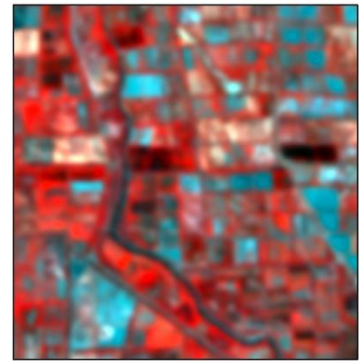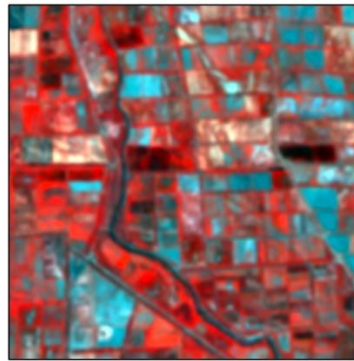# Agricultural crop mapping
from multi-scale remote sensing data

Concepts and applications in heterogeneous
Middle Asian agricultural landscapes

Dissertation

zur Erlangung des naturwissenschaftlichen

Doktorgrades der Julius-Maximilians-Universität Würzburg

Vorgelegt von Dipl. Geogr. Fabian Löw

geboren am 23. Juli 1983 in Heidelberg

Würzburg, November 2013

Eingereicht am: 27. November 2013

von: Dipl. Geogr. Fabian Löw

am: Lehrstuhl für Fernerkundung der Julius-Maximilians-Universität Würzburg, in Kooperation mit dem Deutschen Luft- und Raumfahrtzentrum (DLR)

1. **Gutachter:** Prof. Dr. Christopher Conrad, Würzburg
2. **Gutachter:** Prof. Dr. Ulrich Michel, Heidelberg

1. **Prüfer:** Prof. Dr. Christopher Conrad, Würzburg
2. **Prüfer:** Prof. Dr. Ulrich Michel, Heidelberg

Mündliche Prüfung am: 02. Juli 2014

Mit jedem Schritt, den wir vorwärts machen, mit jedem Problem, das wir lösen, entdecken wir nicht nur neue und ungelöste Probleme, sondern wir entdecken auch, dass dort, wo wir auf festem und sicherem Boden zu stehen glaubten, in Wahrheit alles unsicherer und im Schwanken begriffen ist.

Karl R. Popper

ACKNOWLEDGEMENTS

- First I want to thank my PhD supervisors, Prof. Dr. Christopher Conrad (Würzburg) and Prof. Dr. Ulrich Michel (Heidelberg), for supervising this thesis and the helpful discussions. I am grateful that I was given the opportunity to work at the Department of Remote Sensing in Würzburg, and that I was given the ability to do research on my dissertation in the context of international research projects. Further, I express my gratitude to Prof. Dr. Stefan Dech (DLR/DFD), who was supportive of my work, and also to Prof. Dr. Roland Baumhauer (Würzburg).

- My special thanks goes to Dr. Grégory Duveiller from the JRC in Ispra, Italy. The many discussions we had via E-mail, phone, and in Edinburgh, helped me a lot to get things in the right direction. Thank you, Grégory, for bringing your broad perspective, profound insight, and your support.

- My deep gratitude also goes to Gunther Schorcht for guiding and mentoring me through this 3-years lasting adventure, for the many discussions and his frequent feedback regarding the content of this thesis. His advice, critical questions, and patience in explaining me programming-related issues improved my work a lot. Thank you, Gunther!

- What would these 3 years have been without my fellow colleagues! I want to thank all of them for the unforgettable time we shared in the office, and that we spent in Würzburg. I also want to thank Dr. Hooman Latifi and Dr. Martin Wegmann for their advice in solving R-specific problems and their helpful comments on my work and manuscripts. Acknowledgements are also given to the student assistants Peter Zellner and Moritz Rudloff. I will not forget to express my gratitude to the brave colleagues who enjoyed proofreading and commenting on the beta-versions of my thesis chapters: Sebastian, Sarah, and Gunther.

- I am very grateful to the German National Academic Foundation (*Studienstiftung des deutschen Volkes*) for funding this thesis by way of a PhD grant.

- This dissertation has been carried out in the framework of several projects: I would like to thank Dr. Ralf Peveling and Benjamin Mohr from the *Deutsche Gesellschaft für Internationale Zusammenarbeit* (GIZ) for supporting the field studies in Middle Asia, which were conducted in the scope of the EEWA project. Thanks to this project I had the chance to conduct my research in Uzbekistan and Kazakhstan, where I spent an important amount of time to collect field data. Further, this thesis was conducted in the framework of the CAWA project. I am grateful to Dr. John Lamers (ZEF, Bonn) for enabling me to get in touch with the German-Uzbek Khorezm-Project and to spend some informative weeks in Urgench. I thank all my friends and colleagues in Karakalpakstan (Murat, Dawlet, Mukhabat, Mamanbek) and Kyzyl Orda (Marat), for their patience and their endurance.

- I am also grateful to my parents (Karin and Franz) and to my brother Mario, for their motivation and love. Last but not least I want to thank Elisabeth, for her love, patience, and her mental support.

ABSTRACT

Agriculture is mankind's primary source of food production and plays the key role for cereal supply to humanity. One of the future challenges will be to feed a constantly growing population, which is expected to reach more than nine billion by 2050. The potential to expand cropland is limited, and enhancing agricultural production efficiency is one important means to meet the future food demand. Hence, there is an increasing demand for dependable, accurate and comprehensive agricultural intelligence on crop production. The value of satellite earth observation (EO) data for agricultural monitoring is well recognized. One fundamental requirement for agricultural monitoring is routinely updated information on crop acreage and the spatial distribution of crops. With the technical advancement of satellite sensor systems, imagery with higher temporal and finer spatial resolution became available. The classification of such multi-temporal data sets is an effective and accurate means to produce crop maps, but methods must be developed that can handle such large and complex data sets. Furthermore, to properly use satellite EO for agricultural production monitoring a high temporal revisit frequency over vast geographic areas is often necessary. However, this often limits the spatial resolution that can be used. The challenge of discriminating pixels that correspond to a particular crop type, a prerequisite for crop specific agricultural monitoring, remains daunting when the signal encoded in pixels stems from several land uses (mixed pixels), e.g. over heterogeneous landscapes where individual fields are often smaller than individual pixels.

The main purposes of the presented study were (i) to assess the influence of input dimensionality and feature selection on classification accuracy and uncertainty in object-based crop classification, (ii) to evaluate if combining classifier algorithms can improve the quality of crop maps (e.g. classification accuracy), (iii) to assess the spatial resolution requirements for crop identification via image classification.

Reporting on the map quality is traditionally done with measures that stem from the confusion matrix based on the hard classification result. Yet, these measures do not consider the spatial variation of errors in maps. Measures of classification uncertainty can be used for this purpose, but they have attained only little attention in remote sensing studies. Classifier algorithms like the support vector machine (SVM) can estimate class memberships (the so called soft output) for each classified pixel or object. Based on these estimations, measures of classification uncertainty can be calculated, but it has not been analysed in detail, yet, if these are reliable in predicting the spatial distribution of errors in maps. In this study, SVM was applied for the classification of agricultural crops in irrigated landscapes in Middle Asia at the object-level. Five different categories of features were calculated from RapidEye time series data as classification input. The reliability of classification uncertainty measures like entropy, derived from the soft output of SVM, with regard to predicting the spatial distribution of error was evaluated. Further, the impact of the type and dimensionality of the input data on classification uncertainty was analysed. The results revealed that SMVs applied to the five feature categories separately performed different in classifying different types of crops. Incorporating all five categories of features by concatenating them into one stacked vector did not lead to an increase in accuracy, and partly reduced the model performance most obviously because of the Hughes phenomena. Yet, applying the random forest (RF) algorithm to select a subset of features led to an increase of classification accuracy of the SVM. The feature group with red edge-based indices was the most important for general crop classification, and the red edge NDVI had an outstanding importance for classifying crops. Two measures of uncertainty were calculated based on the soft output from SVM: maximum a-posteriori probability and alpha quadratic entropy. Irrespective of the measure used, the results indicate a decline in classification uncertainty when a dimensionality reduction was performed. The two uncertainty measures were found to be reliable indicators to predict errors in maps. Correctly classified test cases

were associated with low uncertainty, whilst incorrectly test cases tended to be associated with higher uncertainty.

The issue of combining the results of different classifier algorithms in order to increase classification accuracy was addressed. First, the SVM was compared with two other non-parametric classifier algorithms: multilayer perceptron neural network (MLP) and RF. Despite their comparatively high classification performance, each of the tested classifier algorithms tended to make errors in different parts of the input space, e.g. performed different in classifying crops. Hence, a combination of the complementary outputs was envisaged. To this end, a classifier combination scheme was proposed, which is based on existing algebraic operators. It combines the outputs of different classifier algorithms at the per-case (e.g. pixel or object) basis. The per-case class membership estimations of each classifier algorithm were compared, and the reliability of each classifier algorithm with respect to classifying a specific crop class was assessed based on the confusion matrix. In doing so, less reliable classifier algorithms were excluded at the per-class basis before the final combination. Emphasis was put on evaluating the selected classification algorithms under limiting conditions by applying them to small input datasets and to reduced training sample sets, respectively. Further, the applicability to datasets from another year was demonstrated to assess temporal transferability. Although the single classifier algorithms performed well in all test sites, the classifier combination scheme provided consistently higher classification accuracies over all test sites and in different years, respectively. This makes this approach distinct from the single classifier algorithms, which performed different and showed a higher variability in class-wise accuracies. Further, the proposed classifier combination scheme performed better when using small training set sizes or when applied to small input datasets, respectively.

A framework was proposed to quantitatively define pixel size requirements for crop identification via image classification. That framework is based on simulating how agricultural landscapes, and more specifically the fields covered by one crop of interest, are seen by instruments with increasingly coarser resolving power. The concept of crop specific pixel purity, defined as the degree of homogeneity of the signal encoded in a pixel with respect to the target crop type, is used to analyse how mixed the pixels can be (as they become coarser) without undermining their capacity to describe the desired surface properties (e.g. to distinguish crop classes via supervised or unsupervised image classification). This tool can be modulated using different parameterizations to explore trade-offs between pixel size and pixel purity when addressing the question of crop identification. Inputs to the experiments were eight multi-temporal images from the RapidEye sensor. Simulated pixel sizes ranged from 13 m to 747.5 m, in increments of 6.5 m. Constraining parameters for crop identification were defined by setting thresholds for classification accuracy and uncertainty. Results over irrigated agricultural landscapes in Middle Asia demonstrate that the task of finding the optimum pixel size did not have a "one-size-fits-all" solution. The resulting values for pixel size and purity that were suitable for crop identification proved to be specific to a given landscape, and for each crop they differed across different landscapes. Over the same time series, different crops were not identifiable simultaneously in the season and these requirements further changed over the years, reflecting the different agro-ecological conditions the investigated crops were growing in. Results further indicate that map quality (e.g. classification accuracy) was not homogeneously distributed in a landscape, but that it depended on the spatial structures and the pixel size, respectively. The proposed framework is generic and can be applied to any agricultural landscape, thereby potentially serving to guide recommendations for designing dedicated EO missions that can satisfy the requirements in terms of pixel size to identify and discriminate crop types.

Regarding the operationalization of EO-based techniques for agricultural monitoring and its application to a broader range of agricultural landscapes, it can be noted that, despite the high performance of existing methods (e.g. classifier algorithms), transferability and stability of such methods remain one important research issue. This means that methods developed and tested in one place might not necessarily be portable to another place or over several years, respectively. Specifically

in Middle Asia, which was selected as study region in this thesis, classifier combination makes sense due to its easy implementation and because it enhanced classification accuracy for classes with insufficient training samples. This observation makes it interesting for operational contexts and when field reference data availability is limited. Similar to the transferability of methods, the application of only one certain kind of EO data (e.g. with one specific pixel size) over different landscapes needs to be revisited and the synergistic use of multi-scale data, e.g. combining remote sensing imagery of both fine and coarse spatial resolution, should be fostered. The necessity to predict and control the effects of spatial and temporal scale on crop classification is recognized here as a major goal to achieve in EO-based agricultural monitoring.

ZUSAMMENFASSUNG

Landwirtschaftlicher Ackerbau spielt heute eine Schlüsselrolle bei der Nahrungsmittelversorgung der Menschheit. Eine der zukünftigen Herausforderungen wird die Ernährung der stetig wachsenden Erdbevölkerung sein, welche bis zum Jahr 2050 auf neun Milliarden Menschen anwachsen wird. Das Potential zur Ausdehnung von Ackerland ist jedoch begrenzt, so dass die Steigerung der landwirtschaftlichen Produktionseffizienz ein wichtiges Mittel ist, um den künftigen Nahrungsmittelbedarf zu decken. Daher gibt es einen zunehmenden Bedarf an belastbaren, genauen und umfassenden Informationen über die Agrarproduktion. Der Nutzen der Satellitenbild-Fernerkundung ist in diesem Kontext mittlerweile anerkannt. Eine wichtige Voraussetzung für das Agrarmonitoring sind aktuelle Informationen über die Fläche sowie die räumliche Verteilung von Anbaukulturen. Durch die technologische Entwicklung steht heute eine Vielfalt an Satellitenbildsystemen mit immer höherer räumlicher und zeitlicher Auflösung zur Verfügung. Die Klassifikation solcher hochaufgelösten, multi-temporalen Datensätze stellt eine bewährte Methode dar, um Karten der agrarischen Landnutzung zu erstellen und die benötigten Informationen zu erhalten. Jedoch müssen die dabei verwendeten Methoden auf die sehr komplexen Eingangsdaten anwendbar sein. Zudem benötigt man zur Modellierung der Agrarproduktion oft eine hohe Aufnahmefrequenz bei gleichzeitig großer räumlicher Abdeckung. Diese Voraussetzungen schränken jedoch aus technischen Gründen oftmals die zur Verfügung stehenden Pixelgrößen ein, da Sensoren, welche diese Voraussetzungen erfüllen, in der Regel eine gröbere räumliche Auflösung haben. Die Unterscheidung von Pixeln unterschiedlicher Landnutzung als eine Voraussetzung für feldfrucht-spezifisches Agrarmonitoring kann dann erschwert sein, wenn Satellitenbilder über heterogenen Landschaften aufgezeichnet werden. In solchen Fällen kann das im Pixel kodierte Signal von mehreren Nutzungstypen stammen (Mischpixel), was zur Zunahme von Klassifikationsfehlern führen kann.

Hauptgegenstände dieser Studie sind: (i) die Untersuchung des Einflusses der Größe sowie der Art der Eingangsdaten auf die Klassifikationsgenauigkeit und die Klassifikationsunsicherheit in der objekt-basierten Landnutzungsklassifikation; (ii) die Kombination von Klassifikationsalgorithmen zur Steigerung der Klassifikationsgenauigkeit; (iii) die Untersuchung des Einflusses der Pixelgröße auf die agrarische Landnutzungsklassifikation.

Die Genauigkeit einer Klassifikation wird im Allgemeinen mit Hilfe von Gütemaßen ermittelt, welche auf der Konfusionsmatrix basieren. Jedoch berücksichtigen diese Maße nicht die räumliche Variabilität von Klassifikationsfehlern in einer Karte. Maße der Klassifikationsunsicherheit können für diesen Zweck verwendet werden, allerdings ist deren Anwendung in der Fernerkundung bislang nur selten untersucht worden. Klassifikationsalgorithmen wie das Stützvektorverfahren können für jedes Pixel oder Objekt klassenweise Abschätzungen der Klassenzugehörigkeit berechnen, aus welchen dann Maße der Klassifikationsunsicherheit (z.B. Entropie) berechnet werden können. Jedoch wurde noch nicht hinreichend untersucht, ob die damit gewonnenen Informationen zur Abschätzung der räumlichen Verteilung von Klassifikationsfehlern in Karten zuverlässig sind. In dieser Studie wurde das Stützvektorverfahren verwendet, um die agrarische Landnutzung in bewässerten Agrarlandschaften Zentralasiens zu klassifizieren. Fünf Kategorien von Eingangsdaten wurden aus Aufnahmen des RapidEye Systems berechnet und als Grundlage für die agrarische Landnutzungsklassifikation verwendet. Es wurde untersucht, ob Maße der Klassifikationsunsicherheit, welche auf den pixel- bzw. objektweisen Abschätzungen der Klassenzugehörigkeit durch das Stützvektorverfahren basieren, die räumliche Verteilung von Klassifikationsfehlern in Landnutzungskarten zuverlässig schätzen können. Weiterhin wurde der Einfluss sowohl der Art als auch der Größe der Eingangsdaten auf die Klassifikationsunsicherheit untersucht. Die Ergebnisse der Untersuchung weisen darauf hin, dass sich sowohl die getrennte als auch die kombinierte Verwendung der fünf Eingangsdatenkategorien unterschiedlich zur Klassifikation verschiedener Landnutzungsklassen eignen. Die kombinierte

Verwendung aller fünf Kategorien führte zum Teil zu einer Reduktion der Klassifikationsgenauigkeit, was wahrscheinlich auf das Hughes-Phänomen zurückzuführen ist. Durch die Verwendung des „Random Forest" Verfahrens zur Selektion geeigneter Eingangsdaten konnte die Klassifikationsgenauigkeit des Stützvektorverfahrens gesteigert werden. Eingangsdaten basierend auf dem sogenannten „Red Edge" Kanal des RapidEye Systems waren zur Klassifikation von Feldfrüchten am wichtigsten, insbesondere der „Red Edge NDVI". Zwei Maße der Klassifikationsunsicherheit wurden berechnet: die maximale a-posteriori Klassifikationswahrscheinlichkeit und die Alpha-Quadrat Entropie. Die Ergebnisse weisen darauf hin, dass diese beiden Maße verlässliche Prädiktoren für die räumliche Verteilung von Klassifikationsfehlern sind. Korrekt klassifizierte Testfelder waren durch geringe Klassifikationsunsicherheit und inkorrekt klassifizierte Testfelder in der Regel durch hohe Klassifikationsunsicherheit charakterisiert.

Es wurde untersucht, ob die Kombination mehrerer Klassifikationsalgorithmen zu einer Steigerung der Klassifikationsgenauigkeit führt. Zunächst wurde das Stützvektorverfahren mit anderen nicht-parametrischen Verfahren (neuronalen Netzwerken und Random Forest) verglichen. Obwohl die getesteten Klassifikationsalgorithmen gute Gesamt-Klassifikationsgenauigkeiten erzielten, bestanden große Unterschiede in den klassenweisen Genauigkeiten. Daher wurde ein Verfahren entwickelt, um die teilweise komplementären Ergebnisse unterschiedlicher Klassifikationsalgorithmen zu kombinieren. Dieses Verfahren basiert auf der Erweiterung algebraischer Kombinationsoperatoren und kombiniert die Ergebnisse verschiedener Klassifikationsalgorithmen basierend auf den pixel- bzw. objektweisen Abschätzungen der Klassenzugehörigkeit. Zudem wurde jeder Klassifikationsalgorithmus klassenweise bewertet, basierend auf Maßen der Konfusionsmatrix. So konnten Klassifikationsalgorithmen für diejenigen Klassen von der Kombination ausgeschlossen werden, für deren klassenweisen Genauigkeiten bestimmte Kriterien nicht erfüllt wurden. Das vorgestellte Verfahren wurde mit den Ergebnissen der einzelnen Klassifikationsalgorithmen verglichen. Zudem wurde auf räumliche und zeitliche Übertragbarkeit hin getestet und der Einfluss der Auswahl von Trainingsdaten wurde untersucht. Obwohl die einzelnen Klassifikationsalgorithmen genaue Ergebnisse erzielten, konnte das vorgestellte Kombinationsverfahren in allen Gebieten und über mehrere Jahre bessere Ergebnisse mit geringerer Variabilität erzielen. Zudem konnte das Verfahren auch dann genauere Ergebnisse liefern, wenn nur wenige Trainingsdaten oder Eingangsdaten zur Verfügung standen.

In dieser Studie wurde eine Methodik entwickelt, um quantitativ die maximal tolerierbaren Pixelgrößen für die agrarische Landnutzungsklassifikation zu bestimmen. Diese Methodik kann verwendet werden, um den kombinierten Effekt von Pixelgröße und Pixelreinheit im Kontext der Feldfruchtidentifikation mittels überwachter Klassifikation zu untersuchen. Die feldfruchtspezifische Pixelreinheit (definiert als der Grad der Homogenität des in Pixeln kodierten Signals) wurde verwendet um zu untersuchen, wie inhomogen die in gröberen Bildpixeln gespeicherte Information sein darf, um unterschiedliche Anbaukulturen mittels überwachter und unüberwachter Klassifikation unterscheiden zu können. Als Eingangsdaten für die Untersuchung wurden Bilder des RapidEye Systems verwendet. Es wurden Bildgrößen zwischen 13 m und 747.5 m in Schritten von 6.5 m simuliert. Als limitierende Faktoren für die Klassifikation wurden unterschiedliche Schwellenwerte für Maße der Klassifikationsgenauigkeit und Klassifikationsunsicherheit berücksichtigt. Die Ergebnisse zeigen, dass die Werte für tolerierbare Pixelgrößen und Pixelreinheiten sowohl landschafts- als auch feldfruchtspezifisch waren. Zudem konnten Feldfrüchte nicht simultan innerhalb der Wachstumsperiode identifiziert werden und die Voraussetzungen änderten sich in verschiedenen Jahren, was wahrscheinlich auf die unterschiedlichen agro-ökologischen Bedingungen in den untersuchten Landschaften zurückgeführt werden kann. Die Ergebnisse zeigen, dass Klassifikationsgüte in Karten räumlich ungleich verteilt war und von den räumlichen Strukturen bzw. von der Wahl der räumlichen Auflösung abhing. Die vorgestellte Methodik kann auch in anderen Agrarlandschaften

getestet werden. Des Weiteren kann die Eignung bestehender bzw. die Entwicklung künftiger Satellitenbildmissionen unterstützt werden.

In Hinblick auf die Nutzung von Satellitenbild-Fernerkundung für Agrarmonitoring und deren Anwendung in einer Vielfalt von Agrarlandschaften kann festgestellt werden, dass die räumliche Übertragbarkeit von Methoden und die Stabilität der Ergebnisse (z.B. gleichbleibend hohe Klassifikationsgenauigkeiten) weiterhin einen wichtigen Forschungsgegenstand darstellen. So konnte in dieser Studie gezeigt werden, dass herkömmliche Methoden zur Landnutzungsklassifikation bzw. Aussagen zu optimalen Pixelgrößen nicht in allen Fällen auf andere Regionen oder über mehrere Jahre übertragbar sind. In Zentralasien, welches die Fokusregion dieser Studie ist, zeigte sich, dass die Kombination verschiedener Klassifikationsalgorithmen sinnvoll ist, da die Klassifikationsgenauigkeit bei Klassen mit nur einer geringen Anzahl von Trainingsgebieten gesteigert werden konnte. Dies macht die Anwendung dieses Verfahrens im operationellen Kontext interessant. Die Eignung eines einzigen Satellitenbildsystems (mit einer bestimmten Pixelgröße) für die agrarische Landnutzungsklassifikation in mehreren Agrarlandschaften muss in Frage gestellt werden und die synergistische Nutzung von Daten unterschiedlicher räumlicher Auflösung sollte vorangetrieben werden. Dabei ist die Untersuchung des kombinierten Einflusses der räumlichen und zeitlichen Auflösung auf die agrarische Landnutzungs-klassifikation von großer Bedeutung für das erdbeobachtungsgestützte Agrarmonitoring.

CONTENTS

LIST OF ACRONYMS

AA	Average class-wise classification accuracy
ACC	Overall classification accuracy variable
AFM	Area between values in first lag of the semivariogram and its function until first maximum
ANN	Artificial neural networks
APM	Agricultural production monitoring
AQE	Alpha-quadratic entropy variable
ARVI	Atmospherically resistant vegetation index
AUC	Average area under the receiver operating characteristics curve variable
BGI	Blue-green ratio index
BP	Back-propagation
BT	Bagged trees
CA	Class-wise accuracy variable
CART	Classification and regression trees
CAWA	Central Asian Water project
CCD	Charge-coupled device
CCS	Classifier combination scheme
CCWS	China Crop Watch System
CLC	Corine Land Cover Inventory
CMM	Cropland mapping and monitoring
CNES	Centre National d'Etudes Spatiales
CUR	Curvature index
DOY	Day of year
DMF	Difference between MFM and semivariance at first lag
DT	Decision trees
DVI	Difference vegetation index
EO	Earth observation
ERM	Empirical risk minimization
ETM	Enhanced Thematic Mapper sensor
EVI	Enhanced vegetation index
FDO	First derivative near origin
FER	Fergana Valley
FML	First maximum lag value
FNEA	Fractal net evolution approach
FPR	False positive rate
GCP	Ground control point
GEOSS	Global Earth Observation System of Systems
GI	Greenness index
GIFOV	Ground instantaneous field of view
GIS	Geographical information system

GLAM	Global Agriculture Monitoring project
GLCM	Grey-level co-occurrence matrix
GMES	Global Monitoring for Environment and Security
GNDVI	Green normalized difference vegetation index
GPS	Global Positioning System
GRNDVI	Green-red edge normalized difference vegetation index
GSD	Ground sampling distance
GSI	Ground sampling interval
ICWC	Interstate Coordination Water Commission
IFAS	International Fund for the Aral Sea
IFOV	Instantaneous field of view
KAP	Kappa index of agreement variable
KHO	Khorezm
KKP	Karakalpakstan
KYZ	Kyzyl Orda
LAI	Leaf area index
MARS	Monitoring Agriculture through Remote Sensing Techniques
MCARI	Modified chlorophyll absorption ratio index
MCS	Multi-classifier systems
MDA	Mean decrease of OOB accuracy
MFM	Mean of semivariogram values up to first maximum
MLC	Maximum likelihood classifier
MLP	Multilayer perceptron
MODIS	Moderate-Resolution Imaging Spectroradiometer
MSR	Modified simple ratio
MSS	Multispectral Scanner Sensor
NASA	National Aeronautics and Space Administration
NASS	National Agricultural Statistical Service
NDVI	Normalized difference vegetation index
NIR	Near-infrared
OA	Overall classification accuracy
OAA	One-against-all
OAO	One-against-one
OOB	Out-of-bag
OSH	Optimal separating hyperplane
P	Pan-chromatic
PC	Principal component(s)
PCA	Principal component analysis
PDF	Probability density function
PSF	Point spread function
RBF	Radial basis function kernel

RDVI	Renormalized difference vegetation index
REDNDVI	Red edge-red normalized difference vegetation index
REL	Red edge length
RF	Random forest
RFE	Recursive feature elimination
RFS	Random feature selection
RGI	Red-green ratio index
RMM	Ratio between semivariance at first maximum and mean semivariance up to first maximum
RMSE	Root mean square error
RNDVI	Red edge normalized difference vegetation index
ROC	Receiver operating characteristic
RREL	Relative red edge length
RRI1	Red edge ratio index 1
RRI2	Red edge ratio index 2
RSF	Ratio between semivariance at second and first lag
RVF	Ratio variance at the first lag
RVI	Ratio vegetation index
SA	Steepest ascent
SAVI	Soil adjusted vegetation index
SCF	Sampling coverage of agricultural fields variable
SDF	Second-order difference between first lag and first maximum
SDT	Second derivative at third lag
SFFS	Sequential forward floating selection
SOM	Self-organizing feature map
SPOT	Satellite Pour l'Observation de la Terra
SRM	Structural risk minimization
SV	Semivariogram
SVM	Support vector machine
SWIR	Short-wave infrared
TCARI	Transformed chlorophyll absorption ratio index
TM	Thematic Mapper sensor
TOC	Top of canopy reflectance
TPR	True positive rate
TVI	Triangular vegetation index
UTM	Universal Transverse Mercator coordinate system
VAR	Variance from GLCM matrix
VI	Vegetation index
VIGREEN	Vegetation index green
VFM	Variance of semivariogram values up to first maximum
WGS	World Geodetic System

LIST OF SYMBOLS

α	Alpha, parameter that controls the behaviour of alpha quadratic entropy
a_j	Activation level of neurons (in MLP classification)
β	Parameter that weights precision and recall of the F_β-measure
C	Regularization parameter (in SVM classification)
C_f	Cover fraction (share of agricultural fields in the landscape)
cov	Image-specific variance-covariance matrix
D_c	Mean length scale calculated from semivariogram
E	Uncertainty measure
f	Number of features
F_β	F-measure
γ	Gamma (width of RBF kernel in SVM classification)
$\hat{\gamma}$	Semivariance
$\hat{\gamma}_r$	Semivariogram sill (value of the semivariogram at the effective range)
$H(x)$	Shannon entropy (uncertainty measure)
$H_{rel}(x)$	Relative entropy (uncertainty measure)
$H_\alpha(x)$	Alpha-quadrat entropy or quadratic score (uncertainty measure)
$H_\alpha rel$	Relative alpha-quadrat entropy or quadratic score (uncertainty measure)
$I_{f,avg}$	Average importance score of a feature according to the RF OOB method
$I_{g,avg}$	Average importance score of a feature group according to the RF OOB method
$I_{t,avg}$	Average importance score of an acquisition date according to the RF OOB method
k	Cohen´s Kappa coefficient of agreement
l	A fold in l-fold cross validation
m_{try}	Number of feature to split the nodes in RF classification
N_i	Pixel population size (number of pixels) of a specific class
N_0	Pixel population size at the finest investigated spatial resolution
$pr(x)$	Vector that contains membership estimations from a classifier (soft output)
$pr_{max}(x)$	Maximum numerical value in $pr(x)$
π	Pixel purity
σ	Standard deviation
ν	Pixel size
ν_0	Finest investigated pixel size
ν_{max}	Maximum tolerable pixel size
ν_{min}	Minimum tolerable pixel size
x	A sample (pixel or object)

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1. Agriculture in the global perspective

Since its "invention" during the Neolithic revolution approximately 10,000 years ago agriculture became mankind's primary source of food production and today plays the key role for cereal supply to humanity (Nentwig, 2005). A rather small number of the crops that have been cultivated by human mankind constitute about 75 % of the calories that humans consume: rice, maize, soy, and wheat (Lobell et al., 2011). Today, agriculture (crops and pasture) covers approximately 40 % of the earths´ ice-free surface (Ramankutty et al., 2008). The conversion of 15 million km^2 natural land surfaces for human use, in particular cropland (Monfreda et al., 2008) became a force of global significance (Foley et al., 2005).

One of the future challenges will be to feed a constantly growing population, which is expected to reach more than 9 billion by 2050 (UN, 2003). The FAO estimates that by 2030 12 % of the developing country population (more than 800 million people) will still be living in regions with low levels of food consumption (FAO, 2006). The prediction of the global increase of agricultural production demand by 2050 sums up to 60 %, and the strongest increase in demand for crops is expected to occur in developing countries (77 %) (Alexandratos and Bruinsma, 2012). This will lead to an increasing demand for food, which only can be met by boosting agricultural production (Foley et al., 2011).

Critically the potential to expand cropland is limited. Although agricultural production, expressed as yield of major crop groups (cereals, oilseeds, fruits, and vegetables) has almost doubled between 1985 and 2005, agricultural land increased by less than 10 % (FAO, 2013a; Foley et al., 2011). Because of the spatial (physical) limits of arable land it was technological advances like fertilization, chemical pest control, and partly irrigation that achieved this gain in agricultural production (Gleick, 2003). For instance the "Green Revolution" could enhance global crop production by a factor of two between the mid-1950s and the mid-1990s and average yields of wheat, maize, and rice doubled in this time span (Lobell et al., 2011). Although these technological developments provided an effective means to assure food security to the constantly growing human population, they have also had significant ecological impacts (Foley et al., 2005). They are one of the driving forces in the loss of biological diversity through land transformation (Chapin et al., 2000).

Against the background of continuing population growth and limited potential to enlarge suitable cropland, irrigation is suspected to become an increasingly important means to enhance the world´s food supply (Wichelns and Oster, 2006). Irrigated agriculture is a major component of the total agriculture (Howell, 2001). Developing countries alone account for three quarters of the global irrigated area (Bruinsma, 2003). Although irrigated land in these countries covers only approximately one fifth of the world´s total arable land it accounts for 30 to 40 % of the global crop production, and almost 60 % of the global cereal production (Bruinsma, 2003). Yet, it is also agriculture's largest user of freshwater: more than 70 % of the freshwater resources are withdrawn for irrigation alone (Bastiaanssen et al., 2000; Sauer et al., 2010).

Changes in the climate system can further exaggerate the future pressure on the freshwater resources (Pachauri and Reisinger, 2008). Agriculture will be impacted by climate change through reshaping the pattern of water availability and increasing the frequency of extreme weather events that disrupt agricultural production (e.g. droughts). This will offset the observed increase in average yield and change the pattern of crop production in some regions of the world (Godfray et al., 2010). Water requirements in most of the global irrigated areas will increase (Döll, 2002). Whilst changes in precipitation pattern and intensity will impact rain fed agriculture, increasing temperatures are melting mountain glaciers which supply water for irrigation agriculture (Justice and Becker-Reshef, 2007). The positive crop yield trends observed between 1980 and 2008 for maize, wheat, and soy slowed in Asia, and only yield trends for rice sped up, most of which can be attributed to increasing temperatures (Lobell et al., 2011). This

together with rising cost of fertilizers will aggravate the ability of some (and most likely the poorest) countries to satisfy the increasing food demand of a growing population in the future. These trends will contribute to continuing uncertainty in food supply, with the associated tensions reflected in market food prices (Justice and Becker-Reshef, 2007). Extensification of agricultural land use to enhance global food production and to meet the future food demand might not be the best option, but rather agriculture´s footprint must decrease while increasing the cropping efficiency (Foley et al., 2011). An increase of approximately 30 % in production of major cereals like wheat, maize, and rice is possible in the next decades but needs considerable changes in nutrient and water management (Mueller et al., 2012), and intensification is the preferred means (Foley et al., 2011). One solution for mitigating these issues is restructuring agricultural systems, and remote sensing bares a great potential for this task (Bastiaanssen and Bos, 1999; Bastiaanssen et al., 2000; Conrad et al., 2007; Eckhardt et al., 1990; Ozdogan et al., 2010; Wesseling and Feddes, 2006). Thus, increased attention should be given to putting in place reliable agricultural production monitoring (APM) systems (Atzberger, 2013; Béquignon et al., 2010).

1.2. Remote sensing-based agricultural monitoring

The global trends described above suggest an increasing demand for dependable, accurate and comprehensive agricultural intelligence on crop production. APM can support decision-making and prioritization efforts towards ameliorating vulnerable parts of agricultural systems. Remote sensing is a valuable input for APM because it provides variables that are strongly linked with the two main components of crop production, namely crop acreage and yield, in a spatially explicit and temporally frequent manner (Justice and Becker-Reshef, 2007). The value of earth observation (EO) data in environmental (Barrett and Curtis, 1992) and agricultural monitoring is well recognized (Justice and Becker-Reshef, 2007) and a variety of methods have been developed in the last decades to provide agricultural production related statistics (Carfagna and Gallego, 2005; Gallego et al., 1993). However, spatially explicit monitoring of agricultural production requires routinely updated information on the total surface under cultivation, and sometimes the spatial distribution of crops as input (Atzberger, 2013; Justice and Becker-Reshef, 2007; Ozdogan et al., 2010). This underlines the need for developing accurate and effective methods to map and monitor the distribution of agricultural lands and crop types (crop mapping).

1.2.1. Cropland mapping and monitoring – Opportunities and challenges

The principal technique to identify and map crops is image classification with one of the many available classifier concepts (Tso and Mather, 2009). Yet, the classification of crops has become a challenging issue, due to the recent advances in remote sensor technology: nowadays a huge amount of data (optical, radar, hyper-spectral) with ever finer pixel sizes (e.g. RapidEye 6.5 m), radiometric resolutions (e.g. 12–16-bit), and frequent revisit times (e.g. 3–5 days) was made available to the research community to meet the demand for EO data at the appropriate spatial resolution and at critical times during the growing season (see appendix A.1 for an overview of sensors used in crop mapping). It was demonstrated in many studies that multi-temporal data could significantly improve crop classification accuracy. More details including references are given in the following sections, but although such kind of data is preferable to classify crops, it may also undermine the ability of a classifier algorithm to perform accurately and they may even fail with large, high-dimensional data sets and only limited amount of training data (Pal and Foody, 2010; Waske et al., 2010). Data sets of some recent and future sensor systems (e.g. Sentinel-2, TerraSAR-X) will become even more complex. Hence, one of the main challenges for the operationalization of EO-based crop mapping is the development of methods that can

handle such kind of data because errors in crop maps can impact outputs from spatial models (Canters, 1997; Doraiswamy et al., 2004; Stehman and Milliken, 2007). Consequently, these methods must be accurate, applicable over various agro-ecological landscapes, and also transferable among the years in order to be of value in an operational context.

Another difficulty lies in the requirements in terms of revisit times and geographical coverage, which is confronted with inherent limitations of remote sensing systems. As was already mentioned, cropland mapping and monitoring (hereafter called CMM) is a fundamental input to agricultural monitoring. Yet, the term crop mapping can have different meanings depending on what specific application is considered. Some applications require delineating accurately where all crops (and perhaps all other land uses) are located over the entire area of interest, e.g. to produce accurate crop specific masks or to make crop acreage estimations by counting the classified pixels´ surface (Gallego, 2004). Crop monitoring typically requires information over large geographic extents in order to provide information at regional or global scales[1] (Wardlow and Egbert, 2008). However, this often leads to a spatial resolution that is coarser than desired, because sensor systems that fulfil this requirement typically have coarser pixel sizes, e.g. the Moderate-Resolution Imaging Spectroradiometer (MODIS) with 250 m (see appendix A.1). Crop mapping at these scales suffers from this compromise when applied to heterogeneous agro-ecological landscapes around the world, because crop identification in a pixel then relies on a signal that may be composed of reflected radiation coming from different adjacent land uses, thus diluting the information content regarding the target crop (mixed pixels). Consequently, at coarser scales it is probably not possible to make an exhaustive classification of all land uses (or crop types) with the required standard of accuracy. The extent to which crop specific monitoring with coarser EO data is possible was shown to depend on the landscape structure (Duveiller and Defourny, 2010), but no such framework exists to assess EO data requirements over different agricultural landscapes for crop classification. As can be seen from Figure 1-1 coarser EO supports (5 km–500 m) provide more fundamental information like agricultural state or cropland area, rather than crop specific information. Although coarser observation supports are sufficient to meet the information needs in agricultural monitoring (e.g. large geographical coverage and high observation frequency), this is also the scale at which most of the enhancements are needed (Justice and Becker-Reshef, 2007): specifically there is a need to foster and improve the knowledge about the possibilities and limitations to identify crops in coarser pixels (see question marks in Figure 1-1), as a prerequisite for crop specific monitoring.

The necessity to handle complex data sets for accurate crop mapping, plus the aforementioned limitations when using coarse EO data drive the motivation of this thesis: (i) to improve crop classification at the per-field scale and (ii) the identification of EO data requirements for an effective crop identification over heterogeneous agro-ecological landscapes. In the following sections, recent operational APM systems will be presented. Then, major technological trends and research needs related to CMM will be highlighted. This discussion will primarily focus on two aspects: methods for crop classification to create accurate crop masks and regional crop mapping, respectively. Having clarified the research needs, the scope and the specific objectives of this thesis will be highlighted at the end of the introduction chapter.

---

[1]: Throughout this thesis regional scale refers to the extent of the geographic coverage, e.g. national or sub-national, preferably with an administrative delineation in order to link results to official statistics. Local scale means that the geographic extent is a small area and the elementary unit of interest is a single field.
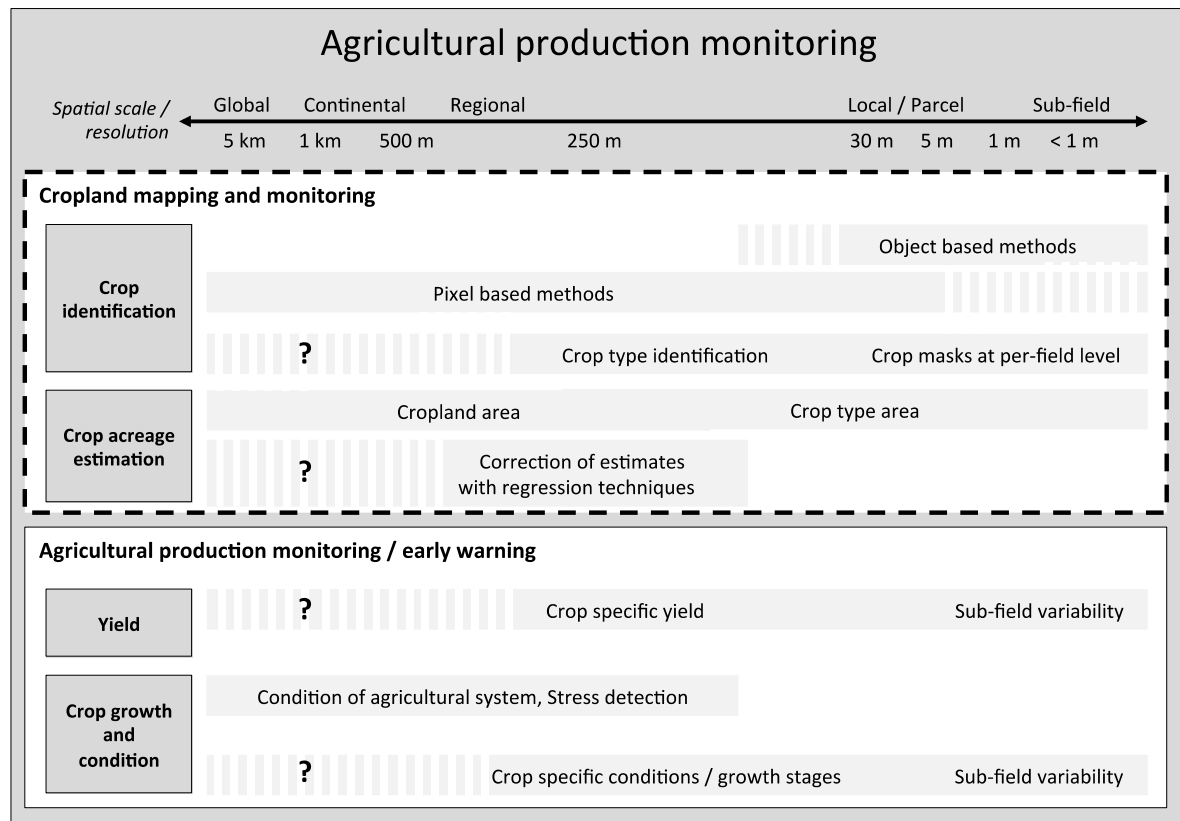
Figure 1-1: Applications related to agricultural monitoring. The emphasized domain (dashed contour), "Cropland mapping and monitoring", relates to this thesis. Precision agriculture is not shown here. Question marks indicate were further research is needed with regard to crop identification for crop specific agricultural monitoring with coarser EO data. Partly adapted from Justice and Becker-Reshef (2007).

1.2.2. Existing operational agricultural monitoring systems

Summarizing the work from Justice and Becker-Reshef (2007) there are three main functional components belonging to a comprehensive operational system for APM:

> Mapping and monitoring of changes in distribution of cropland area (cropped area and crop type distribution) and the associated cropping systems.

> Monitoring of agricultural production (yield).

> Early warning of shortfalls and famine.

Similar, in his review Atzberger (2013) counts crop mapping and acreage estimation as one of the major applications of remote sensing in agriculture (next to monitoring production, stress, crop phenology, and land use change). Also precision agriculture should be mentioned, which provides information for variable management practices within a field (Seelan et al., 2003). Over the last decades several operational or quasi-operational agricultural monitoring systems were developed at a regional scale that use EO data and provide CMM applications. One of these is the Agricultural Resource Inventory Through Aerospace Remote Sensing (AgRISTARS) or the pioneering Large Area Crop Inventory Experiment (LACIE) in the United States (U.S.) (Bailey and Boryan, 2010; Erickson, 1984). Today the National Agricultural Statistical Service (NASS) program provides frequently updated crop area estimates in the U.S. In Europe the Monitoring Agricultural Resources (MARS) of the Joint Research Centre (JRC) was initiated to foster the application of EO for the collection of agricultural production statistics (van Diepen and Boogaard, 2009).

One particular focus of these programs is mapping the extent of croplands (Bailey and Boryan, 2010; Gallego, 1999), crop type mapping, and monitoring of crop production and condition (Justice and Becker-Reshef, 2007). Yield forecasting and climate change related vulnerability analysis and resilience of crop production systems is one crucial element of MARS (www.mars.jrc.ec.europa.eu/mars, last accessed 16-Jun 2013). Another monitoring program is the China Crop Watch System (CCWS) at the Institute of Remote Sensing Applications of the Chinese Academy of Sciences (www.cropwatch.com.cn/en, last accessed 16-Jun 2013). Other national monitoring systems can be found in Brazil, India, and Russia. The Global Agriculture Monitoring (GLAM) Project, a joint initiative of the National Aeronautics and Space Administration (NASA) and other U.S. organizations goes beyond national or supranational scale to global information retrieval on cropland distribution, yield forecasting, and crop growth monitoring, and is the only operational provider of agricultural information at global scale (Becker-Reshef et al., 2010). The European EO program Copernicus, previously known as Global Monitoring for Environment and Security (GMES), is coordinated and managed by the European Commission. It provides, among others, the Corine Land Cover (CLC) inventory. The CLC land cover legend includes various agricultural land use classes (www.land.copernicus.eu/pan-european, last accessed 16-Jun 2013). Through its global land service, which is built on the BioPar component of the FP7 Geoland2 project, it also provides global biophysical parameters in near real time since 2009 (www.copernicus.eu, last accessed 16-Jun 2013). All these projects underline the relevance of EO data as part of agricultural monitoring systems, but specifically the integration of EO data into agricultural monitoring systems of developing countries still must be fostered (Justice and Becker-Reshef, 2007). Existing global systems like GLAM use coarse resolution EO data (250 m–1,000 m) (Becker-Reshef et al., 2010). Yet, information at this scale derived from coarser resolution EO data might not be suited for crop specific applications over heterogeneous landscapes (see Figure 1-1) such as they can be found in many agricultural regions in the world, and EO products used in such monitoring applications may have to be calibrated to local conditions (Fritsch et al., 2012).

1.2.3. Status of agricultural monitoring in Middle Asia

Away from the high level of technology and operational abilities of agricultural monitoring concepts in Europa or North America, the situation in Middle Asia[2] is entirely different. It belongs to the source areas of recent crops like melons or apples (Nentwig, 2005) and is one of the world's core regions of irrigation, with a strong focus on mono cropping of cotton and wheat (Lal et al., 2007). After the collapse of the Soviet Union in the 1990s the five independent republics in Middle Asia have faced the significant loss of management capacities (Cai et al., 2003). The former commonly managed water resources became transboundary and one of the most conflict-prone resources in the region (Glantz, 2005; Micklin, 2002). This dependence on shared resources, redrawn national borders (that do not fully take into account cultural legacies), the associated loss of the Soviet controlled economical compensation mechanisms, and the economies and infrastructure that were skewed to monoculture cotton production aggravated tensions between the countries after 1991 (McKinney, 2004; Micklin, 2002). The ability of the republics to maintain the irrigation infrastructure and to monitor agricultural production has been hampered or this even became impossible. Consequently, water control and distribution, which are critical to face the increasing food demand of the growing population in this region (Asian Development Bank, 2012; UNDP, 2005) through a deteriorating water supply system, are at risk (Cai et al., 2003).

---

[2]: The terms "Middle Asia" and "Central Asia" are frequently used synonymous. Referring to Cowan (2007) the term Middle Asia is preferred in this study because it refers to Kazakhstan, Turkmenistan, Uzbekistan, Tajikistan and Kyrgyzstan collectively, and avoids the ambiguity with the term Central Asia. The latter includes other countries of the former UdSSR, parts of China and Mongolia, and areas of the former Soviet/Chinese border, which are not in the scope of this thesis.

Although an operational agricultural monitoring system in Middle Asia has not been established yet to fully compensate for the disruption of the formerly Soviet controlled land management, attempts were made by the foundation of the Interstate Coordination Water Commission (ICWC) or the International Fund for the Aral Sea (IFAS) (McKinney, 2004), however without operational EO components. Land and water resource management in Middle Asia faces both, inconsistency of information, as well as inadequate and inappropriate tools for analysis, especially at the superordinate irrigation system and watershed level (Chemin et al., 2004). This procedure is time consuming, and limited for a long-term crop mapping, not to speak of regular crop growth monitoring or early warning applications. National states in Middle Asia do not have the financial capacities to maintain expensive transboundary agricultural monitoring systems. Reliable crop maps, the prerequisite for APM, are still lacking in the majority of regions.

