote sensing data in Khorezm, Uzbekistan, Proceedings of the Humboldt Kolleg 2004, Uzbekistan, July 6-10, 2004 Tashkent, Uzbekistan. The use of geographic information systems and simulation models for research and decision support in central Asian river basins. Humboldt Kolleg, July 6th – 10th 2004, Tashkent, Uzbekistan, pp. 118-125.

2003

Colditz, R.R. and Hudson, P.F., 2003, Land use / land cover classification of a large and complex floodplain environment: lower Pánuco basin, Mexico, 99th annual meeting of the Association of American Geographers (AAG), March 4th – 8th 2003, New Orleans, LA. 10 pages.

2002

Colditz, R.R., 2002, Geomorphic mapping of lower Pánuco basin floodplain environments using satellite imagery, Annual meeting of the South Western Association of American Geographers (SWAAG), November 6th – 9th 2002, Laredo, TX, Poster.

Heitmuller, F.T., Hudson, P.F., Colditz, R.R. and Crawford, M.M., 2002, Floodplain morphology of lower Moctezuma and Pánuco basins, Mexico, 98th annual meeting of the Association of American Geographers (AAG), March 19th – 22nd 2002, Los Angeles, CA, Poster.

Reports, Thesis

2004

Rücker, G.R., Conrad, C., Colditz, R.R., Strunz, G. and Dech, S.W., 2004, Remote sensing based mapping and characterization of soil and vegetation quality of potential plantation areas in the desiccated Aral Sea area, ZEF Work Papers for Sustainable Development in Central Asia, Bonn, Germany, pp. 20. http://www.khorezm.uni-bonn.de/downloads/WPs/ZEF-UZ-WP07-Ruecker2.pdf

2003

Colditz, R.R., 2003. Land cover and geomorphological floodplain mapping of the lower Pánuco basin, Mexico, utilizing remote sensing and GIS-methods, University of Wuerzburg, Wuerzburg, Germany, 198 pp.

Würzburg, September 04th 2007 René Colditz