The use of EO data in conjunction with geographical information systems (GIS) can be of significant importance to improve this situation as it is a reliable and increasingly inexpensive means to map and monitor land and water resources (Atzberger, 2013; Bastiaanssen and Harshadeep, 2005). The value of remote sensing in agricultural monitoring context has been recognized in Middle Asia (Chemin et al., 2004; Conrad et al., 2007, 2010, 2011a; Conrad et al., 2013; El-Magd and Tanton, 2003; Kariyeva and van Leeuwen, 2012; Löw et al., 2012; Machwitz et al., 2010; Schorcht et al., 2012) and other parts of Asia (Bastiaanssen and Harshadeep, 2005). First attempts to implement monitoring networks at regional and international scale in Middle Asia exist: the German-Uzbek Khorezm Project (www.zef.de/khorezm.0.html, last accessed 16-Jun 2013) was dedicated to improving the economic efficiency and ecological sustainability of the agricultural sector, but was spatially restricted to the Khorezm region in Uzbekistan. The project Central Asian Water (CAWA) was launched in 2008 and is funded by the German Federal Foreign Office (www.cawa-project.net, last accessed 16-Jun 2013). It is part of the German water initiative for Central Asia (the so-called "Berlin Process") and intends to contribute to a sound scientific and a reliable regional data basis for the development of sustainable water management strategies in Middle Asia, and further aims to provide some agriculture related baseline information (e.g. land use maps, agricultural leaf area index). However, initializing an operational application for Middle Asia is still at its beginning, and (more technically spoken) the requirements for the kind of EO data that is needed to do accurate and operational crop mapping has not been systematically explored, yet. This underlines the need to investigate classification techniques and the EO data requirements for crop mapping in Middle Asia, which was selected as the study region in this thesis.

1.3. Recent trends in crop mapping and research motivation

1.3.1. Creating accurate crop masks

The development of concepts for (crop-) classification is a challenging and on-going topic in remote sensing, scientifically driven by recent technological advances and a growing demand for operational remote sensing systems in support of environmental issues and the application of international agreements (Brachet, 2004; Peter, 2004; Rosenqvist et al., 2003). A variety of methods is employed to produce accurate maps (Lu and Weng, 2007), including the use of non-parametric classifier algorithms that are not constrained to assumptions like parametric distributions of the input data (Brown et al., 2013; Conrad et al., 2011b; Loosvelt et al. 2012a; Mathur and Foody, 2008), object-based rather than pixel-based approaches (Baatz and Schäpe, 2000; Conrad et al., 2010; Duro et al., 2012; Lobo et al., 1996; Peña-Barragán et al., 2011; Schorcht et al., 2012; Turker and Ozdarici, 2011; de Wit and Clevers, 2004), or fusion of multi-source data (Blaes et al., 2005; McNairn et al., 2009; Solberg et al., 1994;

Waske and Benediktsson, 2007). So called machine learning algorithms (Witten and Frank, 2005) like random forest (RF) (Breiman, 2001) and support vector machine (SVM) (Burges, 1998; Cortes and Vapnik, 1995; Vapnik, 1998) are increasingly used in crop classification due to their high classification performance on many types of input data (optical, radar, hyper spectral) and their general superiority in applications with high dimensional data sets (Foody and Mathur, 2004; Loosvelt et al., 2012a; Mathur and Foody, 2008; Waske and Benediktsson, 2007; Waske and Braun, 2009; Waske et al., 2010).

It was shown that by adding more acquisition dates the different periods in the growing season can be accurately captured (e.g. the different sowing and harvesting dates among crops) and usually classification accuracy is high or enhanced compared to mono-temporal data sets, respectively (Brown et al., 2013; Conrad et al., 2011a; Murakami et al., 2001; Sakamoto et al., 2010; Serra and Pons, 2008; Simonneaux et al., 2008; Wardlow and Egbert, 2008; Wardlow et al., 2007). There is evidence that employing time series of multi-spectral, radar, or a combination of both bare great potential in crop mapping (Blaes et al., 2005; Conrad et al., 2013; McNairn et al., 2002, 2009; Van Niel and Mcvicar, 2004; Waske and Benediktsson, 2007). However, these new datasets become increasingly more complex, due to the ever-finer pixel sizes (e.g. RapidEye with 6.5 m), acquisition frequencies, and enhanced radiometric resolutions (e.g. 16-bit). Employing complex multi temporal data sets from different data sources (e.g. spectral and textural) is becoming a key aspect in crop mapping because it was shown to enhance classification accuracy (Balaguer et al., 2010; Berberoglu et al., 2000; Rodriguez-Galiano et al., 2012a). However, adding this amount of information comes at the expense of increasing the complexity of the data set, and can even have negative effects on the performance of classifier algorithms like SVM (Melgani and Bruzzone, 2004; Pal and Foody, 2010; Waske et al., 2010) or RF (Rodriguez-Galiano et al., 2012a).

Different kind of classifier algorithms might not perform equally well for all classes considered, or make errors in different parts of the feature space, e.g. perform complementary for different classes (Benediktsson et al., 2007; Foody et al., 2007; Waske and Benediktsson, 2007). Circumventing this issue by classifier combination was shown to overcome constraints of individual classifiers and has given promising results (Benediktsson and Kanellopoulos, 1999; Benediktsson and Swain, 1992; Benediktsson et al., 2007; Gonçalves, 2011; Licciardi et al., 2009; Liu et al., 2004; Waske and Benediktsson, 2007; Waske et al., 2010; Zhang, 2010). Yet, only few applications of classifier combination exist in multi-temporal time-series applications (Benediktsson and Kanellopoulos, 1999; Briem et al., 2002), and even fewer that dedicate this technology explicitly for multi-temporal crop mapping (Jeon and Landgrebe, 1999; Waske and Benediktsson, 2007; Waske and van der Linden, 2008). The estimates of class memberships from classifier algorithms, also referred to as the "soft output" of a classifier (Foody, 2000; Giacco et al., 2010) is a valuable source of information on the spatial distribution of classification uncertainty in maps (Foody and Atkinson, 2002) and the foundation of classifier combination (Bloch, 1996; Kittler et al., 1998). However, the so-called predictive strength of an algorithm, e.g. its capability to produce soft outputs that can be used to correctly predict the spatial distribution of classification error in maps, has attained little attention (Brown et al., 2009; Giacco et al., 2010; Loosvelt et al., 2012a). Compared with the efforts to produce accurate maps or enhance classification accuracy, considerable less attention has been drawn to the assessment of classification uncertainty. Yet, attempts to derive and use classification uncertainty measures to evaluate the spatial quality of maps were shown to be promising and an informative supplement to traditional accuracy metrics from the confusion matrix (Brown et al., 2009; Foody, 2008; Giacco et al., 2010; Loosvelt et al., 2012a), although there is further need of research, in particular (i) concerning the evaluation of the predictive strength of soft outputs from different classifier algorithms, and (ii) the influence of feature space size on spatial uncertainty in crop maps. Existing studies rather focus on "hard" classification accuracy metrics like overall accuracy (OA) or Cohen´s kappa (Cohen, 1960).

Concluding, best opportunities to handle complex multi-temporal data sets can be expected by combining the strength of advanced concepts like classifier combination and the object-based approaches. However, such approaches must be transferable to sites with different spatial characteristics and over several years, in order to be suitable for operational crop mapping. This presents some difficulties, which pose a further need for research and that drive the motivation of this thesis:

> Crops can be difficult to distinguish because of the differences in environmental conditions, spectral similarities, and different management practices at small scales (seeding and harvesting times, spatially varying complex multi-seasonal signatures). These are the main limitations that forces multi-temporal approaches. However, the high dimensionality of multi-temporal datasets requires techniques that can handle huge input spaces.

> Classifier combination is a promising and proven means, but the predictive strength of the soft outputs from different classifier algorithms, which are the input to classifier combination, has only attained little attention.

> It might be useful to add multi-temporal data of different kind as classification input (spectral, textural, geostatistical), but it is not known beforehand which kind of data is needed to classify specific types of crops, and at which time during the growing season the features are needed to effectively distinguish specific crops in a specific environment.

> The influence of input data on spatial classification uncertainty is not well understood, and specifically there are only few studies that exploited the reliability of uncertainty measures from SVM with regard to correctly predicting the spatial distribution of classification error.

> Methods developed in one place might not be portable from site to site in order to guarantee the spatial transferability of the method and comparability of the results (e.g. not robust in terms of classification accuracy). Ground truth data is often limited and methods must be developed that can cope with unbalanced and small training data sets.

1.3.2. Regional crop mapping

Whilst methods for crop mapping at the local scale (e.g. object-based approaches) suffices for understanding complex cropping patterns, smallholder environments and local applications, they will not always be suited for regional or even global agricultural monitoring, e.g. when crop specific monitoring is required (Figure 1-1). For applications at regional to global scales, there is an extra requirement of having a large swath to have a wide geographic coverage. Yet, the identification of crops in coarser EO supports might be hampered due to mixed pixels. Further, the spatial resolution requirements for an effective crop identification will vary as a function of field size and pattern, and mapping cropland will be specifically difficult in fragmented and very complex agricultural landscapes. Although high-resolution[3] instruments can provide sufficient spatial detail to resolve the spatial frequencies of fragmented agricultural landscapes, they can fail to follow the temporal development of land surface characteristics (crop growth or phenology) and to overcome the observation limitations due to cloud cover or extensive haze. It is technical reasons that cause an intrinsic trade-off between spatial and temporal resolution: the information that can be recorded, processed and stored by the instruments is limited (Schowengerdt, 2007). High spatial resolution imagery (e.g. 2.5–60 m) comes to the expense of low revisit frequency (e.g. Landsat-5 Thematic Mapper (TM): 16 days), whilst high frequent revisit

---

[3]: The convention of what is high and low resolution frequently becomes "outdated" due to continued technological advancements with instruments with higher (finer) resolutions. In this thesis instruments with resolutions lower (coarser) than 250 m (MODIS) will be considered "low", and instruments with resolutions higher (finer) than 250 m as "high".

times (1–3 days) until recently was only possible using instruments with coarse spatial resolution (e.g. MODIS: 250–1,000 m)[4]. For applications at regional to global scales, there is an extra requirement of having a large swath to have a wide geographic coverage. Up to now, a good candidate to satisfy these requirements has been Advanced Wide Field Sensor (AWiFS), which has been used to generate the Cropland Data Layer products in the U.S. (Johnson and Mueller, 2010). Undoubtedly, the new and upcoming satellite EO systems, such as RapidEye, Landsat-8 and Sentinel-2, provide new opportunities for crop mapping. Although they will not entirely satisfy those for crop growth monitoring, which requires higher temporal resolution (Duveiller et al., 2013), they should be well adapted for the purpose of operational crop mapping over a wide scale. However, they might be limited in providing such frequent observations over large geographic extents because of their small observation footprints, and financial costs for delivering such imagery will restrict their usage in developing countries. Consequently, coarser spatial resolution data such as MODIS or the Medium Resolution Imaging Spectrometer (MERIS) should not be discarded. Not only they provided added information with the higher revisit frequency, but also they will retain much importance as a source of long-term historical record, which the new systems won't achieve for decades to come. Deriving archives where crop specific time series have been identified for the past years can be very valuable for agricultural monitoring (Brown et al., 2013; Duveiller et al., 2012), or to analyse past changes in crop production (Fritsch, 2013) or land use (Conrad et al., 2011a), as it can be used to understand the past behaviour of agricultural systems and thereby infer changes in productivity or resilience against increasingly variable climatic conditions (Justice and Becker-Reshef, 2007). To extract such information (either from archive images or from new images in an operational context, given what is currently available) requires performing crop identification based on coarse spatial supports.

The necessity for a continued exploitation of coarser spatial resolution data, plus the growing interest in exploiting multi-scale data synergistically, drive the reasoning for exploring the spatial resolution requirements for the specific task of crop mapping in this thesis. Defining suitable pixel sizes for remote sensing applications has a long tradition of research (Wu and Li, 2009). Attempts to define the spatial resolution requirements can be found for land cover classification (Atkinson and Curran, 1997; Ju et al., 2005; Marceau et al., 1994a; Woodcock and Strahler, 1987), crop growth monitoring (Duveiller and Defourny, 2010), or quantitative remote sensing (McCabe and Wood, 2006; Nijland et al., 2009; Sepulcre-Cantó et al., 2010; Tarnavsky et al., 2008), but only a few explicitly address this issue in the context of CMM (McCloy and Bøcher, 2007; Ozdogan and Woodcock, 2006; Turker and Ozdarici, 2011). Furthermore, Ozdogan and Woodcock (2006) and Duveiller and Defourny (2010) illustrate how, for a given application like crop area estimation, the spatial resolution requirement (e.g. in terms of a maximum tolerable pixel size) differs considerably over different landscapes.

But what type of remote sensing data should be used as classification input for proper crop identification? In general smaller pixels are preferred to assure a good delineation of individual fields in heterogeneous agricultural landscapes with highly variable spatial crop pattern and to reduce the amount of mixed pixels, but increasing the spatial resolution may lead to oversampling, resulting in increased within-feature or class variability. Better classification accuracies may be attained using coarser pixel sizes (McCloy and Bøcher, 2007), but selecting too coarse pixels can deteriorate the quality of the classification due to mixed pixels when the heterogeneity of the land cover class in one pixel increases (Hsieh et al., 2001; Smith et al., 2003).

---

[4]: More details on sensor systems commonly used in crop mapping applicatiosn is given in appendix A.1.

One example demonstrates this: Figure 1-2 shows two datasets, high-resolution RapidEye and low-resolution MODIS. In the left-hand image, which is a subset of the right-hand image, the structures of the fields and even inter-field roads and irrigation canals can be made out. Non-agricultural land cover can be distinguished from agricultural land use, and even single buildings of a nearby village can be discerned visually. In the right-hand image the total extent of the irrigation system can be made out, and even large-scale structures like the irrigation canals or the river can be identified, but the structures of the fields and a fine distinction between cultivated and non-agricultural land is not possible. Nevertheless the literature reports comparatively high classification accuracies for crop classification in Middle Asia using both, high- and low-resolution images (Conrad et al., 2011a, 2011b). But what is the cost in terms of area estimation accuracy when using "accurate" but coarse maps to estimate the area of a given crop in fragmented landscapes like in Figure 1-2? What is the maximum tolerable pixel size for a given application, e.g. identification of a specific crop? What exactly is the loss of information when using coarser satellite images?
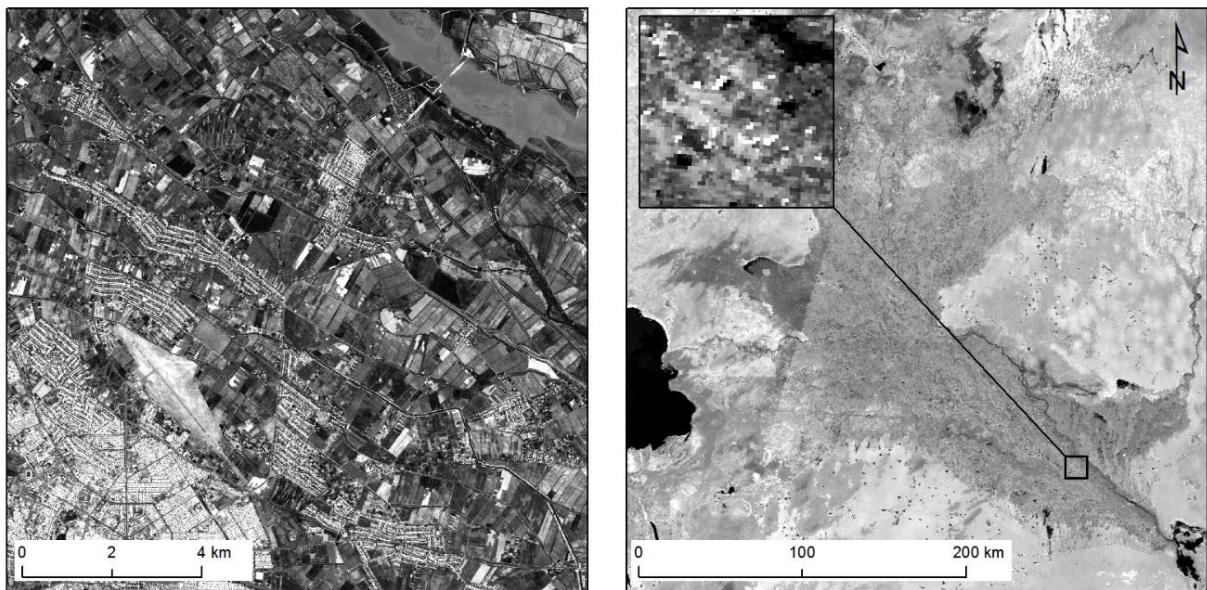
Figure 1-2: Agricultural landscape in Khorezm, Uzbekistan as seen by two different instruments: (left) RapidEye with a nominal ground sampling distance (GSD) of 6.5 m, (right) MODIS with a nominal GSD of 250 m. The small image on the right shows the same extent depicted in the left image but as seen by MODIS (NASA).

To put this into context with CMM: The choice of an appropriate spatial resolution for crop identification is not trivial and multiple challenges arise when using coarser pixel sizes, some of which are still not fully explored in remote sensing research and that further drive the motivation of this thesis:

> Mixed pixels become more abundant when pixel sizes coarsen, and pixels suitable for training a classifier algorithm become less abundant, but few studies exist to link this with landscape characteristics and quantify the degree of mixing ("pixel purity") when using coarser images, and the implications for crop classification.

> A single best pixel size for a given application does not exist. Furthermore, for a given application like crop area estimation, the spatial resolution requirement (e.g. in terms of a maximum tolerable pixel size) differs considerably over different landscapes. What is less investigated in remote sensing applications is in how far different classes (e.g. different types of crops) display individuality regarding the spatial resolution requirement over one particular landscape, and how the requirements for one specific crop type vary over different landscapes.

> There is a lack in knowledge if EO data requirements (in terms of pixel sizes) change along the growing season (e.g. as a function of the number of acquisitions), and how this influences possible early detection applications.

> Little research has been carried out to explore EO data requirements for crop mapping in agro-ecological landscapes in Middle Asia, which is a prerequisite for implementing operational EO-based crop mapping and crop specific agricultural monitoring.

1.4.   Scope and objectives of the thesis

Considering the principal research needs that were highlighted above, this thesis proposes a research framework that addresses the key issues in CMM that were detailed in the sections above, and more specifically focuses on:

> Improving crop classification at the per-field scale and providing information on the spatial distribution of classification uncertainty with the aim of creating accurate and crop specific masks. The challenge here is contributing to methods that can handle high-dimensional data sets from high-resolution time series data. Since techniques that combine different classifier algorithms seems worthwhile, this thesis focuses on concepts that combine the strengths of different classifier algorithms and adopts methods for assessing their reliability for the classifier combination process. Further, the transferability of these techniques among different sites and years, respectively, and their utility under limiting conditions in order to be suitable for potential future operational applications will be evaluated.

> Exploring the EO data requirements for an effective crop identification over heterogeneous agro-ecological landscapes. More specifically, the major challenge can be seen in the way how to define region specific EO data requirements (e.g. in terms of suitable pixel sizes) for crop identification in coarser pixels, as a prerequisite for crop specific monitoring at the regional scale which typically is done with such coarse resolution systems. Research here needs to address the technical implementation of a framework that quantified these requirements. Further, such a framework can allow for a fine-tuning of the EO data acquisition strategies and aid in designing EO-based crop mapping and crop specific agricultural monitoring systems.

Based on these principal research needs and motivations, the overarching objective of this thesis can be formulated as follows:

To develop concepts and techniques for remote sensing-based agricultural crop classification at the per-field scale and a framework to quantitatively define region specific requirements for accurate crop identification that can serve as input for crop specific monitoring.

The following specific research questions are investigated to contribute to this overall objective:

> *What is the influence of feature selection on accuracy and classification uncertainty in object-based crop classification?* In the interest of using crop maps as input to spatially explicit models, it is of interest to explore the spatial distribution of classification error (e.g. classification uncertainty) in the maps, and to explore the influence of feature input size and acquisition dates on the classification uncertainty inherit to the output. For some classifier algorithms like SVM there still exist no study to assess the influence of feature selection on spatial uncertainty in crop classification. The potential usefulness and interpretation of uncertainty measures from SVM is provided, and a contribution to a better understanding of the relationship between uncertainty measures derived from the SVM and the hard result accuracies (from the confusion matrix) is envisaged.

> *How do different classifier algorithms perform in crop classification using multi-spectral time series data at the object-level, and can combining classifier algorithms improve the overall quality of crop masks?* Taking advantage of combining the strength of different classifier algorithms can be reasoned with their different performance in classifying certain crops. In particular when there is high disagreement among the single classifier algorithms, this could lead to a loss of valuable information. The membership estimations (soft output) from different classifier algorithms, which is the input to classifier combination, might not be equally reliable, but this should be considered in the classifier combination process. In this regard classifier combination should have the ability to select the most reliable classifier algorithms for each class investigated. To be suitable for application in crop mapping, such an approach must be transferable among the sites, consistently result in high classification accuracies, be transferrable to other years, and applicable under limiting conditions (e.g. when only few training data is available).

> *What is the spatial resolution requirement for crop identification via image classification, in particular minimum and coarsest acceptable pixel sizes, and how do these requirements change over different landscapes?* Although the answer to this question seems to be simple and just a relation between field size and pixel size, seeking the appropriate pixel size is far from being a trivial task. From a practical point of view it is of interest to find the maximum tolerable pixel size for a given application since coarser images potentially bare attributes that makes them interesting for operational crop specific monitoring (spatial coverage, revisit time, and low financial costs). A framework will be proposed that quantifies, based on user defined constraints, the region specific EO data requirements for crop identification.

1.5. Structure of the thesis

Chapters 2 and 3 summarize information on the study sites, satellite data and the main pre-processing steps involved. Each methodological chapter of this thesis, chapters 4–6 (Figure 1-3), is introduced by providing the state-of-the-art in the corresponding field of research, stating its research goals, and by outlining its relationships with the other chapters as well as the research questions stated before in the introductory chapter. This thesis is structured as follows:

> **Chapter 2** gives a brief description of the four study sites selected in this thesis. The final choice of the study sites was reasoned with geostatistical indicators from the semivariogram, as well as the diversity of crops, management practices, field pattern, and access to the study site and satellite data availability.

> **Chapter 3** summarizes the pre-processing of the satellite data. Relevant data sources (e.g. remote sensing imagery, ground reference data, cropping calendars) are listed. A brief description of the atmospheric and geometric correction of the satellite data is given. The image segmentation procedure for the generation of the field masks is detailed.

> **Chapter 4** explores the influence of feature selection on classification uncertainty and accuracy. This chapter directly answers the first research question and addresses the problem of feature selection and its influence on spatial classification uncertainty in supervised image classification using SVM. An intentionally huge input set of multi-seasonal spectral and geostatistical features is calculated to analyse the relative contribution of different feature types to crop classification and to find optimal feature compositions to distinguish crop classes. The potential value of spatial classification uncertainty measures derived from SVM is evaluated,

and the causes for the spatial distribution of uncertainty in maps, e.g. due to environmental factors, are discussed.

> **Chapter 5** responds directly to the second research question by comparing different classifier algorithms for crop classification at the field level and by addressing the issue of classifier combination and evaluating the transferability of methods among the test sites. In this regard, the outputs of classifier algorithms are combined at the decision level to take advantage of several classifier architectures and to improve classification accuracy. This chapter provides an extension of known algebraic operators by taking into consideration the reliability of different classifier algorithms for each class investigated. The impact of classifier combination on classification uncertainty is analysed, and the applicability of the investigated classifier algorithms is evaluated under limiting conditions (e.g. training data and input data scarcity).

> **Chapter 6** is devoted to the issue of crop mapping at the regional scale by analysing the effect of pixel size on crop identification. A framework is proposed to define the EO data requirements for an effective crop identification. This will be done by simulating how agricultural landscapes, and more specifically the fields covered by a crop of interest, are seen by instruments with increasingly coarser resolving power, based on a generic point spread function that is scaled to coarser pixel sizes. The concept of crop specific pixel purity is used to analyse how mixed the pixels can be (as they become coarser), without undermining their capacity to describe the desired surface properties. The framework lays the technical foundation of a processing chain to retrieve site-specific EO data requirements for crop mapping at the regional scale.

The main findings and synthesis of the results are summarized and discussed in the conclusion chapter at the end of this thesis. The discussions and concluding remarks will point on the capabilities and pathways for future research tasks in the context EO-based crop mapping and agricultural monitoring.
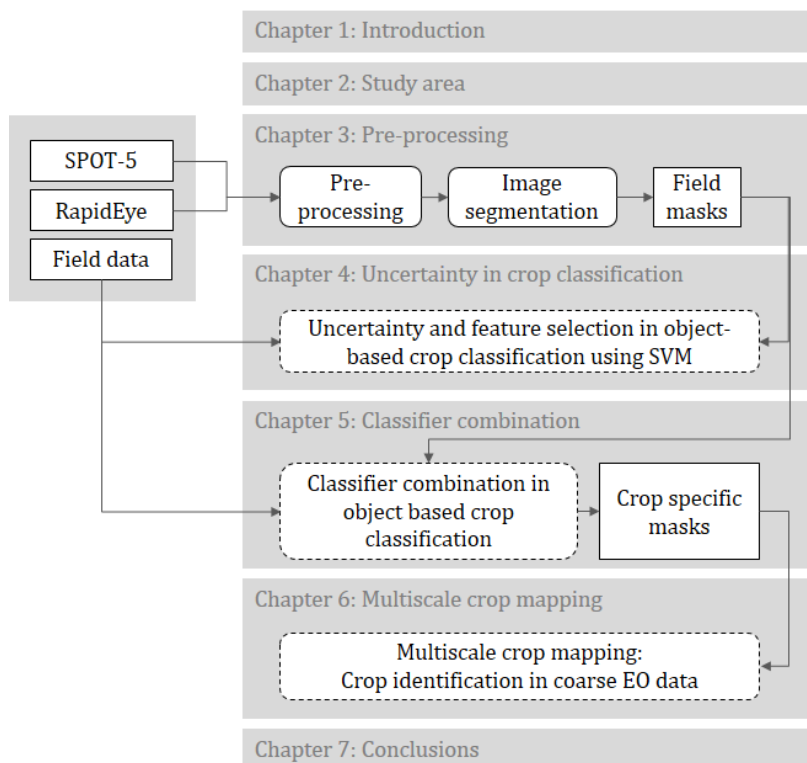
Figure 1-3: Overview of the methodological components of this study and their interactions (arrows). The main methodological approaches, corresponding to the objectives of this thesis, are shown in white boxes with dashed contours.

Chapter 2

DESCRIPTION OF THE STUDY AREA

Abstract

Middle Asia exhibits vast irrigated systems whose origin can be traced back over several millennia. Because of the arid climate agriculture is restricted to irrigation that is practiced along the two major rivers Amu Darya and Syr Darya. Today this region is known for mono cropping of cotton, wheat, and partly rice, which are cultivated on vast agricultural systems that were excessively expanded during the Soviet era since the 1960s. After independence, parts of the irrigation system have undergone dramatic changes, caused by mismanagement, water scarcity, and soil salinization, resulting in widespread land abandonment and degradation processes. In this thesis four sites of 30×30 km were selected from different parts of this irrigation system in order to perform the experiments. The choice of the particular sites was oriented toward fragmented landscapes that represent (i) different spatial patterns of agricultural production systems in Middle Asia, composed by varying field sizes and spatial configurations of fields, (ii) both lowland and mountainous regions, and (iii) different management practices (crop types, levels of mechanization, different and partly state-imposed crop rotation systems). Geostatistics were used to quantify the landscapes characteristics, e.g. the spatial field pattern, and to test if the subsets that were chosen are large enough to characterize the spatial structures of the corresponding landscapes they cover.

2.1. Agriculture in Middle Asia – History and recent challenges

Middle Asia belongs to the source areas of recent crops (e.g. melons, apples) (Nentwig, 2005) and is one of the world's core regions of irrigation. It is a primarily arid region in the centre of the Eurasian continent, characterized by diverse physiographic conditions ranging from semi-arid and arid desert areas to steppe and mountainous regions. Because of the aridity, agriculture in the catchment is dependent on irrigation, so it is concentrated within a dense irrigation network that stretches up to several kilometres on both sites of the two major rivers in Middle Asia, the Amu Darya and Syr Darya (in ancient times called Oxus and Jaxartes, respectively). Both rivers drain into the Aral Sea (Figure 2-1). The drainage basin of the Aral Sea is estimated approximately 1.8 million km² (Micklin, 2007).

Figure 2-1: Main irrigated areas in Middle Asia. Red rectangles indicate the footprints (each 30×30 km) of the satellite images in the four study sites that were selected in this thesis.

Today Middle Asia is known as a region of mono cropping for cotton and wheat (Lal et al., 2007). It exhibits vast agricultural systems (Figure 2-1), whose origin can be traced back over several millennia (Boroffka et al., 2006; Oberhänsli et al., 2007). However, both the recent extent and configuration of the irrigation infrastructure stem from its extension that began in the 1930s and its excessive development during the Soviet era between the 1960s and 1980s (Saiko and Zonn, 2000). The "Virgin Lands" program of Nikita Khrushchev in the 1960s resulted in the transformation of 36 million ha steppe land into cropland (Douglas, 1962). Expansion of irrigated agriculture (especially due to cotton production) in the second half of the 20th century under the aegis of the Soviet leaders was able to enhance the employment opportunities and increase income (Qadir et al., 2009), and made irrigated agriculture in Middle Asia a mainstay for their national economies.

The total area of irrigated land in the Aral Sea basin increased from 5.4 million ha (1950) to around 8 million ha in the 1990s (Saiko and Zonn, 2000). Today Middle Asia exhibits 9.1 million ha of irrigated land (FAO, 2013b). Until the second half of the 20th century agricultural extension had only a limited impact on the environment (Micklin, 1988), but the 70 % increase in irrigated area after the 1970s was not without negative environmental implications (Saiko and Zonn, 2000). Unsustainable withdrawal of water from the two major rivers and inefficient water use led to a drastically decrease of the inflow of

the rivers into the Aral Sea. Inflow was reduced from 55 km³ before 1950 to less than 10 km³ between 2001–2005, and the Aral lost more than 90 % of its pre-1960 volume (Micklin, 2007). The exposed seabed (called Aralkum) became the major source of salt and dust storms in Middle Asia, and complex land cover change has further led to widespread environmental implications and water quality deterioration in the adjacent regions (Glantz, 2009; Kotte et al., 2012; Löw et al., 2013b, 2012; Micklin, 2007, 2010).

Today, agriculture in Middle Asia faces a variety of challenges, which are briefly described in the following[5]. Covering more than 4 million km² the five ex-Soviet countries Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan today are home to more than 55 million people (FAO, 2013b), from which a large fraction (22 million) depends on irrigated agriculture for their livelihoods (Qadir et al., 2009). A major part of the economic output of Uzbekistan, Tajikistan, and Kyrgyzstan is derived from (predominantly irrigated) agriculture (Bucknall et al., 2003). Resource sharing in Middle Asia is complicated due to several factors: dependence on shared but transboundary resources, redrawn national borders (that do not fully take into account cultural legacies), proceeding infrastructure deterioration (that was skewed to cotton monoculture production), and the loss of the centrally managed Soviet economy and compensation mechanisms. The three downstream countries (Uzbekistan, Turkmenistan, and Kazakhstan) are the main freshwater consumers in the Aral Sea basin and use water in the summer season for irrigation. They suffer land degradation and water quality deterioration (Qadir et al., 2009). Some downstream parts of the irrigation system suffer continuously decreasing agricultural production because of deteriorating water supply (JICA et al., 2010). On the upstream site, water resources are mainly generated in the energy-poor yet water-rich mountainous regions in Tajikistan and Kyrgyzstan (Sapper et al., 2007), which use water for hydropower production in the winter season. Hostile reactions from downstream countries on planned or existing dam projects in the upstream countries (e.g. the Rogun dam in Tajikistan) underline this conflict-laden situation. As reminded by (Lioubimtseva and Henebry, 2009) the core regional problem in Middle Asia is not a general lack of water resources, but the way they are managed and distributed: it is rather an issue of ineffective water management systems and lacking coordination among the irrigation systems. For instance it is estimated that Middle Asia loses 1.7 billion U.S. dollar per year caused by poor water management that lowers agricultural yields (UNDP, 2005).

Meanwhile Middle Asia experienced significant increases in population size (with average annual growth rates of 1–2 % in all Middle Asia countries except of Kazakhstan), which puts additional pressure on food production and water resources for irrigation (Asian Development Bank, 2012; Cai et al., 2003; UNDP, 2005). Following the disintegration of the Soviet Union in 1991 agricultural yields declined in Middle Asia by 20–30 % in the last decade, with a loss of agricultural production from salinization alone estimated at 2 billion U.S dollar (UNDP, 2008).

On top of this, climate change will affect future water availability and food production in Middle Asia: precipitation will most probably slightly decrease by 2050 with a daily rate of approximately 1mm/day (Lioubimtseva and Henebry, 2009). Temperature will most probably increase by 3–5 °C by 2080, especially in summer and fall, which will be accompanied with a further increase in aridity and increasing water demand for irrigation (Lioubimtseva and Henebry, 2009), see Figure 2-2. Climate change will affect the agricultural production by shifting the spatial and temporal pattern of irrigation water availability. Melting of the glaciers in the mountainous areas, e.g. in Pamir and Tian-Shan is projected to increase (Glantz, 2005; Lioubimtseva and Henebry, 2009). Since the freshwater resources in Middle Asia are mainly generated in the mountain glaciers, the melting of the ice and snow packs will initially translate into additional runoffs from the Amu Darya and Syr Darya for a few decades.

---

[5]: The interested reader is referred to Bucknall et al. (2003), Glantz (2005), Harris (2010), Létolle and Mainguet (1996), Martius et al. (2012), and Sapper et al. (2007) for more detailed information on this subject.

Ultimately this will be followed by a reduction of the flow after the glaciers have disappeared (Glantz, 2005), and the loss of glacier volume, depending on their location, is predicted to amount 32−73 % until 2050 (Giese and Sehring, 2007). Critically there remain uncertainties in the scenarios for predicting the future climate trends (Mannig et al., 2013; Siegfried et al., 2012). In this regard, climate change can be seen as an additional (potential) threat to the existing vulnerability of the irrigation system in Middle Asia by reducing the region´s future water supply.
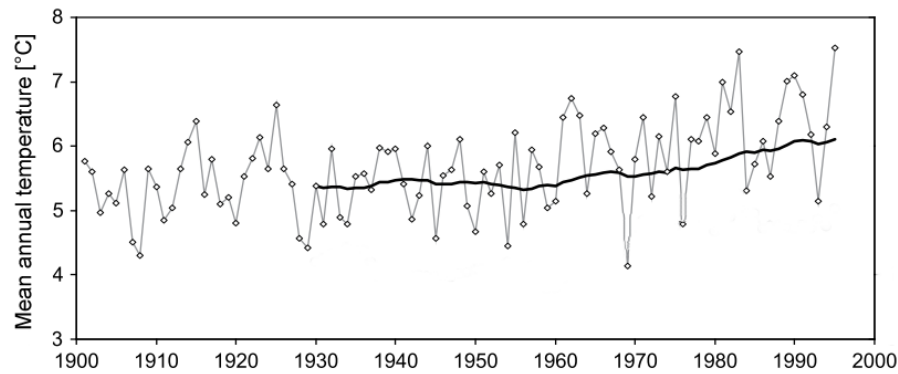
Figure 2-2: Change in the mean annual temperatures over Central Asia during the 20th Century, adapted from Lioubimtseva and Henebry (2009).

2.2. Test site description

Four distinct sites of 30×30 km within the irrigated land of Middle Asia were selected for implementation of the methodology of this thesis (Figure 2-1)[6]. Three sites belong to Uzbekistan: Khorezm (KHO), Karakalpakstan (KKP), and Fergana Valley (FER). One site belongs to Kazakhstan: Kyzyl Orda (KYZ). The choice of the particular study sites was oriented toward fragmented landscapes that represent (i) different spatial patterns of agricultural production systems in Middle Asia, composed by varying field sizes and spatial configurations of fields, (ii) both lowland and mountainous regions, and (iii) different management practices (crop types, levels of mechanization, different and partly state-imposed crop rotation systems). In this regard, accessibility, permissions, and satellite data availability also influenced the choice of these four regions. In the following a brief description of the four selected test sites is given, thereby focusing on major physio-geographical and agricultural characteristics (climate, soils and vegetation). Table 2-1 summarizes the main characteristics of the test sites.

The first site (Figure 2-1) is located in the Khorezm region (KHO) in the north-western part of Uzbekistan, approximately 120 a.s.l. It is located between the deserts Kara-Kum and Kyzyl-Kum, approximately 250 km south from the former shoreline of the Aral Sea in Uzbekistan. Climate is arid with 119 mm precipitation per year (Table 2-1) that mostly falls in winter, and monthly evaporation exceeds precipitation. Natural vegetation cover in the test site is sparse. Near the Amu Darya remnants of the formerly wide spread Tugai vegetation can be found, and reeds (e.g. *Phragmites australis*), cattail (*Typha angustiflora*) and rushes can grow where ground water levels are sufficiently high. Soils in KHO are classified as Calcaric Gleysols and partly Calcaric Fluvisols, according to the FAO-UNESCO Soil Map of the World[7]. The strong influence of groundwater, caused by inefficient irrigation management, and the proximity to the Amu Darya, lead to high salt dynamics and secondary soil salinization, and severe degradation of the irrigated land is reported to occur in Khorezm (Ibrakhimov et al., 2007). Cotton (e.g.

---

[6] : All test sites are focus regions of the German-Uzbek Khorezm-Project, the CAWA-Project, and the GIZ EEWA-Project.

[7]: Source: http://www.fao.org/nr/land/soils/digital-soil-map-of-the-world/en/ (last accessed 03-Feb 2013). The soil type was downloaded as ESRI shapefile and overlain with the study sites to estimate the fractional soil type cover within the four sites.

Gossipum hirsutum), winter wheat (*Triticum aestivum*), and rice (e.g. *Oryza sativa*) fields dominate the agricultural landscape (Ibrakhimov et al., 2007). Beside these, fruit trees, sorghum, and maize are cultivated. Small garden and household plots, so-called *Tamorkas*, can be found but are excluded from the later analysis. A dense network of irrigation and drainage canals demark the fields (Conrad et al., 2007). The agricultural landscape appears fragmented due to a comparatively high diversity of crops, although partly crops are grown on adjacent, regular shaped fields that together exceed up to 25 ha (Conrad et al., 2011a). Among the four test sites KHO exhibits the second largest fields (mean field size is 4.31 ha)[8] and the cover fraction (C_f) defined as the share of agricultural fields in the landscape, is the largest among all sites (0.59).

Table 2-1: Characteristics of the four study sites. Max D_c is the maximum of mean length scale of the normalized vegetation index (NDVI) by Rouse et al. (1974) along the season in [m][9]. Total number of fields, field sizes, D_c, and C_f are based on own calculations based on segmented image objects, see section 3.4.

Study site	Scene center [Lat/Lon]	Annual means of temperature [°C] and precipitation [mm]	Total number of fields	Mean / standard deviation field size [ha]	Median field size [ha]	C_f	Max D_c
KHO	60°69E 41°53N	13.1 / 119 (Meteo Station: Urgench)[10]	22,247	4.31 ± 2.07	3.21	0.59	2268
KKP	59°33E 42°42´N	10.4 / 106 (Meteo Station: Cimbaj)	21,205	2.19 ± 1.86	1.71	0.32	2811
KYZ	64°55E 44°58N	9.8 / 149 (Meteo Station: Kyzyl Orda)	14,561	2.45 ± 1.62	2.14	0.25	1719
FER	71°45E 40°32N	13.7 / 177 (Meteo Station: Fergana)	12,670	6.74 ± 2.25	5.47	0.57	2405

The second site is situated in the autonomous republic of Karakalpakstan (KKP), in the north-western part of Uzbekistan (Figure 2-1). It is located in the lowest reaches of the Amu Darya, approximately 80 km south to the former Aral Sea shoreline, approximately 60 m a.s.l. Climate is arid, with 106 mm annual precipitation (Table 2-1) with peaks in early spring and winter, out of the main vegetation period. Major soil types found in this site are Eutric and Calcaric Gleysols. The natural phanerophyte vegetation in KKP comprises open and dense shrublands (e.g. *Tamarix sp., Haloxylon persicum* and *Haloxylon aphyllum, Calligonum sp.*), and reed areas (e.g. *Phragmites australis*). The only surface inflow of water into KKP, as is in KHO, is from the Amu Darya. Discharge varies significantly from year to year in the downstream site, introducing high uncertainties in water availability and causing severe drought, e.g. in 2000 and 2001 (JICA et al., 2010). The last decades have witnessed dramatic changes of land use, resulting in widespread land abandonment caused by declining water availability, secondary soil salinization, and proceeding desertification. In 2011 more than 50 % of the agricultural fields in the test site were temporally or permanently fallow (see section 4.3.6), or characterized by shrub encroachment and other land degradation related phenomena (Beyer, 2012). Major crops in KKP are cotton, winter wheat, and rice. Other crops like maize, sorghum, watermelons, fruit trees, and alfalfa are cultivated. Identical crop types partly are grown on 2−3 adjacent fields, but the crop pattern in KKP is heterogeneous. Although crop rotation practices were introduced in KKP

---

[8]: Mean and median field sizes were calculated in ArcGIS, see chapter 3.

[9]: Garrigues et al. (2006a) used the mean length scale to quantify the spatial heterogeneity in an agricultural landscape. It is defined as the square root of the SV integral range. Here the maximum value of the (dimensionless) mean length scale in the season is taken to quantify if the scene is large enough to characterize the spatial structure in the scene, as proposed by Garrigues et al. (2006a).

[10]: Meteorological data was taken from climate diagrams, downloaded from www.klimadiagramme.de (last accessed: 20-Jul 2013), data from station "Urgench" is provided by the ClimWat data base of FAO (Munoz and Grieser, 2006).

already in the Soviet era they went out of practice, due to limited supply of water, lack of seeds and knowledge, or replacement of fodder crops with wheat to increase food production (JICA et al., 2010). Regular shaped fields dominate the KKP landscape in the southwest, whilst in the north-eastern direction the landscape becomes increasingly more fragmented with smaller and more irregular shaped fields.

The third site (Figure 2-1) is located in the Kyzyl Orda oblast[11] (KYZ) in southern Kazakhstan, and was chosen to have an example with more regular shaped field structures. The site is located 115 m a.s.l in the flat lowland region at the lower reach of the Syr Darya, approximately 350 km east to the former Aral Sea shoreline. Climate is arid and continental, with 149 mm annual precipitation (Table 2-1) that mainly falls in the spring and late autumn season. Eutric Fluvisols dominate the soil cover, partly interspersed with Eutric Gleysols and Eutric Histosols. Next to natural riverine vegetation near the Syr Darya open shrublands and patches of reed areas dominate the landscape in KYZ. Only two crops are dominating the agricultural landscape: rice and alfalfa. Beneath them the Kyzyl Orda oblast is known for its melon production, but melons are absent in the selected test site. Nowadays a crop rotation system is applied in KYZ, where rice and alfalfa are combined in a distinct pattern: after two years of rice cultivation the fields are left fallow and legume crops, in particular alfalfa, is grown for up to three consecutive years. Then other crops (wheat, vegetables) are grown for two other consecutive years before rice is cultivated again. This rotation pattern is mandatory in Kyzyl Orda oblast since 2009 and is intended for soil fertility conservation and regeneration, although still not applied region wide (personal communication Dr. Ulikbanuli Nurlibai, vice-director of the Rice Institute in Kyzyl Orda). Large and regular shaped agricultural fields of approximately 2–3 ha each characterize the landscape, e.g. the same crop is grown on adjacent fields, which together exceed the area of between 500×500 m and 1,000×1,000 m (25–100 ha). The fields that are aggregated to blocks together make up a relatively small proportion of the total test site area (C_f = 0.25). From the four test sites KYZ is the most regular in terms of the field structure (size, shape, arrangement) and cropping pattern (e.g. high spatial aggregation of few crop types). The remaining areas between these blocks are filled with sparse shrublands and, to a lesser extent, reed areas.

The fourth site (Figure 2-1) is located in the Fergana Valley (FER), in the eastern part of Uzbekistan. It is situated in the intermountain basin between the foothills of the Tian-Shan Mountain in the north, and the Gissar-Allay in the south. It is the only mountainous site and situated in the upstream region of the Syr Darya. The central part of the FER test site is flat and approximately 450 m a.s.l. Rivers deposit sediments (deposits of gravel) thereby forming deltas in the valley. In the FER site Calcaric Gleysols cover a major part, partly interspersed with Gleyic Solonchaks and Calcic Xerosols. Climate in FER is arid continental, with an average annual temperature of 13.7 °C and 177 mm annual precipitation (Table 2-1). However up to 60–70 % of the precipitation falls outside the main vegetation period (Hakimov et al., 2007). Fergana Valley is one of the most fertile agricultural regions in Middle Asia, with a strong focus on cotton, winter wheat, and fruit trees. Comparatively large and regular shaped fields, and a variety of crops including cotton, rice, maize, watermelon, sunflowers, alfalfa, winter wheat, and fruit trees, characterize agricultural production in Fergana Valley. Like in KKP and KHO, the production of cotton (*Gossypium hirsutum L.*) and the staple winter wheat (*Triticum aestivum L.*) is under "state-order", hence the annual cultivation area for cotton is state determined. Rice in Fergana Valley is mainly cropped for private demands and income. FER and KHO have similar landscape characteristics, but different cropping pattern. The mean field sizes in FER and KHO are largest among the four sites (6.74 and 4.31 ha, respectively). C_f in FER is 0.57.

---

[11]: The term "oblast" (Russian) describes an administrative territorial division in countries of the former Soviet Union. The next subordinate administrative unit is the „Rayon".

Figure 2-3 shows the crop calendar for the major crops. This calendar details the average information of regular crop stages in the four test sites, although the exact timing of planting, sowing, and harvesting can differ due to factors related to water availability, management practices, or farmer's decisions. In all sites the main part of the growing period is within the low precipitation period. In general, the growing season is between April and October. Onset of green (start of season) is earlier within the year in FER than in the three downstream sites (Bohovic et al., 2011).
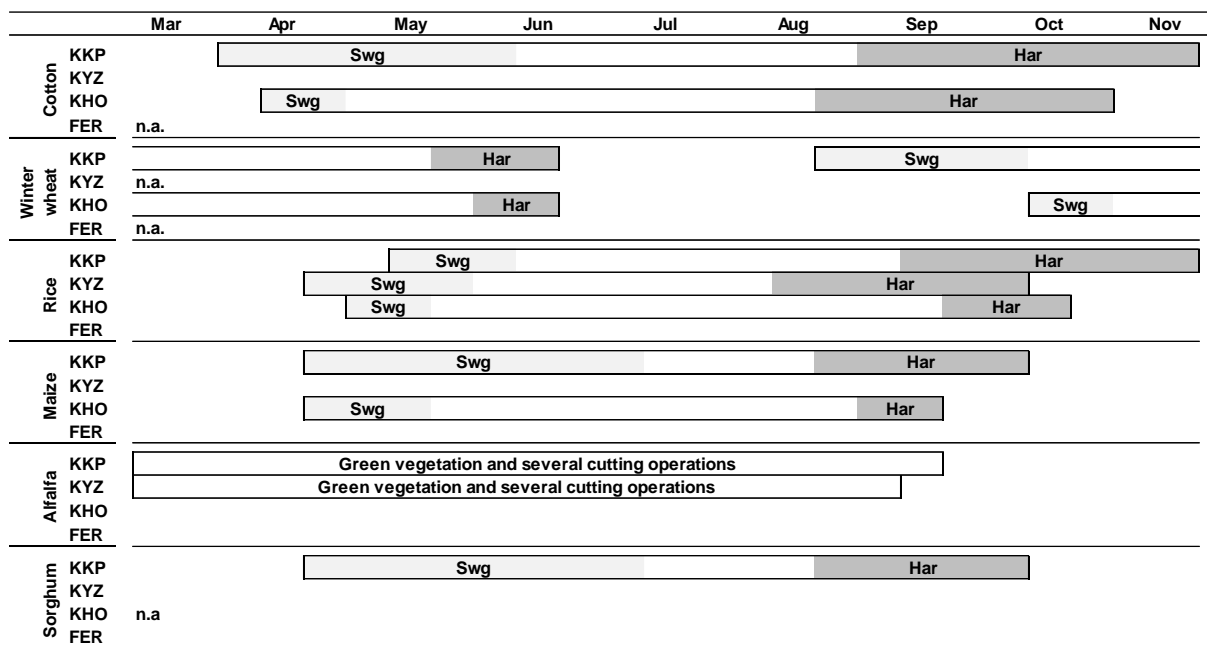
Figure 2-3: Cropping calendar of major crops in the four test sites. Note: "n.a." means that no data is available, but the crop is present in the test site selected for this study. Swg = sowing time, Har = Harvest time. Source: GIZ, German-Uzbek Khorezm-Project, CAWA project, Rice Institute in Kyzyl Orda (Kazakhstan), and JICA et al. (2010).

In the following a short description of the main characteristics of the crops found in the test sites is given:

> Alfalfa, which is used as fodder, is characterized by several cutting operations within the growing period, which results in heterogeneous pattern and indistinct NDVI profiles (Conrad et al., 2013), with high spatial and temporal variability of its dense canopy and crown structure within a year.

> Winter wheat is sown from end of August and beginning of October, and harvested between end of May and beginning of June. In KHO, KKP, and FER a second crop is sometimes grown on the same field after winter wheat was harvested (intra-annual crop rotation), for instance the second crop can be sorghum, maize, or rice.

> Cotton is grown in KHO and KKP and seeded in April and April–May, respectively. Harvest of cotton begins in end of August and beginning of September.

> Rice is grown in all test sites expect for FER, and is seeded from April–March. If rice is cultivated in a sequence with winter wheat in one year in the same field (crop rotation), then seeding time of rice is a couple of weeks later than on rice fields without rotation (Conrad, 2006).

> Maize and sorghum are sown in April, and harvest is between August–September. They are partly cultivated on the same field.

> Fallow fields can be found in all four sites. In the KYZ site fields are intentionally left fallow for 2–3 years after rice cultivation, for soil regeneration purposes. In the Uzbek sites however this can also be a result of temporal water shortage or land abandonment, especially in KKP and KHO.

The four test sites are subsets of larger agricultural landscapes, which means that more than the aforementioned crops can be found in the corresponding region, e.g. melons are cultivated in Kyzyl Orda, but absent in the test site KYZ in 2011. Also smaller horticultural plots or private gardens (e.g. Tamorkas) provide a wide range of agricultural and horticultural products: vegetables and fruits that were not classified in this thesis because of limited access and the relatively small size and amount of the corresponding field plots, respectively (e.g. potatoes, tomatoes, beans, sunflower, carrot, grapes, and others).

2.3. Geostatistical characterization of the test sites

Geostatistical experiments[12] were conducted to answer the following questions concerning the landscape structure in the test sites: (i) what are the differences in spatial structures (agricultural fields) across the studied landscapes with respect to agricultural fields? (ii) Are the selected image subsets large enough to fully characterize the spatial surface heterogeneity (i.e. variance of the remote sensing image)? Answering both questions is of particular importance for an understanding of the impact of landscape characteristics on crop monitoring with coarser pixel sizes, which will be recalled in chapter 6. Details on the characteristics and pre-processing of the satellite images, which are used to perform the geostatistical experiments, are given in chapter 3 of this thesis.

2.3.1. What are the differences in spatial structures across the studied landscapes?

One approach to answer this question is using geostatistical indicators to quantify the landscape heterogeneity. Applications of geostatistics to quantify landscape heterogeneity and to characterize the spatial variability of vegetation amount over a scene can be found in Duveiller and Defourny (2010), Garrigues et al. (2006a), and Ozdogan and Woodcock (2006), and for characterizing the temporal evolution of landscape heterogeneity along the season in Garrigues et al. (2008a). Semivariograms (SV) were shown to provide comprehensive characterization of image spatial structures (Curran, 1988; Garrigues et al., 2006a). To quantify the heterogeneity of the test sites SVs were calculated for the field masks that were generated of each study site (see section 3.4). These masks take on the value of "1" if the location is cultivated, and "0" otherwise (Figure 2-4). A classical SV estimator was applied to all image pixels in these binary masks:

$$\gamma(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[z(x_i) - z(x_i * h)]^2 \qquad\qquad Eq.\ 2\text{-}1$$

where $z(x_i)$ represents the value of the variable at the location x_i, h the separation (the so-called lag, given in the measurement unit of the analysis, e.g. in km) between elements in a given direction, and $N(h)$ the number of data pairs occurring at locations x_i and $x_i * h$. An experimental SV is obtained by computing the mean variance of the variable under consideration obtained from pairs of points that are separated by an increasing distance, the so-called lag. The $z(x_i)$ in the formula above is used here to

---

[12]: Geostatistics were calculated in the software package gstat (Pebesma and Graeler, 2013) in the freely available software environment R (R Development Core Team, 2012).

represent an indicator function, which takes on the value of "1" if the location x that is cultivated is present and "0" otherwise (Collins and Woodcock, 1999; Ozdogan and Woodcock, 2006). Following this idea the area of cultivated lands is defined as all points for which $z(x_i) = 1$. In this study omni-directional SV were calculated, which means that they were computed by pooling together all directions.

The SV describes the degree of spatial dependence of a regionalized variable. The two most commonly used parameters from the SV are the sill ($\hat{\gamma}_r$) and the SV range. The former is the overall variance of the image and is defined as the maximum level of the SV. The latter is the distance to $\hat{\gamma}_r$ (e.g. the lag distance in km at which the SV reaches its maximum value, the sill) (Curran, 1988). The sill parameters is related to the proportion of agricultural fields that covers an image, and the range to the size and shape of the fields (Ozdogan and Woodcock, 2006). Assuming that the variable has a second-order stationarity, the SV will reach its maximum (e.g. a "stable" value, the sill $\hat{\gamma}_r$) at a given lag (the range). $\hat{\gamma}_r$ is equal to the variance of the variable over the entire scene. These two parameters were used to quantify the landscape heterogeneity (Figure 2-4).

Some of the differences of the test sites already become evident from their SV (Figure 2-4). In KHO the rise of the SV is very steep at comparatively small lags, and reaches $\hat{\gamma}_r$ at the highest value of all sites. KHO has the highest C_f value (0.59) and the highest standard deviation of field sizes (±2.07 ha) that almost exceed the mean field size. This indicates a large amount of small scale and irregular structures (fields), despite the large mean field size. The SV reaches $\hat{\gamma}_r$ at 9–10 km. The SV in KKP rises in a similar fashion like in KHO, but stabilizes in a shorter range of approximately 5–6 km. This indicates, similar to KHO, the presence of smaller-scale structures and more irregular shaped fields. In KYZ the rise of the SV is the lowest, and the overall variance is low as compared with the other sites. This indicates the presence of large and regular structures (e.g. the regular and aggregated field complexes that are clearly spatially separated from each other). The semivariance $\hat{\gamma}$ decreases after it reaches $\hat{\gamma}_r$ in approximately 5.2 km distance, which is called the *hole-effect semivariogram* (Pyrcz and Deutsch, 2003). This indicates that the studied variable has a periodic spatial behaviour. Critically this *hole-effect* is not completely registered by the corresponding SV graph due to the fact that the analysis radius distance used to compute the SV is not sufficient to capture the complete periodic spatial pattern defined by the larger distance of the field blocks. The SV of the FER test site looks similar to the SV of the KHO site, and it also reaches $\hat{\gamma}_r$ in the range of approximately 9 km, but rises more slowly. FER exhibits large objects (6.74 ha on average) that are arranged in a much more regular pattern than in KHO. In FER the SV reaches $\hat{\gamma}_r$ in comparatively high ranges. This indicates the presence of large objects (fields).
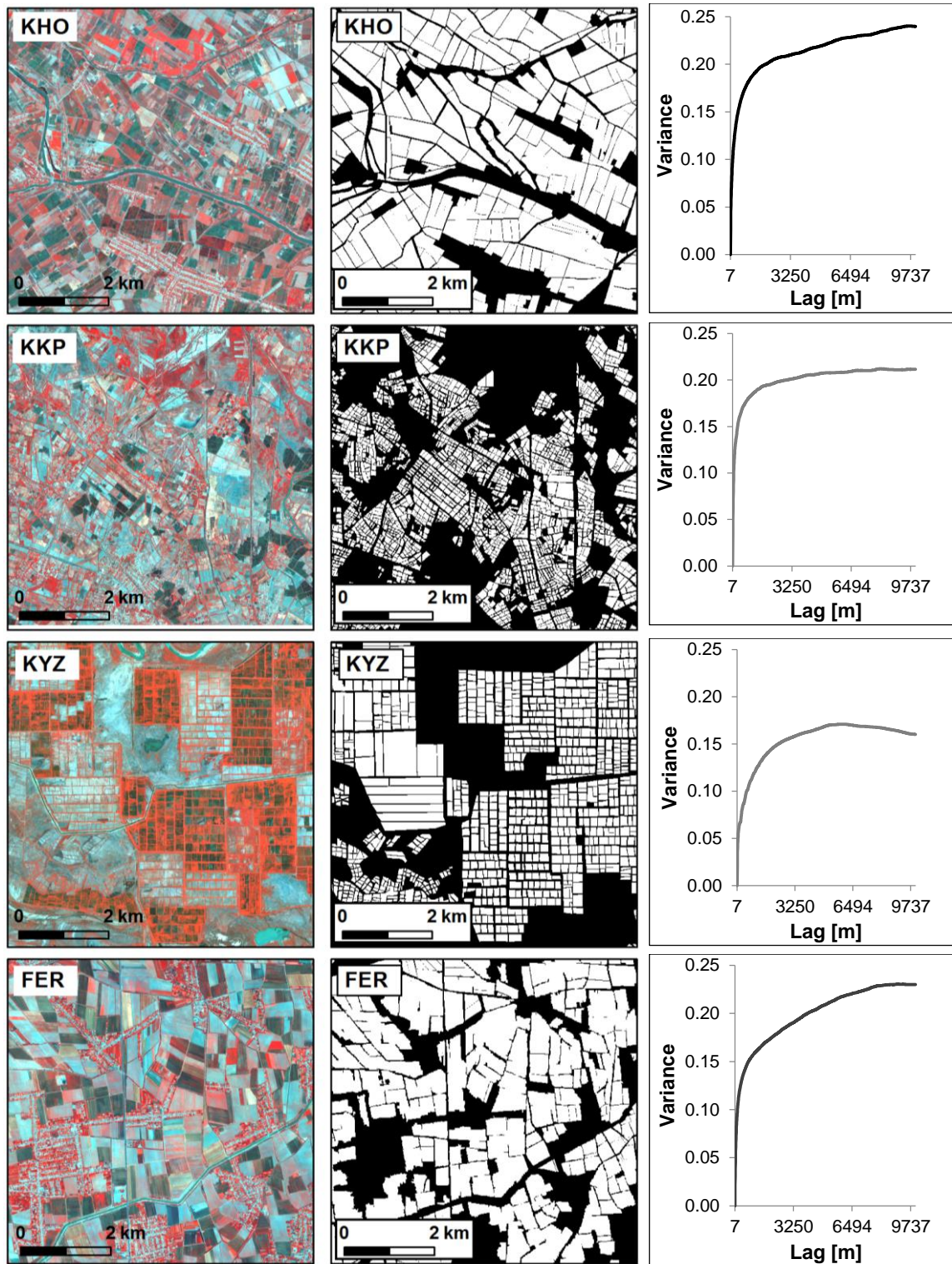
Figure 2-4: Subsets (6.5×6.5 km) of the satellite imagery and crop masks illustrating the typical cropping patterns within the four test sites. The imagery is displayed using a near-infrared-green-blue band combination of the RapidEye sensor. Contrast is adjusted to each band separately. The corresponding experimental semivariograms are shown right to the binary masks.

2.3.2. Are the image subsets large enough to characterize the spatial surface heterogeneity?

Having characterized the general landscape structure of the four tests sites with help of SV that were calculated on the field masks, the second question is answered by calculating experimental SV for the NDVI from the RapidEye image stacks (see section 3.1 for more information on the RapidEye images) over each site and eight acquisition dates, respectively.

An exponential model was fitted over each of the eight experimental SV curves (Figure 2-5). None of the SV exhibit nugget effects, which is an indication that the spatial variations at small scales (Curran, 1988), e.g. smaller than the GSD of the RapidEye sensor (6.5 m) in the test sites, are small relative to environmental variations, and can be neglected (Garrigues et al., 2008b). All SV show a quick rise and a stabilization of the semivariances within 10 km distance. From the modelled SV the sills $\hat{\gamma}_r$ (overall spatial variability of the image) were calculated (Figure 2-5). In all study sites the SVs reach their $\hat{\gamma}_r$ below a distance of 30 km (corresponding to the test site size), which indicates that the images can be considered large enough to characterize the spatial structure within the landscape they cover (Guissard et al., 2004). Interestingly the pattern of the SV in KYZ becomes periodic between June–August, which reflects the repetitive pattern of blocks of fields in that site (Figure 2-4). The SV reach $\hat{\gamma}_r$ at larger lags, showing some further spatial dependency up to 5–6 km after which $\hat{\gamma}_r$ is reached. This larger range could indicate the approximate scale in which growing conditions (e.g. management practices, water availability, soil properties) are similar. The SV of the field masks in FER and KHO exhibit a similar behaviour, indicating a comparable spatial field structure, but the trajectories of their NDVI sills differ distinctly (Figure 2-5).
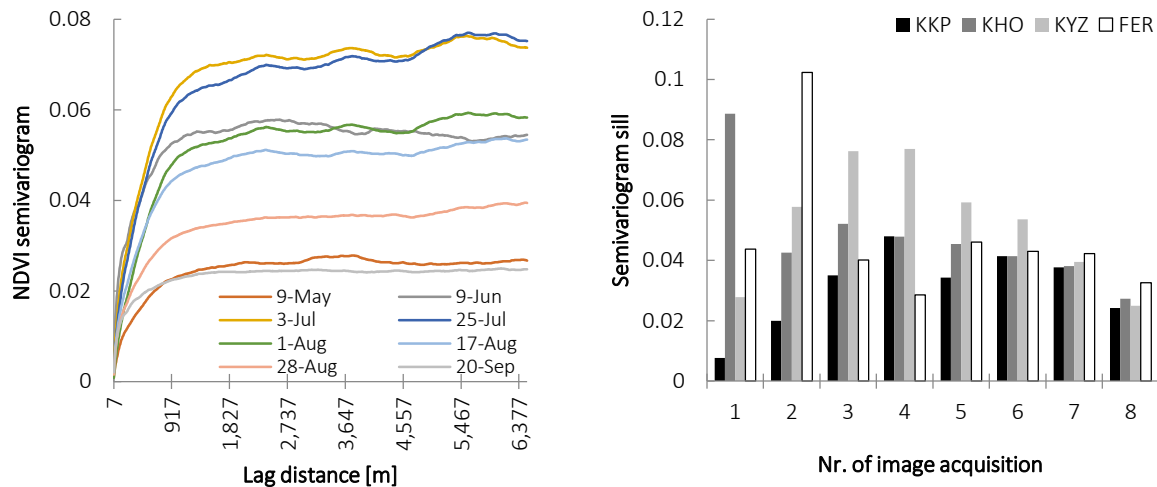
Figure 2-5: Left: Experimental semivariograms of NDVI for each acquisition date over the KYZ test site (left). The y-axis shows the NDVI semivariance, the x-axis shows the lag distance [m]. Right: Temporal trajectory of the spatial variability of the time series quantified by the semivariogram sill $\hat{\gamma}_r$ for the test sites.

In a second experiment, the mean length scale D_c of Garrigues et al. (2006a) was computed from these SV. This parameter D_c was shown to be suitable to assess if an image is large enough to characterize the spatial structures within the landscape: Garrigues et al. (2006a) propose that an image is large enough if the integral range of the SV is smaller than 5 % of the image surface. Hence, the corresponding D_c for a 30×30 km image must be below 6.7 km. To test if this hypothesis is fulfilled the maximum of all D_c values for the NDVI along the season was calculated, and the subsets were considered large enough (Table 2-1). The results of the experiments in this section confirm that the four image subsets can be considered adequate in size to characterize and quantify the different spatial structures in the landscapes they cover.

Chapter 3

DATA PRE-PROCESSING

Abstract

This chapter summarizes the main pre-processing steps of the satellite imagery, starting with the geometric and atmospheric correction. Multispectral time series from the German RapidEye mission with five reflectance bands were available over the four test sites. Atmospheric correction was done with the ATCOR module. SPOT-5 images were the input for object-oriented image segmentation to retrieve the agricultural field boundaries, since no cadastre information was available for the test sites. A three-level segmentation approach was implemented using the eCognition Developer 8.7, 64-bit software from Trimble. The quality of the segmented objects was evaluated using an empirical discrepancy method. Then a comprehensive set of features was calculated at the object level for each acquisition date, which were used as input to the classification experiments. Particular focus was on features that are based on RapidEye´s red edge canal and on the use of multi-temporal geostatistical features from each objects semivariogram.

3.1.   Satellite Data

For this study, multispectral data from the RapidEye system (Tyc et al., 2005) was available for the four study sites. The German RapidEye satellites, launched in 2008, represent the first fleet of multi-spectral satellites operationally providing the red edge spectrum and observations with high temporal and spatial resolution for agricultural monitoring applications (Tyc et al., 2005). A constellation of five identically constructed satellites sample the earth surface with swath of 77 km and an at-nadir revisit time of 5.5 days, using a multi-spectral push-broom imager (RapidEye AG, 2012). The images have a nominal GSD of 6.5 m at nadir with five spectral bands: blue (440–510 nm), green (520–590 nm), red (630–685 nm), red edge (690–730 nm), and near-infrared (NIR) (760–850 nm). The signal is encoded in 16-bit unsigned integers. RapidEye offers a product resampled to 5 m (L3A product), however because of potential distortions in the radiometric signal this product was not considered in this thesis. The wavelength between the visible red and the near infrared, the so-called red edge spectrum is particularly focused on, which is one peculiarity of the RapidEye system. Because of its sensitivity with regard to the chlorophyll content of plants, it can be used for identifying vegetation stress or crop nutrition diagnostic (Eitel et al., 2007). RapidEye has successfully been used to characterize land surface phenology and for multi temporal crop classification (Conrad et al., 2013; Ehammer et al., 2010; Löw et al., 2012, 2013a; Schorcht et al., 2012; Schuster et al., 2012).

Table 3-1 shows the number and year of level 1B products, which were already radiometric and sensor corrected by the vendor. Satellite imagery of the different study sites was acquired from the RapidEye satellites in 2009 (KHO), 2010 (KHO), 2011 (KKP, KYZ, FER), and 2012 (FER). Images completely covering the 30×30 km study sites were selected for eight dates within the growing period, each based on the pre-study of each site´s crop calendar (see section 2.2). This ensured to include key growth stages between early spring in March and early fall in October.

Level 1B data from SPOT-5 (Système Pour l'Observation de la Terre) of the French space agency CNES (Centre national d'études spatiales) was used to derive the field boundaries by image segmentation. SPOT-5 is a push-broom scanner that operates since May 2002. Its sensor consists of an array of CCDs (charge-coupled device) that are arranged along a vertical line to the satellite orbit track (Richards and Jia, 2005). Three types of instruments are installed on SPOT-5: the High Geometric Resolution (HRG), the High-Resolution Stereoscopic imaging (HRS), and the VEGETATION instrument. HRG has the ability of panchromatic (P) and multispectral (XS) recording (SPOT IMAGE, 2012). The images, acquired in 2006 (KHO) and 2007 (KYZ), have one band with a GSD of 2.5 m (P) and four XS bands with a GSD of 10 m (Table 3-1).

Clouds and cloud shadows present in the Khorezm image acquired on 2–Jun 2009 were filled with spectral information from scenes recorded two days before (31–May 2009). Further, some acquisition dates in Table 3-1 are composed of two scenes, acquired at different dates, in order to achieve full coverage for the study site: KKP (2011): 07–Jun and 16–Jun, FER (2011): 23–Jun and 24–Jun, KHO (2009): 20–Sep and 22–Sep, and KHO (2010): 01–Oct and 03–Oct. All other scenes were cloud free or could be covered with one image.

Table 3-1: Overview of the sensor configurations and image acquisition dates of RapidEye and SPOT-5.

Study site	Acquisition date	Year	Spectral bands and spectral range [nm]	GSD	Bit depth
RapidEye					
KKP	04-Apr, 08-May, 07/16-Jun, 14-Jul, 27-Jul, 13-Aug, 28-Aug, and 09-Oct	2011			
KYZ	09-May, 09-Jun, 03-Jul, 25-Jul, 01-Aug, 17-Aug, 28-Aug, and 24-Sep	2011	1: Blue (440–510) 2: Green (520–590) 3: Red (630–685) 4: Red edge (690–730) 5: NIR (760–850)	Bands 1–5: 6.5 m	16-bit
FER	08-Apr, 13-May, 31-May, 23/24-Jun, 29-Jul, 07-Aug, 08-Sep, and 27-Sep	2011			
FER	15-Apr, 23-May, 01-Jun, 17-Jun, 02-Jul, 01-Aug, 01-Sep, 02-Oct	2012			
KHO	31-May/02-Jun, 14-Jul, 01-Aug, 13-Aug, 08-Sep, 20/22-Sep, 03-Oct, 13-Oct	2009			
KHO	16-Apr, 25-Apr, 13-May, 13-Jun, 04-Jul, 23-Jul, 01-Aug, 01/03-Oct	2010			
SPOT-5					
KHO	19-Jun and 22-Jun	2006	1: Panchromatic (480–710) 2: Green (500–590) 3: Red (610–680) 4: NIR (780–890) 5: SWIR (1,580–1,750)	Band 1: 2.5 m Bands 2–5: 10 m	8-bit
KYZ	04-Jun, 18-Jun, and 31-Jul	2007			

3.2. Reference database from ground surveys

A GIS database was generated that contained information on infrastructure, soil types, and administrative boundaries as vector, and high-resolution satellite data from SPOT-5 and ALOS PRISM. This information aided in correctly identifying the actual positions of agricultural fields during the field campaigns, which was controlled by the Global Positioning System (GPS) with an approximated location of less than 2–3 m.

Ground reference data for the training and testing stage of the classification experiments in this thesis was acquired during extensive field surveys in the study sites: a two-time random sampling was performed in KKP, KHO, and FER because of the presence of crop rotations: multiple cropping (growing sequentially two or more crops in the same field within a single growing season) is sometimes practiced, typically starting with winter wheat and following with another crop. Because of the absence of rotations in KYZ a one-time random sampling was performed in this site. KYZ was visited in the summer season in 2011. In KKP (2011) and KHO (2009) the random sampling was conducted in the spring season (April–May) and before harvest in the summer seasons. In KKP and KHO the same fields were visited in the spring and summer seasons, respectively.

During the spring sampling period only winter crops (winter wheat) were sampled that were in the stage of growing (stem elongation and shoot development), whilst all other fields were fresh ploughed or fallow, except for alfalfa fields that were already in the growing stage. In the summer season the same fields were visited again to check if a second crop was cultivated on the corresponding winter wheat field (crop rotation). In addition, several other fields were visited in the summer season. At the time of the field visits in summer, crops were in the final stages of their growing season, and fruits were partly fully developed (e.g. melons or fruit trees). In FER only one field sampling in the summer season was

performed (for logistical reasons), and crop rotations were identified afterwards by checking the fields´ NDVI signatures on-screen. During the field sampling the crop type, phenological stage, field homogeneity, and crop height was recorded at each visited field and geo-located with a GPS-handheld from Trimble. At each site several hundreds of reference plots (agricultural fields) were visited and geo-located. Figure 3-1 shows examples of crops found in the study sites. In KYZ two stages of alfalfa can be distinguished: one year old alfalfa fields (hereafter this class will be called alfalfa-1y) characterized by high amounts of weed flora, and 2−3 year old fields (hereafter called alfalfa-3y) that are mainly composed of alfalfa. In the latter case the weed flora cannot compete with the alfalfa after the first year and several cutting operations (harvests), and the temporal signature of 2−3 year old fields become very distinct from the 1-year-old fields. Sorghum and maize, which partly are grown at the same field in KHO and KKP and were merged to one class: sorghum/maize.

(a) Fallow field   (b) Melons   (c) Rice

(d) Alfalfa-3y   (e) Cotton   (f) Winter wheat (harvested)

(g) Fruit trees   (h) Sorghum/maize   (i) Alfalfa-1y
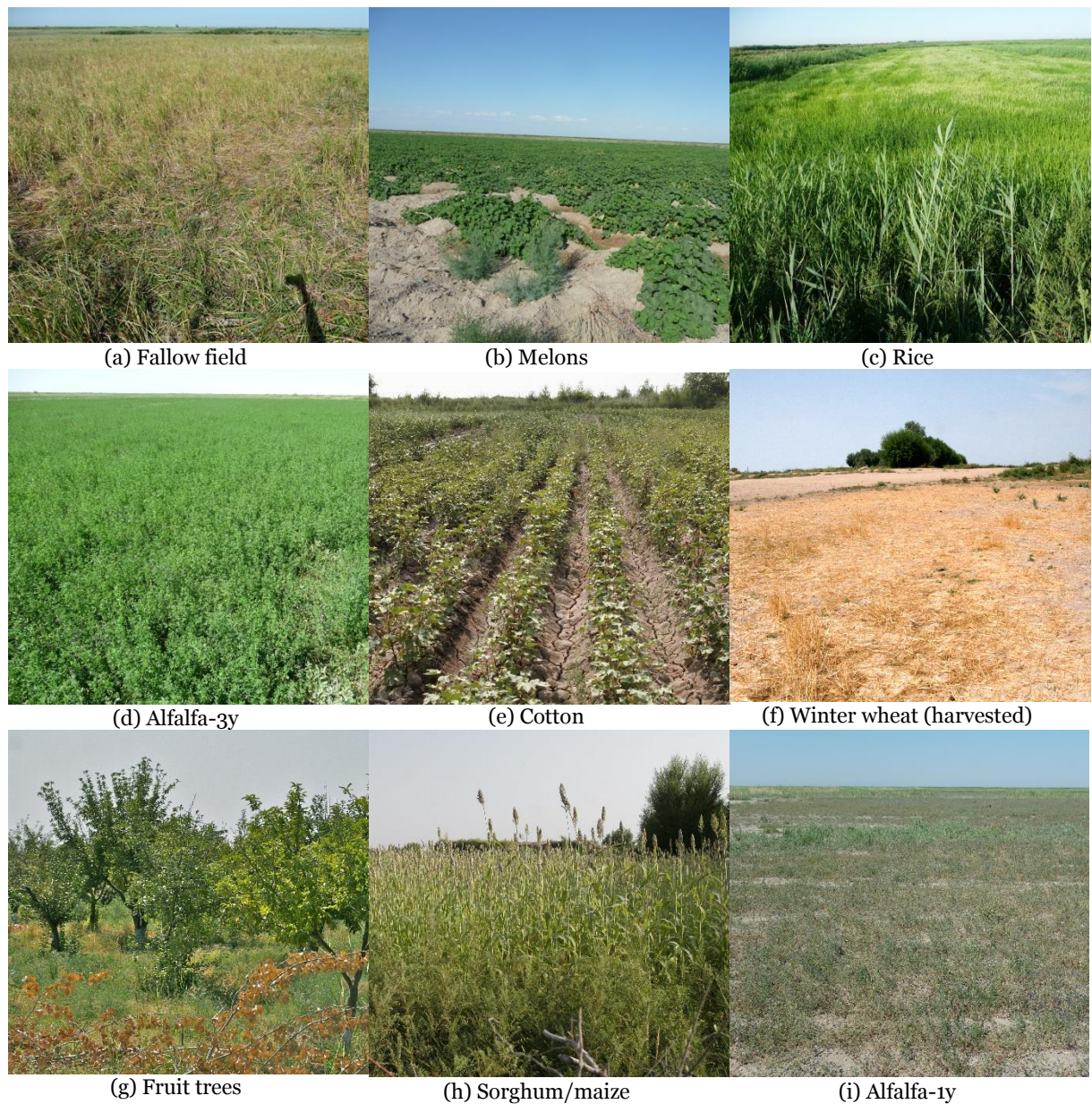
Figure 3-1: Major crop types in the four study sites that were classified in this study (line by line from left to right): fallow field in KYZ, melons in KYZ, rice field in KYZ, three year old alfalfa field in KYZ, cotton field in KKP, harvested winter wheat field in KKP, fruit (apple) trees in KKP, sorghum in KKP, and one year old alfalfa in KYZ.

Table 3-2: Classification scheme for crops in the study sites. "x" indicates that the class was not included in the legend. The crop map in KHO for the year 2010 was taken from Conrad et al. (2011b).

Class name	KHO 2009	KKP 2011	KYZ 2011	FER 2011	FER 2012
Cotton	113	46	X	120	110
Rice	70	118	116	x	X
Fallow	75	122	115	29	38
Fruit trees	65	x	X	77	18
Sorghum/maize	70	21	X	x	X
Winter wheat	34	137	28	55	24
Winter wheat - other	103	21	X	128	108
Alfalfa-1y	x	36	117	x	X
Alfalfa-3y	x	x	29	x	X
Melons	x	25	X	x	X
Total number of classes	7	8	5	5	5
Total number of samples	530	526	405	409	298

Statistical tests for accuracy assume a random distribution of the samples from the total population (Jones and Vaughan, 2010). In general training and validation data sets can be generated in different sampling strategies e.g. simple random sampling, clustered and systematic sampling, stratified random sampling, or sometime transect-wise sampling (Jones and Vaughan, 2010; McCoy, 2005). In simple random sampling each sample has an equal chance to be selected without operator bias and gives theoretically statistically optimal results (Jones and Vaughan, 2010), but at the expense of under-sampling of small classes, whilst the systematic approach selects samples with an equal interval over the test sites (McCoy, 2005). Random sampling might be prohibitive due to physically restricted access to all locations in the test sites and might have to be relaxed in some cases (Foody, 2002). Stratified random sampling combines a-priori knowledge about the test site, for instance pre-existing land cover maps or environmental conditions, and combines it with the simple random sampling within each of the a-priori defined categories (Congalton and Green, 2009), which can be defined by cluster analysis or existing land use maps (Conrad et al., 2010). It reduces the chance of small categories to be under-sampled, whilst clustered sampling may reduce field work time and be preferred for pragmatically reasons (Jones and Vaughan, 2010). Not only does the sampling design has an impact on the classification accuracy (Plourde and Congalton, 2003) but also the reference unit (e.g. pixel, aggregates of pixel, or objects) can impact the result, but a universally best solution does not exist (Stehman and Wickham, 2011; Stuart et al., 2008). The initially targeted stratified random sampling scheme had to be relaxed for several reasons, e.g. access to fields was partly restricted because of private ownership, flooded canals without crossings, or bad road conditions. Spatial distribution of the sample sites is shown in Figure 3-2.
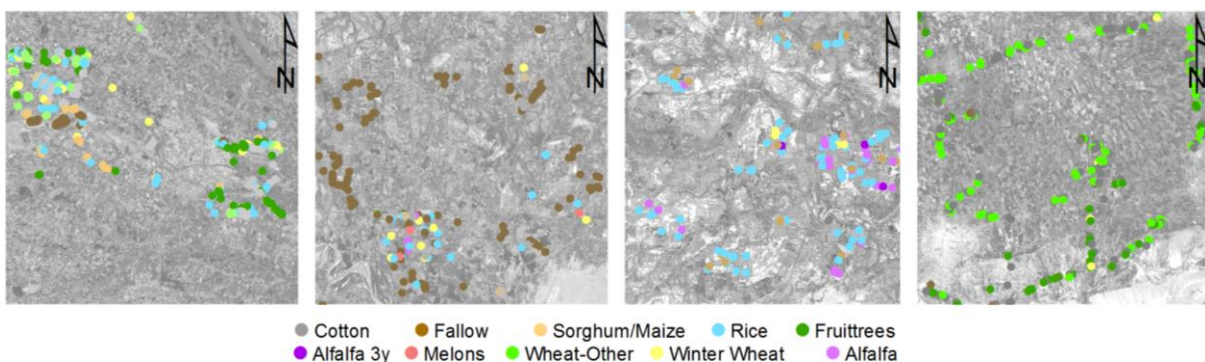
Figure 3-2: Location of the sampling sites in (from left to right) KHO, KKP, KYZ, and FER. Each site is 30×30 km.

Because of insufficient field samples for some of the classes in KHO, additional samples were selected from very high resolution Quickbird time series data (Conrad et al., 2013). Additional samples could be extracted based on sub-field segmentation techniques using NDVI time-series, thereby splitting existing objects according to the similarity of NDVI profiles (Schorcht et al., 2012). The resulting number of field samples and the classes included in the legends are summarized in Table 3-2. Outliers in the reference database were deleted based on visual inspection of NDVI temporal profiles from RapidEye time series, resulting in 298–530 samples in the sites. The final number of the reference fields was variable among the different crop types and study sites, causing the number of samples per crop to vary between 18 and 137. The field sampling was conducted at the per-field level with preference towards homogeneous fields (one cop per field, homogeneous growing state).

3.3. Satellite imagery pre-processing

Before the image classification, geometrical and atmospheric pre-processing of RapidEye data was performed (Figure 3-3). Agricultural fields were derived by segmentation of SPOT and RapidEye data. The quality of the segments will be evaluated with an empirical discrepancy method.

Geometrical and atmospheric correction

Geometrical errors in satellite images are due to earth rotation skew, panoramic distortion, pixel size distortions, and variations in platform speed and elevation (Richards and Jia, 2005). The vendor corrects these errors prior to the image delivery, which is generally denoted as "Level 1B" product. These level 1B products then need to be transferred to a common reference projection system and geocorrected. The SPOT-5 scenes have been geocorrected in ERDAS Imagine using ground control points (GCP) that were collected and located in the study sites (GCP_f), and then projected to Universal Transverse Mercator (UTM), WGS-84 datum. Then each RapidEye scene was co-registered and corrected to the SPOT-5 reference scenes with the ERDAS AutoSync module (ERDAS, 2010) which automatically generates hundreds of ground control points (GCP_{aut}) per image pair and conducts the resampling. In KKP and FER, where no SPOT images were available, one RapidEye image that was formerly geocorrected by using GCP_f was taken as the master scene for the remainder RapidEye images.

For the image-to-image registration, GCP_{aut} were used to calculate the transfer functions between the RapidEye scenes and the geocorrected master images. During the image-to-image processing the RapidEye time series data was transferred into the corresponding zone of the UTM: UTM 40N (KKP), UTM 41N (KHO, KYZ), and UTM 42N (FER). Selecting a nearest neighbour routine, with a polynomial model (2nd degree) resulted in sub-pixel accuracies for all scenes, e.g. the root mean square errors (RMSE) was ≤ 0.9 RapidEye pixels, which is considered of minor impact for the experiments in this thesis. An additional ortho-rectification, which can be done with a digital elevation model, was not performed because of the overall flat terrain of the study sites.

Atmospheric correction and radiometric calibration was performed to obtain top of canopy reflectance (TOC) from at-aperture radiance. Because several factors affect the at-satellite radiance, such as atmospheric absorption and scattering by gaseous and particle components, it is worthwhile to perform an absolute atmospheric correction in most remote sensing applications (Song et al., 2001). Atmospheric correction was done using the ATCOR-2 module for flat terrain, version 7.1 for IDL, which is based on the MODTRAN radiative transfer code (Richter, 2011). Model parameters included a mid-latitude summer rural atmosphere model with scene visibilities set between 10 and 50 km. Scene visibilities were calculated automatically in ATCOR-2. The default value of the adjacency range in ATCOR-2, which accounts for the influence of neighbouring background effects, e.g. atmospheric

scattering of radiance between adjacent fields of different reflectance, which is scattered into the view direction (Richter, 2011), was left unchanged (1.0 km). Only for some scenes these values were adjusted to 0.5 km in order to have better results from the atmospheric correction.
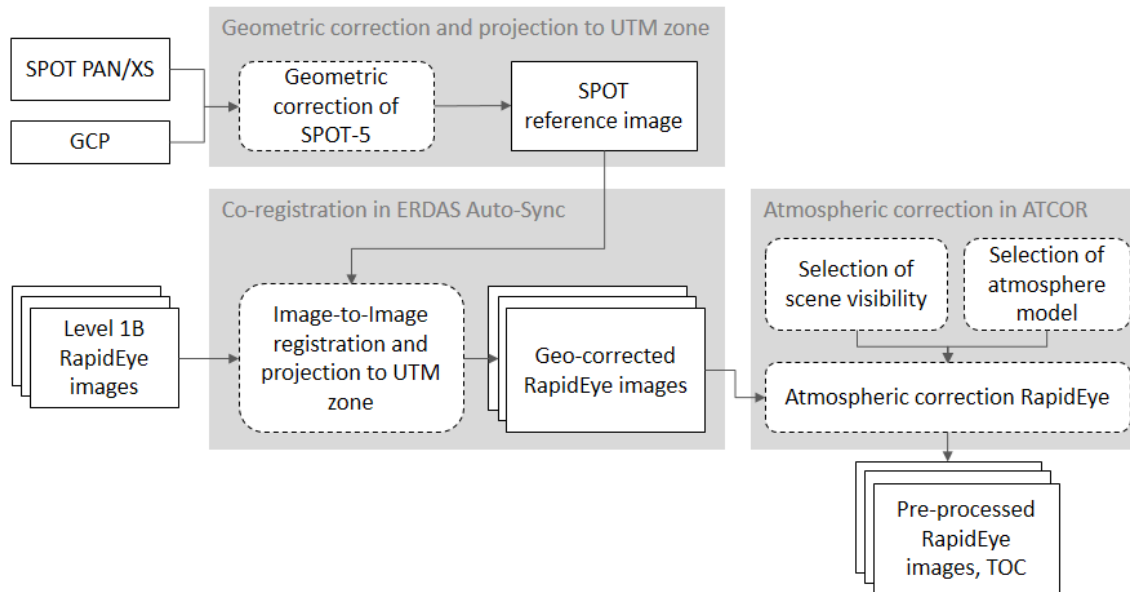
Figure 3-3: Schematic overview of the pre-processing.

3.4.   Field parcel delimitation by image segmentation

3.4.1. Algorithm description

The agricultural fields were extracted using object-based image segmentation and supervised classification to identify the target (field vs. non-field). The multi-resolution and spectral difference segmentation algorithms, which are included in the 64-bit version of the commercial software eCognition Developer 8.7 (Trimble Germany GmbH, 2011), were used. This approach is based on the Fractal Net Evolution Approach (FNEA) segmentation algorithm from Baatz and Schäpe (2000). In this bottom-up process spatially adjacent and spectrally similar pixels are aggregated together in image objects. This region-growing algorithm seeks an optimal solution after a series of iterations: as the objects grow a decision whether to merge adjacent objects or not is done based on minimizing the average heterogeneity of the pixels within the objects. This is accomplished by calculating a variance-covariance matrix of all spectral bands or indices of the input used to separate the image objects. Adjoining objects are fused if the spectral heterogeneity of the object resulting from the fusion does not exceed a certain threshold, which determines the maximum heterogeneity. Consequently the so-called scale parameter greatly influences the size of the resulting objects. Other user defined parameters affect the object growing, e.g. colour, smoothness, compactness, and shape (Benz et al., 2004; Trimble Germany GmbH, 2011).

For KHO and KYZ, SPOT-5 data recorded in spring of 2006 and 2007, respectively, were used for segmentation. SPOT-5 multi-spectral XS bands were pan-sharpened with the P-band to a spatial resolution of 2.5 m using a high pass filter algorithm (Pohl and van Genderen, 1998) implemented in ERDAS Imagine. The five pan-sharpened XS bands and a variance texture image (size: 5×5 pixels) were used as input. Agricultural fields in KKP and FER were delineated following the same approach but

based on RapidEye 6.5 m images because no SPOT images were available for this study. The characteristics of the fields differ from site to site, e.g. in terms of field sizes and intra-field heterogeneity. Hence, for each site different segmentation parameters (scale, colour, shape, smoothness, compactness, and band weights) have been tested. Segmentation and generation of a field mask were done in several consecutive steps, which were designed as a hierarchical framework (Figure 3-4). In particular the scale parameter was focused, which has the largest influence on the segmentation. The characteristics of the objects, e.g. their size that is controlled by this scale parameter, can have an impact on the classification accuracy (Myint et al., 2011). The selection of the optimal numerical values for the parameters was guided by an iterative "trial-and-error" procedure as no ready-to-use accuracy assessment method is implemented in eCognition. In an initial step the image was over-segmented using a scale parameter of 5 to have objects that are smaller than the fields (*level 1*), thereby creating objects that are sufficiently small to fit to the field borders of larger fields. A merging process by segmenting with a higher scale factor (20–35) created the smaller fields and some of the larger fields (*level 2*). Then, adjacent objects from level 2, which have similar spectral properties, were merged on basis of their spectral similarity, below a given threshold and allowed for defining the larger fields (*level 3*). This 3-level procedure gave better results over the four study sites than an initially tested two-level approach, in that *level 2* was skipped.

A (unsegmented raw image) **B (Level 1)**

C (Level 2) **D (Level 3)**
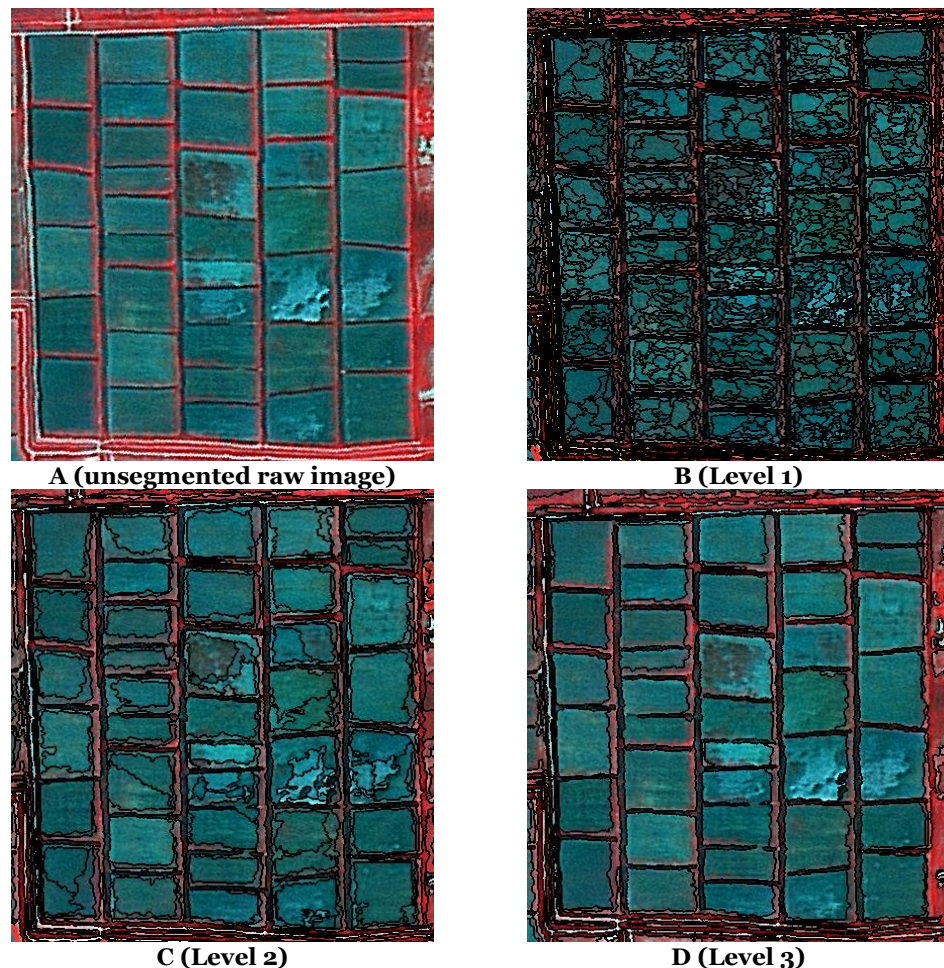
Figure 3-4: Exemplary output of the object-oriented image segmentation using a mid-spring SPOT-5 image in KYZ (red, green and near-infrared bands): A) portion of the SPOT-5 image (2.5 m pan-sharpened), B) over-segmented output at level 1 to create small objects that fit to the field borders, C) output at level 2 with a larger scale level to create objects that are approximate the size of the real field borders, and D) the output of level 3 after spectral similarity merging of adjacent objects to eliminate with-in field patches from level 2.

In doing so the object's size is no longer limited by a single static segmentation scale parameter, which is a particular weakness of using only one hierarchy level for segmentation with eCognition, and both large and small fields can be delineated simultaneously in a same landscape. Strictly speaking the segmentation approach used in this study results in clustering of adjacent pixels of similar spectral characteristics, which might differ from "real" field borders. Objects containing less than 20 pixels were not considered for classification to assure that sufficient pixels are contained for calculating the objects spectral mean values and geostatistics (section 3.4.3).

3.4.2. Evaluation of segmentation quality

An empirical discrepancy method was applied to measure the quality of the segmented objects (Ortiz and Oliver, 2006; Zhang, 1996). This was accomplished by assessing the similarity of the outputs from the image segmentation (section 3.4.1) with 400 randomly distributed reference polygons that were digitized on-screen. The reference polygons represent fields from all studied crop types. The following characteristics were used to assess the similarity between reference and the corresponding segmented object: shape area, shape perimeter, and shape index[13]. The resulting average discrepancies between reference and segmented fields, indicated by these three characteristics, were: 16.5 % (KHO), 9.0 % (KKP), 5.6 % (KYZ), and 4.7 % (FER). The fields in KHO were derived in a previous study by Conrad et al. (2010). Although their approach only slightly differs from the one chosen in this thesis, the quality of the segmentation in KHO differs considerably from the other three sites. Hence, a cadastre database with on-screen digitized fields was used instead that was available through the Khorezm-Project. No such kind of cadastre database was available in the other sites.

3.4.3. Calculation of object features

Since the objective of chapters 4 and 5 is object-based image classification, a variety of features were computed as input to the classification. Features at the object level have an advantage over pixel-based features because they suffer less from the pixel heterogeneity or reflectance variability within agricultural fields, e.g. by calculating the mean or standard deviation of reflectance values of pixels belonging to one object (Blaschke, 2010). After the segmentation a variety of features like the pixel reflectance values, textural information, object features or neighbourhood relationships can be utilized.

A set of 71 spectral and geostatistical features was calculated for each field and acquisition date, respectively that can be categorized into five groups (here denoted with letters A−E, see Table 3-3). The choice of the features was oriented toward their usefulness for land use classification as reported in the literature (Balaguer et al., 2010; Glenn et al., 2008; Peña-Barragán et al., 2011; Rodriguez-Galiano et al., 2012a). More detailed descriptions and applications of the selected features in remote sensing are given in the references in Table 3-3. The feature groups are briefly characterized in the following.

---

[13]: $S_i = P_i / \left(4 * \left(\sqrt{A_i} \right) \right)$, where S_i is the shape index, P_i the shape perimeter, and A_i the shape area (Neubert et al., 2006). Further details about assessing image segmentation quality can be found in Neubert et al. (2008).

Group A: RapidEye bands

The first group contains the reflectance information from the five RapidEye bands (Tyc et al., 2005), see Table 3-3 for details.

Group B: Spectral indices

A total of 13 broadband spectral indices were computed for each object. Such vegetation indices (VI) are able to enhance spectral features and to reduce background effects, e.g. the soil reflectance. They characterize the spectral properties of the vegetation and plant canopy and can be directly estimated by satellite sensors (Glenn et al., 2008; Jones and Vaughan, 2010).

This group includes some ratio VIs, e.g. the Ratio Vegetation Index (RVI) (Jordan, 1969) or the NDVI (Rouse et al., 1974), which are based on the red and NIR part of the spectrum and one of the most often used index in remote sensing-based vegetation studies. These indices are characterized by a constant vegetation amount that diverge from the origin in the red–NIR space and increase at different slopes (Jones and Vaughan, 2010). Further, some alternative formulations of the basic NDVI are included, like the Green-NDVI (GNDVI) where green is substituted for red to improve the sensitivity when dense vegetation with high leaf area index (LAI) is sensed (Jones and Vaughan, 2010). Another type of modifications aims to reduce the influence of atmospheric effect, e.g. the Atmospherically Resistant Vegetation Index (ARVI) (Kaufman et al., 1992) or the Enhanced Vegetation Index (EVI) that takes into account the effect of atmospheric aerosols by adding the blue spectrum which is more affected by scattering (Huete et al., 2002).

Another group of VIs aims at reducing soil background effects, e.g. the Difference VI (DVI) (Richardson and Everitt, 1992). Those VIs are sometimes referred to as orthogonal indices, as they do not converge on the origin of the red-NIR space and stay parallel to the soil line (Huete, 1988). Hybrid VIs combine characteristics of ratio and orthogonal VIs, e.g. the Soil Adjusted VI (SAVI) from Huete (1988) that corrects the effect of soil background and aims a positive intercept of the soil line relating to zero vegetation cover. The SAVI is more closely related to LAI over a wide range of soil reflectance than NDVI (Jones and Vaughan, 2010).

Group C: Red edge indices

The indices in group-C focus on the red edge spectrum (approximately 680 nm to 740 nm), which is one peculiarity of the RapidEye instrument. Because of its sensitivity with regard to the chlorophyll content of plants, it is used for identifying vegetation stress or crop nutrition diagnostic (Eitel et al., 2007). The red edge refers to the reflectance rise between red and NIR, caused by strong chlorophyll absorption (low reflectance in the red spectrum) and high reflectance in the NIR spectrum because of internal leaf scattering (Jones and Vaughan, 2010). These attributes make information from the red edge spectrum of great value for vegetation status and crop differentiation. For instance Schuster et al. (2012) found an increase in classification accuracy when incorporating red edge from RapidEye into the input data in SVM classification, when substituting red edge for NIR in the NDVI formula in a mono-temporal application.

In this study several VIs from group-B were re-calculated based on the information of the red edge band by substituting the red or NIR canal, see Table 3-3. In a similar fashion Ehammer et al. (2010) proposed two indices based on red edge (the Red Edge Ratio Indices (RRI) 1 and 2), and showed the superiority of red edge for statistical derivation of biophysical parameters (e.g. LAI).

Group D: Curvature indices

Group-D comprises features describing the curve properties between red and NIR utilizing the angle and Euclidian distances between red, red edge and NIR. These experimental indices have potentials for differentiation of dense vegetation canopies, where other VIs often shows saturation effects (Conrad et al., 2012). Another application of these curvature VIs is given in Löw et al. (2012).

Group E: Geostatistical and GLCM texture measures

Group-E consists of features derived from each objects semivariogram (SV) (Balaguer et al., 2010) or grey-level co-occurrence (GLCM) matrix (Haralick et al., 1973). Several researchers (Balaguer et al., 2010; Berberoglu et al., 2000; Chica-Olmo and Abarca-Hernandez, 2000; Lloyd et al., 2004; Rodriguez-Galiano et al., 2012a) demonstrate the general usefulness of geostatistical or GLCM texture measures in remote sensing-based classification of land cover. The SV quantifies the spatial associations of the values of a variable, and measures the degree of spatial correlation between different pixels in an object (Ruiz et al., 2011). Whilst SV estimators have successfully been used in pixel-based approaches using moving windows (Buddenbaum et al., 2005; Chica-Olmo and Abarca-Hernandez, 2000), object-based approaches are less exploited. For instance Berberoglu et al. (2000) used semivariances as additional input in a pixel-based classification that was converted into the object level via majority vote. Rather than using raw SV values, which might be redundant or lacking interpretability, Balaguer et al. (2010) proposed a rigorous object-based approach that included several features that were extracted from the objects SV.

In this study omni-directional SVs were calculated for each object and along the season, respectively. Thus, only pixels inside a field were considered. From each object´s SV 11 features were calculated using the software FETEX 2.0 (Ruiz et al., 2011). The features are fully described in Balaguer et al. (2010), formulas are given in Table 3-3. The SV were calculated for NIR reflectance values.

Image texture measures from the GLCM, proposed by Haralick et al. (1973), are also called co-occurrence or second order measures and characterize the relative frequencies between two pixel brightness values linked by a spatial relation (Clausi, 2002). Such measures are widely used in land cover classification (Beekhuizen and Clarke, 2010; Rodriguez-Galiano et al., 2012a). However, there is evidence that many of the measures are correlated (Clausi, 2002), and their calculation is computationally demanding. Therefore only one measure of object textural information based upon the GLCM, namely the variance, is included in group-E (among other possible measures like homogeneity, contrast, dissimilarity, second moment, or entropy), and was also calculated in FETEX 2.0.

Table 3-3: Spectral and textural features explored in this study. Formulae show the RapidEye bands used for calculation of the indices. References and/or applications for the indices are given.

Group	Feature number	Index	VI adapted to RapidEye	Reference
A	1–5	Reflectance values of bands of RapidEye	$RE1, RE2, RE3, RE4, RE5$	Tyc et al. (2005)
B	6	Greenness Index (GI)	$RE2/RE3$	-
	7	Ratio Vegetation Index (RVI)	$RE5/RE3$	Jordan (1969)
	8	Blue-Green Ratio Index (BGI)	$RE1/RE2$	-
	9	Red-Green Ratio Index (RGI)	$RE3/RE2$	-
	10	Difference Vegetation Index (DVI)	$RE5 - RE3$	Richardson and Everitt (1992)
	11	Normalized Difference VI (NDVI)	$(RE5 - RE3)/(RE5 + RE3)$	Rouse et al. (1974)
	12	Green NDVI (GNDVI)	$(RE5 - RE2)/(RE5 + RE2)$	Gitelson et al. (1996)
	13	Green Vegetation Index (VIgreen)	$(RE2 - RE3)/(RE2 + RE3)$	Gitelson et al. (2002)
	14	Soil adjusted Vegetation Index (SAVI)[a]	$\dfrac{(1 + R) * (RE5 - RE3)}{RE5 + RE3 + R}$	Huete (1988)
	15	Enhanced Vegetation Index (EVI)[b]	$2.5 * \dfrac{(RE5 - RE3)}{RE5 + c1 + (L * RE3 - c2 * RE1)}$	Huete et al. (2002)
	16	Atmospherically Resistant Vegetation Index (ARVI)[c]	$(RE5 - rb)/(RE5 * rb)$	Kaufman et al. (1992)
	17	Renormalized Difference Vegetation Index (RDVI)	$\dfrac{RE5 - RE3}{\sqrt{RE5 + RE3}}$	Roujean and Breon (1995)
	18	Modified Simple Ratio (MSR)	$\dfrac{(RE5/RE3) - 1)}{\left(\left(\frac{RE5}{RE3}\right)^2\right) + 1}$	Chen (1996)
C	19	Red edge NDVI (RNDVI)	$(RE5 - RE4)/(RE5 + RE4)$	Vina and Gitelson (2005)
	20	Green-Red edge NDVI (GRNDVI)	$(RE4 - RE2)/(RE4 + RE2)$	Löw et al. (2012)
	21	Red edge -RED NDVI (REDNDVI)	$(RE4 - RE3)/(RE4 + RE3)$	-
	22	Modified Chlorophyll Absorption Ratio Index (MCARI)	$((RE4 - RE3) - (0.2 * (RE4 - RE2))) * (RE4/RE3)$	Daughtry et al. (2000)
	23	Transformed CARI (TCARI)	$3 * (((RE4 - RE3) - (0.2 * (RE4 - RE2))) * (RE4/RE3))$	Haboudane et al. (2002)
	24	Triangular Vegetation Index (TVI)	$0.5 * (120 * (RE4 - RE2) - 200 * (RE3 - RE2))$	Broge and Leblanc (2000)
	25	Red edge Ratio 1 (RRI1)	$RE5/RE4$	Ehammer et al. (2010)
	26	Red edge Ratio 2 (RRI2)	$RE4/RE3$	

	#	Feature	Formula	Reference
D	27	Red edge Length (REL)	$\sqrt{(RE5-RE4)^2 * (d\_RE5\_RE4)^2 *} \sqrt{(RE4-RE3)^2 * (d\_RE3\_RE4)^2}$	Conrad et al. (2012)
	28	Relative Red edge Length (RREL)	$\dfrac{REL}{\sqrt{(RE5-RE3)^2 * (d\_RE3\_RE5^2)}}$	
	29	Red edge Curvature (CUR)[d]	$\dfrac{\dfrac{(RE5-RE4)}{(d\_RE5\_RE4)} - \dfrac{(RE4-RE3)}{d\_RE3\_RE4}}{d\_RE3\_RE5}$	
E	30	Ratio variance at the first lag (RVF)	$\dfrac{Variance}{\hat{\gamma}_1}$	
	31	Ratio between semivariance at second and first lag (RSF)	$\dfrac{\hat{\gamma}_2}{\hat{\gamma}_1}$	
	32	First derivative near origin (FDO)	$\dfrac{\hat{\gamma}_2 - \hat{\gamma}_1}{h}$	
	33	Second derivative at third lag (SDT)	$\dfrac{\hat{\gamma}_4 - 2\hat{\gamma}_3 + \hat{\gamma}_2}{h^2}$	
	34	First maximum lag value (FML)	max_1	
	35	Mean of SV values up to first maximum (MFM)	$\dfrac{1}{max\_1} \sum\limits_{i=1}^{max\_1} \hat{\gamma}_i$	
	36	Variance of SV values up to first maximum (VFM)	$\dfrac{1}{max\_1} \sum\limits_{i=1}^{max\_1} \left(\hat{\gamma}_i - \hat{\gamma}_{max\_1}^{mean}\right)^2$	Balaguer et al. (2010)
	37	Difference between MFM and semivariance at first lag (DMF)	$\hat{\gamma}_{max\_1}^{mean} - \hat{\gamma}_1$	
	38	Ratio between semivariance at first local maximum and mean SV values up to this maximum (RMM)	$\dfrac{\hat{\gamma}_{max\_1}}{\hat{\gamma}_{max\_1}^{mean}}$	
	39	2nd order difference between 1st lag and 1st maximum (SDF)	$\hat{\gamma}_{max\_1} - 2\hat{\gamma}_{\frac{max\_1}{2}} + \hat{\gamma}_1$	
	40	Area between SV values in the first lag and the SV function until 1st maximum (AFM)	$\dfrac{h}{2}\left(\hat{\gamma}_1 + 2\left(\sum\limits_{i=2}^{max\_1-1} \hat{\gamma}_i\right) + \hat{\gamma}_{max\_1}\right) - (\hat{\gamma}_1(h_{max\_1} - h_1))$	
	41	Local variance (VAR), moving window	$\delta_i^2 = \sum\limits_{i,j=0}^{N-1} P_{i,j}\,(i - \mu_i)^2$ $\delta_j^2 = \sum\limits_{i,j=0}^{N-1} P_{i,j}\,(j - \mu_j)^2$	Woodcock and Harward (1992), Haralick et al. (1973)

Notes:

Mean and standard deviation were calculated for features 1–29 and 41 for each acquisition date and object, respectively. One SV derived statistic (features 30–40) was calculated per object and acquisition date, respectively. In total 568 features were calculated for each object (71 features times 8 acquisition dates)

RE1 = RapidEye blue band (440–510 nm)
RE2 = RapidEye green band (520–590 nm)
RE3 = RapidEye red band (630–685 nm)
RE4 = RapidEye red edge band (690–730 nm)
RE5 = RapidEye NIR band (760–850 nm)
[a] L= 0.5 (soil adjustment factor for SAVI)
[b] L=1, c1=6.0, c2=7.5, coefficients for MODIS EVI were taken
[c] rb = RE3-β(RE1-RE3), β=1
[d] d_RE5_RE4, d_RE3_RE4: distances between RE5 and RE4, and RE3 and RE4, respectively. Distances were calculated between the corresponding centre wavelengths of bands [μm], and the distance between NIR (RE5) and red (RE3) was normalized to 1.

Chapter 4

IMPACT OF FEATURE SELECTION ON THE SPATIAL UNCERTAINTY AND ACCURACY OF PER-FIELD CROP CLASSIFICATION USING SUPPORT VECTOR MACHINES[14]

Abstract

Remote sensing derived crop maps are frequently being used in spatially explicit agricultural production models, e.g. yield or water use, or up-scaling methods. The map quality is crucial and influences the model outputs. Traditional measures to assess the classification accuracy of such maps stem from the confusion matrix and do not consider the spatial variation of error. However, measures of classification uncertainty that can be used for this purpose attained much less attention in remote sensing studies although they are informative supplements to the traditional accuracy assessments. Classifiers algorithms like the support vector machine (SVM) can estimate class memberships for each classified pixel or object (soft output). From these estimates, uncertainty measures can be derived. Hence, the major goal of this chapter is to assess the value of classification uncertainty measures derived from SVM with regard to predicting the spatial distribution of error in the maps (the so-called predictive strength), and to evaluate the impact of the kind of features and feature space size on the classification uncertainty of SVM. The SVM is applied for the classification of agricultural crops in irrigated landscapes in Middle Asia at the object-level. Five types of features, defined in chapter 3, were calculated from the RapidEye time series data as input to the classification. The experimental results revealed that SMVs applied to the five groups separately performed different in classifying different types of crops. Incorporating all five groups of features into a multi-type data set by concatenating them into one stacked vector did not lead to an increase in accuracy, and partly reduced the model performance most obviously because of the Hughes phenomena. Yet, applying the random forest (RF) algorithm to select a subset of features led to an increase of more than 4 % classification accuracy of the SVM. Overall the SVM applied to a subset of input features from all five groups composed of the most informative features resulted in highest accuracies. The feature group with red edge-based indices was the most important for general crop classification, and the red edge NDVI had an outstanding importance for classifying crops. Further, the RF variable importance tool was employed to define and extract different feature subspaces from the multi-type data set, which followed different rationales: these subspaces were oriented toward practical considerations, e.g. computation time (or efficiency), classification accuracy, and the use of few acquisition dates. The relationship between the hard result accuracy and the soft output from the SVM is investigated by employing two measures of uncertainty, the maximum a-posteriori probability and the alpha quadratic entropy. The experimental results indicate a decline in classification uncertainty when a dimensionality reduction is performed, and that the uncertainty of correctly classified test cases decreases.

4.1.   Background

This chapter addresses the first out of the three research questions, which was formulated in the introduction, namely analysing the impact of feature space size on classification uncertainty, and the relative contribution of different types of features in crop classification. Crop identification via supervised image classification provides important baseline information for numerous agricultural decision support and monitoring applications. The quality of the crop maps is crucial when taken as input to crop water use calculations (Stehman and Milliken, 2007), area estimation (Canters, 1997; Gallego, 2004), or crop specific yield estimation (Doraiswamy et al., 2004). Yet, whilst the traditional way to assess map quality is via the confusion matrix (Congalton, 1991), some applications might require information on the spatial distribution of classification uncertainty in the maps in addition (Foody, 2002). Knowledge on the spatial variability classification uncertainty might perhaps be used to help re-direct fieldwork to refine the classification, to mask out regions of high uncertainty from later modelling, or filter undesirable pixels from algorithm training (Gonçalves et al., 2009).

Classification uncertainty can be defined as a quantitative measure of doubt or mistrust in the class allocation of a classifier algorithm at the per-case basis (pixel or object), and can be used as an indicator of the spatial distribution of classification error in maps (Foody, 2002)[15]. Information on the spatial distribution of classification uncertainty in maps can be generated by modelling estimates of class membership at the per-case basis (per pixel or per object). Such membership estimates are usually provided as a vector that contains an estimation of the membership degree of a classified pixel or object to the classes under investigation, also called the soft output of a classifier (Giacco et al., 2010). Many parametric and non-parametric classifiers can generate such membership estimates, but the way these are calculated differs from algorithm to algorithm. Membership estimates have been evaluated for a series of algorithms, e.g. RF (Loosvelt et al., 2012a), maximum likelihood classifier (MLC) (Foody, 1995a), or artificial neural networks (ANN) (Brown et al., 2009). For SVM this is explained in more detail in section 4.2.4 of this chapter, and for RF and ANN in section 5.2.2 of this thesis. In contrast to the accuracy measures derived from the confusion matrix, which apply to the hard result when a pixel or object is associated with one single class, membership estimates do provide spatially explicit information on the strength of class membership at the pixel or object level. Measures of classification uncertainty like entropy can be calculated from each of these membership vectors, and be used to indicate a doubt or mistrust on a per-case basis of belonging to the categories of the classes investigated. They can help to better understand the allocation of local error in maps, which may be much larger than reported by the confusion matrix or by intermediate approaches like spatially constrained confusion matrices (Foody, 2005). Measures of uncertainty such as membership estimations of MLC (Foody et al., 1992), Shannon entropy (Foody, 1995a; Maselli et al., 1994), or the magnitude of activation levels in the output layer of ANN as a measure of class membership (Foody, 1996a) were used on a per-pixel basis to indicate the uncertainty of the class allocation. Existing research demonstrates that lower levels of classification uncertainty are associated with higher classification accuracies, which was shown for decision tree classifiers (DT) (Colditz et al., 2011; Liu et al., 2004; McIver and Friedl, 2001), RF (Loosvelt et al., 2012a), or ANN (Brown et al., 2009). These studies demonstrated the so-called predictive strength of these algorithm with regard to correctly predicting the spatial distribution of classification error (classification uncertainty), and demonstrated that such uncertainty measures can be a useful supplement to traditional accuracy metrics.

---

[15]: Here the term uncertainty referes to the membership estimations from classifier algorithms for each classified case (pixel or object) with regard to each class investigated. For example, in remote sensing, land cover is often classified using hard classification (in which each pixel or object is allocated to only one out of several classes). Yet, each hard class allocation is made with some ambiguity, which can be expressed as a probability (Foody and Atkinson, 2002). In the literature the term uncertainty itself is not consistently defined or used, see discussion in Foody and Atkinson (2002).

The spatial pattern of uncertainty can be non-stationary (Comber et al., 2012), and not uniformly distributed over a map, but rather correlated with the boundaries of classes (Liu et al., 2004), e.g. mixed pixels (Loosvelt et al., 2012b) or different land cover types (Liu et al., 2004). Many more factors can affect the spatial variation classification uncertainty, which can be related to landscape characteristics (patch sizes, elevation, slope) and land use type (Steele et al., 1998), reference data density, reliability and spatial composition (Comber et al., 2012; Yu et al., 2008), membership of pixels or objects to classes, or object features[16] and algorithm performance. Steele et al. (1998) point out that on a larger scale there may even be responses to shifts in climate conditions.

A variety of algorithms have been employed to map crops, and to model classification uncertainty. More recent non-parametric classifier algorithms like RF have successfully been applied for crop classification and where shown to provide results that are more accurately than conventional classifiers like parametric MLC or DT (Waske and Braun, 2009). In addition they provide meaningful information on classification uncertainty that can be used to evaluate spatial variability of map quality and to indicate the location of classification error in the maps (Loosvelt et al., 2012b). Another non-parametric classifier algorithm is SVM (Cortes and Vapnik, 1995), a well-established machine learning technique that has not only given promising accuracies in crop classification (Foody and Mathur, 2004; Mathur and Foody, 2008), but also was shown to give membership estimations that can be interpreted as estimates of the membership degrees to the different classes (Giacco et al., 2010). Like RF or ANN, the SVM is not constrained to assumptions concerning the statistical frequency distribution of the input data, and can perform well with only small training sets even if high dimensional data is classified. In remote sensing applications SVM were shown to perform more accurately than other classifiers in land cover (Huang et al., 2002a) and crop classification (Waske and Benediktsson, 2007). The libsvm approach of Chang and Lin (2011) for training of a SVM is a fast and easy-to-use implementation of the most popular SVM formulations (see more details on the different SVM strategies in appendix A.2). Another feature that is of relevance for this study includes the computation of decision and probability values for class predictions (Wu et al., 2004), based on Platt´s posterior probabilities (Platt, 2000). This makes it an interesting approach for modelling spatial uncertainty in classification tasks.

Although attempts to derive uncertainty measures from SVM-based classifiers are promising (Foody, 2008; Giacco et al., 2010), there is further need of research, in particular (i) concerning the evaluation of the predictive strength of uncertainty measures from SVM, and (ii) the influence of feature space size on the spatial uncertainty in crop maps. Concerning the first issue, there exist no studies on the relationship of soft outputs from SVM (and specifically from libsvm) and the hard result accuracy (e.g. from the confusion matrix). Although it was shown that SVM can generate a soft output, it was hardly analysed how uncertainty measures derived from this soft output of SVM relate to the "hard" accuracy, e.g. the strength of such uncertainty measures in predicting classification error in the maps has attained only little attention (Giacco et al., 2010). Second, it is argued that SVM is resistant against the dimensionality of the input feature space and classifies well even with small training sets (Cortes and Vapnik, 1995), which was demonstrated in Foody and Mathur (2006). Yet, existing studies point to a possible impact of feature space size on the classification accuracy (Pal and Foody, 2010; Waske et al., 2010). There is still no study showing the impact of feature space size on classification uncertainty in the context of SVM classification, like it was demonstrated for the RF (Loosvelt et al., 2012a).

---

[16]: Following Kuncheva (2004), objects (here: on-screen digitized or segmented objects representing agricultural fields) are described by certain characteristics called "features". The features might be qualitative or quantitative (here based on EO imagery), calculated as means or standard deviations of reflectance values from pixels within each object. The pattern recognition techniques applied in this study operate with these numerical feature values, which are arranged as an n-dimensional vector: $x = [x_1, \ldots x_n]^T \in \Re^n$. $\Re^n$ is called the feature space, each axis corresponding to a physical feature. The information to design a classifier is usually in the form of a labelled data set $Z = \{z_1, \ldots z_N\}, z_j \in \Re^n$, where the class label of z_j is denoted by $l(z_j) \in \Omega, j = 1, \ldots, N$, assigned by a classifier function: $D: \Re^n \to \Omega$. More details can be found in Kuncheva (2004).

Concerning the input data for crop classification, studies discuss the application of SVM in mono-temporal (Mathur and Foody, 2008), hyper-spectral (Pal and Foody, 2010), or multi-sensor approaches that combine optical and radar data (Waske and Benediktsson, 2007). Multi-temporal radar images (Tan et al., 2007; Waske and Benediktsson, 2007) and MODIS time series (Carrão et al., 2008; Shao and Lunetta, 2012) have been accurately classified using SVM. Recent technological advances like the multispectral RapidEye system (Tyc et al., 2005) allow for monitoring crop growth and classification of crop types at the field level using its better spatial resolution and frequent revisit times. However against the backdrop of recent findings on the influence of huge feature space sizes on the performance of SVM (Pal and Foody, 2010; Waske et al., 2010), and due to the potential high dimensionality of such high-resolution multi-temporal data sets, limitations to the efficiency and accuracy of the classification using SVM may be experienced. One reason is that with a limited number of training data the classification accuracy can degrade when the number of features increase, also known as Hughes effect (Hughes, 1968). It would therefore be beneficial to evaluate the impact of the feature space size on SVM accuracy and further develop strategies to reduce the number of features, by only withholding those features or acquisition dates, respectively with highest impact on the classification result. Further a reduction of feature space size may even reduce the uncertainty of the resulting maps, and speed up computation time. Critically, knowing the most suitable set of features is not possible beforehand, and most often exploited post-classification (Carrão et al., 2008; Hüttich et al., 2009; Peña-Barragán et al., 2011; Rodriguez-Galiano et al., 2012a). In this study the RF algorithm was chosen as the feature selection strategy. RF is an increasingly used statistical method introduced by Breiman (2001). RF used as a classifier gives very accurate results in crop mapping applications (Loosvelt et al., 2012b; Waske and Braun, 2009). In addition, it computes a feature importance score that can be used to select and reduce variables (Hüttich et al., 2009; Loosvelt et al., 2012a; Rodriguez-Galiano et al., 2012a), which makes it very interesting in multi-source studies where data dimensionality is very high. It was demonstrated that RF could compete with other feature selection algorithms in selecting the most important features for another classifier, which was shown to enhance classification accuracy (Pal and Foody, 2010).

Concluding from the discussion above, there is still a gap in understanding the interplay between feature space size and the predictive strength of uncertainty measures derived from SVM, in particular concerning the widely used libsvm approach. This study strives to fill this gap by analysing the influence of feature space size on SVM performance in the context of object-based crop classification with multi-date optical satellite data. The feature importance score tool of the RF algorithm was chosen as feature selection strategy. Specifically the effect of feature selection on classification uncertainty is focused on. Four distinct irrigated sites in Middle Asia were chosen to evaluate the generalizability of the findings within a broader range of irrigated landscapes.

The overall objective of this study can be split in a series of partial objectives, defined as follows:

> Increasing the knowledge if the incorporation of different spectral and geostatistical features enhances classification accuracy, and how different feature groups contribute to crop classification.

> Determining and selecting the most informative features for the classification of crops using the RF as feature selection strategy, and a comparison of the RF feature selection strategy with the principal component analysis (PCA) as an alternative means to reduce the feature space size.

> Assessing the sensitivity of libsvm toward high dimensional multi-seasonal data sets, and specifically the effect of feature space size reduction on the spatial uncertainty in crop maps derived from the SVM classifier.

> Evaluating the relationship of soft outputs from `libsvm` with the hard result accuracy (e.g. the predictive strength) to assess if uncertainty measures from `libsvm` can be used to indicate spatial distribution of error in maps.

4.2. Data and Methods

In the following, dimensionality reduction strategies will be highlighted with emphasize on the PCA and the RF method, which will be compared. Then the classifier concepts used in the experiments will be highlighted, and the metrics used to evaluate classifier performance and uncertainty will be summarized. A brief description of the derivation of uncertainty measures from SVM will be given and finally the experimental setup will be presented.

4.2.1. Dimensionality reduction and feature selection

Dimensionality reduction is a commonly recommended pre-processing step when complex and high-dimensional data sets like hyper-spectral data is used as classification input (Benediktsson et al., 2007). In particular parametric algorithms that assume a Gaussian distribution (e.g. MLC) can suffer from small ratios of trainings set size to the number of features, as it will not be able to correctly estimate the first- and second-order statistics, i.e. covariance and mean (Tadjudin and Landgrebe, 1999). The Hughes phenomenon has been observed in many remote sensing studies on classification using different parametric and non-parametric algorithms (Lu et al., 2007; Pal and Foody, 2010). This effect can lead to decreasing generalization performance of the algorithms as the number of input variables, sometimes highly correlated, increases. A number of strategies can be employed to counter the negative effects of the Hughes phenomena: increasing the training set size, using classifiers that are insensitive to huge data spaces, or dimensionality reduction. Because increasing the training set size can be prohibitive (due to financial reasons or constrained physical access to the study site), and there is evidence in the literature that even non-parametric classifier algorithms can be sensitive to the Hughes effect, the latter strategy is focused in the following.

The issue of dimensionality reduction is commonly addressed in hyper-spectral applications (Chan et al., 2008; Licciardi et al., 2009; Melgani and Bruzzone, 2004; Pal and Foody, 2010), however multi-temporal datasets can exhibit significant temporal autocorrelation, and it has been shown that feature space size in multi-spectral time series can have negative effects on the classification accuracy (DeFries et al., 1995). In particular the incorporation of multi-seasonal geostatistical, curvature and spectral indices that is proposed in this study may imply a large increase in the dimensionality and complexity of the datasets being used (e.g. hundreds of variables), and this large volume of data can exceed the ability of the algorithm to accurately deal with it. Adding more information may be useful for the classification process, e.g. the distinction of specific classes (Rodriguez-Galiano et al., 2012a), but in general it is not known beforehand which type of information is needed to most accurately classify a specific crop category (Peña-Barragán et al., 2011). Beneath possible positive effects on the classifier performance (accuracy), performing feature selection is an interesting means to deepen the understanding of which input features are the most suitable for specific applications. For instance Rodriguez-Galiano et al. (2012a) incorporated up to 330 variables from multi seasonal spectral and geostatistical functions to assess the performance of Breiman´s RF classifier, and Peña-Barragán et al. (2011) employed 336 multi seasonal spectral, textural and other object features using a DT classifier. Gessner et al. (2013) employed RF to assess the importance of features for regression applications.

The aim here is to purposefully provide a large number of features and to select the most informative variables for crop classification. In doing so feature selection helps to increase accuracy by alleviating the effect of high dimensionality, to identify the most informative features, to speed up processing time, and to enhance the generalization capability of the algorithm. Dimensionality reduction is achieved either by (i) combining features to find a data space with lower dimensionality for representing the data while satisfying a given criterion, called feature extraction, or (ii) reducing the dimensionality of the input data by selecting a subset of relevant features (e.g. for classification), called feature selection. Both aim at retaining the most relevant representation of the input data in a lower-dimensional space. One basic example for the former is calculating VIs from spectral reflectance information, like the NDVI. Unsupervised feature extraction algorithms work without additional training and seek to find a data space with lower dimensionality for representing the data while satisfying a given criterion, not class discrimination as in supervised feature selection. The PCA, sometimes referred to as eigenvector, Karhunen-Loeve, or Hotelling transformation, considers second-order statistics and aims to de-correlate variables of a given pixel or object vector. It belongs to the multivariate statistical techniques that seek to select uncorrelated linear combinations of variables. This is done by a coordinate transforming the original input data into an uncorrelated space where most of the data variance is contained along the first axis, the principal components which maximize the variance represented by each component (Richards and Jia, 2005). The transformation consists of three steps and is based on the image-specific variance-covariance matrix cov of the input data (Tso and Mather, 2009). cov is calculated by

$$cov = \frac{\sum_{i=1}^{N}(x_i-m)*(x_i-m)^T}{N-1} \qquad\qquad Eq. \ 4\text{-}1$$

where x_i is the individual pixel or object vector, N the total number of pixel or object vectors, and m the mean pixel or object vector of a pixel or object x_i:

$$m = \frac{1}{N}\sum_{i=1}^{N}x_i \qquad\qquad Eq. \ 4\text{-}2$$

The eigenvectors and eigenvalues of cov are extracted by solving equation:

$$(cov - \lambda_j I) * G_j = 0 \qquad\qquad Eq. \ 4\text{-}3$$

where $G_j = (g_1, g_2, ..., g_k)^T$ is the eigenvector corresponding to the eigenvalue λ_j, k number of feature space dimensions, and I the identity matrix. The new coordinate system is finally computed using the normalized eigenvectors of cov and each value of x is then projected into the coordinate system (Tso and Mather, 2009).

For the latter strategy for dimensional reduction, namely feature selection, there exist different methodological approaches, most typically divided into three categories (Blum and Langley, 1997): filters (e.g. PCA), wrappers (e.g. recursive feature elimination for SVM (SVM-RFE) by Guyon et al. (2002)), and embedded approaches that fully integrate feature selection and classifier design together[17]. Filters are independent from the classifier design, and feature selection is performed before the classification. Wrapper methods on the other site search an optimal subset of features for a given classifier based on criteria that express the accuracy or generalization capability of the considered classifier. The wrapper algorithm need not be the same as the classifier. Finally the embedded methods are a class of feature selection strategy where feature selection and classification design are fully integrated together (Bazi and Melgani, 2006), e.g. in the RF framework. In general filters and embedded

---

[17:] A more detailed survey of different feature selection strategies can be found in Guyon and Elisseeff (2003).

approaches are computationally efficient and frequently used for feature ranking (Blum and Langley, 1997). Wrappers in turn evaluate features by performing internal classifications with a given test algorithm. Wrappers are attractive because they evaluate features with respect to an inductive algorithm and the resulting feature subset might be better suited to this algorithm, and ultimately achieve better accuracies. But there is controversy in this issue, as feature selection methods independent from the classifier can result in even better accuracies (Bazi and Melgani, 2006; Pal and Foody, 2010).

Different feature selection strategies will result in different feature subsets, and there exist SVM specific algorithms like the SVM-RFE, a wrapper approach specifically developed for SVM. Its ranking is based on the objective function $\|w\|^2/2$ (see appendix A.2) as a feature-ranking criterion to produce a list of features ordered by apparent discriminatory ability (Guyon et al., 2002). However, Pal and Foody (2010) demonstrate the influence of different feature selection algorithms (among them the RF and SVM-RFE) on the performance of SVM in pixel-based classification of hyper-spectral data. They demonstrated that, for a wide range of different training set and feat space sizes, applying different feature selection methods does not necessarily result in significant differences in SVM classification performance, and the resulting feature sets that yielded highest accuracies do not significantly differ in size (number of features), albeit different kind of features are selected. The use of feature selection methods that are independent from the SVM are frequently reported (Fassnacht et al., 2012) and can even result in better accuracies than methods that exploit criteria intrinsically related to the SVM classifier properties, e.g. SVM-RFE (Bazi and Melgani, 2006; Pal and Foody, 2010). Further, the problem of hyper-parameter selection in conjunction with feature selection when using SVM-RFE remains an issue (Guyon et al., 2002).

Among the more recent non-parametric algorithms that can be used as embedded strategy in feature selection are DT (Peña-Barragán et al., 2011), bagged trees (BT) (Prasad et al., 2006), or RF (Rodriguez-Galiano et al., 2012a). The RF can be used to select a smaller subset of features from a huge input data space that can result in equal (Chan et al., 2008) or improved classification accuracy of another classifier (Löw et al., 2013a; Pal and Foody, 2010). In this regard an important feature of the RF is the calculation of predictor variable or feature importance ranking (Breiman, 2001). A more detailed description of the RF algorithm can be found in appendix A.2. Several measures for feature importance with RF exist: (i) counting the frequency that each feature is selected during the tree-building process by all n trees in the ensemble, and (ii) the permutation importance is calculated as the difference in prediction accuracy before and after randomly permuting each of the f features while keeping all other features unchanged and running all trees on the out-of-bag (OOB) sample again. The result is a decrease of accuracy as a consequence of the permutation process. For each tree the votes for the correct class in the changed data set (without the permuted fth feature) is counted, and this number is subtracted from the number of votes for the correct class from the unchanged data set. This value is averaged over all trees in the ensemble to generate the raw importance score for a given feature, hereafter denoted I_f. A scaled version of I_f, the so-called Z-score can be derived from this number by normalizing I_f by its standard error, which can be used to assign an importance level to the features and to generate a ranked list of the features (Breiman, 2001); iii) the Gini importance is calculated as mean improvement in the Gini gain splitting criterion, which is produced by each feature over all n trees.

4.2.2. Classifier description and parameter tuning

In the following, the SVM and the RF algorithms are briefly described with focus on their implementation and parameterization. A more detailed description on the algorithms including a brief review on their application in remote sensing studies is given in appendix A.2.

The implementation of SVM in the package e1071 (Meyer et al., 2012) in R provides an interface to the libsvm approach from Chang and Lin (2011). SVM is based on the notion of separating classes in a higher dimensional features space by fitting on optimal separating hyper-plane (OSH) between them, focusing on those training samples that lie at the edge of the class distributions, the so-called support vectors (Foody and Mathur, 2004). Multi-class problems are solved using the one-against-one (OAO) voting approach. The multi-class problem is split into $n(n-1)/2$ binary classification problems (where n = number of classes) and finally the outputs from the individually trained binary classifiers are combined by a majority voting strategy (Hsu and Lin, 2002). The seconds approach, one-against-all (OAA) splits the multi-class problem into n individual cases (where each single class competes against the rest). Training of the SVM includes choosing the kernel parameter γ ("gamma" in e1071) and the regularization parameter C (termed "cost" in e1071), which are sometimes referred to as hyper-parameters, where γ determines the width of the kernel, and C controls the penalty associated with misclassified training samples. Both have an impact on the classification accuracy (Burges, 1998), although SVM were also shown to be partly robust to changes in C (Belousov et al., 2002). Taking larger and larger values for C increases the magnitude of penalties that are associated with training samples on the wrong site of the hyper-plane, but can result in a model over-fitted to the training data and limit generalization capability (Belousov et al., 2002; Foody and Mathur, 2006). Tuning of the hyper-parameters C and γ was done using a systematic grid search in 2-D space that is spanned by γ and C. The range of γ was [0.00125, 2], the range of C was finally set to [1, 200], after a wider range of values tested did not achieve higher accuracies, but taking a smaller range of C resulted in much shorter training times. The widely used radial basis function (RBF) kernel was selected in this study, since linear or polynomial kernels were tested but resulted in lower accuracies (not reported here). It was reported that the accuracy of SVM with RBF kernel can degrade if the input data is not normalized (Ali and Smith-Miles, 2006), e.g. attributes with large original scale can bias the solution. Hence, the data space was normalized to a common scale [0, 1].

The implementation of Breiman´s RF in the randomForest package in the R programming environment was used for variable selection in this study (Breiman, 2001). Two free parameters can be optimized in the RF. The number of trees T at which an optimal accuracy level is achieved varies with the number of samples and features, and with the variability of feature values. The number of trees that usually achieves good results and therefore is considered adequate is 500 (Gislason et al., 2006), but a smaller number leads to faster computation and can suffice if the OOB error converges (see section 5.2.1). The second free parameter relevant for accurate classifications is the number of features m_{try} to split the nodes (Breiman and Cutler, 2007). As is considered adequate in literature (for categorical classification) the number of features at each node was set to the square root of the total number of input features, $\sqrt{f}$, where f is the number of predictor variables (features) within a dataset (Gislason et al., 2006).

4.2.3. Measures of accuracy and uncertainty

For the classification evaluation, a set of accuracy metrics was employed to evaluate the performance of the SVMs on the subspaces resulting from the RF feature selection. Accuracy metrics were derived from the confusion matrix to assess the accuracy of the hard results (class labels) of the SVMs (Congalton, 1991): overall classification accuracy (OA), defined as the total proportion of correctly classified fields per total number of test fields, is one of the most common measures of classification performance in remote sensing (Foody, 2002).

A class-wise measure of accuracy (CA$_i$), the F_β-measure of Van Rijsbergen (1979), was employed for each class i under investigation, defined as as

$$CA_i = F_\beta = (1 + \beta^2) \frac{pr_i * tp_i}{\beta^2 * pr_i + tp_i}$$ *Eq. 4-4*

where tp_i is the true positive rate (TPR) which gives the proportion of samples classified into class i among all samples which truly have class i, and pr_i the precision which gives the proportion of samples which truly have class i among all samples which were classified as class i. The fomer determined the error of omission (false exclusion), the latter the error of commission (false inclusion). The traditional F_β-measure, which was used in this study, gives equal importance to tp_i and pr_i ($\beta = 1$), and is sometimes referred to as F_1 measure. All CA_i may be averaged to get the average class-wise accuracy (AA):

$$AA = \sum_{i=1}^{n} CA_i$$ *Eq. 4-5*

Cohen´s kappa is another measure that takes into account agreement occurring by chance (Cohen, 1960). It is defined as

$$k = \frac{pr_o - pr_e}{1 - pr_e}$$ *Eq. 4-6*

where pr_e is the proportion agreement expected by chance, and pr_o the proportion that was observed by agreement. A perfect agreement between all observations and predictions results in $k = 1$. If there is no agreement except by chance then $k = 0$.

Several measures of uncertainty can be employed to assess the influence of feature reduction on the spatial classification uncertainty and for comparison of the feature selection strategies. Depending on the algorithm used for classification different approaches exist to model or generate information on the uncertainty in classification (Breiman, 2001; Foody, 1995a, 1996a; Foody et al., 1992; Giacco et al., 2010; Maselli et al., 1994). Beneath the final class label, algorithms like SVM can compute, for each classified object or pixel x, a soft output (Giacco et al. 2010) that provides an estimation of the membership degrees of x to the investigated classes. This soft output is in the form of a vector, which contains the class membership estimations associated with x:

$$pr(x) = \{pr_1(x), pr_2(x), \dots, pr_i(x), \dots, pr_n(x)\}$$ *Eq. 4-7*

where $pr_i(x)$ is the membership degree of a pixel or field x to class i, and n the number of classes. Each of the elements in $pr(x)$ can be interpreted as a degree of belief or probability that a case x actually belongs to a class i. These probability estimates generally take their values in a closed interval (e.g. $pr_i(x) \in [0,1]$). The maximum value in $pr(x)$, hereafter denoted $pr_{max}(x)$, belongs to the class that is usually taken as final class when the soft results are transformed into a hard one, e.g. the final class $C_{fin}(x)$ that is most likely to occur:

$$C_{fin}(x) = arg\ max(pr(x))$$ *Eq. 4-8*

From the probability vector $pr(x)$ the following uncertainty measure can be calculated for x as

$$E = 1 - pr_{max}(x)$$ *Eq. 4-9*

E is an expression of the strength of the class assignment and can be used to quantify map uncertainty. A potential disadvantage of E is that it fails to capture the entire distribution of probabilities

in $pr(x)$, and thus is limited in giving information of possible confusion with other classes. Let in a hypothetical four-class example the first element in $pr(x)$ be 0.51 (here giving 51 % probability that a class is correctly classified). $pr_{max}(x)$, which is 0.51, does not take into account the subsequent ranked probabilities in $pr(x)$ which could be 0.20, 0.15, and 0.14 for the second, third, and fourth class, respectively, resulting in limited doubt about the final class allocation. However the probability of the second element in $pr(x)$ could be 0.49. In the latter case the secondary ranked class has a probability that is very close to $pr_{max}(x)$ and there is considerable doubt about the final class allocation. A solution to handle this is using measures of entropy, which take into account all elements in $pr(x)$. Among the most-widely used uncertainty measures in remote sensing applications that summarizes all information contained in $pr(x)$ is the Shannon entropy (Brown et al., 2009; Maselli et al., 1994; Zhu, 1997), originating from information theory (Shannon, 1948). The Shannon entropy H can be calculated from $pr(x)$ for each classified pixel or object x as:

$$H(x) = - \sum_{i=1}^{n} pr(x_i) \, log_2 pr(x_i) \qquad\qquad Eq.\ 4\text{-}10$$

where $H(x)$ is the entropy of x, n the number of classes, and $pr_i(x)$ the probability of occurrence of class i. $H(x)$ can assume values in the interval [0,log$_2$n], and is therefore dependent on the number of classes n used, a field or pixel with a maximum value of $pr_{max}(x)$ has an entropy equal to zero. The ratio of the observed to the maximum possible entropy may be used to compare uncertainties from maps with different number of classes (Maselli et al., 1994). It is calculated as:

$$H_{rel}(x) = \frac{H(x)}{H_{max}(x)} * 100 \qquad\qquad Eq.\ 4\text{-}11$$

where $H_{max}(x)$ is the maximum possible value of $H(x)$ (when all $pr_i(x) = 1/n$). $H_{rel}(x)$ will reach zero if an object is found to have maximum probability of belonging to one class ($pr_{max}(x)$ = 100 %) and reaches its maximum level if $pr_i(x)$ of each classes is partitioned evenly between the classes (e.g. $pr_i(x)$ = 0.25 for each of four possible classes). The advantage of $H_{rel}(x)$ is that it captures the entire distribution of the probabilities within each element in $pr(x)$, and takes into account the differences between the maximum and all subsequent ranked probabilities of the classes within the vector.

The α-quadratic entropy (Pal and Bezdek, 1994), also termed alpha score, is another measure of uncertainty and is defined as

$$H_\propto(x) = \frac{1}{n*(2^{-2\propto})} \quad * \quad \sum_{i=1}^{n} pr_i^\propto(x)\,(1 - pr_i(x)\,)^\propto \qquad\qquad Eq.\ 4\text{-}12$$

where $pr(x)$ is the vector that contains the soft outputs, n the number of classes, and α an exponent that determines the behaviour of $H_\alpha(x)$. The advantage of this measure is that it summarizes all the information contained in $pr(x)$ and commits the probabilities of the other classes in the uncertainty evaluation. Further it has a higher sensitivity as compared with the Shannon entropy $H(x)$ when the components of $pr(x)$ change. When values of α increase from 0 to 1, $H_\alpha(x)$ will become more and more selective if the components in $pr(x)$ tend toward equalization. In this study α = 0.5 was chosen. The ratio of the observed to the maximum possible $H_\alpha(x)$ was calculated to normalize $H_\alpha(x)$ to the same scale as E, [0, 1], hereafter denoted $H_\propto rel$. If a field is found to have maximum probability of belonging to one class ($pr_{max}(x)$ = 1) then $H_\propto rel$ will reach zero, and if each element in $pr(x)$ is partitioned evenly (e.g. $pr_i(x)$ = 0.25 for each of four possible classes) then $H_\propto rel$ will reach one. In this case the identification performance of the classifier with respect to different crops is low, resulting in higher classification uncertainty, e.g. because of heterogeneous fields, the presence of subfields, etc.

Figure 4-1 shows the influence of α on $H_\alpha(x)$. The uncertainty measure becomes very selective as α increases from 0 to 1. Critically with α close to zero $H_\alpha(x)$ will be close to one and thus the same degree

of uncertainty, so the uncertainty measure is not sensitive to changes in the elements within $pr(x)$. The lower the value of α the more selective becomes $H_\alpha(x)$ when $pr_i(x)$ deviates from zero or one. Hence, a value of $\alpha = 0.5$ was chosen as a good trade-off between these two extremes. Finally, two measures of uncertainty were chosen for the experiments: E that was calculated from the maximum value in $pr(x)$, and $H_\alpha rel$ that was calculated by including all elements in $pr(x)$.
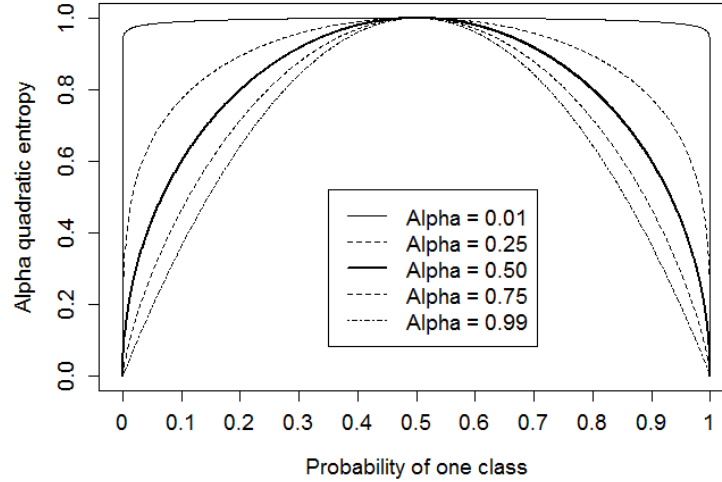
Figure 4-1: Influence of α on $H_\alpha(x)$ of a pixel or object in a hypothetical two-class problem. The x-axis shows the probability $pr_i(x)$ of one class, scaled to [0,1].

4.2.4. Derivation of uncertainty measures from libsvm

For the SVM there has been considerable research to extend the two commonly used architectures, OAO and OAA, to retrieve and to use multi-class soft outputs (Foody, 2008; Giacco et al., 2010; Waske and Benediktsson, 2007; Wu et al., 2004). The standard formulation of a SVM that adapts the OAO strategy does not provide estimations of class memberships. One example for deriving soft outputs from SVM that employs the OAO approach can be found in Hastie and Tibshirani (1998). Giacco et al. (2010) provides an example to extend SVM to produce multiclass soft outputs based on the statistical Bradley-Terry model and pairwise coupling (Hastie and Tibshirani, 1998).

libsvm by Chang and Lin (2011) can produce a soft output $pr(x)$ next to the hard class label. libsvm employs the OAO strategy and builds $n(n-1)/2$ binary SVMs, where n is the number of classes. Each of these binary SVMs separates two classes and computes, for each sample x, a distance that this sample has to the OSH. The class probability is then calculated from these $n(n-1)/2$ distances using an improved implementation (Lin et al., 2007) of Platt's a posteriori probabilities (Platt, 2000) where a sigmoid function is fitted to the decision values $f_j(x)$ of each of the binary SVM classifiers. The soft output of each binary SVM classifier is then combined as proposed in Wu et al. (2004). Further information on producing soft outputs with libsvm can be found in Karatzoglou et al. (2006).

4.2.5. Selection of feature subspaces

In this study SVM classifications using feature subspaces of different sizes were undertaken in which the dimensionality of the input data set was varied. This was done by incrementally adding features in order suggested by the ranked feature list generated with the RF, where the importance of the features is assessed by means of the MDA. The feature were added in groups of fifteen, beginning with the 10 most important features from the ranked list, then continuing with the second 25 most important and so on. OA and two uncertainty metrics, E and $H_\alpha rel$ were computed to evaluate the classification result

at each incremental step. Beyond this a more detailed evaluation of classification accuracy and uncertainty was restricted to four specific SVM classifications representing feature subspaces with distinct properties that might be defined for practical considerations, similar to the experimental analysis of uncertainty for other classifier algorithms (Foody, 1995a; Loosvelt et al., 2012a; McIver and Friedl, 2001):

> FS_1: SVM applied to all features. The features from the five groups (see chapter 3) were concatenated into one stacked vector before the SVM classification, without feature selection.

> FS_2: SVM applied to the subset of the features that yielded the highest OA. The features were ranked according to the RF importance score I_f of each feature on each acquisition date. The features were then removed from the concatenated stacked vector according to their importance score. In doing so the removed features were not necessarily the same for all acquisition dates. For each classification the OA statistic was calculated, and finally the SVM with the optimal feature subspace that achieved highest OA was selected.

> FS_3: SVM applied to as few features as possible that resulted in OAs that are not substantially lower than the FS_2 (e.g. more than 5 % deviation), and still reach the commonly recommended accuracy of 85 % (Anderson et al., 1976). It has to be noted that in remote sensing studies no clearly defined levels for the evaluation of classification accuracy exists (Foody, 2002). This strategy aims at further reducing the number of features, but maintaining a classification accuracy that is non-inferior to that derived from the FS_2 strategy. The reason to do so is that features selection can have positive impacts beyond accuracy, e.g. reduced computation time or data storage requirements (Bazi and Melgani, 2006; Pal and Foody, 2010).

> FS_4: From a pragmatic point of view selecting only relevant acquisition dates can be advantageous both for financial and computational reasons. Selecting "optimal" acquisition dates is not straightforward and most often based on previous empirical analysis of the impact of varying acquisition date compositions on classification accuracy (Murakami et al., 2001; Van Niel and Mcvicar, 2004; Peña-Barragán et al., 2011). Critically, a few acquisition dates can suffice to fulfil a certain level of accuracy, or adding more and more dates does not significantly improve the classification accuracy beyond a certain level (Murakami et al., 2001; Van Niel and Mcvicar, 2004). Here it is proposed to adapt and further develop the idea of Loosvelt et al. (2012a) and to define the relative importance of the acquisition dates by computing the average importance score $I_{t,avg}$ of all features within each single time step. This procedure yields a ranked list of acquisition dates, starting with the most important date, from which single dates are subsequently removed in order of increasing importance until a certain threshold of classification accuracy is achieved (e.g. 85 %). In doing so only the most important acquisition dates are withheld and classified with the SVM.

To enhance the reliability of the experiments each classification with SVM was repeated 50 times using different training and testing sets, respectively, by means of a repeated random subsampling. Finally the accuracy metrics and uncertainty measures defined above were averaged over the 50 model runs for each feature subspace that was classified.

4.3. Results and discussion

In the following sections the results from the experiments are given. First the classification performance of the five feature groups is compared separately. Afterwards the feature groups are concatenated into one stacked vector and classifier performance is compared using the total data set, before and after dimensionality reduction. Second the results of the classifications of the different feature subspaces (FS_{1-4}) defined above for the SVM are detailed. Third the effect of feature selection on spatial classification uncertainty is illustrated, and the causes for the spatial distribution pattern of uncertainty are discussed.

4.3.1. Classification of spectral and spatial features

The first experiment is based on the combination of the information from the five feature groups that were defined in chapter 3. First, each group is classified separately with the SVM. In the second experiment all features were concatenated into one stacked vector, and then the RF based variable importance analysis is performed with ten repetitions, and the SVM that yielded highest accuracy (FS_2, see section 4.2.5) was selected. For the sake of clarity only the RF feature importance score and the PCA are compared for the dimensionality reduction, but theoretically a variety of feature selection methods could be tested instead, like the sequential forward floating selection (SFFS) (Pudil et al., 1994), steepest ascent (SA) techniques (Serpico and Bruzzone, 2001), or kernel-PCA-based techniques (Fauvel et al., 2008).

The results are summarized in Table 4-1 – Table 4-4. As can be seen from these tables, there was no clear "winner", e.g. feature groups A–D gave complementary results, however using geostatistical features (group E) gave the least accurate results. When the feature groups were concatenated into one stacked vector (column "All combined" in Table 4-1 – Table 4-4), the overall classification accuracy decreased in three sites (except for KHO) as compared with the best result from the SVM classifications based on the single feature groups. Yet, when the RF-based feature selection was performed on this stacked vector, the SVM classification on this feature selected input set outperformed all other feature groups in terms of OA, Cohen´s k, and AA (column "FS_2" in Table 4-1 – Table 4-4). The better classification accuracy was achieved with fewer features than 568, partly less than 15 % of the total feature set: KHO: 85, KKP: 100, KYZ: 145, respectively, and FER: 115. The classification performance on the non-reduced input in KYZ and FER were already at a comparatively high level, and the differences after feature selection were less pronounced. An interesting feature is that CA_i of sorghum/maize in KKP was 60 % after feature selection applied to the data set with all features (but 0% for the single feature groups). Further, winter wheat fields tended to be classified more accurately when applying the SVM after the dimensionality reduction in KKP and KHO. These results implicate that there is complementary information in the five feature groups. In Table 4-1 – Table 4-4 it can be seen that the classes for which each feature group performs best were complementary. For instance, in KHO group-C features performed best for cotton, sorghum/maize, fruit trees, and wheat-other, whilst group-A features were better for fallow fields, winter wheat, and rice. Similar, in KYZ features from group-C were best in distinguishing winter wheat fields, whilst group-A features perform better for fallow fields (Table 4-3). Making use of the complementary information in the different groups had advantages and gave more stable results (e.g. increased accuracies for most crop classes), and all of the classes were better classified than the worst case of the single feature groups. This indicates that the SVM actually could make use of the complementary information of the single feature groups, and providing the input data with all 568 features in combination with a feature selection strategy yielded best results. In general terms this effect was less pronounced in KYZ and FER, with only 5 classes.

Table 4-1: Summary of class-wise (CA$_i$), average class-wise (AA) and overall accuracies (OA) [%] on the test set for the SVM classification in KHO. The Kappa coefficient of variation k is also given. The crop classes include cotton (C), fallow (FA), rice (R), sorghum/maize (S), fruit trees (F), winter wheat (W), and wheat-other (WO).

	Group A	Group B	Group C	Group D	Group E	All combined	FS_2
OA	86.03	79.21	84.87	84.39	63.60	87.65	91.02
AA	78.55	64.10	74.78	70.43	44.51	74.54	84.03
k	82.74	74.19	81.39	80.66	54.42	82.80	89.20
No. features	80	208	128	48	104	568	85
C	91.67	85.00	93.68	89.44	75.54	90.50	93.99
FA	92.68	87.89	89.61	87.84	78.83	95.70	98.06
R	71.08	58.24	63.91	69.85	21.77	71.40	75.00
S	29.64	0.00	29.26	0.00	0.00	9.00	36.39
F	90.02	84.41	90.95	90.37	72.11	87.20	90.66
W	87.97	47.76	65.56	66.14	0.00	85.70	100.00
WO	86.78	85.40	90.51	89.36	63.31	91.30	94.15

Table 4-2: Summary of class-wise (CA$_i$), average class-wise (AA) and overall accuracies (OA) [%] on the test set for the SVM classification in KKP. The Kappa coefficient of variation k is also given. The crop classes include cotton (C), fallow (FA), rice (R), sorghum/maize (S), alfalfa-1y (A1), melons (M), winter wheat (W), and wheat-other (WO).

	Group A	Group B	Group C	Group D	Group E	All combined	$FS2$
OA	88.74	84.98	89.49	89.88	63.08	89.02	93.33
AA	70.00	63.20	73.66	74.64	40.11	78.57	86.41
k	85.62	80.81	86.60	87.11	50.14	84.70	92.3
No. features	80	208	128	48	104	568	100.00
A1	94.98	81.99	97.52	94.51	66.53	100.00	100.00
C	81.42	86.64	87.09	87.44	57.28	76.60	93.60
FA	87.62	82.10	91.06	91.06	47.71	86.00	93.20
M	52.81	26.54	53.19	57.54	0.00	57.10	84.20
R	94.63	94.05	91.56	93.55	83.17	94.40	94.40
S	0.00	0.00	0.00	0.00	0.00	30.80	60.00
W	96.62	92.86	97.34	96.86	66.17	97.20	98.90
WO	51.90	41.43	71.48	76.14	0.00	54.50	75.00

Table 4-3: Summary of class-wise (CA$_i$), average class-wise (AA) and overall accuracies (OA) [%] on the test set for the SVM classification in KYZ. The Kappa coefficient of variation k is also given. The crop classes include fallow (FA), rice (R), alfalfa-1y (A1), alfalfa-3y (A3), and winter wheat (W).

	Group A	Group B	Group C	Group D	Group E	All combined	FS_2
OA	92.86	92.01	91.06	92.19	82.59	91.51	94.23
AA	77.23	72.78	75.95	68.27	54.56	70.98	87.20
k	86.36	84.76	83.01	85.06	66.40	87.00	89.90
No. features	80	208	128	48	104	568	145
FA	89.24	81.01	72.75	82.27	52.38	74.80	83.00
R	80.72	83.07	80.82	82.62	64.69	99.60	99.80
A1	78.41	76.68	77.35	76.75	60.64	76.90	92.50
A3	99.66	99.62	98.97	99.70	95.09	83.60	90.70
W	38.11	23.52	49.83	0.00	0.00	20.00	70.00

Table 4-4: Summary of class-wise (CA$_i$), average class-wise (AA) and overall accuracies (OA) [%] on the test set for the SVM classification in FER. The Kappa coefficient of variation k is also given. The crop classes include cotton (C), fallow (FA), fruit trees (F), winter wheat (W), and wheat-other (WO).

	Group A	Group B	Group C	Group D	Group E	All combined	FS_2
OA	93.51	94.22	91.59	93.96	76.27	91.41	94.60
AA	90.62	89.03	87.63	83.55	82.50	92.00	92.44
k	91.67	92.67	88.99	92.30	66.31	90.10	93.60
No. features	80	208	128	48	104	568	115
C	96.97	97.57	96.62	96.88	88.81	100.00	100.00
FA	93.24	93.24	84.96	93.25	0.00	96.00	96.00
F	95.78	94.52	93.72	92.68	66.41	85.70	92.00
W	74.60	78.00	60.70	80.98	0.00	85.70	80.00
WO	92.50	93.67	91.97	93.90	78.66	92.60	94.20

In Table 4-5 the SVM is compared to the MLC, which assumes Gaussian distribution in the data, and to the RF as classifier algorithm. Further, the RF-based feature selection was confronted with another strategy for feature reduction, the PCA. The RF is assumed resistant to the input space size, and therefore was directly applied to the non-reduced data set[18]. In all sites, the SVM that was applied to the dimensionality-reduced data sets outperformed the MLC and RF, respectively. Considering PCA, the feature reduction that retained 85 % of the variance in the principal components (PC) achieved the best results for SVM. The number of transformed features retained was: 19 (KYZ), 18 (FER), 22 (KKP), and 18 (KHO) (not shown in Table 4-5).

Those PCs that accounted for 95 % of the variance in the input data achieved best classification accuracies with MLC. The number of transformed features (the number of PCs retained) that corresponds to 95 % of the total variance criterion was (not shown in Table 4-5): 53 (KYZ), 75 (KKP), 49 (KHO), and 48 (FER).

In all sites except for FER did the RF-based feature selection yield higher accuracies than the PCA when using the MLC or SVM as classifier algorithm. Another interesting aspect revealed in Table 4-5 is the variability of the results among the test sites, e.g. the absolute difference of OA was lowest for the SVM after the RF feature selection was applied (3.58), as contrasted with the MLC (17.50) or the RF (5.33). Overall, the SVM (after feature selection) outperformed the RF and the MLC as classifier algorithms. The RF-based feature selection strategy had a better ability to select relevant features for higher classification accuracies than the PCA. Both classifier algorithms (MLC and SVM) showed a susceptibility to a huge input space, and could benefit from dimensionality reduction.

Table 4-5: Overall accuracies [%] using different classifier algorithms, and the SVM on the dimensionality reduced input data. RF and PCA were tested for feature reduction. Lowercase abbreviations "_PCA95/85" and "_RF" indicate the selected feature reduction method.

	MLC	MLC$_{PCA95}$	MLC$_{RF}$	RF	SVM	SVM$_{PCA95}$	SVM$_{PCA85}$	SVM$_{RF}$
KHO	58.24	75.00	77.65	89.50	87.65	80.85	87.76	91.02
KKP	53.93	73.01	77.30	89.02	89.02	87.69	88.09	93.33
KYZ	66.41	85.42	87.06	92.71	91.49	88.94	91.95	94.23
FER	71.43	88.46	84.87	94.35	91.41	90.38	92.94	94.60
Abs. difference	17.50	15.45	9.76	5.33	3.84	9.53	5.18	3.58

---

[18] : This assumption is confirmed with experimental evidence presented in chapter 5.

4.3.2. Feature importance

This section details the composition of features that achieved the peak accuracy, corresponding to the FS_2 subspace (see section 4.2.5). From Figure 4-2 some general trends in the importance pattern can be revealed. The predictive capacity of the feature groups (here indicated by the number of features used) varied as a function of time. Spectral features from group B and C dominated over all time steps and in all sites, but its relative importance is highest during the two peaks in spring and late summer in KKP, KHO, and FER. More than 50 % of the top-ranked features defined in May and the beginning of September were the most used in KKP and KHO. This bi-modal distribution reflects the dominating presence of crop rotations in KKP and KHO, and their absence in KYZ where features contribute almost equally over all time steps with slightly more features in the first half of the growing period. Group C features contribute the highest portion in the critical time steps in KHO and KKP over the growing season, whilst the influence of the other groups dominates in the other time steps. For instance, geostatistical features (group E) were used the most between 14–Jul and 13–Aug in KKP and KHO. Geostatistics appear to have higher importance around mid to late summer, when crops reached the final stages of crop development in the flowering stage, although their importance was not obvious in FER. In contrast, the relative importance of all groups was more equalized in KYZ for all time steps.

The range of the median NDVI values of all agricultural fields per acquisition date showed important variations through the vegetation period (Figure 4-2, dashed line). The difference of NDVI reflects the phenological variability of the crops in the test sites, and the presence of winter and summer crops. For instance the difference of crop median NDVI values was maximum in spring in KHO and FER (maximum of greenness of winter crops and bare soils of summer crops), and late summer (maximum greenness of summer crops). In spring, as winter crops were in growing stage, the fields started to differentiate from the bare soil fields of the summer crops, and difference of NDVI increased. At that time the variation in crop phenology was high and geostatistical features became more important, albeit they still played a minor role as compared with other groups. The relatively high importance of spring acquisition dates in May (KHO and FER) and early July (KKP) was indicated by the higher number of features used. The difference was minimum when winter and summer crops were still bare soil (April), in between the vegetation peaks of winter and summer crops (June–July), and after harvest. This pattern was less pronounced in KYZ, where winter crops were absent and the NDVI difference was lowest in April (bare soil) and September (after harvest).
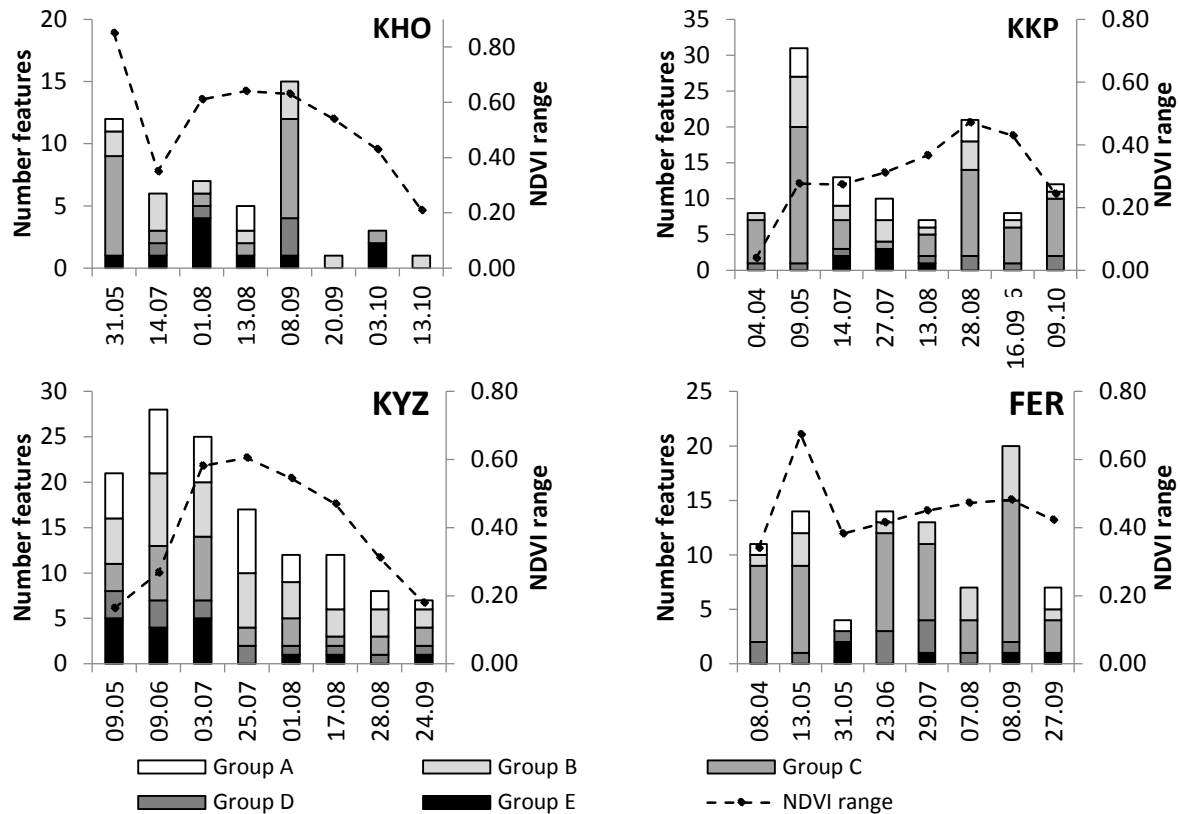
Figure 4-2: Group wise number of top ranked features per acquisition date used in the SVM model based on the FS_2 feature subset. The x-axis shows the acquisition date of the corresponding satellite images. The range of median NDVI values for all crop classes is shown as a dashed line.

There is indication that the relative importance of the acquisition dates reflects this pattern, e.g. when NDVI difference is higher the acquisition dates tends to be more important. Pronounced rates of change in the NDVI difference are observed in between the two peaks of winter and summer crops (except for KYZ) when crop state changes rapidly. Group-C features, which contain red edge information, were found to be most prominent in all experiments. In particular they strongly influenced the success of RF when NDVI difference was highest. This indicates the predictive strength of the red edge canal in discriminating crops during the critical phases of the growing period. The relative importance of geostatistics tends to be highest in the period of rapidly changing crop stages after winter crops have been harvested and before the summer peak. This points to the advantage of geostatistical features over the spectral indices when seasonal state differences between the crop types.

The MDA measure from the RF, here denote as I_f, is used to assess the relative importance of variables. After reducing the number of features, according to the targets defined above for the subspaces FS_{2-4}, the relative contribution (expressed as number of features) of the five feature groups A-E was analysed for FS_2 and FS_3 (Figure 4-3), as in these subspaces the number of feature types retained differs, whilst in FS_1 and FS_4 always 71 different feature types per acquisition date are present. When looking at the proportions of features from the groups that contribute to the FS_2 and FS_3, one striking characteristic is the general dominance of vegetation indices that contain red edge information in all study sites (group-C).
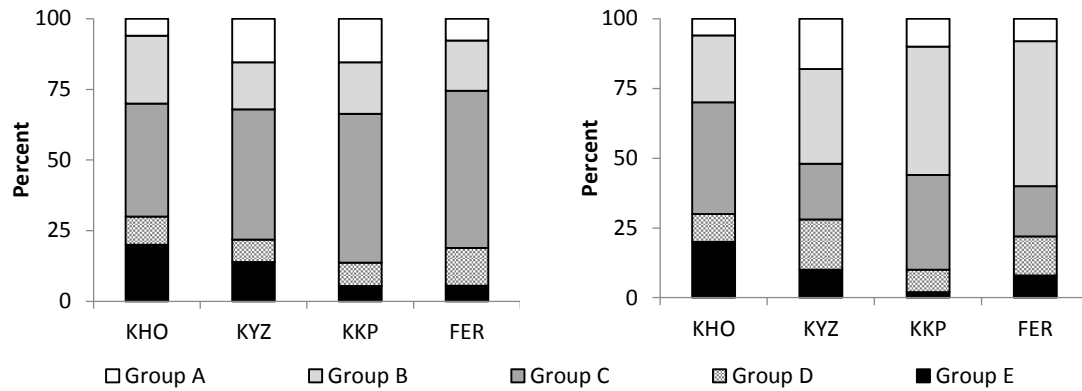
Figure 4-3: Relative contribution of the feature groups in the FS_2 (left) and FS_3 (right) models, expressed in terms of the percentage of the total number of features in FS_2 and FS_3, respectively, irrespective of the vegetation season.

Interestingly when using the FS_3 instead of FS_2 subspace the relative contribution of group B and group C almost reversed, except for in KHO. Geostatistics from group E tend to be irrelevant, most obviously because of the relatively homogeneous nature of the objects (crop fields). Assuming that averaging the importance scores of all features within a group gives an indication of the relative importance of this group, $I_{g,avg}$, a ranking of the groups was established (Table 4-6). In the majority of cases group C has a larger relative importance than the rest of the groups, and group D tends to be the second most important. This is also striking because the absolute number of feature types in group C is smaller (8) than in group B (13). Critically the standard deviations of $I_{g,avg}$ reveal a partial overlap in importance in some of the groups, so this figure rather gives a rough approximation of "group performance".

Table 4-6: Mean importance score averaged over all features within a group ($I_{g,avg}$) and standard deviation of $I_{g,avg}$.

	KHO				KYZ			
Group	FS_1	FS_2	FS_3	FS_4	FS_1	FS_2	FS_3	FS_4
A	0.07 ± 0.03	0.12 ± 0.01	0.22 ± 0.01	0.09 ± 0.03	0.07 ± 0.04	0.12 ± 0.02	0.13 ± 0.01	0.07 ± 0.03
B	0.10 ± 0.04	0.09 ± 0.06	0.33 ± 0.06	0.10 ± 0.03	0.04 ± 0.03	0.09 ± 0.03	0.10 ± 0.02	0.04 ± 0.02
C	0.10 ± 0.05	0.11 ± 0.02	0.46 ± 0.02	0.11 ± 0.02	0.05 ± 0.03	0.12 ± 0.04	0.14 ± 0.04	0.05 ± 0.02
D	0.09 ± 0.03	0.12 ± 0.05	0.34 ± 0.04	0.09 ± 0.06	0.05 ± 0.03	0.11 ± 0.03	0.13 ± 0.02	0.09 ± 0.02
E	0.06 ± 0.04	0.09 ± 0.04	0.30 ± 0.03	0.06 ± 0.02	0.03 ± 0.03	0.09 ± 0.01	0.09 ± 0.01	0.05 ± 0.04
	KKP				FER			
Group	FS_1	FS_2	FS_3	FS_4	FS_1	FS_2	FS_3	FS_4
A	0.07 ± 0.04	0.17 ± 0.04	0.17 ± 0.04	0.08 ± 0.04	0.04 ± 0.03	0.15 ± 0.03	0.18 ± 0.03	0.05 ± 0.03
B	0.08 ± 0.03	0.19 ± 0.02	0.19 ± 0.02	0.10 ± 0.03	0.06 ± 0.04	0.16 ± 0.04	0.17 ± 0.04	0.08 ± 0.03
C	0.10 ± 0.04	0.23 ± 0.04	0.23 ± 0.05	0.12 ± 0.04	0.06 ± 0.03	0.14 ± 0.03	0.18 ± 0.02	0.09 ± 0.04
D	0.09 ± 0.03	0.18 ± 0.03	0.18 ± 0.03	0.12 ± 0.04	0.08 ± 0.04	0.14 ± 0.02	0.16 ± 0.02	0.07 ± 0.06
E	0.03 ± 0.02	0.13 ± 0.01	0.16 ± 0.01	0.03 ± 0.02	0.03 ± 0.03	0.15 ± 0.02	0.17 ± 0.02	0.03 ± 0.02

Figure 4-4 shows the relative importance of the five feature groups in the FS_2 subspaces with regard to the individual crop categories, irrespective of the vegetation season. In general terms, it can be seen from this figure that group C (red edge features) affects the classification the most, indicated by the overall high mean group variable importance ($I_{g,avg}$). This confirms the general relevance of the red edge spectrum in land cover classification (Schuster et al., 2012). Group A, B and D tend to be the second most important groups. Curvature indices defined in group D were decisive in the classification of melons in KKP, sorghum/maize in KHO, alfalfa-3y in KYZ, and fruit trees in FER. Geostatistics from group E tend to be irrelevant, most obviously because of the relatively homogeneous nature of the objects (crop fields). Yet, they were important for classifying cotton in KHO and fallow fields in KYZ. The relatively coarse resolution of the input data, and the relatively small field sizes (as compared with field sizes reported in Europe or the U.S., e.g. in Wardlow et al. (2007)) could be one reason, because the spatial extent of within field heterogeneity might be smaller than the pixel size.
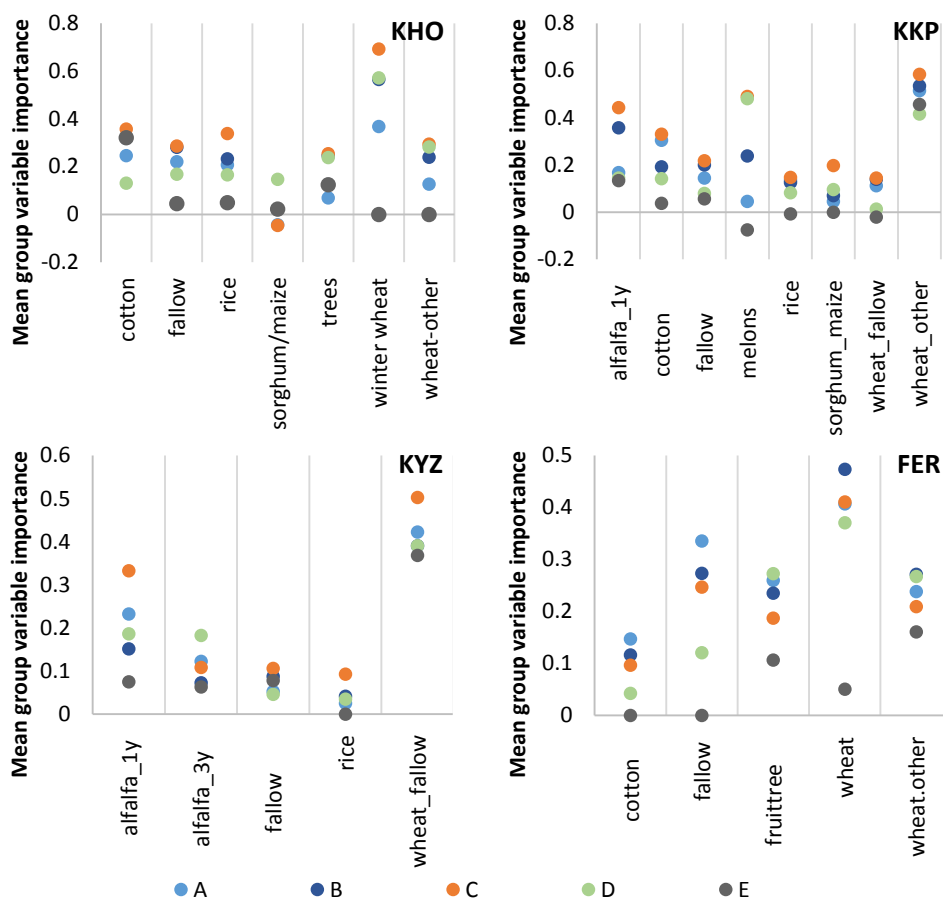
Figure 4-4: Mean relative importance of feature groups ($I_{g,avg}$) considering the FS_2 subspace, per crop type.

The analysis of the ten most informative features is shown in Figure 4-5. The importance score reveals that the composition of feature types differs among the four sites. Interestingly, variants of the NDVI, like REDNDVI or GRNDVI occurred as the most important predictors for crop types over two or more acquisition dates (e.g. GRNDVI in KYZ was ranked the most important feature in three acquisition dates). In general approximately 50 % of the 10 most important features are group C features. In three sites a feature from group C was ranked the most important, e.g. REDNDVI and GRNDVI. Most of the features are mean values, only in KYZ and KHO standard deviation values from group D were important. The acquisition dates of the single features reflect the average importance $I_{t,avg}$, shown in Figure 4-8. These findings offer positive perspectives towards the upcoming Sentinel-3 mission and its utility in

crop mapping, because it will be equipped with two bands covering the red edge spectrum, and its high repetition rate is suitable for multi-seasonal crop mapping (Donlon et al., 2012). These findings, strictly speaking, do not necessarily mean that the same order relation of variable importance exists for the SVM, but it gives insights into the relative importance of different types of features for crop classification in the RF OOB context.
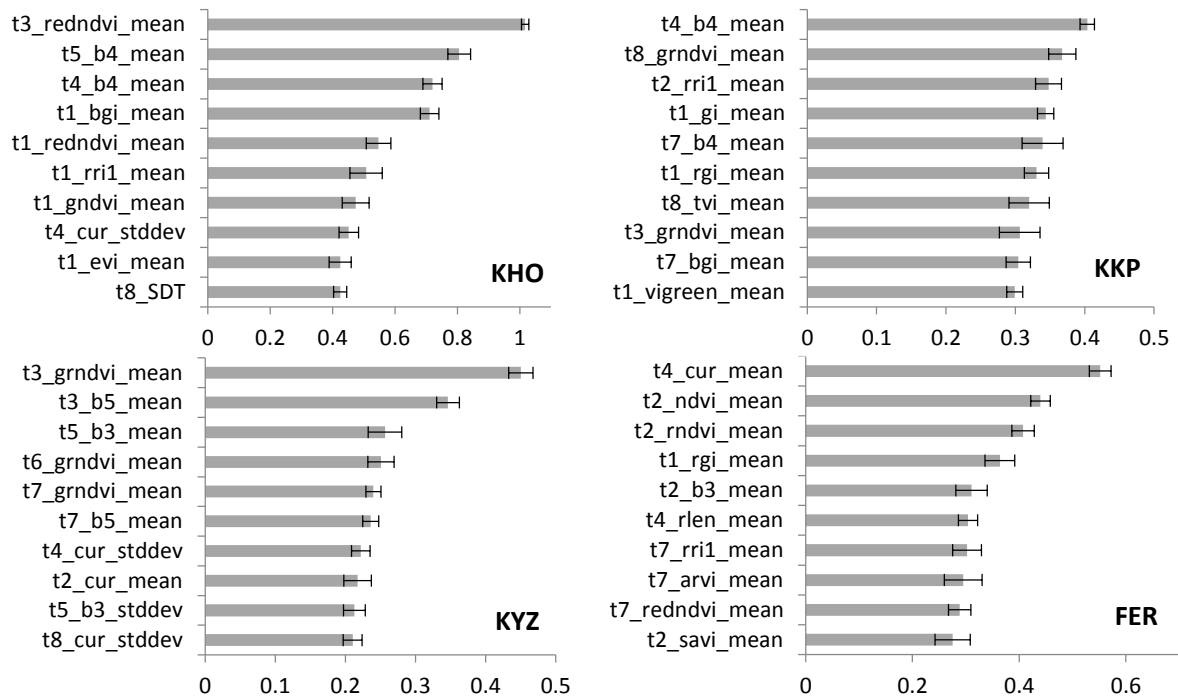
Figure 4-5: Mean relative importance of the 10 most important features. Error bars indicate standard deviation of the variable score (as mean decrease accuracy) over the 10 model runs. "t1−t8" in the feature names indicate the acquisition date, abbreviations of the variables are explained in chapter 3.

4.3.3. Principal components for detecting highly correlated features

PCA was performed using the software package FactoMineR (Husson et al., 2013) in R in order to group and interpret the redundancies between the different features and feature groups, respectively. After computation of the PCA over the complete set of features most of the variability in the data is concentrated in the first few components. For better understanding of the redundancies the data was visualized on the first two principal component (PC) factor-planes. The results shown in Figure 4-6 indicate a strong, albeit heterogeneous grouping tendency, which indicates that the clustered features provide similar, correlated information. Features located in a symmetrical position of the PCA planes are highly correlated. For instance when looking at the correlation structure between the feature groups in KHO, group A features on the one, and group E features on the other site tend to be less correlated, whilst there is a strong correlation between group B and C features.

Looking at the individual features allows for a more detailed assessment. In general the positions of the ten most important features (Figure 4-5) in the PC factor planes (Figure 4-6) tend to be dispersed over the total PC factor plane space. A small grouping tendency is noticed among a few features, which means that these groups provide similar information. For instance in KHO this is the case of the group of features "t1_bgi_mean" and "t1_redndvi", both of them being calculated at the same acquisition date. Another group of features is that composed of "t3_redndvi_mean" and "t4_b4_mean", both of them being red edge features calculated in the same part of the growing season. Similar in FER the features "t7_arvi_mean", "t7_redndvi_mean", and "t7_rri1_mean" form a group of correlated features from the

same acquisition date. Concluding, the RF tends to select both, highly and lowly correlated features. For instance "t3_redndvi_mean" was the most important feature in KHO and highly correlated with a couple of other features, which were not automatically selected as most important too. The following features were the five most correlated with "t3_redndvi_mean" (not shown in Figure 4-6), but they were not among the 85 features defined in FS_2 (Table 4-7) that achieved highest SVM classification accuracy: "t4_redndvi_mean", "t3_len_stddev", "t4_vigreen_stddev", "t3_var_stddev", and "t6_rlen_mean". Consequently, strong correlations were not the motive behind feature selection.
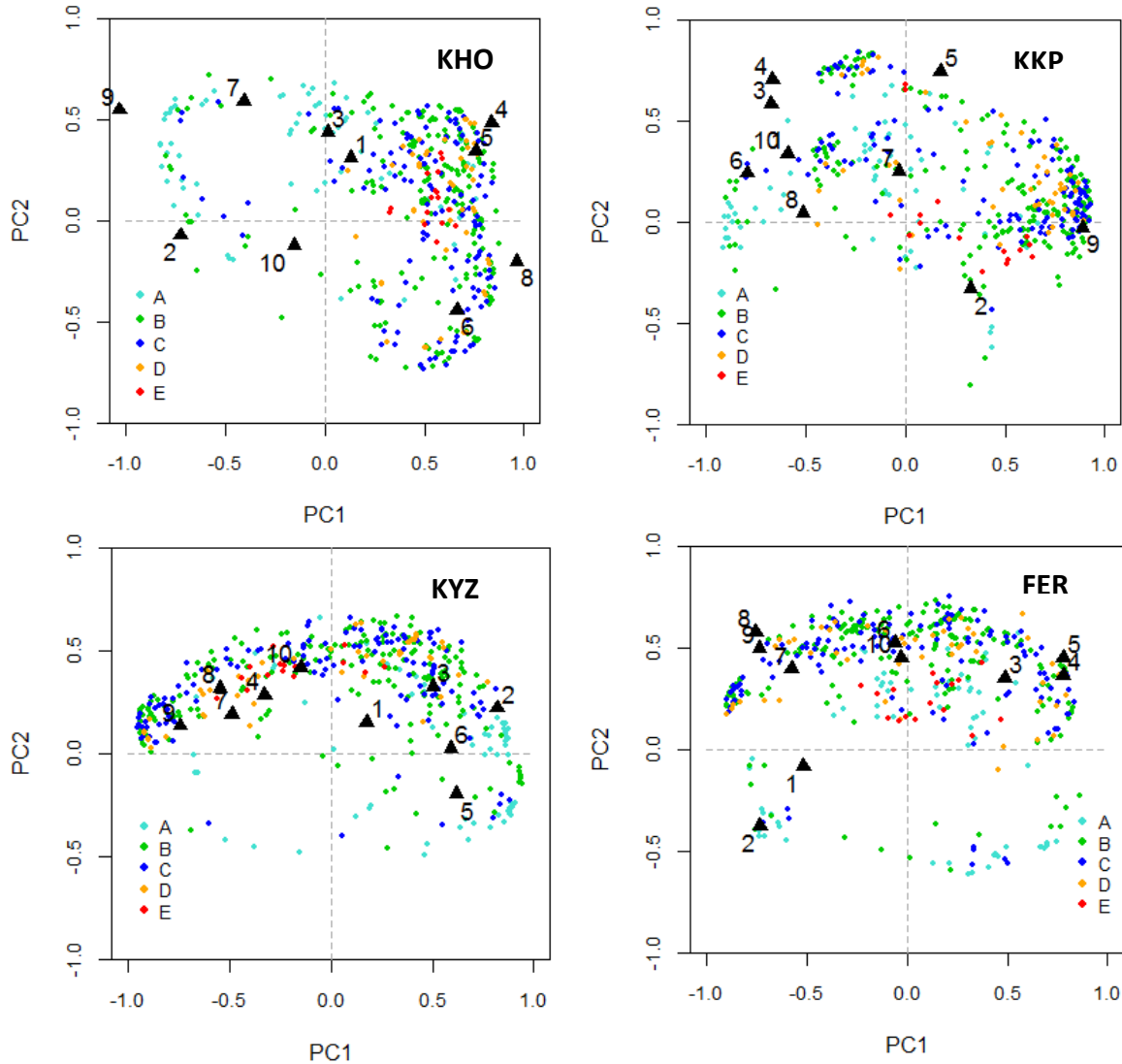
Figure 4-6: Projection of the 568 proposed features in principal component planes. Letters A–E correspond to the five feature groups. Black triangles indicate the position of the ten most important features, numbers from 1 to 10 correspond to their position in the feature ranking list (Figure 4-5), with "1" being the most important feature.

4.3.4. Influence of dimensionality reduction on classification accuracy

It was shown that feature selection after concatenating all feature groups improves classification performance, and the composition of features that yielded the highest classification accuracy with SVM was presented. Here the influence of feature selection on the classifier performance is explored in more detail, focusing on the four specific subsets (FS_{1-4} subspaces) defined in section 4.2.5. A strong influence of the feature selection procedure on SVM performance can be observed (Figure 4-7), e.g. classification

accuracy and uncertainty by the SVMs varied as a function of the number of features. In general terms, OA tended to increase with increasing number of features until a peak, which gave rise to the definition of the FS_2 subspace. In KYZ and FER the accuracy rose to a peak at 145 features (94.23 %) and 115 features (94.60 %), respectively, and then declined and stabilized with the addition of more features. The largest difference between the peak accuracies and that obtained using all features was 2.74 % and 3.19 % for KYZ and FER, respectively. The average standard deviation of the accuracies over the 50 model runs was below ± 1 % in all sites (Figure 4-7). Similar trends were observed for KKP and KHO, but the decrease in accuracy with increasing number of features is more pronounced. Classification accuracy declined with the addition of further features after peak accuracies of 93.33 % (100 features) and 91.02 % (85 features) in KKP and KHO, respectively. Differences between peak accuracies and that obtained using all features were 4.31 % (KKP) and 3.37 % (KHO) (Table 4-7). These results indicate a negative effect of the dimensionality of the feature space (Hughes effect) and is in unison with recent studies on that phenomenon using hyper-spectral data and SVM classification (Pal and Foody, 2010; Waske et al., 2010).
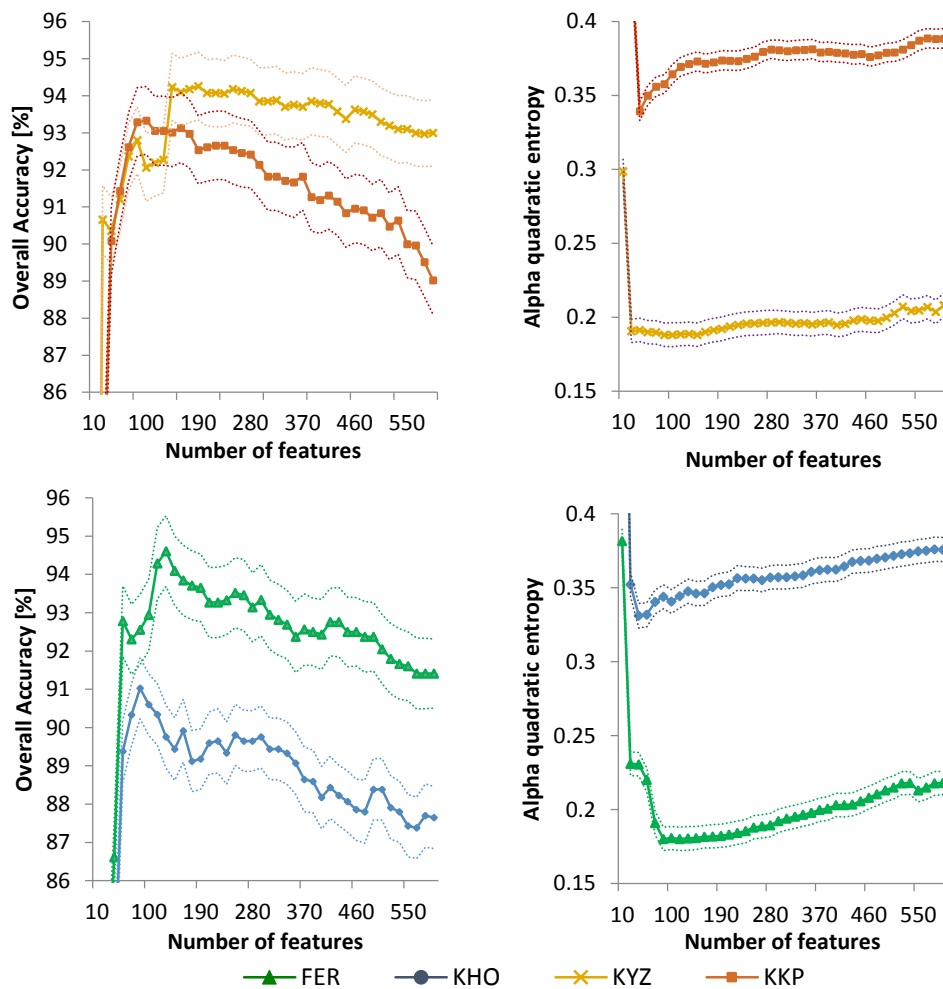
Figure 4-7: Overall accuracy [%] (left) and median uncertainty of the test fields (right) achieved by 50 model runs of SVM using different input features suggested by the RF feature score. Dotted lines indicate the standard deviation of overall accuracy and alpha quadratic entropy, respectively, over 50 model runs.

Having established that the accuracies vary as a function of the number of features used, the four feature subspaces defined above were evaluated in more detail. Table 4-7 summarizes the accuracy statistics, computational time, architecture complexity, and dimensionality for the SVMs. Most notably

in all sites the feature selection was able to produce more accurate maps as compared with the regular SVM using all features (FS_1), and FS_2 was the most accurate and FS_4 the least accurate. In all sites except for KYZ 55 feature would suffice to derive classifications that fulfil the targets defined for the FS_3 subspace (see section 4.2.5), which provided additional advantages in terms of computation time and architecture complexity (e.g. fewer support vectors), respectively. Only for the KYZ site this would have to be relaxed to 70 features.

Selecting FS_2 resulted in highest values for all evaluation measures from the confusion matrix, which indicates a better transferability and stability of this feature selection strategy among the four test sites, although it still relied on a considerable number of features (between 85 and 145). There are two possible explanations for the relatively high number of features needed: first there is a high degree of redundancy present in the total of 568 features (see section 4.3.3), which could diminish the contribution of the single features to the crop characterization. In this respect the RF might not optimally order the features on the rate of convergence of the classification performance, and FS_3 might theoretically be composed of even less features when using an alternative feature selection algorithm. Further, this study did not seek to realize the fastest possible convergence toward highest accuracies, and choosing another feature selection strategy suitable for SVM in combination with a much smaller input feature space might suffice to achieve sufficiently high accuracies, as was shown by Carrão et al. (2008) and Pal and Foody (2010).

Concerning the class-wise accuracies (CA$_i$), their lower bounds increased when using FS_2 compared to the other SVMs. Although the OA of the other SVMs compared to FS_2 were relatively high, there was a higher discrepancy between the lower bounds of CA$_i$, partly reaching 9 %. This indicates that these models failed to attain similar high accuracies for all crop classes. This difference is most pronounced in KKP and KHO, where this can be attributed to the failure of these SVMs to correctly classify the class sorghum/maize, and in KYZ the class fallow, which was substantially confused with alfalfa fields (alfalfa-1y and alfalfa-3y).

Table 4-7: Summary of overall classification accuracy (OA), class wise accuracy (CA$_i$), and Cohen´s k for the SVMs over 50 model runs. For CA$_i$ the range of for all classes in the study site is shown. "Feature types" refer to the 71 different types of features defined in chapter 3.

Site	Sub-space	OA SVM	OA RF	CA$_i$ SVM	k SVM	Dimen-sion	Feature types	Number of SVs	Training time SVM [sec]	Variable score calculation time RF [sec]	Variable score calculation time BT [sec]
KHO	FS_1	87.65		[9,95.7]	82.8	584	71	189	989		
	FS_2	91.02	89.5	[36.3,100]	89.2	85	38	134	124	31	39
	FS_3	89.37		[25.5,97.8]	87.3	55	31	112	97		
	FS_4	85.04		[9,95.8]	78.4	284	71	179	367		
KKP	FS_1	89.02		[30.8,100]	85.2	584	71	251	1,821		
	FS_2	93.33	89.0	[60,100]	91.0	100	47	195	248	42	60
	FS_3	91.43		[30.8,97.8]	89.2	55	40	161	172		
	FS_4	86.09		[40,100]	82.6	213	71	211	731		
KYZ	FS_1	91.49		[20,99.6]	85.7	584	71	412	1,856		
	FS_2	94.23	92.7	[70,99.8]	89.9	145	52	191	256	58	72
	FS_3	92.36		[55.6,99.6]	87.0	70	33	138	112		
	FS_4	87.81		[9,98.6]	74.7	71	71	141	711		
FER	FS_1	91.41		[80,96]	89.6	584	71	153	635		
	FS_2	94.60	94.3	[85.7,100]	92.7	115	37	99	83	19	28
	FS_3	92.78		[66.7,96]	90.9	55	25	74	63		
	FS_4	85.25		[42.1,87.8]	86.6	213	71	138	209		

Notably none of the eight acquisition dates was dismissed after feature selection in FS_2 or FS_3, e.g. features from all of the eight acquisition dates were retained. This challenges previous suggestions of selecting few optimum acquisition dates for optimal crop classifications (Murakami et al., 2001; Van Niel and Mcvicar, 2004). Critically the underlying premise of FS_{1-3} was that multi-date data is available (e.g. in all three models features from all eight acquisition dates were being used). It might be feasible, however, to assess if relatively few acquisition dates are sufficient to fulfil certain requirements for crop mapping. The relative importance of the features as derived by the feature importance ranking from the OOB method under the RF provides insights on what features are most useful with respect to classifying crops. In this study the importance score was used to evaluate the relative importance of the acquisition dates, and then to maintain only the most relevant dates for classification (see definition of the FS_4 subspace section 4.2.5). According to the average importance $I_{f,avg}$ of the dates, a ranked list can be derived that shows their relative contribution to the SVM classification. In general terms, the importance of the acquisition dates is smaller than the importance of the feature groups, indicated by the smaller scores. This is because the importance scores of the groups partly consider the input data after feature selection that withholds only the most important features. Hence, eliminating one feature group leads to a more pronounced impact than removing an acquisition date that contains all features, irrespective of their importance.

The SVM was then trained based upon the date-reduced feature space (FS_4), and OA was calculated. Figure 4-8 shows the effect of removing images in order of decreasing importance of each acquisition date ($I_{t,avg}$). After a slight increase of OA in all sites there is a drop in accuracy that was most pronounced in KHO and KKP when removing more than 4–5 images, whilst in KYZ there was almost no significant effect on OA. In KHO images in the beginning of June and August / September are sufficient to achieve accuracies of more than 85.0 %. These four acquisition dates correspond well with the results from Conrad et al. (2013). In their experiment, Conrad et al. (2013) tested each possible combination of acquisition dates in KHO using the same satellite images as in this study, and exploited the best performing combination of dates for object-based crop classification using RF. Being much more intensive from the computational point of view these experiments resulted in similar OA when using four instead of eight acquisitions (84.8 %). In FER and KKP 85.0 % OA was achieved when only three dates were retained. The preference of acquisition dates in spring and summer, respectively reflect the presence of crop rotations and the necessity to have images at these critical periods for accurate crop mapping in all sites except for KYZ, where one early summer scene sufficed to classify all crops in the site with sufficient accuracy, and where crop rotations were totally absent. Figure 4-8 reveals that the time of acquisition of the most important dates reflected the presence of crop rotations in the study sites, and also coincides with the importance pattern shown in Figure 4-2. The highest ranked dates were: 31–May (KHO), 09–May (KKP), 9–Jun (KYZ), and 13–May (FER). The higher importance of features from late spring to early summer seasons nicely reflect the observed high variance in NDVI values (Figure 4-2) at the time when crops start to differentiate.
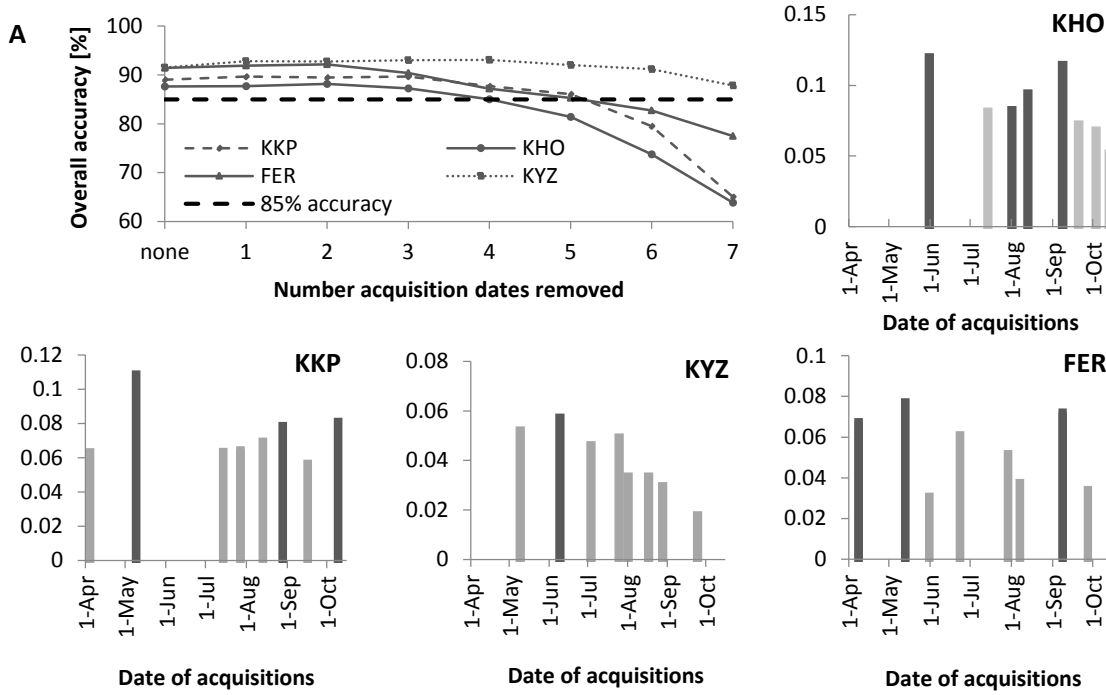
Figure 4-8: Overall accuracy as a function of removing acquisition dates in order of their decreasing average importance scores $I_{t,avg}$ (A). The average importance scores $I_{t,avg}$ of each acquisition date (shown on the y-axis) are given for each site. Dark grey bars indicate the acquisitions actually used in FS_4 to achieve more than 85 % OA.

The accuracy of the SVM was further compared with the accuracy of the RF algorithm. The maximum achieved accuracy of the RF is presented in Table 4-7. The RF was only slightly affected by the feature space size, as compared with the SVM, so it could be applied to the total data set. Although operating at a comparatively high level of accuracy, the RF was generally outperformed by the SVM applied to the FS_2 subspace. The difference between RF and SVM was the most pronounced in KKP (4.33 % on average). Another positive aspect of feature selection is reducing computational demand. Computation time was calculated for each experiment using a PC with 2 CPUs (each 2.7 GHz) and 8 GB RAM. After feature selection the time for model building was reduced (Table 4-7), and the architecture complexity of the resulting SVMs decreased, e.g. fewer support vectors were needed. The built-in variable selection of the RF, although iterated 10 times, only needed 19–58 seconds to generate a ranked feature score list. Another tree-based method that can be used to assess the feature importance, namely BT, needed 28–72 seconds (Table 4-7). Using a feature selection strategy before the SVM classification resulted in an improvement in processing time, when applied to the data set used in this study.

4.3.5. Impact of feature selection on classification uncertainty

Having analysed the impact of feature selection on classification accuracy, the performance of the SVMs were then compared by looking at their soft outputs that were computed on the independent test sets, which were not used for training. Classification uncertainty in the final crop maps, as quantified by the median $H_\propto rel$ of the test samples behave similar to OA, but with fewer fluctuations (Figure 4-7): first the uncertainty decreases when adding more features until it reaches a minimum, where the highest proportion of fields is classified with low uncertainty, and thereafter increases again with the addition of more features. The uncertainty values of KKP and KHO on the one site, and KYZ and FER on the other site are in the same order of magnitude.

One peculiar distinction is that the lowest values of uncertainty are achieved with fewer features than necessary to achieve the peak accuracy (FS_2): 40 features in KKP and KHO (corresponding to OA of 90.0 % and 83.1 %, respectively), and 85 features in KYZ and FER, respectively (corresponding to OA of 92.8 % and 92.5 %, respectively). Being less accurate than FS_2 these SVMs classified crops with less uncertainty, albeit this was more pronounced in KKP and KHO. Two aspects can explain this: the feature subspace FS_2, which achieved peak accuracy, might still contain some highly correlated features since the RF does not necessarily eliminated correlated features first. In this regard even further reducing the input space size (FS_3), albeit at the expense of the accuracy, and might lead to more confident results, indicated by decreasing uncertainty. This is similar to findings of Loosvelt et al. (2012a), who found decreasing map uncertainties as a result of feature reduction, albeit at the expense of decreasing accuracies. Second, when further reducing the feature space size, the proportion of incorrectly classified pixels increases (because OA decreases, see Figure 4-7). However, the uncertainty of these incorrectly classified cases was also found to decrease, which might explain the general decrease of $H_\propto rel$ over the total image. Consequently an increase in the number of incorrectly classified fields could lead to a further decrease in uncertainty as long as the incorrectly classified fields are associated with lower uncertainties, in particular in KKP and KHO. This is why FS_3 was associated with lower uncertainties than FS_2 in KKP and KHO (Figure 4-7). In FER and KYZ however this phenomena was less pronounced and FS_3 is associated with higher uncertainties than FS_2.

This demonstrates how feature selection can have positive effects beyond increasing accuracy by decreasing classification uncertainty, and that crop maps with higher confidence (e.g. reduced uncertainty) can be produced if a certain drop in accuracy can be tolerated, which further supports the definition of FS_3.

To get an idea of the predictive strength of the SVMs a simple test can be computed to assess if the correctly classified cases are characterized by low uncertainty values, and vice-versa. If this assumption holds true the uncertainty values of correctly classified cases should be equal to or lower than the uncertainty values of the misclassified cases (Zhu, 1997). To this end empirical frequency distribution of E and $H_\propto rel$ were calculated for correctly and incorrectly classified test fields (Figure 4-9 and Figure 4-10). The shape of these distributions gives an indication for the reliability of the classifier algorithm in predicting the spatial distribution or error (predictive strength) and the quality of the crop maps, as was proposed and demonstrated for boosted DT by McIver and Friedl (2001) and exemplarily shown for pixel-based RF classification in Loosvelt et al. (2012a). It is evident that correctly classified cases tend to display relatively low values of E and $H_\propto rel$, respectively. Critically, the degree to which this is the case differs from site to site and from subspace to subspace, respectively.

In general this confirms the high prediction strength of the SVMs over all sites. As the dimensionality of the feature space decreased, a larger proportion of correct predictions were made with lower uncertainty, e.g. lower values of E and of $H_\propto rel$. In general the SVMs applied to FS_1 and FS_4 had the lowest prediction strength. The feature reduction in KYZ only slightly affected the shape of the distributions of E and $H_\propto rel$ for correctly classified fields, although the frequency of fields lying within the interval [0,0.2] of FS_2 was highest. This indicates that the SVM model in KYZ and, to a lesser extent, FER already produce results with low classification uncertainty, without further potential for significant improvements. In KKP and KHO the SVM of FS_2 showed the highest frequency peak in that interval for E. Looking at the frequency distributions of E an exponential decrease toward higher E values can be found. This trend is less pronounced when looking at $H_\propto rel$: in KHO the frequency distribution peaks within the [0.2,0.4] interval for all reduced subspaces. Further $H_\propto rel$ seems to be the stricter measure of uncertainty as the frequencies are more balanced among all intervals, in contrast to the peaks of E in the [0,0.2] interval.

Incorrectly classified cases mostly had moderate (KKP) or high (KHO) uncertainties, which mean that if the SVM voted for an incorrect class, the prediction was uncertain due to confusion between other classes. Critically the shape of the distribution of misclassified cases in FER and KYZ was found to be different. In KYZ and FER the misallocated cases lie in the lower intervals of both, E and $H_\propto rel$. This confirms that the classification errors in these sites are confident misallocations, in which the fields were allocated to the wrong class but with little classification uncertainty. For instance from the misclassified cases in KYZ more than 70 % occurred in the [0,0.2] interval of E, and more than 50 % in the [0,0.2] interval of $H_\propto rel$, irrespective of the model used. The majority of these cases were fallow fields being misallocated to both of the two alfalfa classes (not shown here). It is assumed that the presence of mixed fields (e.g. alfalfa and other herbaceous vegetation) can result in misallocations in KYZ.

Based on these results it can be concluded that correct predictions in all sites and using all models are confident and with limited confusion with other classes (e.g. low $H_\propto rel$ values). The FS_2 showed lower uncertainties in all sites, indicated by a shift of the frequency of correct predictions toward lower values of E and $H_\propto rel$, respectively. These results indicate that feature reduction is able to provide more confident crop maps, and to increase the number of cases (fields) that were classified with low uncertainty. It would have been expected that incorrect cases were classified with high uncertainties, yet this was not the case in FER and KYZ where errors were confident misallocations, irrespective of the feature subspace considered: in these cases the SVM had little doubt about the class allocation, albeit the wrong class was chosen.
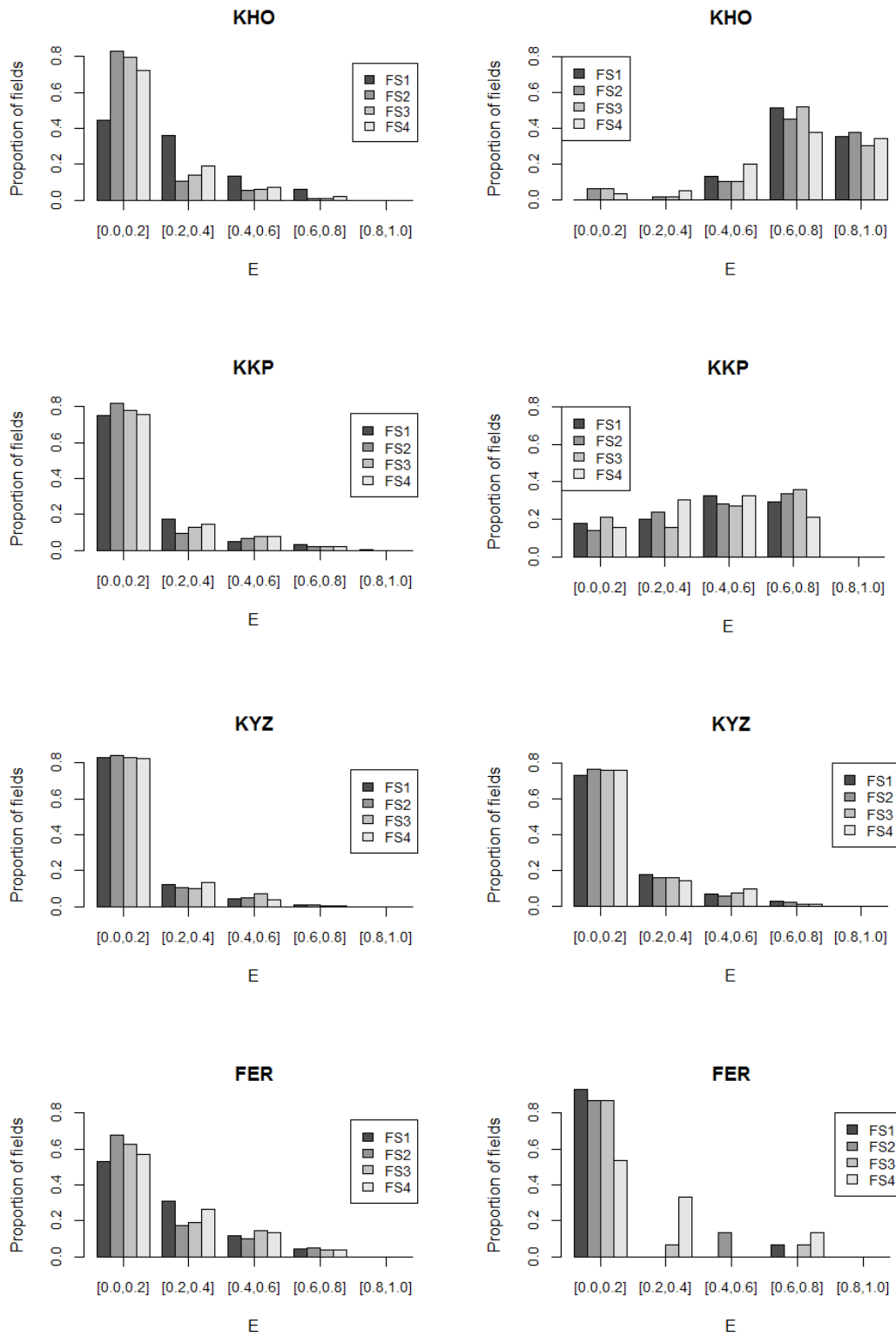
Figure 4-9: Distributions of *E* for correctly (left column) and incorrectly (right column) classified test fields as resulted from classifying the FS_{1-4} subspaces, calculated over 50 model runs. The x-axis shows intervals of *E*, the y-axis the percentage of test fields within the interval.
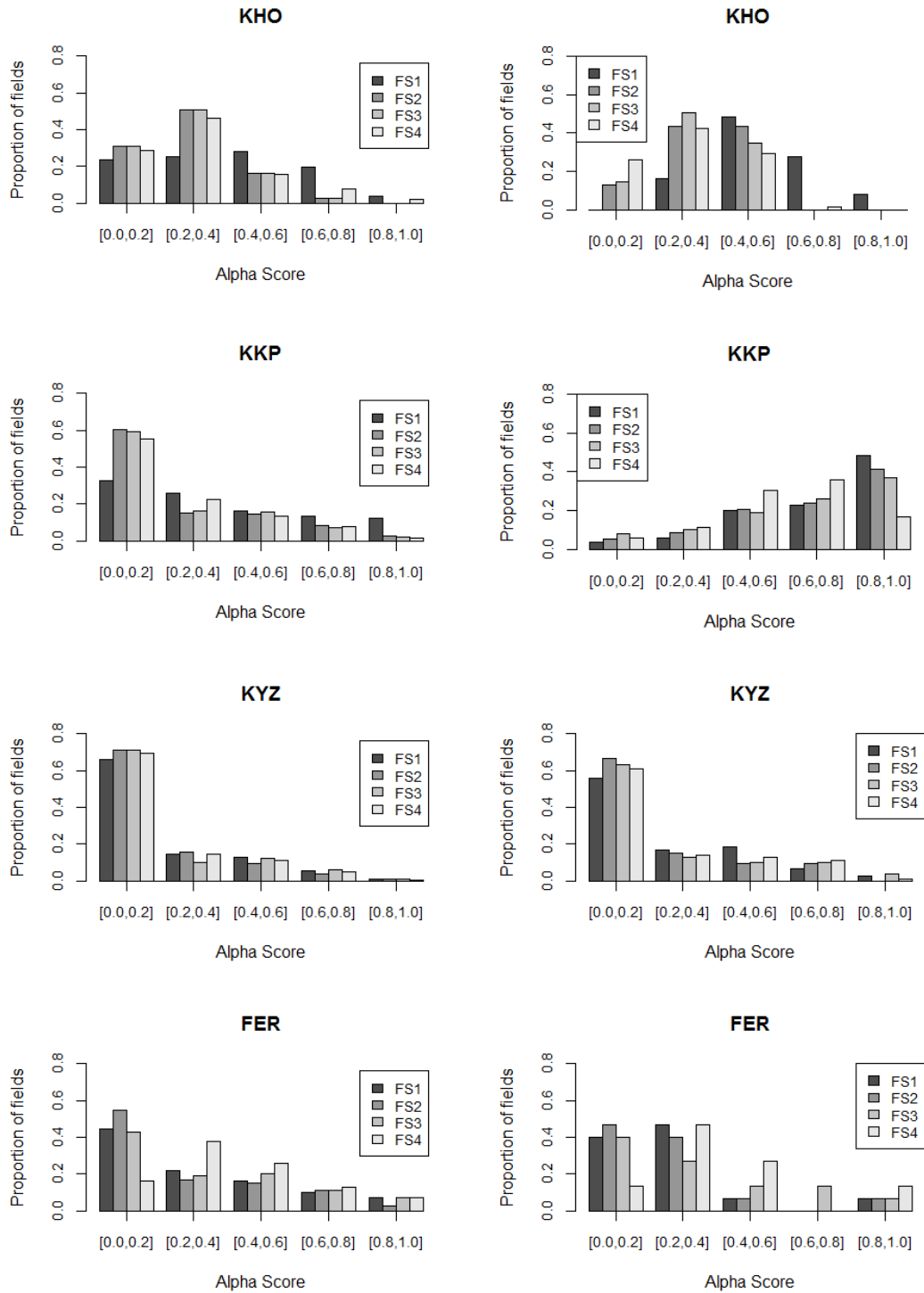
Figure 4-10: Distributions of $H_\alpha rel$ for correctly (left column) and incorrectly (right column) classified test fields as resulted from classifying the FS_{1-4} subspaces, calculated over 50 model runs. The x-axis shows intervals of $H_\alpha rel$, the y-axis the percentage of test fields within the interval.

The results shown in Figure 4-9 and Figure 4-10 are summarized in Table 4-8, by the median uncertainty values (E and $H_\alpha rel$) for correctly and incorrectly classified test fields, respectively. Based on this, it can be concluded that correct predictions by FS_{1-4} were made with limited doubts about the prediction and confusion with other crops was of minor importance, as indicated by low median values of E and $H_\alpha rel$, respectively. Lowest uncertainties were found for the correct predictions for FS_2 in all sites, and highest uncertainties were associated with FS_1.

Table 4-8: Median maximum probability (E) and median alpha quadratic entropy ($H_\alpha rel$) of correct and incorrect classified test fields, as results from the four feature subspaces.

Site	Measure	FS_1		FS_2		FS_3		FS_4	
		Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
KHO	E	0.26	0.75	0.15	0.71	0.16	0.68	0.18	0.67
	$H_\alpha rel$	0.41	0.55	0.27	0.36	0.28	0.34	0.31	0.32
KKP	E	0.17	0.46	0.12	0.46	0.13	0.46	0.14	0.42
	$H_\alpha rel$	0.40	0.72	0.22	0.67	0.23	0.64	0.24	0.59
KYZ	E	0.09	0.13	0.08	0.11	0.09	0.12	0.11	0.12
	$H_\alpha rel$	0.18	0.24	0.13	0.19	0.16	0.22	0.15	0.21
FER	E	0.13	0.20	0.12	0.19	0.13	0.18	0.22	0.23
	$H_\alpha rel$	0.35	0.54	0.25	0.28	0.29	0.31	0.40	0.42

In contrast to this, incorrect predictions were made with moderate values of E and $H_\alpha rel$, respectively, and with relatively low uncertainties in KYZ and FER (see discussion above). In general the uncertainty decreases from FS_1 to FS_4, which means that incorrect predictions are characterized with lowest uncertainties in FS_4. In KYZ and FER the characteristics of incorrect predictions is different, and the incorrect predictions made by FS_2 tends to be the least uncertain. The characteristics of the incorrect predictions made by FS_4 tend to be somewhat in between FS_2 and FS_3. These results confirm the earlier findings from the analysis of the uncertainty distributions and could explain why Figure 4-7 reveals a further decrease in classification uncertainty of the test fields in KHO and KKP when using fewer features than in FS_2, as discussed above.

Another investigation was undertaken by evaluating the uncertainty of the whole image, e.g. all classified fields in the produced maps. The class-specific uncertainties by means of E and $H_\alpha rel$ show distinct differences (Table 4-9). In general terms FS_1 was outperformed by all other SVMs, indicating the benefits of feature selection irrespective of the final SVM chosen. Further the class-wise uncertainties among the four subspaces were complementary and, depending on the target class of a given application, an appropriate subspace might be selected. For instance all SVMs showed high values of E and $H_\alpha rel$, respectively for melons and sorghum/maize, irrespective of the subspace chosen for classification. By contrast rice and winter wheat was correctly allocated with low uncertainty using any of the subspaces. On the other site the uncertainty in allocating cotton fields was reduced after feature selection in KHO and KKP.

In KHO the FS_4 achieved the best results (e.g. most classes with lowest uncertainty) when looking at E and the second best when taking $H_\alpha rel$ into account. Interestingly when using only few acquisition dates (FS_4) the classification uncertainty quantified by $H_\alpha rel$ for crop rotations tends to be higher in KKP and FER, although the most critical acquisition dates were maintained (Figure 4-8). This indicates the complexity of the class wheat-other, and highly varying crop schedules could be the reason for the demand of more acquisition dates to achieve lower uncertainties.

Table 4-9: Median values of E (A) and $H_\alpha rel$ (B) of the four feature subspaces and for each of the investigated crop classes separately. The crop classes include cotton (C), fallow (FA), rice (R), sorghum/maize (S), alfalfa-1y (A1), alfalfa-3y (A3), melons (M), fruit trees (F), winter wheat (W), and wheat-other (WO).

A: median E

	Subspace	C	FA	R	S	A1	A3	M	F	W	WO
KHO	FS_1	0.40	0.36	0.50	0.64				0.41	0.48	0.25
	FS_2	0.21	0.29	0.47	0.62				0.39	0.44	0.24
	FS_3	0.19	0.28	0.45	0.55				0.38	0.43	0.24
	FS_4	0.17	0.27	0.42	0.40				0.38	0.42	0.23
KKP	FS_1	0.28	0.21	0.06	0.59	0.55		0.58		0.09	0.56
	FS_2	0.31	0.20	0.07	0.62	0.56		0.55		0.10	0.52
	FS_3	0.25	0.23	0.05	0.56	0.09		0.55		0.09	0.53
	FS_4	0.30	0.21	0.09	0.54	0.53		0.50		0.11	0.51
KYZ	FS_1		0.31	0.02		0.35	0.35			0.42	
	FS_2		0.31	0.02		0.33	0.32			0.44	
	FS_3		0.29	0.01		0.32	0.28			0.39	
	FS_4		0.26	0.02		0.26	0.30			0.48	
FER	FS_1	0.04	0.32						0.15	0.29	0.04
	FS_2	0.05	0.29						0.14	0.26	0.07
	FS_3	0.03	0.36						0.15	0.31	0.10
	FS_4	0.10	0.31						0.29	0.52	0.17

B: median $H_\alpha rel$

	Subspace	C	FA	R	S	A1	A3	M	F	W	WO
KHO	FS_1	0.63	0.61	0.70	0.80				0.63	0.73	0.44
	FS_2	0.48	0.54	0.56	0.65				0.53	0.51	0.48
	FS_3	0.41	0.53	0.61	0.64				0.54	0.59	0.44
	FS_4	0.37	0.51	0.64	0.78				0.60	0.66	0.41
KKP	FS_1	0.88	0.48	0.24	0.87	0.86		0.90		0.34	0.67
	FS_2	0.46	0.39	0.16	0.78	0.79		0.72		0.16	0.68
	FS_3	0.43	0.43	0.12	0.78	0.77		0.77		0.21	0.75
	FS_4	0.46	0.40	0.21	0.75	0.72		0.71		0.24	0.74
KYZ	FS_1		0.53	0.03		0.52	0.53			0.66	
	FS_2		0.47	0.04		0.44	0.43			0.47	
	FS_3		0.53	0.03		0.52	0.48			0.70	
	FS_4		0.48	0.02		0.50	0.48			0.73	
FER	FS_1	0.10	0.62						0.32	0.51	0.11
	FS_2	0.13	0.39						0.29	0.38	0.15
	FS_3	0.11	0.66						0.33	0.55	0.25
	FS_4	0.26	0.52						0.51	0.69	0.34

4.3.6. Crop pattern and area statistics in the study sites

The spatial crop pattern in the four sites is revealed in the crop maps in Figure 4-11, the crop acreages are shown in Figure 4-12. In general the spatial pattern that was detailed in section 2.2 is revealed in these maps. The most prominent feature in all sites is the dominance of relatively few crops: cotton and/or winter wheat (KKP, KHO, FER), and rice (KYZ). For instance, cotton and wheat-other fields share more than 60 % of the total cropped area in KHO and FER, whilst in KKP more than 40 % of the fields are fallow. In contrast, the KYZ landscape is dominated by rice fields (more than 50 % coverage), which form large clusters of 25–100 ha, but winter wheat fields share less than 1 % of the area. Although the highest crop diversity can be found in KKP, the highest proportion of fallow fields characterizes this site. This is a result of low water availability and bad environmental conditions, e.g. soil salinization

(JICA et al., 2010), whilst in KYZ the high proportion of fallow fields is part of the management strategy where alfalfa (alfalfa-1y and alfalfa-3y) is cultivated in order to preserve or regenerate soil fertility (see section 2.2). The relatively high share of cotton in all sites except for KYZ reflects the overall importance of cotton as the dominant export commodity in Uzbekistan.

In the KHO landscape cotton and wheat-other fields share more than 60 % of the cultivated area. They are scattered throughout the region, whilst larger and more homogeneous areas of rice areas (8 %) are located in a narrow stripe next to the Amu Darya or along the major irrigation canals and small lakes, possibly due to relatively good water availability (Conrad et al., 2011a). Clusters of fallow fields are concentrated within a narrow corridor in the western part of the test site, where fields are marginal and soils sandy (Conrad et al., 2010). Sorghum/maize, fallow fields, or winter wheat without rotation together share 10 %.
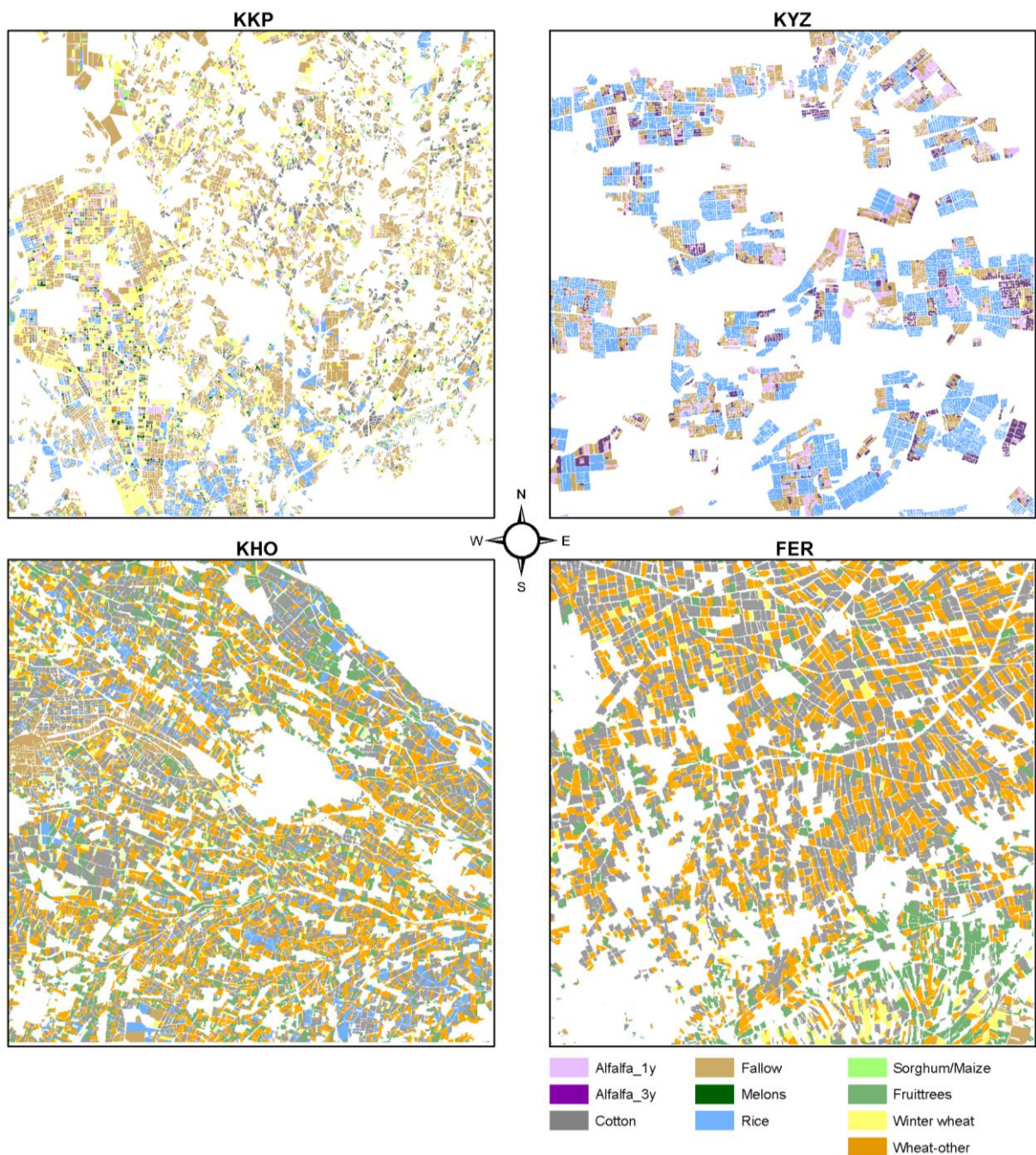
Figure 4-11: Crop maps from the SVM classification based on the FS_2 feature subset. Each site is 30×30 km.

In KKP there is a spatial trend in northern and north-eastern direction, which is characterized by increasing fraction of fallow fields in the landscape. This is due to the limited water supply and availability when moving northwards from the intake point in the south. Rice fields are preferably located among the main irrigation canals that spread northwards. Winter wheat and fallow fields dominate the northern half of the site, whilst rice, cotton, and to a lesser degree, melons and alfalfa-1y fields dominate the south. The intensity and diversity of land use was highest in the south of the irrigation system, where partly rotations with winter wheat were found. This pattern reflects the decreasing water supply to the north (JICA et al., 2010).

FER is dominated by cotton and winter wheat fields with crop rotation, which are heterogeneously distributed in the site. To the southeast fruit trees and winter wheat fields without rotation dominate on the ascending slopes of the Fergana Valley.

A cautious note must be given concerning the area estimates: Figure 4-12 reveals that over the 50 model runs some of the classes exhibit marked deviations from the mean, e.g. fallow fields in KHO. The results from the RF differ in some cases, but given the overlap of the standard deviations there remains uncertainty.
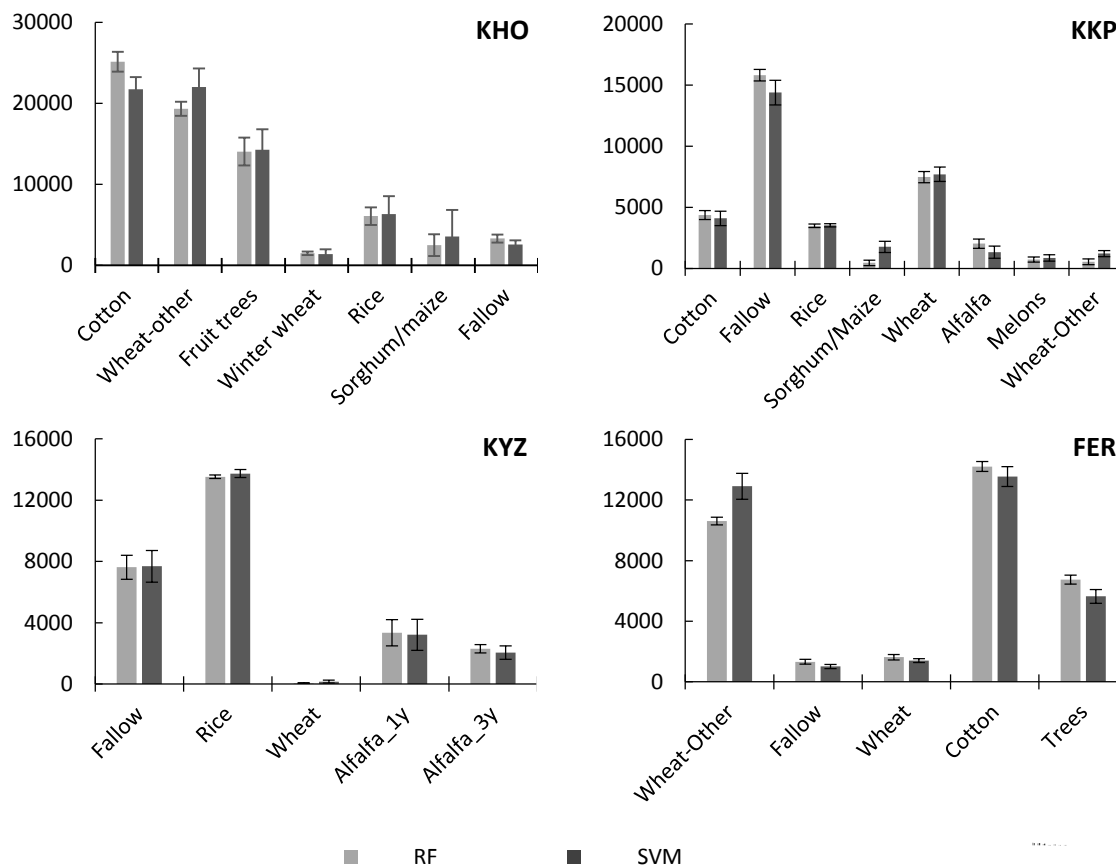
Figure 4-12: Crop acreages in hectares for the four study sites, resulting from the SVM applied to the FS_2 subspace. RF-based crop acreages are given for comparison, error bars indicate the standard deviation of the area estimation over the 50 model runs.

4.3.7. Classification uncertainty pattern in the study sites

The spatial representation of the classification uncertainty as quantified by $H_\propto rel$ is shown in Figure 4-13. The map can be used to allocate and discuss the causes of possible errors in the map. There are different and distinct spatial patterns among the four sites: KYZ is characterized by a chessboard like pattern of high and low $H_\propto rel$ values, which reflects the cropping pattern in the study region, where rice fields are cultivated on adjacent fields (see section 2.2) and classified with lowest $H_\propto rel$ values. There was more confusion among the fallow fields and alfalfa-1y, which results in higher uncertainties in clusters composed of these two classes. Alfalfa-3y is generally allocated with less uncertainty than alfalfa-1y, caused by the distinct temporal profiles of the homogeneous fields.
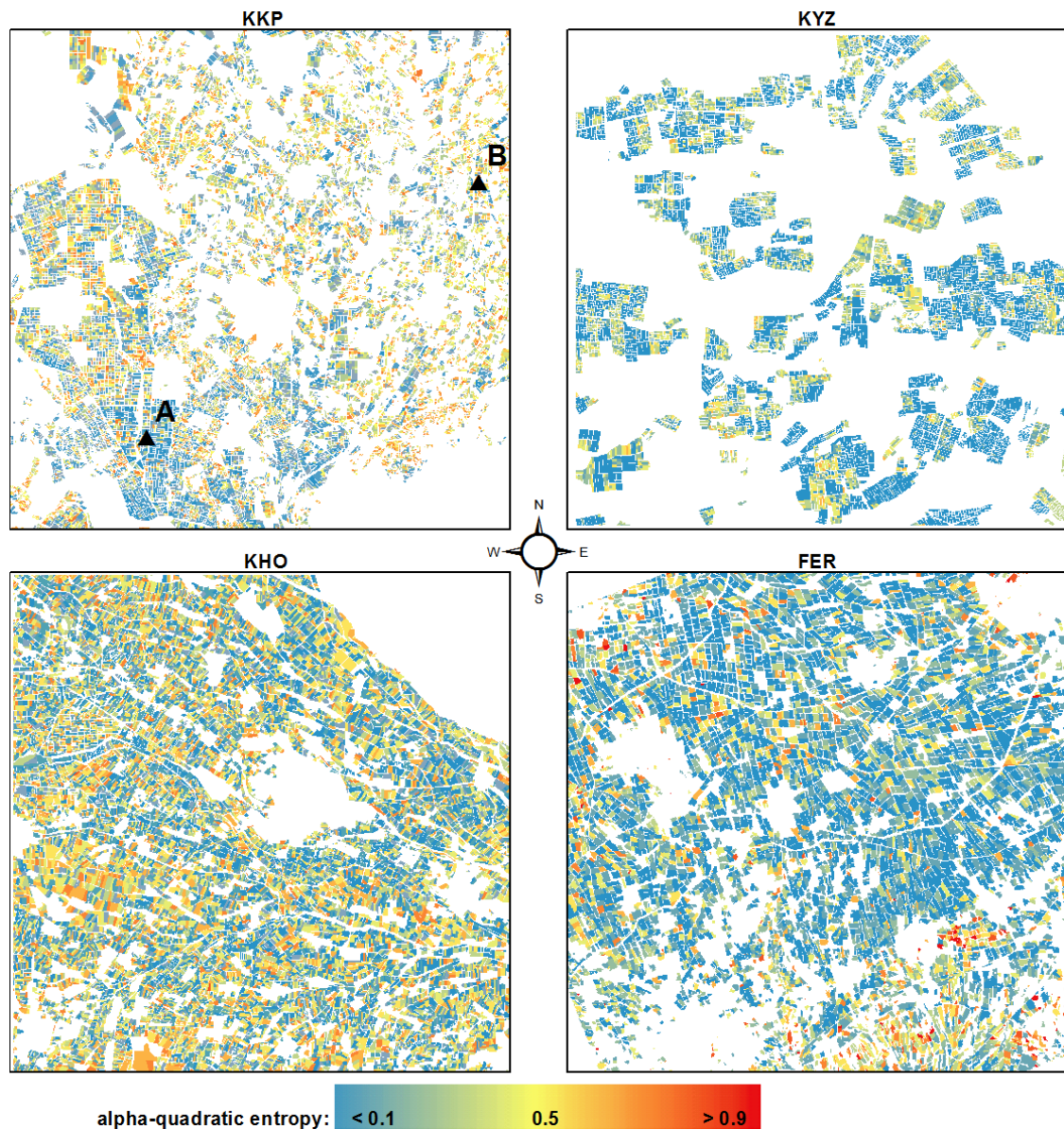
Figure 4-13: Classification uncertainty maps from the SVM applied to the FS_2 subspace. The maps display $H_\propto rel$. Each site is 30×30 km. "A" and "B" indicate locations of sample photos from rice fields, shown in Figure 4-14.

In contrast to KYZ the pattern of $H_\propto rel$ in KHO and KKP is more diffuse, where low uncertainties are mainly associated with sorghum/maize or melon fields. Despite the heterogeneous pattern of uncertainty there are spatial trends of declining $H_\propto rel$ values. In KKP this trend is visible and goes in north-eastern direction when moving northwards from the intake points in the south where water is

distributed into the irrigation system in KKP. In this region, the least irrigation water, which is discharged in the south, reaches the fields, and which makes this region the most drought affected in the Karakalpakstan region (JICA et al., 2010). The more regular shaped fields in the southern half of the site are characterized by low uncertainties and are located among the large irrigation canals that run from south to north. In KHO the uncertainties tend to increase in south-western direction toward the desert fringe, where water availability and supply was shown to be suboptimal (Conrad et al., 2007). These parts of the irrigation system are only under irrigation in years of high water availability (Conrad et al., 2013), and land degradation advances (Dubovyk et al., 2012). The spatial trends indicate locations in the map in which the classifier is unable to provide accurate discrimination, e.g. the class identification performance of the algorithm is low. Both sites have almost twice as many classes and are located in the downstream locations of the Amu Darya, where land degradation as a result of unreliable availability of irrigation water is widely reported.

It is assumed, but not proven quantitatively in this study, that these spatial trends are caused by declining water availability, resulting in reduced biomass production, and ultimately a flattening of the temporal profiles (e.g. NDVI) which can lead to increasing confusion in class allocation. As an example, Figure 4-14 shows the NDVI temporal profiles of all rice fields in KKP. Fields classified with high uncertainties are obviously characterized by low biomass production, as indicated by the flattening of the NDVI profiles. The typical negative NDVI values which indicate the field leaching at the beginning of the growing period is absent when uncertainties raise above 0.2. Exemplary photographs from two rice fields in KKP demonstrate how reduced water availability results in increased uncertainty values (Figure 4-14 A and B).
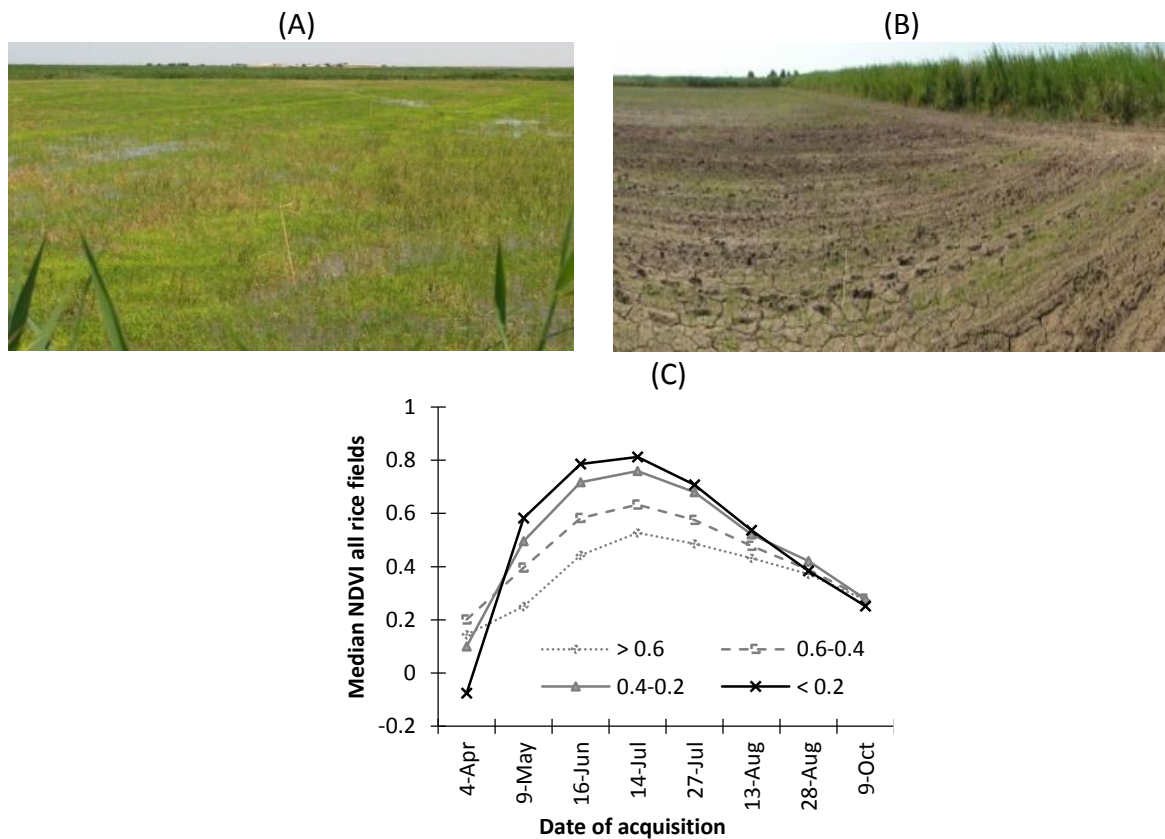
Figure 4-14: Exemplary photographs were taken from two different rice fields on 27-Jul 2011, which were classified correctly as rice, but with different values of $H_\propto rel$: 0.25 (A) and 0.72 (B). Median NDVI profiles of all fields in KKP classified as rice are shown in (C). Different line types indicate different levels of uncertainty.

Figure 4-15 shows the NDVI profiles of fallow fields. Here high uncertainties are indicated by more irregular NDVI profiles and increased biomass production. The latter can be explained with higher surface heterogeneity and biomass production when high amounts of weed flora or reeds grow within the fallow fields. This contrasts the flat profiles of fallow, ploughed fields with flat NDVI profiles, and results in disturbed spectral information of such fields, which most likely increases classification uncertainty. In FER the pattern of uncertainty is more random, with a cluster of higher uncertainty in the south-eastern part. This coincides with the preferred cultivation of fruit trees and winter wheat on the gently ascending hill slopes, which were classified with higher uncertainty compared to the other crops. FER is located in the upstream region of the Syr Darya and is characterized by relatively high water availability, as is KYZ located in the mid region of the Syr Darya. In both sites only five crops are cultivated, and class allocation is assumed to be less uncertain.
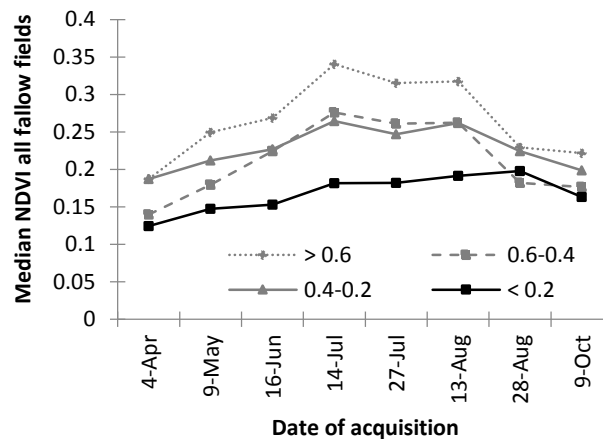
Figure 4-15: Median NDVI profiles of all fields in KKP classified as fallow. Different line types indicate different levels of uncertainty.

As can be seen from these observations, the spatial pattern of uncertainty can be explained by means of some observable spatial phenomena. In all study sites no significant correlation between field sizes and $H_\propto rel$ was detected (all correlation coefficients $r^2 < 0.0031$), whereas for instance Yu et al. (2008) found a positive relation of classification uncertainty and larger object sizes in vegetation mapping, where objects of smaller size occurred in areas of spectrally heterogeneous areas and were classified with higher uncertainty. Water availability seems to be one explanatory factor, but further research on this issue is needed.

4.4. Conclusions

This study put emphasis on the issue of spatial classification uncertainty in the context of SVM classification, which has attained only few attention, yet. Although crop classification with SVM has been shown to be advantageous in comparison to other classifiers a potential disadvantage is the feature space size and the effect of feature reduction on SVM classification using multispectral time series data has attained little attention. This study provided experimental evidence that SVM is affected by the feature space size. In particular reducing the feature space size decreased classification uncertainty and saved computation time. In comparison, the performance of the RF was not affected by the feature space size and performed almost equally well. Recent technological advances like RapidEye combine both, high spatial resolution and frequent revisit time, which is well suited for crop identification. However the high dimensionality of such multi-temporal datasets can be a potential problem (Hughes phenomena). The presented work revealed that SVMs appeared as a computationally effective and very accurate

means for agricultural crop identification. As it is unknown beforehand which features are useful for crop mapping, an intentional very large input space was provided in this study by calculating a variety of spectral and geostatistical indices. The RF variable importance score is a powerful tool (and computationally light) to explore the best acquisition date combination. The different feature types provided diverse information and different acquisitions contribute unequally to the classification of the various crop classes. Hence, the different information sources were not necessarily equally reliable. One specific feature or feature group can be more applicable to describe one specific class but not be appropriate for another class. Consequently the use of different type of information as well as the integration of multi-seasonal data sets is valuable. This statement was confirmed by the results of the presented applications.

It is concluded that RF selected subsets featuring predictor variables containing red edge information from the RapidEye sensor, acquired early and late summer, and can be suggested for timely-efficient, accurate classification of crops in the study sites. It was demonstrated that using only a few instead of all acquisition dates sufficed to achieve accuracies better than the oft-recommended target accuracy of 85 %. The choice of the most important acquisition dates was oriented by the average importance score of all features at a date, and reflected the cropping pattern of the study sites (e.g. the presence or absence of crop rotations). Further it was shown that feature reduction resulted in reduced map uncertainties in all study sites, and that the SVM that used fewer features than the best performing SVM resulted in reduced uncertainties for many crops, despite a small drop in accuracy.

The approach presented in this study is easy to implement and computationally light, so its transferability to other (and potentially larger) sites would be possible. A potential disadvantage of the proposed approach can be the computation time of the huge data set, in particular the geostatistical indices that were shown to be of subordinate importance. A recommendation for large scale applications, e.g. regional crop mapping, could be a pre-selection of small area subsets and the definition of the best performing features for crop classification, based on the methodology proposed in this manuscript applied to these subsets, and before calculating this selected feature subset over the total region. A similar attempt is recently being implemented in the context of the CAWA project, in which this study was carried out, for regional crop mapping in irrigated agricultural landscapes in Middle Asia. Further the proposed feature groups could also be applied to another classifier like DT or RF, possibly further reducing computation time.

This study demonstrates that the uncertainty measures derived from the SVM are informative and useful to characterize the spatial map quality. Two uncertainty measures, E and $H_\alpha rel$, were computed from the soft output of SVM and were evaluated with respect to their ability to correctly quantify classification uncertainty and consequently to predict classification error in maps (the predictive strength). It was shown that both measures performed well but different in the four sites. Yet, because $H_\alpha rel$ summarizes the information of the entire membership estimations in $p(x)$, it is to be preferred. Another advantage of $H_\alpha rel$ is that its behaviour can be controlled by the parameter α. The uncertainty maps allowed an interpretation of the causes of the observed spatial pattern of classification uncertainty by environmental factors. This information might be used in agricultural resource management as an additional means to identify water stressed areas in an irrigated agricultural environment, or to improve the reliability of modelling results and backward-tracing regions in crop maps with erroneous class allocations.

Chapter 5

A COMPARISON OF MACHINE LEARNING ALGORITHMS AND CLASSIFIER COMBINATION FOR OBJECT-BASED CROP CLASSIFICATION[19]

Abstract

This chapter deals with the comparison of different classifier algorithms for crop classification. The issue of combining the results of different classifiers to increase classification accuracy is also addressed. Specifically the support vector machine (SVM) approach employed in chapter 4 is compared with two other non-parametric classifier algorithms, the multilayer perceptron neural network (MLP), and the random forest (RF). Despite their comparatively high classification performance, each of the tested classifier algorithms tended to make errors in different parts of the input space, e.g. performed different in classifying crops. Hence, a combination of the complementary outputs was envisaged. To this end, a classifier combination scheme is proposed, based on and extending existing algebraic operators. It combines the outputs of different classifier algorithms at the per-case basis, based on their membership estimations, and assesses the reliability of a classifier with respect to classifying crop classes before the final combination, thereby excluding less reliable classifier algorithms at the per-class basis before the final combination. The experimental setup aims at comparing the proposed classifier combination scheme with its standalone classifier algorithms and other existing algebraic operators, respectively. Further, emphasis is put on evaluating the tested classification approaches under limiting conditions by applying the classifier algorithms to small input data sets and to reduced training sample sets, respectively. Further, the applicability to datasets from another year is demonstrated to assess the temporal transferability of the tested approaches. Although the single classifier algorithms perform well in all sites, this research demonstrates that the classifier combination scheme provides consistently high accuracies over the four landscapes and over different years, respectively. This makes this approach distinct from the single classifier algorithms, which perform different over the four landscapes and show grater variability in class-wise accuracies. Further, the classifier combination scheme performs better when using very small training set sizes and very small input data sizes, respectively. The selection of a classifier algorithm is further oriented by a desire to generate information on classification uncertainty and to indicate the spatial variation in classification quality in maps. Hence, the strengths of the tested classifier algorithms and the classifier combination, respectively in providing such spatially explicit information on classification uncertainty was evaluated, and it was found that the classifier combination provided the least uncertain results.

5.1.   Background

In the previous chapter the SVM was introduced as an effective and accurate means to classify crops at the object level. The algorithm was applied to different subsets of a redundant input space that contained multi-temporal spectral, textural, and geostatistical features. Particular attention was paid to the impact of feature selection on both spatial uncertainty and model prediction accuracy, respectively. It was demonstrated that one specific classifier algorithm like SVM cannot perform equally well on classifying all crop classes investigated, irrespective of the features selected. Chapter 5 addresses the second out of the three research questions that was formulated in the introduction, namely comparing different classifier algorithms for crop classification, and to evaluate if combining the results from different kind of classifier algorithms can enhance classification accuracy. Emphasis is put on evaluating the transferability and stability of the tested approaches among different agricultural landscapes and different years, and their performance under limiting conditions (e.g. training data scarcity and small input data sizes). Finally, it will be evaluated how uncertainty information computed from the soft outputs of the tested classifier algorithms relates to the hard result accuracy.

Whilst there is consensus that object-based classification approaches are in general more accurate than pixel-based, a single "best" classifier algorithm seems not to exist, e.g. no clear recommendation for a specific algorithm can be given when looking at several comparative studies (Gislason et al., 2006; Huang et al., 2002a; Pal and Mather, 2005; Pal, 2005). Despite the general high performance of recent machine learning algorithms (see appendix A.2), each of these classifier algorithms has its own advantages and disadvantages, and it might not be possible to identify the most suitable algorithm to classify a specific class or all classes in advance (Foody et al., 2007), and one classifier algorithm might perform different over different landscapes. Capitalizing the strengths of several algorithms can overcome the constraints of individual classifiers and enhance classification accuracy (Benediktsson et al., 2007). Classifier combination has received considerable attention in the past decades and is becoming an established pattern recognition technique (Benediktsson et al., 2007; Kittler et al., 1998; Kuncheva, 2004; Zhang, 2010). Combining the results of different classifier algorithms is sometimes referred to as "classifier ensembles" (Kittler et al., 1998).

The underlying precondition for classifier combination is that different classifier algorithms are accurate but perform unequal, e.g. produce errors in different parts of the feature space and perform different in classifying specific classes (Jain et al., 2000; Liu et al., 2004). In this way the combination of results from different classifier algorithms aims to make a better decision than taking only one source into account, thereby reducing the classification uncertainty of the final output (Bloch (1996), citing Bloch and Maître (1994)). One central issue in classifier combination is the way the different classification outputs are combined. This is done post-classification, either by aggregating the soft outputs or the class labels (the so-called hard outputs) from different classifier algorithms and then by applying a suitable combination rule (operators) to achieve a consensus decision (Kittler et al., 1998). The soft outputs are modelled in different ways, depending on the algorithm (see section 4.2.4 for SVM), e.g. as posterior probabilities, membership degrees, based on Bayesian theory or as fuzzy sets (Bloch, 1996). Various classifier combination schemes have been developed and it has been experimentally demonstrated that some of these techniques can consistently outperform a single best classifier (Benediktsson and Kanellopoulos, 1999; Kuncheva, 2004; Waske and Benediktsson, 2007), and that they are useful when classifying multi-source or multi-temporal remote sensing data (Benediktsson and Kanellopoulos, 1999; Briem et al., 2002; Jeon and Landgrebe, 1999; Solberg et al., 1994). There exist two broad strategies to generate such ensembles of classifiers: (i) combining the results from variants of the same classifier algorithm, or (ii) considering the combination of the results from multiple classifier algorithms.

The first strategy can be realized by modifying the input data (training data or input features). Iteratively resampling (or rather: reweighting) the training data by boosting strategies (Freund and Schapire, 1996) or bootstrap aggregating (bagging) of Breiman (1996) are well established ensemble strategies of this type. Random feature selection (RFS) is another concept for generating variants of a classifier, for instance Waske et al. (2010) performed RFS in SVM-based classifier ensemble by randomly choosing and classifying subsets of the feature space with SVM. The results were finally combined with another SVM applied on the rule images of the results. RF is another well-known machine learning technique, which combines bagging and RFS strategies (Breiman, 2001) and was shown to have good performances in crop classification (Pal, 2005; Waske and Braun, 2009).

The second strategy is to combine the outputs from different independent classifier algorithms using multi-classifier systems (MCS), which have been introduced in remote sensing applications to overcome the constraints of individual algorithms. By combining two or more independent classifier algorithms and performing classification separately on the input data different outputs are created and combined after classifying the input data separately (Steele, 2000; Waske and Benediktsson, 2007). For instance, Benediktsson and Kanellopoulos (1999) combined neural networks and statistical modelling to classify multi-source data.

The combination of outputs from different classifier algorithms, using either of the two aforementioned strategies, can be done by a diversity of methods, including averaging the predictions, statistical methods, majority voting strategies, weighted majority voting for boosting techniques, or fuzzy-logic-based methods if the classifier outputs are interpreted as fuzzy membership values (Bloch, 1996). Another powerful approach is combining the outputs of different classifier algorithms with another, independent classifier algorithm that is trained on the soft outputs of the different classifier algorithms (Waske and Benediktsson, 2007; Wolpert, 1992). A more general introduction to classifier combination and its application can be found in Benediktsson et al. (2007), Jain et al. (2000), Kuncheva (2002, 2004), and Bloch (1996) gives an overview of advanced fusion operators.

Different classifier algorithms have been tested for classifier combination. To have an example Giacco et al. (2010) combined the results from different SVM architectures and self-organizing feature maps (SOMs). Although the increase in accuracy was only up to 1 %, it has to be noted that the individual classifier algorithms already performed with accuracies of more than 90% and the class-wise improvement can be in the order of several percent increase. Benediktsson and Kanellopoulos (1999) combined neural network and statistical classifiers, where in conflicting situations, e.g. when classifiers disagree, another neural network classifier was used to assign the final class. Jeon and Landgrebe (1999) give an example for multi-temporal classification: they proposed two strategies to combine the results from multi-temporal classifiers applied to Landsat TM data that achieved 10 % increase in OA. Liu et al. (2004) combined the results from a tree structured and a neural network classifier (ARTMAP). Doan and Foody (2007) tested four approaches to combine two neural network and one statistical classifier algorithm and achieved improvements in OA of up to 4.45 %. Many of these concepts were developed based on hyper-spectral image data (Licciardi et al., 2009), but the fusion of different type of data (Benediktsson and Kanellopoulos, 1999; Briem et al., 2002; Waske and Benediktsson, 2007) and multi-temporal datasets (Bruzzone et al., 1999; Doan and Foody, 2007; Udelhoven et al., 2009) also gave convincingly results. Waske et al. (2010) proposed the combination of SVMs that were trained on different, randomly selected feature subspaces, which yielded an improvement of more than 5 % classification accuracy, depending on the training set size. An even more noteworthy result in that study was the class-wise improvement of accuracy that was more than 10 % for some classes, as compared with a regular SVM. Although the concept of classifier combination in remote sensing is attracting more and more attention it has to be noted that increasing classification accuracy is not granted (Foody et al., 2007), or might be statistically marginal (Giacco et al., 2010).

The classifier combination scheme (hereafter called: CCS) that is proposed in this thesis is designed in a generalizable fashion so that a combination of two or more classifier algorithms is possible, as long as the algorithms are capable of computing membership estimations (soft outputs) at the per-case level. The performance of the CCS is investigated and compared with other classifier algorithms and existing combination operators. Three classifier algorithms, RF, SVM, and multilayer perceptron neural network (MLP) were selected for the experiments in this study because of their general good performance in crop classification and because they have different technical properties (more details are given in appendix A.2):

> They are all non-parametric, that means that they are not constrained to assumptions like parametric distributions of the input data.

> RF and SVM appear effective when only few training samples for heterogeneous classes are available (Shao and Lunetta, 2012; Waske and Braun, 2009; Waske et al., 2010).

> The OAO strategy of SVM results in a larger number of classifiers, but therefore the classification problem is divided into many binary classifications (two classes) that are much simpler to solve.

> RF permutes the training data randomly and reduces correlation between the unpruned trees in the ensemble, which makes RF potentially robust toward redundancy in the input space.

> They follow different strategies, e.g. SVM implements structural risk minimization (SRM), the neural network classifier implements empirical risk minimization (ERM), see appendix A.2.

The overall objective of this research is to compare the performance of different classifier concepts under varying conditions and to compare the results with the CCS. This overall objective can be split in a series of partial objectives, defined as follows:

> Increasing the knowledge of the suitability of different classifier algorithms to correctly predict the classification uncertainty in maps, and consequently to use them in classifier combination.

> Evaluating the stability and spatial transferability of the selected classifier algorithms and the CCS by applying it to the four test sites with different field sizes and comparing the performance for another year, respectively.

> Investigating the applicability of the selected classifier algorithms and the CCS with reduced training sample sizes and small input data sets, respectively.

In the following, some background information on how soft outputs can be modelled by different classifiers is given and the parameterization (parameter tuning) of the selected classifier algorithms is highlighted. After a description of the CCS, it is implemented in realistic conditions and applied to EO data over the four test sites (the same data as described in chapter 3). The results are compared with the individual classifier algorithms. Emphasis is also put on comparing the accuracy of the CCS with existing combination operators, and the impact of classifier combination on the spatial distribution of uncertainty in the maps, respectively.

5.2. Methods

5.2.1. Classifier algorithms and parameter tuning

Configuring the classifier algorithm includes finding adequate values for the so-called free parameters (see section 4.2.2) that can have a considerable influence on the learning of an algorithm, and finally on the generalizability of the resulting classifier model. Depending on the algorithm there can be several free parameters that have to be configured. One straightforward way to optimize these parameters, even though not efficient in all cases, is parameter tuning. During tuning a range of values for one or more parameters is permuted and the accuracy of the classifier is tested. The final choice values of the parameters are oriented toward high performance of the model, e.g. the resulting model accuracy. This is done by a repeated l-fold cross-validation on the training set, l being the number of folds. In cross-validation, the training set is partitioned into l subsets of approximately equal size. Sequentially one subset is tested using the classifier trained on the remaining $l - 1$ subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data that are correctly classified. This process ensures that all subsamples are used as parts of the training and testing sets. Results for each of the l folds are then combined and the model with the highest average accuracy is selected. Examples of using cross-validation in remote sensing can be found in Brenning (2009), Friedl et al. (1999), and Huang et al. (2002a). In this study cross-validation for parameter tuning is performed with $l = 5$ folds. Stretching the data range can be another useful pre-processing step, when attributes with large original scale bias the solution, which can lead to degrading accuracies, e.g. in SVM classification (Ali and Smith-Miles, 2006). Hence, the data range of all input features was stretched into [0,1].

Random forest

The implementation of Breiman´s RF (Breiman, 2001) in the `randomForest` package (Liaw, 2013) in R (R Development Core Team, 2012) was used for the experiments. As was discussed in section 4.2.2, two free parameters can be optimized: the number of trees in the ensemble and the number of features m_{try} to split the nodes in the trees (Breiman and Cutler, 2007). As is considered adequate in literature (for categorical classification) the number of features at each node was set to the square root of the total number of input features, $\sqrt{f}$, where f is the number of predictor variables (features) within a dataset (Gislason et al., 2006). The final number of trees was set to a relatively high number of 300 for each study site, in order to allow for convergence of the OOB error statistic (see Figure 5-1). In this figure it can be seen that adding more than 300 trees, irrespective of the number of random features (m_{try}) at each split node tested, does not affect the error rates, neither measured from the OOB, nor from an independent test set.
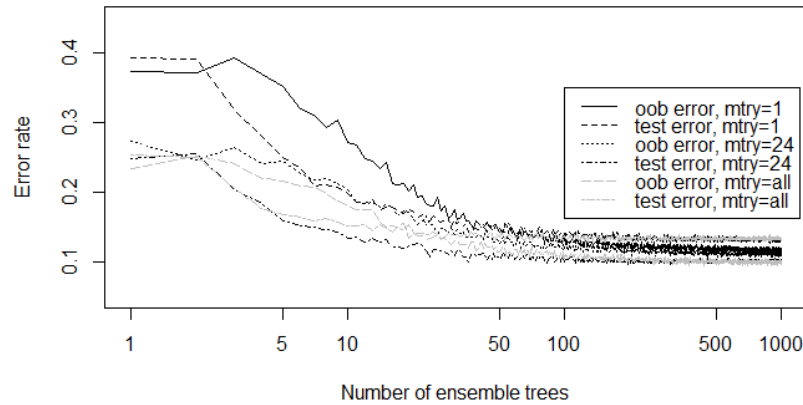
Figure 5-1: Impact of number of trees in the ensemble on OOB and test set error rates, for different numbers of random features (m_{try}) at each split node tested.

Support vector machine

The implementation of SVM in the package `e1071` (Meyer et al., 2012) in R provides an interface to the `libsvm` library from Chang and Lin (2011). Training of the SVM includes choosing adequate values for the kernel parameter γ and the regularization parameter C (see section 4.2.2 for further details). Two different kinds of SVMs were considered: classification with a linear SVM (SVM_L, without kernel transformation, see appendix A.2) and classification with a nonlinear SVM based on the widely known RBF kernel (SVM_R). For both SVMs, C must be tested, and for the SVM_R adequate values for γ must be determined in addition. Tuning C and γ was done using a systematic grid search in 2-D space that is spanned by γ and C. The range of γ was [0.00125, 2], the range of C was finally set to [1, 200]. While some studies employ a grid search similar to that described before (Foody and Mathur, 2006; Shao and Lunetta, 2012; Waske et al., 2010), others used fixed parameter values while testing different input space sizes (Pal and Foody, 2010). The range of γ was [0.00125,2], the range of C was finally set to [1,200].

Multilayer perceptron neural network

In MLP classification the most critical parameter is the number of hidden layer nodes, which has a significant impact on the classifier performance, e.g. selecting too many nodes will cause over-fitting of the model to the training set and reduce the generalizability of the model, while too few nodes cannot identify the internal structure of the input data. If the learning rate is too low the model might end up finding a local minimum instead of the global minimum (Kavzoglu and Mather, 2003), which results in a loss of its so-called generalization ability, e.g. to classify data correctly that was not used in the training stage (Atkinson and Tatnall, 1997). In this thesis three parameters were tuned: the number of hidden layer nodes, the learning rate, and momentum. Following the recommendations of Kavzoglu (2001) the learning rate was chosen between 0.1 and 0.2, and the momentum factor between 0.5 and 0.6, the learning rates a set of discrete values {0.1, 0.125, 0.15, 0.175, 0.2} was tested, and for the momentum the set {0.5, 0.525, 0.55, 0.575, 0.6} was tested. The number of hidden layers was set to one, and the number of hidden layer nodes examined was in a range of $[n, \sqrt{n + f}]$, where n is the number of classes, and f the number of input features, an approach also used by Hu and Weng (2009). The number of iterations determines the ability of the MLP to model the complex input data, however increasing the number of iterations can also decrease the ability of the model to classify the unseen data correctly (Atkinson and Tatnall, 1997). After some trials the number of iterations was set to 200 to reduce the risk of over-fitting. The number of output nodes corresponds to the number of classes n, the number of input nodes correspond to the number of input features f.

Feature selection

As was demonstrated in chapter 4 the performance of classifier algorithms can be negatively affected by a huge input space that contains redundant information. Hence, the RF was used again as feature selection strategy and several classifications were performed by incrementally adding features to the single classifiers in order suggested by the RF feature score. This was done in groups of fifteen, beginning with the 10 most important features from the ranked list. The RF, each of the SVMs, and the MLP models were finally built on the reduced feature subsets that yielded the highest OA.

5.2.2. Classifier combination scheme

In the following the CCS is describe in more detail. Emphasize is put on modelling the soft output from different classifier algorithms, as these are the input to the classifier combination. Finally the combination operator that is proposed in this study is described. Figure 5-2 illustrates the processing steps involved in the proposed CCS, which may guide the reader throughout the following descriptions. After the calculation of the input features from the RapidEye time series the information was aggregated to a stacked vector, containing all 568 multi-temporal features described in chapter 3. A feature reduction is then performed for each classifier algorithm separately, based on the RF feature importance score, and the final outputs of each classifier algorithm tested (class labels and soft output for each classified object) are combined according to the combination operator defined for the CCS.
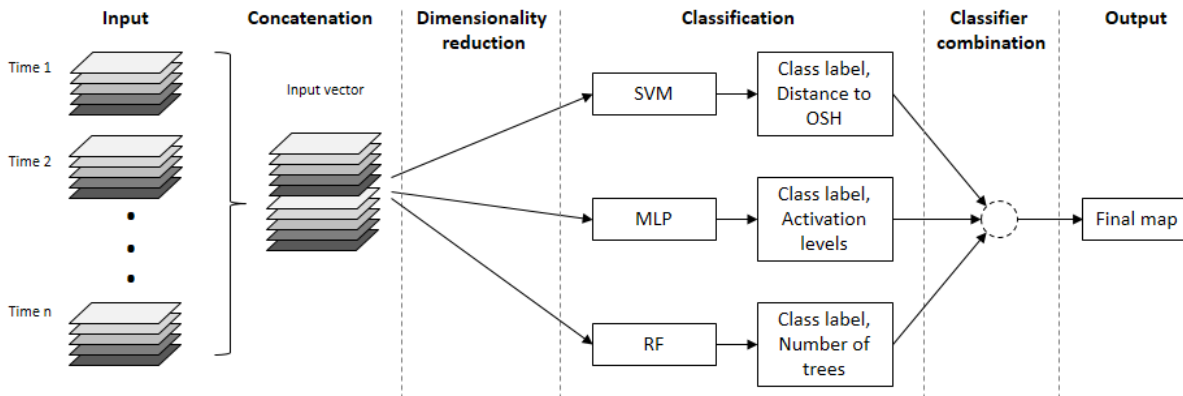
Figure 5-2: Schematic diagram of the CCS proposed in this thesis. "Distance to OSH", "Activation levels", and "Number of trees" refer to the soft outputs computed from the corresponding classifier algorithm, and is explained in the next section.

Modelling soft outputs from different classifier algorithms

Many parametric and non-parametric classifier algorithm can generate a soft output (e.g. estimations of class memberships) that allow for a quantification of the classification uncertainty, but the way this soft output is modelled depends on the classifier algorithm. In RF classification the soft output is calculated as frequencies of observed class values occurring at the leaves of the trees within the forest:

$$pr_i(x) = \frac{T_{(i)}}{T} \qquad \qquad Eq.\ 5\text{-}1$$

where $pr_i(x)$ is the estimated posterior probability for class i, $T_{(i)}$ is the number of trees voting for one class i, and T the number of trees in the ensemble.

Despite its simple computation the evaluation of spatial uncertainty information computed from the soft output of RF has not been fully evaluated in remote sensing applications yet, e.g. the strength of the estimated class probabilities of RF in predicting classification uncertainty (Loosvelt et al., 2012a).

In SVM classification a post-processing method for mapping the soft outputs of a single SVM into posterior probabilities has been proposed by Platt (2000) and improved by Lin et al. (2007), who's SVM implementation can be found in their libsvm library that is used in this study. More details on this issue were given in section 4.2.4 of this thesis.

The output of neural network classifiers can be softened in order to output, for each classified case x, an activation level for each of the classes investigated (Foody, 2000), see appendix A.2 for further details. The magnitude of activation levels may be used to indicate the strength of membership that may be associated with a particular class allocation (Foody, 1996a), although this information is not a direct measures of posterior probability but rather a "by-product" of the algorithm (Brown et al., 2009). MLP provides an activation level a_j of each class at the per-case basis. The transformation of this information to a hard result is accomplished by assigning each case to the class with the highest activation level. a_j may additionally be used to produce soft labels. In this study all a_j are normalized to a common scale [0,1], so that all a_j for a case x sum to one. It is assumed that a large activation level has a larger probability of correct class allocation (Foody, 1996a, 2000; Gong et al., 1996). Brown et al. (2009) and Hu and Weng (2009) used the soft output from neural network classifiers, e.g. maps of activation levels to quantify the classification uncertainty at the pixel-level. Critically, the predictive strength of the soft outputs from neural network classifiers has attained only few attention in the literature (Brown et al., 2009), hence it will be evaluated before considering MLP for its application in the classifier combination.

Measures of classification uncertainty

As was already detailed in section 4.2.3, measures of classification uncertainty like entropy can be computed from soft outputs of classifier algorithms. Here it is considered that several classifier algorithms m assign membership estimations to a classified case x. Let for a given data set consider n classes using j different classifier algorithms. For each case x (agricultural field or pixel) a classifier algorithm j calculates a soft output in form of a vector that contains the estimated membership probabilities that x is classified into a class i,

$$pr_j(x) = \{pr_{j,1}(x), pr_{j,2}(x), \dots, pr_{j,i}(x), \dots, pr_{j,n}(x)\} \ , \ i = 1 \dots n, j = 1 \dots, m \qquad \text{Eq. 5-2}$$

where $pr_{j,i}(x)$ is the estimated membership degree of x to class i, according to a classifier j, and n the number of classes. Each element $pr_{j,i}(x)$ in $pr_j(x)$ is a class membership estimation by j, e.g. that x actually belongs to a class i. The elements in $pr_j(x)$ generally take their values in a closed interval, e.g. $pr_{j,i}(x) \in [0,1]$, thereby adding up to 1, and are the input in classifier combination or decision fusion methods (Bloch, 1996; Kuncheva, 2002, 2004).

As was demonstrated in section 4.3, the higher these values, the more likely does x belong to class i. Low values indicate a dubious result that might occur in the presence of mixed pixels, or be related to sub-pixel cover proportions (Foody, 1996b) or heterogeneous classes. When m classifiers are used to classify x, then m vectors $pr_j(x)$ are calculated for each x:

$$\{pr_1(x), pr_2(x), pr_j(x), \ldots, pr_m(x)\}. \hspace{3cm} \textit{Eq. 5-3}$$

For each $pr_j(x)$ a measure of uncertainty can be calculated. In this study, E_j was selected to evaluate classification uncertainty.

Combination rule

There exist a variety of methods to combine the outputs of individually trained classifier algorithms, e.g. from simple algebraic operators (Kittler et al., 1998) to more sophisticated fuzzy rules like context dependent operators (Bloch, 1996) or Dempster-Shafer techniques that consider the reliability of each source (Bloch, 1996; Kittler et al., 1998). Whilst some combining rules, e.g. the majority vote, operates on the class labels of the algorithms (the hard output), the soft output of classifier algorithms can be combined using rules like the algebraic SUM, MIN, and MAX operators or the product, median, or mean rules (Policar, 2006). Those and other rules are detailed and compared for classification in Kittler et al. (1998), Kuncheva (2002), and Policar (2006).

Combination rules based on algebraic operators were shown to be very effective and accurate (Kittler et al., 1998). In general, these operators are not trainable (Policar, 2006), as they do combine the soft outputs using more or less simple algebraic expressions, e.g. the maximum value out of all elements in $pr_j(x)$. Consequently the training and parameter tuning effort is focussed on the single classifier algorithms, whilst the combination operators apply to their outputs. In general, for a classifier j the final class i becomes the class with the largest support $\mu_j(x)$, after one of the aforementioned rules is applied to summarize the membership probability estimations (Benediktsson et al., 2007; Policar, 2006):

$$\mu_j(x) = arg\, max\{pr_{j,i}(x)\}\,,\ i \in [1, n]. \hspace{3cm} \textit{Eq. 5-4}$$

The final class support can be computed with one of the widely known combination rules, e.g. the MAX rule (Kittler et al., 1998) by determining the highest support among all m classifier algorithms:

$$C_{fin} = arg\, max\, \{\mu_j(x)\}\,,\ j \in [1, m]. \hspace{3cm} \textit{Eq. 5-5}$$

The SUM rule applies a maximum value selector to the averaged soft outputs of each class, thereby averaging the potential variability of the per-class soft outputs (Policar, 2006). A potential drawback of such algebraic operators is that they do not explicitly take into account the reliability of the classifier algorithms, e.g. their performance with regard to classifying a particular class i. An extended combination rule is proposed here that quantifies the reliability of each classifier algorithm with respect to classifying a specific class i, quantified by CA_i from the confusion matrix, and excludes it from the classifier combination if a certain criteria is not fulfilled (see below). The results will be compared with two existing algebraic operators (SUM and MAX).

The proposed combination rule works as follows: for each unseen (unclassified) case x a given classifier algorithm j calculates a soft output vector $pr_j(x)$ and suggests a hard class label (i) assignment. As was already mentioned, the combination rule is based on the per-case estimations of class memberships, e.g. the decision on the final class is made for each pixel or object. Yet, each classifier

algorithm´s reliability is assessed before the final combination by its class-wise accuracies CA_i. For a given class i the classifier j with highest CA_i is selected first, and only the classifier algorithm whose CA_i does not differ more than 5 % from this best classifier was combined (a range of thresholds was tested but 5 % achieved reasonable results: 1 %, 2 %, 10 %, 15 %, and 20 %). This results in $\dot{m}$ classifiers that can compete ($\dot{m}$ being at least 1 classifier that can compete for a class i). This was done because a classifier algorithm j might be reliable for one specific class, but unreliable for another class (see results in section 5.3.2 and 5.3.3). A classifier algorithm is reliable if its class-wise accuracy does not deviate more than 5 % from the single best classifier´s CA_i. As a result from this, the number of classes n for which one classier algorithm j can compete becomes $\bar{n} \subseteq n$. The final class i then becomes the class with the largest support $\bar{\mu}_j(x)$:

$$\bar{\mu}_j(x) = arg \ max \ \{pr_{j,i}(x)\} \ , \ i \in [1, \bar{n}]. \qquad\qquad\qquad Eq. \ 5\text{-}6$$

The final combination rule is rather intuitive and assigns the final class to x by the classifier j with the largest support $\bar{\mu}_j(x)$:

$$C_{fin}(x) = arg \ max \ \{\bar{\mu}_j(x)\} \ , \ j \in [1, m]. \qquad\qquad\qquad Eq. \ 5\text{-}7$$

E_j can be computed for each classified case x as an uncertainty measure (see section 4.2.3), but is defined here as:

$$E_j = 1 - \bar{\mu}_j(x), \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad Eq. \ 5\text{-}8$$

with $\bar{\mu}_j(x)$ being the largest support in Eq. 5-7. In the following, E_j as defined in Eq. 5-8 will be used to assess the spatial uncertainty of the CCS, whilst E_j as defined in 4.2.3 (e.g. computed considering all elements in the soft outputs of each classifier algorithm) will be used to evaluate classification uncertainty of the single classifier algorithms.

5.3. Experimental results

In this section experimental results from the selected classifier algorithms and the CCS are presented and confronted with two additional combination operators, SUM and the MAX, that were shown to be a suitable combination rule in classifier combination (Kittler and Alkoot, 2003; Kittler et al., 1998). The impact of feature selection on classification accuracy is briefly discussed. An evaluation of the soft outputs, computed on the test sets, is also given to evaluate their potential usefulness with regard to classifier combination, similar to the tests in section 4.3.5.

5.3.1. Impact of feature selection on classification accuracy

As already discussed in chapter 4 different classes might require different types of input features to have an accurate result. As it is unknown in beforehand which type of features are suitable, the input data set used in chapter 4 is used again as input. Yet, given the possible negative impacts of the feature space size on the classifier performance (see section 4.3), a feature reduction based on the RF is performed for each classifier separately.

Figure 5-3 plots OA of the classifier algorithms (RF, SVM_R, SVM_L, and MLP) as a function of the number of selected features. In all sites the number of features influences the classification accuracy, but the magnitude of this impact differs. Among the classifier algorithms the RF is the least affected.

The accuracy of RF reached a peak accuracy after approximately 40-70 features were added and then stabilized at this high level with quite low fluctuations of ± 1 %, in all test sites. This demonstrates the superiority of the RF in handling huge input spaces that contain redundant information (see section 4.3.3). The MLP showed similar patterns in all sites expect for KHO, where accuracy decreases after adding 40-50 features. In all sites the accuracies of SVM_L tended to be lower than the SVM_R. In FER and KYZ the accuracy of SVM_L increased with the addition of features up to a certain level between 235 and 265 features, and stabilized thenceforward. Interestingly in all sites the accuracies of the SVMs tended to converge when more and more feature are added. Similar to the SVMs the accuracy of the MLP stabilized in KKP and FER at comparatively high levels after an initial number of features was added.
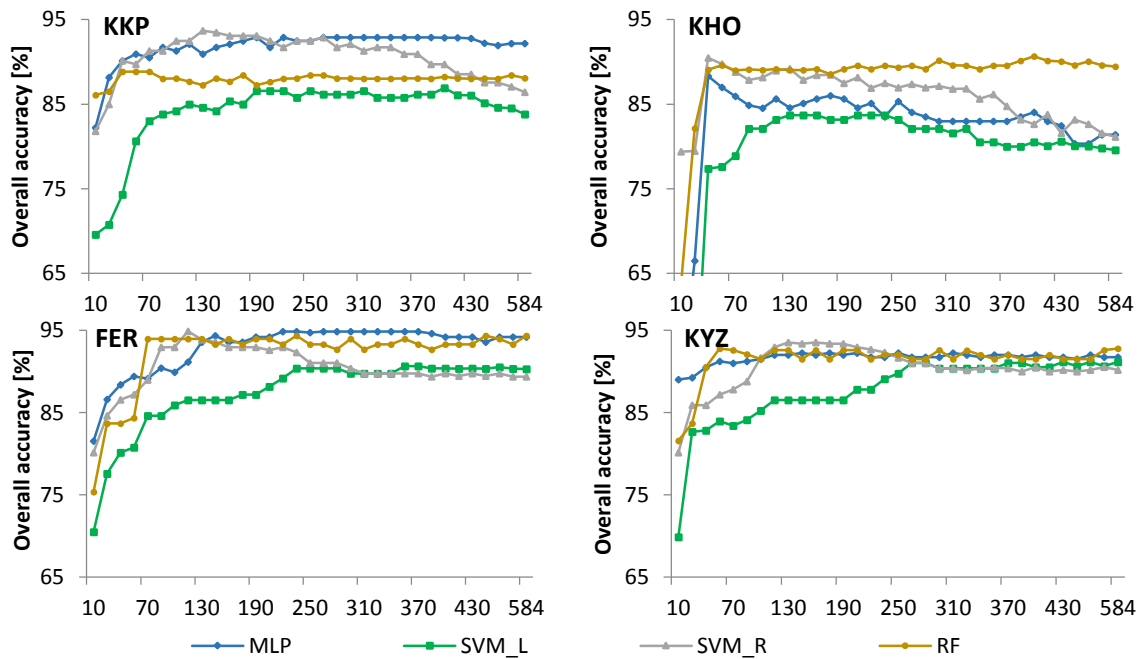
Figure 5-3: Overall accuracy (y-axis) as a function of the number of features (x-axis), obtained on the test set by the four different classifier algorithms considered (RF, SVM_L, SVM_R, MLP). Note: in chapter 4 the experiments were repeated 50 times, hence the results here deviate from the results shown in chapter 4.

5.3.2. Evaluation of classifier uncertainty

As was already mentioned at the beginning of this chapter, different classifier algorithms might assign different levels of uncertainty to their final class decisions. This poses the risk that overly optimistic estimations outperform other classifier algorithm during the classifier combination process. In order to test this, and to limit a possible influence of overly optimistic classifier algorithms in the classifier combination, two tests were undertaken before the final classifier combination. First, the frequency distributions of correctly and incorrectly classified testing cases lying within intervals that were defined on the magnitude of the E_j values are tested, as was done in section 4.3.5 for the SVM, to assess if the uncertainty measures can be indicative of correct classifications. The shape of these distributions gives an indication for the reliability of the classification and the potential usefulness of uncertainty measures to indicate correct or incorrect class allocations in the final map. In a second test the frequency distributions of uncertainty measures for all classified fields were evaluated.

Frequency distributions of test cases

From the frequency distributions (Figure 5-4) it is evident that correctly classified test cases tend to display relatively low E_j values. In all sites their frequency peaks within the lowest interval [0.0, 0.2]. This is followed by a decrease toward higher uncertainty values. This confirms the high prediction strength of the four classifiers over all sites, although it has to be noted that the RF has the highest prediction strength among all algorithms. More than 85 % of the correctly classified test fields (by the RF) are lying within the interval [0.0, 0.2], and only a very limited fraction of the correctly classified pixels was associated with the three lower intervals of E_j values above 0.2.
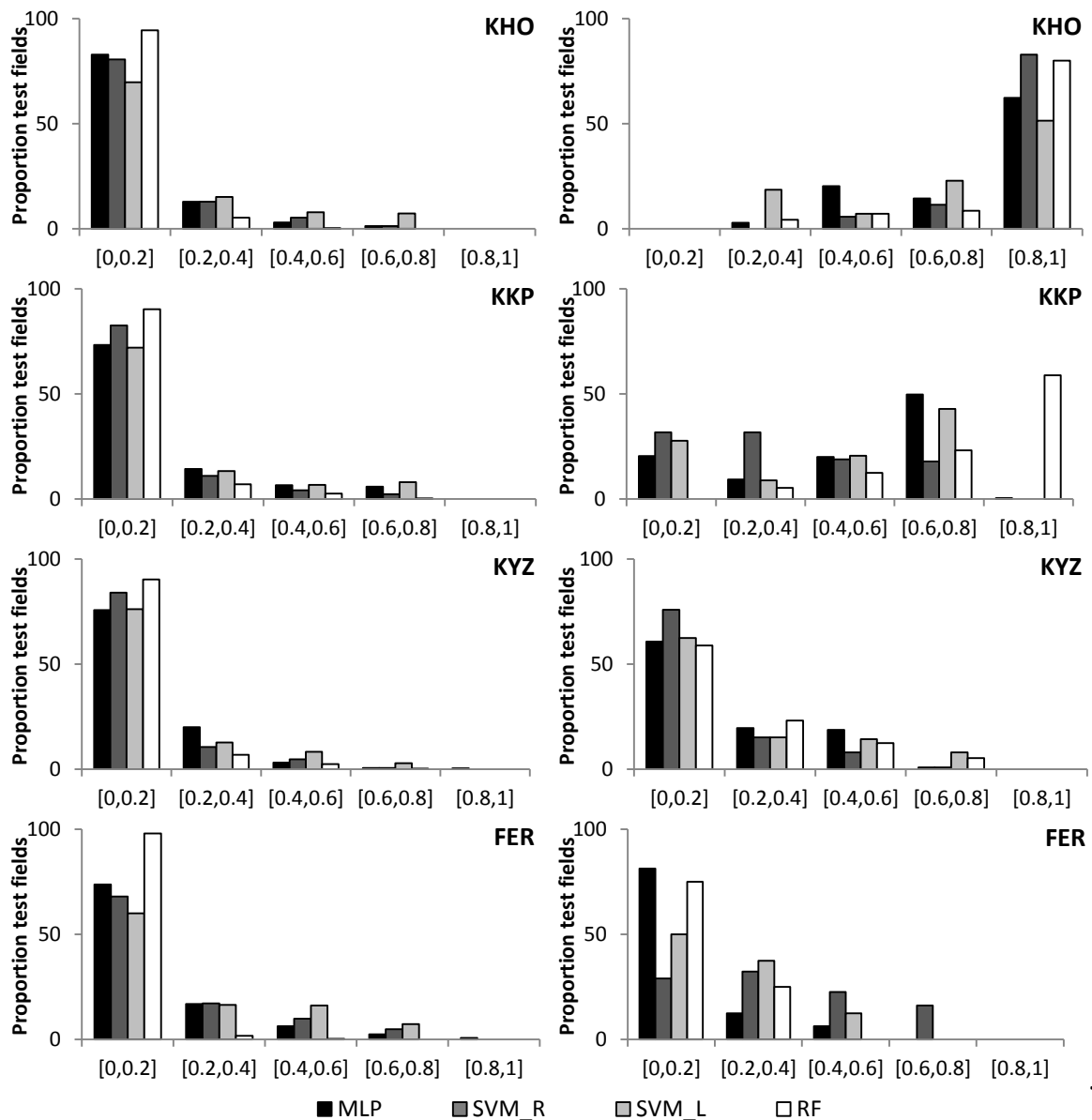
Figure 5-4: Frequency distribution [%] of E_j for the correct (left column) and incorrect (right column) classified test cases (fields) as resulted from the single classifier algorithms over the four test sites.

For the incorrectly classified test cases the pattern is different. E_j tends to increase toward higher values in KKP and KHO only. This indicates that incorrect predictions were made with higher uncertainties, and the prediction was uncertain due to confusion between other classes. Yet, the shape

of the distribution of misclassified cases in FER and KYZ was found to be different (similar to the results in section 4.3.5). Here the misallocated cases lie in the lower intervals of E_j. This confirms that the errors in these sites are confident misallocations, in which the fields were allocated to the wrong class but with little classification uncertainty, e.g. there was little confusion between crop types. An explanation for this could be the presence of relatively few classes in FER and KYZ (five), compared with KKP and KHO (eight and seven, respectively). For instance from the misclassified cases by the SVM_R in KYZ more than 70 % occurred in the [0.0, 0.2] interval of E_j. Based on this results it can be concluded that correct predictions in all sites were made with limited confusion with other classes. Incorrect classifications were made with moderate or low values of E_j in KKP and KHO. The RF outperformed the other classifier algorithms with regard to the predictive strength, e.g. the RF tended to better associate correct class predictions with lower uncertainties. From the two SVMs considered the SVM_R assigned higher portions of correct predictions in the lower uncertainty intervals and assigned higher uncertainty to incorrectly classified cases in KHO and FER.

Frequency distributions of uncertainty from all classified fields

In a second experiment the uncertainty of the produced crop maps was evaluated. Figure 5-5 exemplarily shows the cumulative relative frequency distribution of E_j for all classified fields in FER (class-by-class). These curves give an indication of how uncertainty is distributed among the crop classes in the final map and for each classifier algorithm investigated.
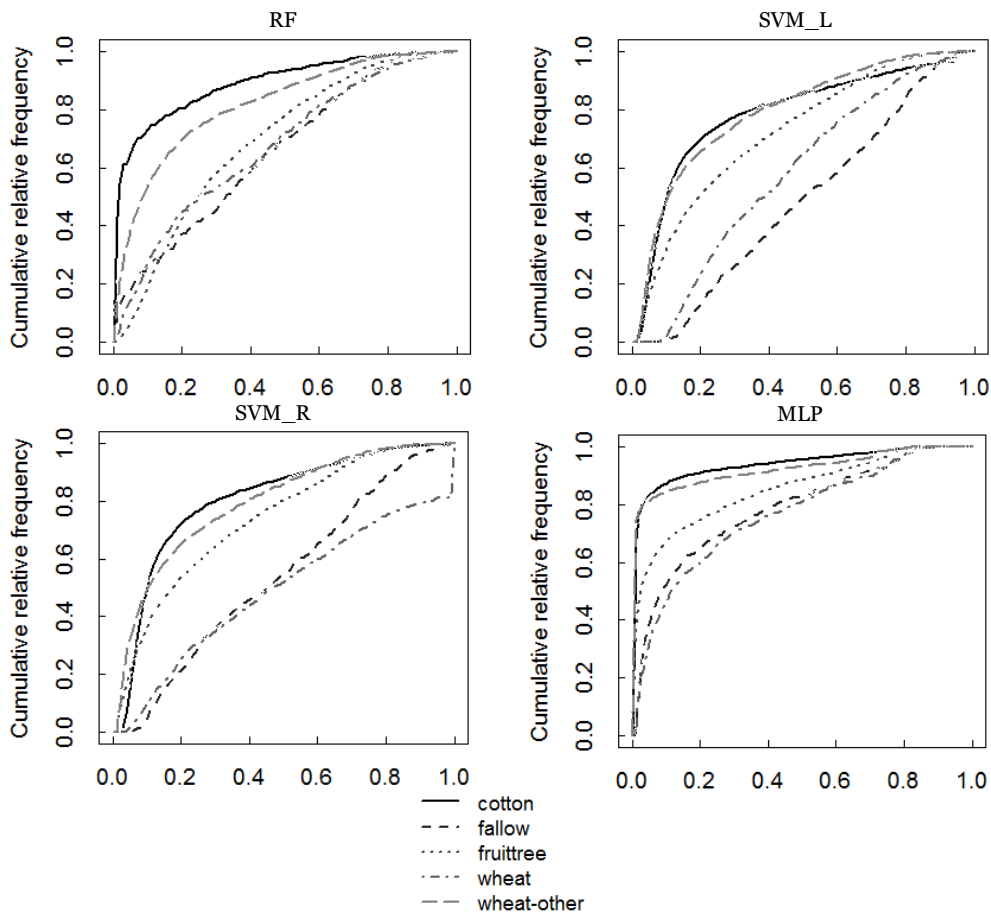
Figure 5-5: Cumulative relative frequency distributions of classified fields in FER, class by class, as a function of E_j

Looking at the curves in Figure 5-5 it can be concluded that the RF assigned E_j values of less than 0.2 to almost 80 % of all classified cotton fields in FER. On contrast more than 50 % of fields classified as fallow had E_j values of more than 0.4, which indicates that there was more confusion with other classes. The shape of the distributions of SVM_R and SVM_L differed only slightly, however the SVM_L results tended to be more uncertain, reflected in the curves of fallow fields and winter wheat. From the curves it can be seen that the MLP had the least classification uncertainty of all classifier algorithms concerning cotton and wheat-other, the two dominating classes in FER. Compared with the other classifier algorithms the MLP also tended to assign lower uncertainty values to fruit trees. The MLP tended to have the most cases with zero uncertainty (e.g. cases where there is no confusion with other classes, resulting in $E_j = 0$), e.g. reflected by the steep curves of cotton and wheat-other at lower uncertainty values. Further, the membership values were allocated in a different way as compared with the other classifier algorithms, e.g. a majority of the cases were allocated within a narrow range of uncertainty values (approximately between 0 and 0.8, whilst for the SVM_R some cases had uncertainty values higher than 0.90. In summary the MLP assigned the lowest classification uncertainty to the results in FER, and RF tended to assign the second lowest uncertainty values.
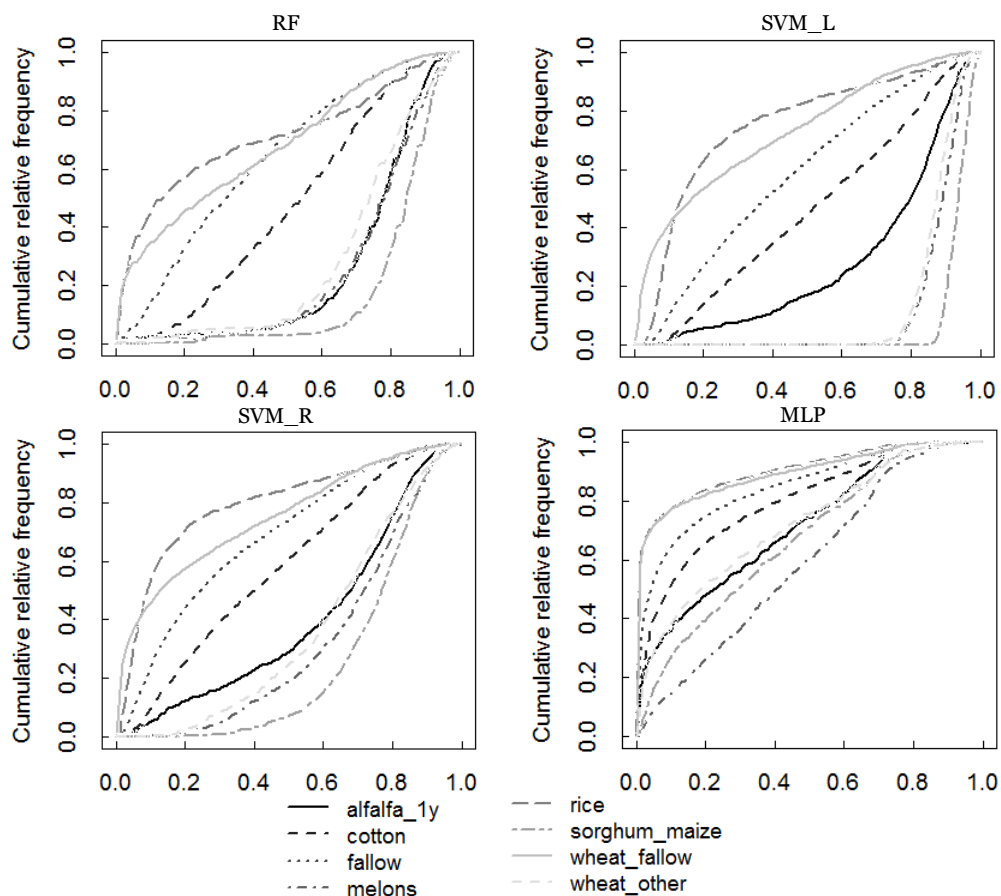
Figure 5-6: Cumulative relative frequency distributions of classified fields in KKP, class by class, as a function of E_j.

To exploit the uncertainty of a map with more classes the cumulative relative frequencies were computed for KKP (Figure 5-6). Similar to FER the MLP in KKP exhibited a different behaviour with respect to the other classifier algorithms: MLP assigned almost no cases with uncertainty values higher than 0.8, with a strong accumulation of cases classified with low uncertainty (e.g. $E_j < 0.4$), and the range of classification uncertainty is larger for the remainder classifier algorithms. Recalling Figure 5-4 it seems that the uncertainty values of the MLP could be too optimistic in KKP, because RF and SVM_R

can better associate correct predictions with lower uncertainty values than the MLP. Looking at the curves in Figure 5-6 it can be concluded that the class allocations of alfalfa-1y, sorghum/maize, melons, and wheat-other were characterized by the highest classification uncertainty, reflected in the high proportion of cases with uncertainty values above a threshold of 0.5 (specifically for the RF and SVM_R), where most of the corresponding curves rose quickly (expect for the MLP). This fits well to the fact that alfalfa in general has a very indistinct temporal growth pattern (due to the multiple and irregular cutting operations within the growing period, see section 2.2). From all curves it can be concluded that winter wheat, rice, cotton, and fallow fields are classified with lowest uncertainty. SVM_R produced maps with lower classification uncertainty than SVM_L, specifically for the classes sorghum/maize, melons, and wheat-other.

5.3.3. Evaluation of classification accuracy

Performance of the single classifier algorithms

In this section the results of the single classifier algorithms, RF, SVM, and MLP are presented. Having performed the feature reduction for each algorithm separately, the classification performances were compared (Table 5-1). In this table the class-wise accuracies achieved on the test sets reveal the complementary behaviour of the classifier algorithms, which makes them so interesting in the context of classifier combination. As an example, this diametrical behaviour was evident in the KKP landscape where the MLP algorithm was better to accurately classify wheat-other (CA_i = 75.0 %), as compared with the SVM (CA_i = 66.6%), but the SVM produced better results for sorghum/maize (CA_i = 60.0 %), as compared with 46.1 % when using the MLP. In the FER landscape the RF could best assign the class winter wheat, and was more accurate (> 5.0 %) then the next best classifier for this class, the MLP. SVM could best assign fallow fields, and the MLP was the most accurate algorithm in correctly assigning wheat-other fields. Comparing Figure 5-3 and Table 5-1 it can be seen that after a feature reduction was performed the SVM_R could outperform the RF in all sites, although this difference was less evident in KYZ and FER. Comparing SVM_R and the MLP after the feature reduction, it became evident that the former performs more accurate, expect for in FER, although the differences were marginal. The higher accuracies for the KYZ and FER sites can be attributed to the smaller number of classes and, in KYZ, well-managed fields resulting in more homogeneous structures in this data set.

Final choice of classifier algorithms for classifier combination

The choice of the final classifier algorithms to be used in the CCS can be reasoned as follows: given that the SVM_R performed best in all sites, the RF best assigned low classification uncertainty to correctly classified test cases, the MLP created the maps with the least uncertainty, and because SVM_R performed better in assigning low classification uncertainty to correctly classified test cases than SVM_L (Figure 5-4), the SVM_L was discarded and the remainder three algorithms were selected in the CCS. In the next section it will be assessed if these differences can be used to achieve better classification performances. Remembering the rules of the CCS defined in section 5.2.2, and when looking at Table 5-1, obviously there was no clear "winner" expect for the SVM_R, which was allowed to compete for all classes in KYZ (because it performed the most accurate for the classes), and for most of the classes in FER, KYZ and KHO. Interesting in the context of uncertainty evaluation and classifier combination is that the MLP algorithm in FER only achieved best accuracies for one class (wheat-other), and achieved lowest OA. Yet, from the frequency distributions of the classified fields (Figure 5-5) this was not obvious because the MLP assigned more fields to low levels of uncertainty than the RF or SVM_R. This is similar

to KKP, where MLP only performed best for three classes (Table 5-1). Most probably the membership estimations (soft output) from the MLP are, to a certain degree, optimistically biased, as they did not in any case reflect the accuracy statistics calculated from the independent test set (Table 5-1). In this regard the strength of the proposed CCS becomes obvious because it can exclude those classifier algorithms from the classifier combination for those classes they erroneously classify, but with little uncertainty.

Table 5-1: Average class-wise (AA), class-wise (CA$_i$), and overall accuracies (OA) of the single classifier algorithms on the test set. The crop classes include cotton (C), fallow (FA), rice (R), sorghum/maize (S), alfalfa-1y (A1), alfalfa-3y (A3), melons (M), fruit trees (F), winter wheat (W), and wheat-other (WO).

	Class	RF	SVM_R	MLP		Class	RF	SVM_R	MLP
FER	C	100.0	100.0	99.1	**KHO**	C	87.8	93.9	89.8
	FA	94.7	96.0	94.7		FA	100.0	98.0	97.6
	F	84.6	92.0	91.2		R	82.7	74.9	74.6
	W	88.8	80.0	83.3		S	9.0	36.3	9.0
	WO	94.3	94.2	95.0		F	91.7	90.6	90.5
	AA	92.5	92.4	92.7		W	100.0	100.0	81.8
	OA	94.3	94.5	94.2		WO	94.0	94.1	93.7
						AA	79.4	84.0	75.4
	Class	RF	SVM_R	MLP		OA	89.4	90.4	88.6
KKP	A1	72.7	100.0	100.0					
	C	75.8	93.6	93.8					
	FA	87.1	93.2	86.9		Class	RF	SVM_R	MLP
	M	66.6	84.2	71.4		A1	84.6	92.5	81.2
	R	100.0	94.4	97.2		A3	85.7	90.7	87.3
	S	22.2	60.0	46.1	**KYZ**	FA	76.2	83.0	74.0
	W	94.5	98.9	97.7		R	99.6	99.8	99.6
	WO	57.1	66.6	75.0		W	9.0	70.0	42.1
	AA	72.0	86.3	83.5		AA	69.2	87.2	76.8
	OA	88.0	93.6	92.4		OA	92.7	93.5	92.1

Performance of the classifier combination scheme (CCS)

Table 5-2 shows the OA of the classifier combination schemes tested, and compares them with the OA of the individual classifier algorithms. The RF achieved OA between 88.0 % and 94.3 %, the SVM_R achieved OA between 90.4 % and 94.5 %, and the MLP between 88.6 and 94.2%. Although the overall classification accuracy is comparable high, the classifier algorithms showed a certain degree of disagreement (Table 5-2) regarding the final class decisions, e.g. they disagreed for some cases and gave different results. The predictions of the single classifier algorithms agreed in 65.7 %–87.1 % of all classified cases, whereas in sites with more classes the disagreement was higher. Disagreement for cases that were classified with low certainty ($E_j < 0.2$) only occurred in less than 4% of all classified cases (not shown in Table 5-2), indicating the high reliability of the classifier algorithms as stand-alone methods for crop classification in the test sites. High-level agreement among the individual classifier algorithms, e.g. when the same prediction was made with $E_j < 0.2$, was found for 49 % (KHO), 40 % (KKP), 64 % (KYZ), and 75 % (FER) of all classified cases (not shown in Table 5-2). This emphasizes the difference between test sites with many classes (KKP and KHO) and test sites with fewer classes (KYZ and FER). Another important point is the observed improvement in OA after the CCS was applied (Table 5-2). This improvement was most pronounced in KHO, and the second most in FER with only five classes. The CCS gave less evident results in KYZ, also with five classes. Accuracies of the CCS ranged between 93.9 % and 96.1 % in the test sites, achieving the lowest range of OA over all sites, compared with the single classifier algorithms (2.2 %). This indicates a better stability of the results and transferability of the CCS compared to the single classifier algorithms, irrespective of the number of classes. Further, the CCS outperformed the two other classifier combination rules, based on the *MIN* and *SUM* operators, respectively, most obvious because it considers the reliability of the single classifier algorithms.

Table 5-2: Overall accuracies [%] of the individual classifier algorithms and the classifier combination schemes tested. "Agreement" is the percentage of fields for which all three classifier algorithms classified the same class. The difference in overall accuracies between the CCS and the single best classifier algorithm is given below.

Test site / Methods	KHO	KKP	KYZ	FER	Range	Combination rule
Single classifier algorithms:						
RF	89.4	88.0	92.7	94.3	6.3	
MLP	88.6	92.4	92.1	94.2	5.6	
SVM_R	90.4	93.6	93.5	94.5	4.1	
Agreement [%]	71.3	65.7	79.0	87.1	-	-
Classifier combination schemes:						
	89.4	92.8	93.2	94.0	4.6	*MIN* operator
All classifiers	93.4	92.4	92.3	94.7	2.4	*SUM* operator
	94.4	94.6	93.9	96.1	2.2	CCS
Difference between CCS and single best classifier algorithm	4.0	1.0	0.4	1.6	-	-

The relative importance of each classifier algorithm in the CCS was assessed by computing the magnitude that each considered classifier algorithm assigned the final class label (Table 5-3). The RF won the most cases in KHO and FER. This might at first glance contradict the general superiority of the SVM_R when regarding the classification performance (e.g. in terms of OA) that is summarized in Table 5-2. For instance, in KHO the SVM_R assigned the final class label to only 39.4 % of the fields, and RF more than 41 %. But the RF was able to provide more confident outputs than the SVM_R (e.g. with lower classification uncertainty), as indicated by the frequency distributions of E_j on the test cases (Figure 5-4). To conclude this, although the SVM_R achieved higher OA than the RF or the MLP, it was the RF that had the least doubt about its predicted classes. A similar behaviour was observed in FER, although the differences in OA were less evident. RF assigned the most classes and SVM_R the least. In KKP the percentages of fields won by the classifier algorithms reflect the order relation of classification OA. The MLP never assigned the most cases in any study site, but more than RF in KKP and KYZ, and more than SVM_R in FER.

Table 5-3: Percentage of fields whose class was assigned by the classifier algorithms under consideration.

Classifier algorithm	Study site			
	KHO	KKP	KYZ	FER
RF	41.2	18.4	18.4	50.2
SVM_R	39.4	55.5	55.5	23.7
MLP	19.4	26.1	26.1	26.1

As was demonstrated in Table 5-2, in some test sites the OA achieved by the CCS was only marginally better than the single best classifier algorithm, but the increase in terms of user´s and producer´s accuracy from the confusion matrix (Congalton, 1991) was high for some classes (Figure 5-7). The differences between the CCS and the single classifier algorithms tended to the positive in most cases, while some classes showed almost no differences, such as rice in KYZ or fallow in KHO. In KYZ the CCS performed markedly better in classifying the class winter wheat, thus outperforming any single classifier algorithm. This also applies for sorghum/maize in KHO and melons in KKP.

Interestingly the improvements in user´s and producer´s accuracies were more pronounced for classes with fewer training samples (compare Table 3-2 in section 3.2): sorghum/maize and winter-wheat in KHO, melons, alfalfa-1y, and sorghum/maize in KKP, winter wheat in KYZ and FER. An exception of this effect can be found for wheat-other in KKP. This finding is important in practice: a

large test set through field sampling in Middle Asia can be prohibitive to assemble for logistical (access to fields) or financial reasons, which might reduce the generalization ability of the individual classifier models when classifying unseen cases. The positive gain in class accuracy after the classifier combination could compensate for limited availability of training data for such classes, e.g. sorghum/maize or melons. In their investigation, Waske et al. (2010) found that the class-wise increase in accuracy in a SVM-based classifier ensemble depended on the number of training pixels per class, where the highest gain was achieved when fewer pixels were taken for training, which is in line with the findings of this study.
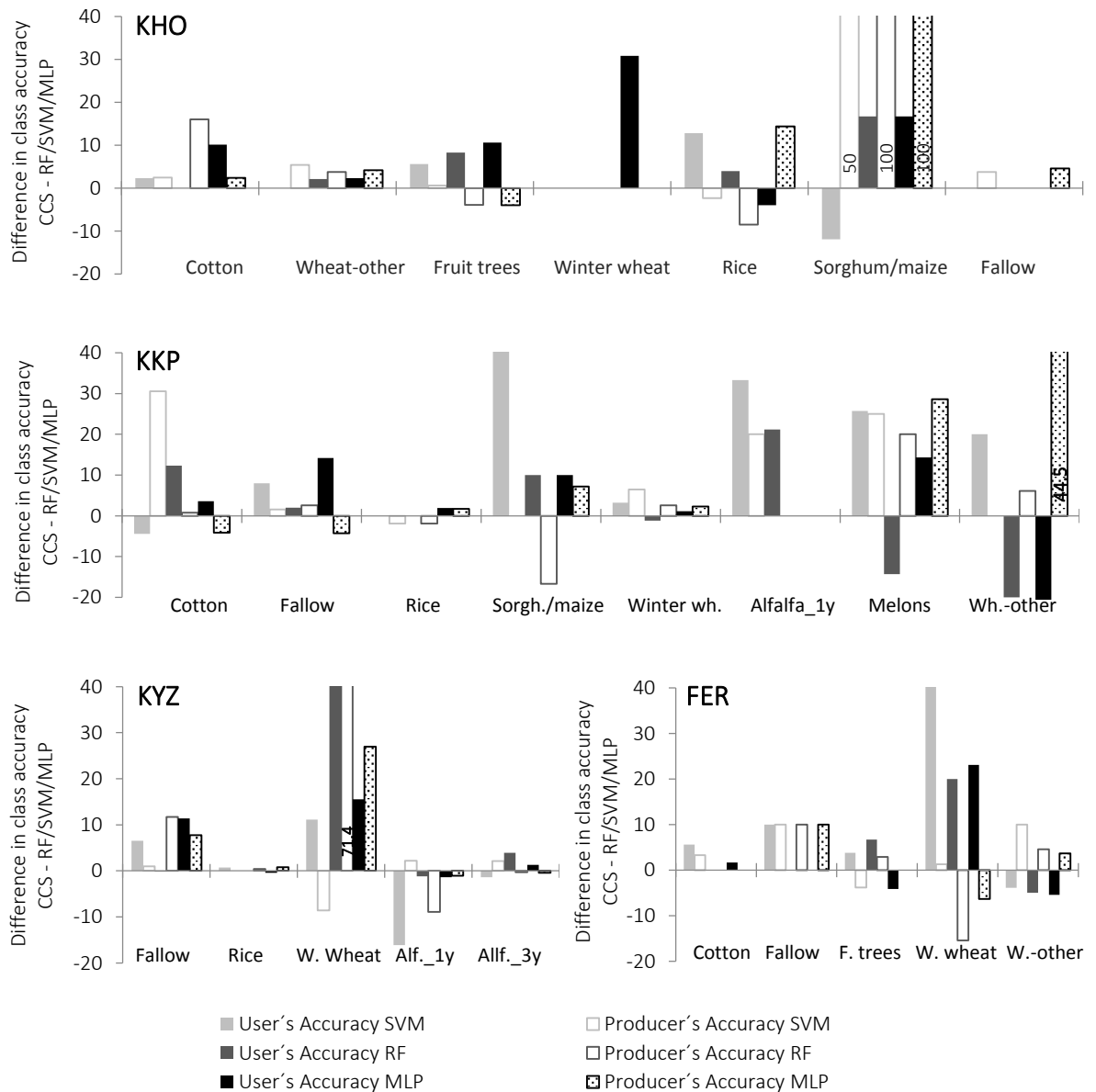
Figure 5-7: Bar plots showing the differences between the user´s and producer´s accuracies [%] achieved by the CSS and the single classifier algorithms tested in the four test sites. Positive values indicate that the CCS outperformed the corresponding classifier algorithm, e.g. in KYZ the user´s accuracy of the CCS for class fallow was 6.5% higher than the user´s accuracy of SVM for that class.

5.3.4. Impact of classifier combination on classification uncertainty

Having compared the classification accuracy of the single classifier algorithms with different classifier combination schemes, the classification uncertainty after classifier combination was assessed and compared with the single classifier algorithms. Emphasis here is put on two aspects of classification uncertainty: (i) the ability to correctly predict classification error in the maps (the predictive strength), and (ii) the spatial distribution of classification uncertainty in the crop maps.

Impact on the predictive strength

To analyse the former aspect, the mean uncertainty for each class, quantified by E_j, was calculated on the independent test set. The class-wise accuracy CA_i, which gives the proportion of test fields per class correctly allocated, was then compared with the class-wise uncertainty by calculating the RMSE between these two measures. This approach was employed for ANN by Brown et al. (2009) to test for the relationship between classification uncertainty and classification accuracy. In this regard, low RMSE values would indicate that low classification uncertainty values correlate with high classification accuracy values (class-by-class). In doing so the RMSE can be used to assess which classifier algorithm better predicts classification error and if the CCS could even increase the predictive strength, compared to the single classifier algorithms. Table 5-4 reveals that the CCS had the smallest RMSE values. This indicates that the uncertainty information from the CCS was a more effective predictor of the per-field uncertainty for all classes, as compared with the single classifier algorithms. RMSE values ranged from 0.133–0.240, with the lowest average RMSE value over all sites achieved by the CCS (0.175).

Table 5-4: RMSE of class-wise uncertainty (E_j) and class-wise accuracy (CA_i). "Mean" is the average of the RMSE values over all sites achieved by one method.

Method	Study site				
	KHO	KKP	KYZ	FER	Mean
RF	0.270	0.244	0.342	0.157	0.253
SVM_R	0.309	0.296	0.393	0.196	0.299
MLP	0.301	0.270	0.137	0.106	0.203
CCS	0.207	0.240	0.136	0.133	0.175

Figure 5-8 shows the relative change of the proportions of correctly classified test cases lying within intervals that were defined on the magnitude of E_j. It is evident that correctly classified cases displayed a positive shift toward lower uncertainty after the CCS was applied. This confirms the higher prediction strength of the CCS over all test sites. For instance, in KYZ the proportion of correctly classified test cases in the interval [0.0, 0.2] increased up to 20 %, as compared with the MLP. Concerning the correct predictions more than 90 % of the fields were lying within the interval [0.0, 0.2] after the classifier combination, compared to 85 % before the classifier combination (for the RF). The highest positive gain in prediction strength was observed in KYZ and FER.
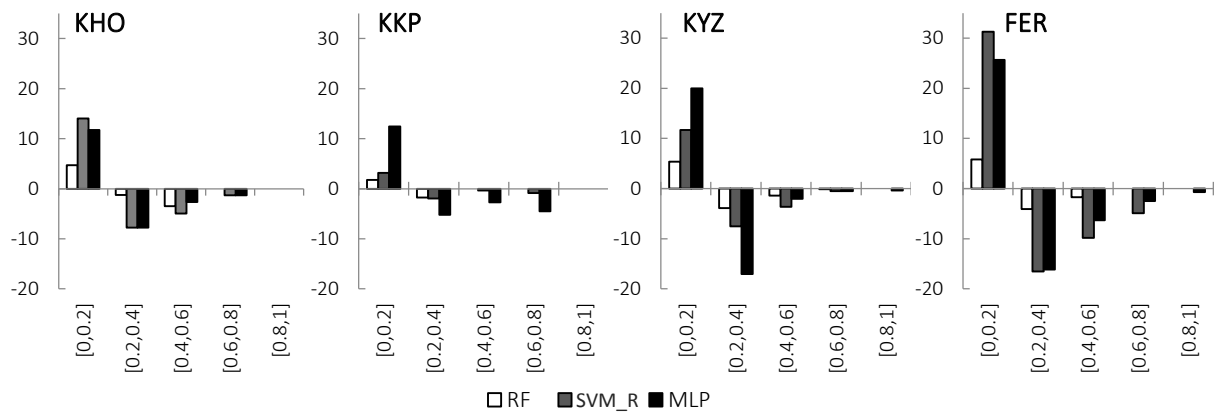
Figure 5-8: Bar plots showing the change of the proportions [%] of correctly classified test cases lying in intervals of E_j after the application of the CCS.

Impact on the spatial variability of uncertainty

As reminded by Bloch (1996), citing from Bloch and Maître (1994), one advantage of combining information from different sources can be reducing uncertainty in the final decision. In the context of classifier combination this can be resumed to reducing spatial classification uncertainty. To test for a possible reduction of uncertainty, and to assess where in the final maps uncertainty was reduced, E_j for each classified field was plotted in a map (Figure 5-9). In this figure, the corresponding uncertainty maps of the classifier algorithms and the CCS in KKP and KYZ, respectively is illustrated. It can be seen that the application of the CCS levelled out the contrasted spatial differences in classification uncertainty, which resulted from the single classifier algorithms in KKP, and that the spatial uncertainty trend described in section 4.3.7 became less obvious after the classifier combination (right most image in Figure 5-9 for the KKP landscape). In KYZ however, the blocky pattern of uncertainty, which traces the aggregation of rice and fallow fields, still existed after application of the CCS.
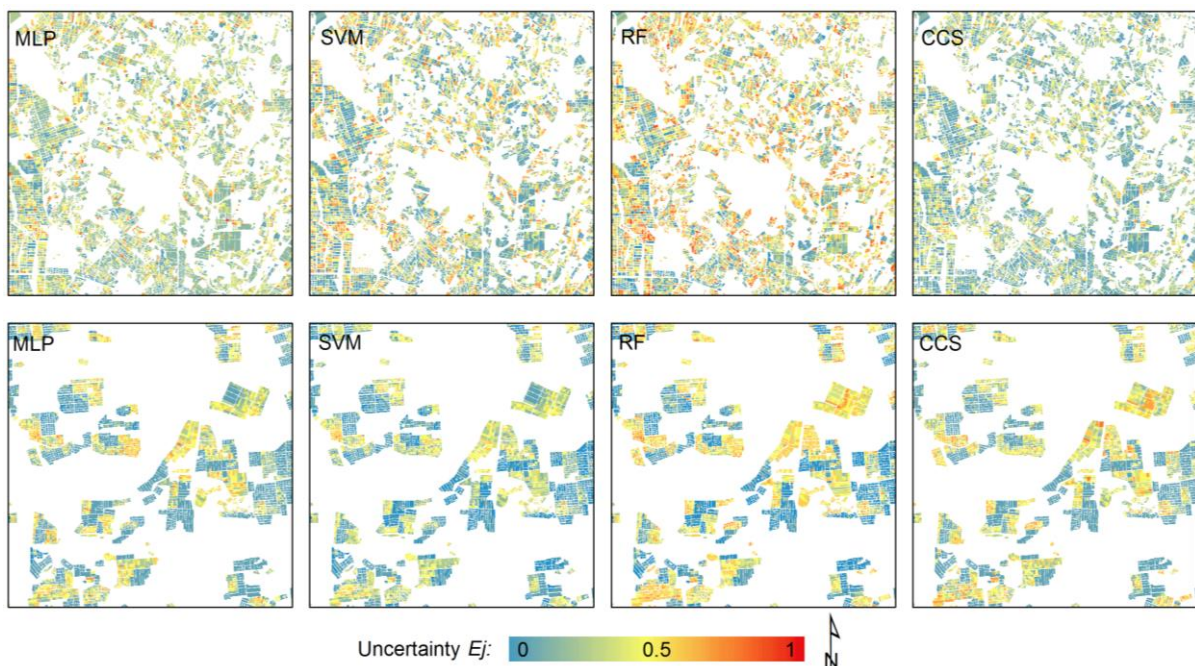
Figure 5-9: Spatial distribution of uncertainty output of the individual classifier algorithms and the CCS in KKP (top row) and FER (bottom row). Uncertainty is quantified by E_j, which was used in the best performing CCS.

This reduction of uncertainty was further quantified by plotting experimental SV of E_j for each study site. In section 4.3.7 this pattern was already detailed, and its causes and relationship with environmental conditions was discussed. Here the focus rather lies on the change of the pattern itself as a result of the classifier combination. To this end, the SV sills ($\hat{\gamma}_r$) were calculated for the uncertainty maps produced by each classifier algorithm and the CCS, respectively (Figure 5-10). Here $\hat{\gamma}_r$ describes the overall variance of classification uncertainty, quantified by E_j. In all sites the classifier combination resulted in a decrease of the $\hat{\gamma}_r$ values. This observed decrease in spatial variability after the classifier combination, quantified by $\hat{\gamma}_r$, is an indication of the decrease in spatial heterogeneity of the observed parameter (Garrigues et al., 2006a). The classification uncertainty resulting from the classifier combination was more homogeneously distributed. In this regard the decreasing value of $\hat{\gamma}_r$ characterizes the loss of spatial variability or more precisely the loss of variability in classification uncertainty, after the CCS was applied in the test sites.
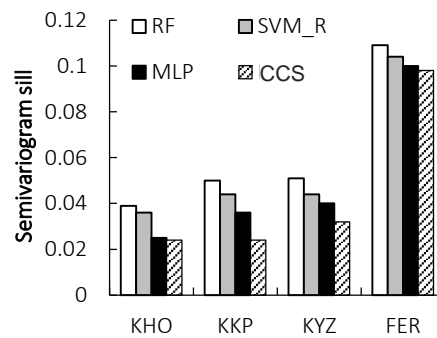
Figure 5-10: Semivariogram sill $\hat{\gamma}_r$ from exponential semivariograms of classification uncertainty maps (quantified by E_j) from the single classifier algorithms and the CCS.

5.3.5. Stability and temporal transferability

Transferability is one important aspect with regard to operational applications, e.g. the results delivered by an approach to classify crops should be comparable over different landscapes and years, e.g. produce maps with comparatively high classification accuracies. Another issue that is important for operational application is stability. This means that the classification approach should be resistant to the variability in the training and test sets, which might be critical in object-based approaches when only few reference fields are available for some classes. Hence, the experiments were repeated 20 times, and OA statistics were averaged over the 20 model runs. This resulted in 20 different randomly split training and test sets, and allowed to examine the stability of the tested classifier algorithms in response to different training data sets. Further, the experiments were repeated on a second dataset in the FER landscape for 2012. Each classification was performed using the corresponding classifier algorithms and combination rules described in section 5.2.2.

Noteworthy the high performance of the CCS could be reproduced over all sites (compare with Table 5-2), and the range of OA among all test sites was lowest when using the CCS (3.6 %), which indicates its better spatial and temporal transferability (over the test sites and years, respectively) compared to the single classifier algorithms. Among the single classifier algorithms the RF had the lowest resulting range of OA over all test sites (4.5 %). The results in Table 5-5 reveal that the CCS still outperformed all other classifier algorithms. Critically the order relation of the accuracies tended to be stable, e.g. the SVM_R still outperformed all other classifier algorithms (expect for KYZ, where the MLP performs slightly better on average), and the proposed CCS performed better than the *MIN* or *MAX* operators,

respectively after 20 trials. The classification accuracy of the FER test site was comparable in both years with slightly better values in 2012. One reason could be that the 2012 data set contains a scene from beginning of July, which might be an important state in crop development when different types of crops differentiate, and hence might support the separation of crop classes.

Table 5-5: Average overall accuracies [%] of the single classifier algorithms and the classifier combination over 20 trials. Difference is given between the CCS and the single best classifier algorithm. The range of OA statistics resulting from one classifier algorithm over all test sites and years is given.

Test site / Classifier algorithm	KHO 2009	KKP 2011	KYZ 2011	FER 2011	FER 2012	Range	Combination rule
RF	89.8	89.9	92.6	92.9	94.3	4.5	
MLP	87.7	91.7	92.8	93.5	94.5	6.8	No classifier combination
SVM_R	89.8	92.3	92.7	94.1	94.7	4.9	
	91.5	92.5	93.0	94.5	95.2	3.7	*MIN* operator
All classifiers	91.2	93.7	93.0	94.5	95.2	4.0	*SUM* operator
	92.1	93.9	93.8	94.9	95.7	3.6	CCS
Difference	2.3	1.6	1.0	0.8	1.0	-	-

5.3.6. Application under limiting conditions

For operational applications, it is also important that a classifier algorithm is applicable under limiting conditions, e.g. when only few training data or acquisition dates are available. Hence, limiting conditions were "simulated" by (i) decreasing the training sample and (ii) selecting small feature sets as input data. To this end, training sets with different numbers of samples were generated by randomly selecting reduced versions of the original training set (per crop class): 100 %, 80 %, 60 %, 40 %, and 20 %. Further, much smaller sets of features were calculated for the acquisition dates, hereafter referred to as FS1–6: FS1) RapidEye bands (5 bands times 8 dates = 40 features), FS2) REDNDVI and GRNDVI (16 features), FS3) NDVI and EVI (16 features), FS4) RapidEye bands plus NDVI and EVI combined (56 features), FS5) RapidEye bands plus REDNDVI and GRNDVI combined (56 features), and FS6) all indices and RapidEye bands combined (72 features). The experiments were repeated 20 times over all study sites, and the accuracy measures were averaged over the 20 model runs.

Figure 5-11 shows the results for these experiments in KKP. It demonstrates the positive effect of increasing the number of training samples on the one hand, and the stability and the high performance of the CCS on the other hand. For instance, the accuracy improved when doubling the number of training samples from 20 % to 40 % of the total available training set size, and when further increasing to 60 %. It slightly improved when using 100 % instead of 80 %. The CCS consistently outperformed the individual classifiers, irrespective of the training set size or input data set chosen. SVM_R tended to be the most accurate of the single classifier algorithms. It is noteworthy that the differences between OA achieved on smaller training set sizes were larger than on larger training set sizes. From the three classifier algorithms tested, RF performed best when using only 20 % of the training data, and consequently could be attributed the least sensitive towards limited training data. Concerning the input data, best results for the CCS were achieved with FS5 or FS6, which indicates that red edge information in addition to bands from RapidEye was best suited to distinguish different types of crops in the test sites. With the results obtained it can be concluded that the CCS consistently achieved better results when applied to limiting conditions (here reduced training data and small feature sets, respectively). Selecting two red edge indices and the RapidEye bands (56 features) gave satisfying results (FS5). Interestingly, using NDVI and EVI (FS4) instead of red edge features resulted in decreasing OA, as compared to FS5.

The results in FER (Figure 5-12) reveal a similar pattern, but the choice of different input data sets had a less obvious impact on the accuracy when taking 80 % or 100 % of the training data, as compared with KKP. In FER, the RF tended to be the best performing algorithm. The MLP algorithm was weaker when applied to FS1-3 with fewer training data (e.g. 40 % or 20 %), and needed more input features to achieve high accuracies when only few training samples were provided. As was observed in KKP, the differences between OA achieved on smaller training set sizes were larger than on larger training set sizes, e.g. when taking only 20 % of the training data, the difference between MLP and CCS became almost 10 % (when using FS1), compared to less than 1 % when using 100 % of the training data.
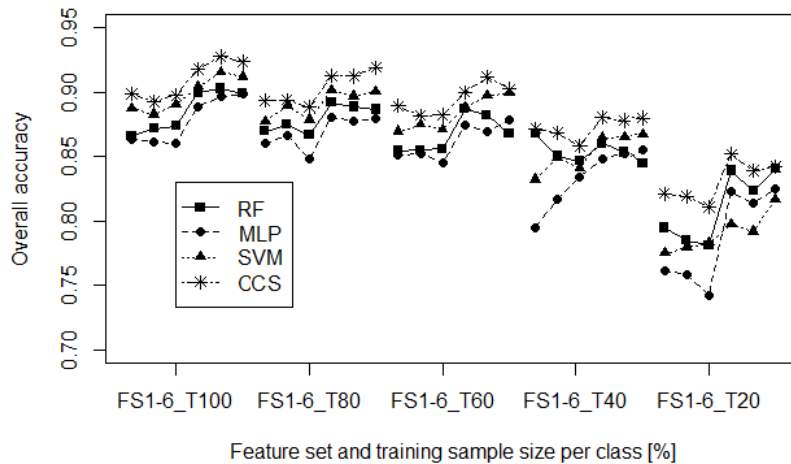
Figure 5-11: Overall accuracy [%] for RF, SVM_R, MLP, and the CCS on the KKP data set, using different numbers of training sample per class and different numbers of features, respectively. FS1-6 refers to the different input data sets, T100-20 to the size of the training data set (as percentage of the total available training set per class).
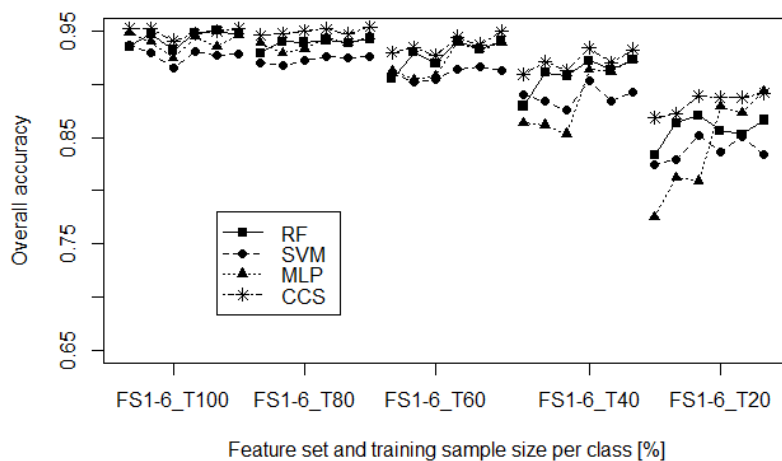
Figure 5-12: Overall accuracy [%] for RF, SVM_R, MLP, and CCS on the FER data set, using different numbers of training sample per class and different numbers of features, respectively. FS1-6 refers to the different input data sets, T100-20 to the size of the training data set (as percentage of the total available training set per class).

Table 5-6 compares the classification performances for 100 % and 20 % of the training samples per class, respectively, applied to FS5, which tended to give very good results. Comparing these results, it can be seen that the classification accuracies (OA and AA) of the CCS tended to be higher in all sites, whilst some classes displayed individuality regarding this, e.g. cotton in KKP for which classification accuracy decreased after the classifier combination when using 100 % of the training data. Some remarkable examples of increasing class-wise accuracies as a result of using the CCS were found for

sorghum/maize in KKP and KHO (approximately 20 % and 41 % increase), wheat-other in KKP (approximately 25 % increase), and winter wheat in KYZ (approximately 50 % increase), where CA_i was much improved as compared with the single best classifier algorithms. Even higher increases were found for wheat-other in KKP when only using 20 % of the training data (approximately 60 % increase). Given the complexity of some of the classes (e.g. wheat-other, alfalfa-1y) the achieved class-wise accuracies (more than 80 % for most classes when using 100 % of the training data), measured with a rigorous independent validation of the CCS are very promising.

Table 5-6: Overall, class-wise, and average class-wise accuracies [%] for the classifier algorithms and the CCS on the test set using 100 %/20 % of the training samples per class, applied to FS5. The crop classes include cotton (C), fallow (FA), rice (R), sorghum/maize (S), alfalfa-1y (A1), alfalfa-3y (A3), melons (M), fruit trees (F), winter wheat (W), and wheat-other (WO).

	Class	RF	SVM_R	MLP	CCS		Class	RF	SVM_R	MLP	CCS
FER	C	98.5/99.2	97.9/96.6	99.9/98.7	98.4/91.5	KHO	C	92.7/87.7	93.1/96.6	94.4/88.7	90.7/84.4
	FA	99.6/66.5	95.8/51.0	92.5/81.5	94.4/94.0		FA	97.1/75.0	94.1/79.7	98.1/100	88.2/76.2
	F	90.6/62.6	91.9/72.7	93.1/75.3	90.0/87.4		R	75.0/75.0	70.4/50.5	71.1/0.0	85.9/100
	W	86.1/39.4	75.4/43.9	74.7/39.9	92.1/88.0		S	17.4/18.7	1.8/0.0	9.0/9.0	61.1/50.0
	WO	94.1/93.4	92.5/93.4	91.8/91.8	94.4/86.8		F	92.0/80.5	93.2/81.9	87.1/76.3	86.2/72.2
	OA	95.0/85.3	92.7/85.1	93.5/87.3	95.0/89.2		W	85.5/39.0	46.5/0.0	19.5/0.0	95.0/75.0
	AA	93.8/72.2	90.7/71.52	90.4/77.4	93.9/89.5		WO	95.4/90.7	95.8/90.0	96.8/83.1	91.6/89.2
							OA	88.3/78.7	85.0/76.1	83.5/69.4	88.7/81.9
							AA	79.3/66.6	70.7/56.9	68.0/51.0	85.5/78.1

	Class	RF	SVM_R	MLP	CCS
KKP	A1	87.3/55.0	94.7/62.7	89.0/36.6	97.4/90.3
	C	87.0/61.7	89.7/73.6	97.0/62.2	81.7/65.2
	FA	91.9/76.1	91.1/79.9	95.0/82.9	90.5/76.4
	M	51.5/36.6	69.7/10.6	2.0/0.0	89.4/44.2
	R	97.4/98.2	98.4/96.7	99.8/99.8	96.0/90.0
	S	14.7/7.0	28.8/5.0	9.0/0.0	48.3/11.1
	W	96.7/90.0	97.9/96.0	97.6/96.2	96.1/88.4
	WO	43.3/16.8	62.7/14.3	19.0/0.0	87.3/77.1
	OA	88.0/76.6	91.3/79.7	87.6/76.7	92.0/81.9
	AA	71.2/55.1	79.1/54.8	63.5/47.2	85.8/67.8

	Class	RF	SVM_R	MLP	CCS
KYZ	A1	78.9/59.2	77.2/32.1	88.3/55.6	79.4/69.6
	A3	76.1/44.6	76.1/63.5	78.7/70.8	91.8/72.0
	FA	76.0/62.7	83.0/78.4	76.2/68.3	72.5/55.1
	R	99.7/99.2	99.7/99.6	99.6/99.3	99.8/99.7
	W	18.6/23.5	37.5/4.1	9.0/0.0	87.2/44.4
	OA	90.7/83.8	92.0/85.4	91.2/86.1	92.5/86.2
	AA	69.8/57.8	74.7/55.5	70.4/58.8	86.1/68.1

5.4. Discussion and conclusions

In this chapter the classification of multi-spectral time series data using different classifier algorithms and classifier combination was performed. The single classifier algorithms tested (RF, SVM_R, and MLP) produced accurate crop maps, but gave complementary results considering different crop classes. In order to capitalize the strength of each classifier algorithm, their results were combined with a CCS that combines the outputs of the single classifier algorithms at the per-case basis, and that employs class-wise measures of accuracy to exclude unreliable classifier algorithm at the per-class basis. Although the application of the CCS outperformed all single classifier algorithms in the test sites, based solely on OA, it appeared to be no significant advantage in selecting a CCS approach, e.g. the results were only marginally better than the single best classifier algorithm. However, there are other compelling reasons for classifier combination: the strength of the proposed CCS strategy was its transferability and robustness when applied over different sites with different classes. For instance, the range of OA achieved over all test sites was smallest when applying the CCS. Another peculiarity of the CCS was that the increase in accuracy became more pronounced when fewer training data were used. Another potential strength of the CCS is its generalizability, as theoretically any kind classifier algorithm could be tested as long as it can create soft outputs (e.g. estimates of class memberships).

The predictive strength of the single classifier algorithms, e.g. their ability to correctly predict the spatial distribution of classification error in maps, was assessed. In general, all classifier algorithms assigned low classification uncertainty values to correctly classified test fields. Yet, it was found that higher classification accuracies did not necessarily coincide with good abilities to correctly predict classification error, e.g. the classifier algorithm that achieved the highest accuracies is not necessarily the best uncertainty predictor. From the frequency distributions of the test fields it could be seen that the RF best assigned low classification uncertainty values to correctly classified test cases. One possible explanation is that for the MLP, in contrast to the RF, the estimates of the class-memberships (soft output) are rather a by-product of the classifier algorithm. In their study, Brown et al. (2009) concluded that the ability of a MLP to correctly predict classification uncertainty decreases as the complexity of the MLP increases (and the activation level outputs tend toward "1" or "0"), but this relationship might not be directly linked with the classification accuracy. Figure 5-13 reflects this behaviour, e.g. when the discrepancy between OA and the classification uncertainty (here measured with the alpha quadratic score) increases with increasing network complexity (number of iterations).
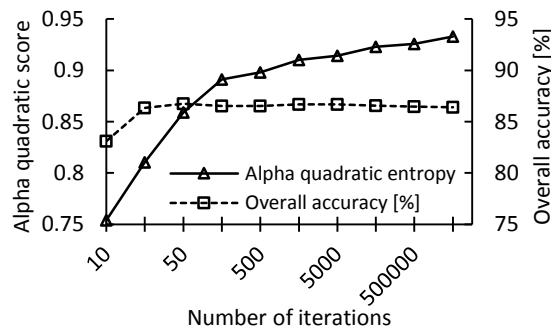
Figure 5-13: Influence of network complexity (number of iterations) in a MLP on classification accuracy (here: OA). Alpha quadratic entropy gives the median classification uncertainty of the test fields.

The proposed CCS is capable of handling such situations by excluding classifier algorithms from competing for crop classes during the classifier combination process when they give too optimistic membership estimations (soft outputs), but at the same time having only low class-wise accuracies. Further, the predictive strength of the CCS, measured as the relationship between class-wise uncertainties and class-wise accuracies, was higher than for the single classifier algorithms. Compared to the single classifier algorithms, the CCS allocated a larger proportion of correctly classified test cases to low classification uncertainty values.

A potential improvement to the proposed classifier combination approach could be to rewrite the rules with fuzzy data fusion rules, which might allow for a finer definition and weighting of the classifier algorithm reliability and finally enhance classification accuracy. From a pragmatically point, combining classifiers exhibits some interesting features: the observed increase in class-wise accuracies was the most pronounced for classes with only few training data (sorghum/maize and wheat-other in KKP and KHO). This is of interest when access to field data is prohibitive for logistical (physical access to fields) or financial reasons. The CCS seems well suited for applications under limiting conditions, e.g. when only few training samples are available or when applied to smaller input data sets, which makes it interesting for operational monitoring, because it is rather simple in implementation and, due to its nature as a non-trainable operator, fast in computation (e.g. the application of the combination rule to the outputs from the single classifier algorithm only needed few seconds). Yet, conclusions drawn from the results presented here refer to the specific landscapes with the specific constellation of crop classes, and therefore the CCS should be tested over a wider range of landscapes with different crop types, respectively.

Chapter 6

DEFINING THE SPATIAL RESOLUTION REQUIREMENTS FOR CROP IDENTIFICATION VIA IMAGE CLASSIFICATION USING REMOTE SENSING DATA[20]

Abstract

Mapping the spatial distribution of crops is one fundamental input for agricultural production monitoring using remote sensing. A relatively high revisit frequency is needed to properly characterise the temporal evolution of the land surface characteristics, which further helps to discriminate crops both from other land covers, and among themselves, based on differences in the seasonal behaviour and on the agro-management practices applied in a region. However, this multi-temporality (e.g. time series of satellite images) plus the desire to have a large geographical coverage (swath) for agricultural monitoring applications at the regional to global scale comes at the expense of coarser observation supports, and can lead to erroneous class allocations caused by mixed pixels. A framework is proposed to analyse spatial resolution requirements for accurate crop identification via image classification by simulating how agricultural landscapes, and more specifically the fields covered by a crop of interest, are seen by instruments with increasingly coarser resolving power. The concept of crop specific pixel purity, defined as the degree of homogeneity of the signal encoded in a pixel with respect to the target crop type, is used to analyse how mixed the pixels can be (as they become coarser), without undermining their capacity to describe the desired surface properties. In this case, this framework has been steered towards answering the question: "What is the spatial resolution requirement for crop identification via image classification and how do these requirements change over different landscapes?" The framework is applied over four contrasting agro-ecological landscapes in Middle Asia. Inputs to the experiment were eight multi-temporal images from the RapidEye sensor, the simulated pixel sizes range from 6.5 m to 747.5 m. Constraining parameters for crop identification were defined by setting thresholds for classification accuracy and uncertainty. The experimental results revealed that different types of crops displayed marked individuality regarding the pixel size requirements, depending on the spatial structures and cropping pattern in the test sites. The coarsest acceptable pixel sizes and corresponding purities for the same type of crop were found to vary from site to site, and some crops could only be identified within a narrow range of pixel sizes. The main causative factors that determined the pixel size requirements, namely field sizes, cover fraction of fields in the landscape, and environmental factors like water availability were discussed. The practical implications of the framework presented in this study include guidance (i) for selecting appropriate imagery (e.g. suitable pixel sizes) for applications focussing on crop mapping, (ii) regarding the selection of thresholds for minimum required pixel purity for the effective identification of crops in coarser pixels, or (iii) for EO data requirements for early crop inventory.

6.1.   Background

In chapters 4 and 5 methods were presented for object-based crop classification. In this chapter the interest is steered towards the exploration of the limitations for crop identification in coarser EO data, which is one prerequisite for crop specific agricultural monitoring at regional or even global scales (see Figure 1-1 in chapter 1). Hence, this chapter focusses on pixel-based approaches and directly addresses the third research question that was formulated in introductory chapter. A framework is proposed and tested to define region specific EO requirements in terms of pixel size and purity, respectively for crop identification via image classification.

The traditional way to retrieve information on the spatial distribution of crops in a landscape, termed crop mapping, using remote sensing is through image classification with one of the widely known classifier concepts and algorithms that are currently available (Tso and Mather, 2009). Crop mapping is a term that can have different meanings depending on what specific application is considered. Some applications require delineating accurately where all crops (and perhaps all other land uses) are located over the entire area of interest. This is necessary for spectral un-mixing of coarse spatial resolution pixels (Verbeiren et al., 2008), to produce accurate crop specific masks (Conrad et al., 2013; Peña-Barragán et al., 2011), or to make crop acreage estimations via counting the classified pixels´ surface (Gallego, 2004). Generally, it is only feasible to make such an exhaustive mapping over a limited geographical coverage in order to keep a decent standard of classification accuracy. For other applications, such as crop monitoring at regional to global scales, it is probably not even possible to make such a spatially exhaustive ("wall-to-wall") classification of all land uses with the required accuracy and in a timely manner. Past studies have shown how an adequate cropland mask can considerably improve classification accuracy (Fritz et al., 2008) or yield estimations (Genovese et al., 2001; Kastens et al., 2005). Further, research has shown how focusing on a population (e.g. a sample of pixels from the total pixel population) of crop specific time series by choosing only those falling adequately into the fields allows the correct characterization of the crop behaviour even in fragmented landscapes (Duveiller et al., 2011, 2012). Since this study targets crop-monitoring applications, the interest is geared towards this notion of crop identification rather than exhaustive crop mapping.

What type of remote sensing data should be used as classification input for proper crop identification? In general spatial resolution[21] should be high enough to resolve the spatial frequencies of fragmented agricultural landscapes. Multi-spectral imagery is suitable to better identify the spectral signature of specific crops. Furthermore it has to be noted that the application of radar data bears great potential (Loosvelt et al., 2012b; Waske and Braun, 2009), but is not in the focus of this thesis. A relatively high revisit frequency is needed to properly characterize the temporal evolution of the land surface characteristics, in particular crops. This helps to discriminate crops both from other land covers, and among themselves, based on differences in the seasonal behaviour and on the agro-management practices applied in a region. For applications at regional to global scales, there is an extra requirement of having a large swath to have a wide geographic coverage. Up to now, a good candidate to satisfy these requirements has been AWiFS, which has been used to generate the Cropland Data Layer products in the U.S. (Johnson and Mueller, 2010). Undoubtedly, the new and upcoming satellite EO systems, such as RapidEye, Landsat-8, and Sentinel-2 provide new opportunities for agricultural applications. Although they will not entirely satisfy those for crop growth monitoring, which requires higher temporal

---

[21]: Here the terms "scale" and "spatial resolution" of remote sensing data is synonymously used to "pixel size", which is equal to the nominal pixel size of the image as defined by the size of the sampling step of the sensor at nadir, also called "nominal" ground sampling distance (GSD) (Schowengerdt, 2007). Other definitions include the ability to separate point targets (point spread function), the ability to measure periodicity of repetitive targets (modulation transfer function) and the ability to measure spectral properties of small objects (effective resolution element), see Aplin (2006), citing Cracknell (1998), Fisher (1997), Forshaw et al. (1983), and Townshend (1981).

resolution (Duveiller et al., 2013), they should be well adapted for the purpose of operational crop identification over a wide scale. However, coarser spatial resolution data such as MODIS or MERIS should not be discarded. Not only do they provide added information with the higher repetitivity, but also they will retain much importance as a source of long-term historical record, which the new systems won't achieve for decades to come. Deriving archives based on which crop specific time series have been identified for the past years can be very valuable for agricultural monitoring (Brown et al., 2013; Duveiller et al., 2012), as it can be used to understand the past behaviour of agricultural systems (Fritsch, 2013) and thereby infer changes in productivity or resilience.

The necessity for a continued exploitation of coarser spatial resolution data, plus the growing interest in exploiting multi-scale data synergistically, drive the reasoning for the subject of this study: *exploring the spatial resolution requirements for the specific task of crop identification.* Past and recent authors have pointed out how spatial resolution is a complex concept that depends on the instrument's spatial response (Cracknell, 1998; Duveiller and Defourny, 2010; Fisher, 1997; Forshaw et al., 1983; Garrigues et al., 2006a; Kaiser and Schneider, 2008; Schowengerdt, 2007). Huang et al. (2002b) demonstrated how the spatial response of MODIS 250 m bands can impact classification accuracy and sub-pixel land cover fraction estimation. Defining suitable pixel sizes for remote sensing applications like image classification has a long tradition of research (Atkinson and Aplin, 2004; Atkinson and Curran, 1997; Duveiller and Defourny, 2010; Marceau et al., 1994a, 1994b; McCloy and Bøcher, 2007; Ozdogan and Woodcock, 2006; Pax-Lenney and Woodcock, 1997; Woodcock and Strahler, 1987). Moreover, attempts to define the spatial resolution requirements can be found for quantitative remote sensing (McCabe and Wood, 2006; Nijland et al., 2009; Sepulcre-Cantó et al., 2010; Tarnavsky et al., 2008), but only a few explicitly address this issue in the context of crop identification via image classification or crop area estimation (McCloy and Bøcher, 2007; Ozdogan and Woodcock, 2006; Turker and Ozdarici, 2011).

Although smaller pixels are preferred to assure a good delineation spatial structures like agricultural fields in fragmented agricultural landscapes and to reduce the amount of mixed pixels, increasing the spatial resolution may result in oversampling, resulting in increased within-feature or class variability, e.g. because of redundancy of data within fields. Such variation can lead to errors in feature identification (Atkinson and Aplin, 2004; Cushnie, 1987; Hsieh et al., 2001), and better classification accuracies may be attained using coarser pixel sizes (McCloy and Bøcher, 2007; Woodcock and Strahler, 1987). On the other side, selecting too coarse pixels can deteriorate the quality of the classification due to mixed pixels when the heterogeneity of the land cover in one pixel increases (Hsieh et al., 2001; Smith et al., 2003). Furthermore, a bias may be introduced when estimating crop related parameters from such coarser and more heterogeneous pixels, e.g. when estimating the characteristics of a region like means, medians, or totals of crop related parameters (Duveiller and Defourny, 2010; Duveiller et al., 2011; Haack and Rafter, 2010; Ozdogan and Woodcock, 2006). Based on this reasoning, it has been questioned if selecting one single spatial resolution is appropriate, e.g. in Atkinson and Aplin (2004), because there is convincing evidence that natural processes operate at a range of spatial and temporal scales rather than at a single, fixed scale (Aplin, 2006; Levin, 1992). Different land cover types may have large ranges of object sizes and there may be significant differences in the local spatial variation observed at different wavelengths (Atkinson and Aplin, 2004; Duveiller and Defourny, 2010). Also, it might not be appropriate to select a single spatial resolution for a single remotely sensed image (Atkinson and Aplin, 2004; Ju et al., 2005). High resolution images support several scales at the same time, allowing for object-based image analysis that exploits and segments images at multiple scales simultaneously (Blaschke and Hay, 2001; Blaschke, 2010), yet with coarser pixels such object-based applications are generally not feasible and pixel-based analyses prevail. Furthermore, Ozdogan and Woodcock (2006) and Duveiller and Defourny (2010) illustrate how, for a given application like crop area estimation, the spatial resolution requirement (e.g. in terms of a maximum tolerable pixel size) differs considerably over different landscapes.

Another sometimes neglected issue when defining observation requirements for crop mapping is that agricultural landscapes are characterized by heterogeneous surface characteristics (Ozdogan and Woodcock, 2006) that can change considerably along the season (Garrigues et al., 2008a). The optimal timing of observation to characterize the crop growth stages or to perform accurate crop classification can vary over different landscapes (see section 4.3.4), which complicates the definition of what is the "optimal" observation scale. Only few studies analysed possible impacts of acquisition timing on the definition of pixel size requirements (Duveiller and Defourny, 2010). Yet, although reconciling both scales (spatial and temporal) seems inevitable in this context, most of the remote sensing investigations solely consider spatial or temporal scale exclusively, or are without explicit focus on crop identification (Bradley and Millington, 2006; Stellmes et al., 2010).

To analyse the spatial resolution requirements for crop identification, this study builds upon and extends a conceptual framework established in a previous work of Duveiller and Defourny (2010). That framework allows defining quantitatively the spatial resolution requirements based on simulating how agricultural landscapes, and more specifically the fields covered by a crop of interest, are seen by instruments with increasingly coarser resolving power. The concept of crop specific pixel purity, defined as the degree of homogeneity of the signal encoded in a pixel with respect to the target crop type, is used to analyse how mixed the pixels can be (as they become coarser) without undermining their capacity to describe the desired surface properties. Duveiller and Defourny (2010) used this approach to identify the maximum tolerable pixel size for crop growth monitoring and crop area estimation, respectively. In the present study, it is proposed to revisit this framework and steer it towards answering the question: "*What is the spatial resolution requirement for crop identification via image classification, in particular minimum and coarsest acceptable pixel sizes, and how do these requirements change over different landscapes?*" A methodology is presented for defining landscape specific pixel size requirements for a specific application, based on objective criteria. The present study further differs from the framework in Duveiller and Defourny (2010) by explicitly incorporating the temporal dimension and working with higher resolution images (RapidEye with 6.5 m) that allow for a finer diagnostic in heterogeneous landscapes like in Middle Asia with relatively small field sizes (see Table 2-1).

Several experiments will be performed in order to provide a more comprehensive understanding of how crop identification in medium to coarse satellite image time series depends on pixel purity and pixel size, respectively. The objectives can be formulated as follows:

> Characterizing how crop identification depends on pixel size and purity, respectively. Specifically it will be analysed how the requirements of one specific crop differ in the four sites, and how different crops display individuality regarding this over one particular landscape.

> Comparing the impact of different classifier algorithms, specifically the RF and SVM, on the definition of suitable pixel sizes. The unsupervised K-means algorithm is also tested to evaluate the potential of identifying natural groupings of pixels with similar spectral signatures, but without training data.

> Assessing how different input variables (e.g. different kind of spectral features or vegetation indices) affect the identification of crop specific signals in the time series, and consequently the definition of acceptable pixel sizes. In particular, the red edge spectrum will be focussed on.

> Exploiting how the suitability of pixel sizes for crop identification change along the season, by analysing the effects of successively increasing the observation length (e.g. the number of satellite images in the annual time series) within the growing-period of a given year.

> An additional experiment will assess if results are stable by repeating the approach on the same site, but in different years.

6.2. Test site characteristics and crop masks

This study is based on the four test sites described in chapter 2. Recalling the site description in section 2.2, the KHO and FER landscapes are similar in terms of field sizes, but a different number of crops is cultivated. Crop mapping using coarser resolution imagery can be expected to be better suited in the FER landscape because many pixels can be suspected to fall within the large fields that are covered by a small number of crop types, whilst in KHO the crop pattern is more heterogeneous. The KKP landscape is complex with small field sizes. Crop mapping using coarse resolution imagery might be hampered because of the relatively small field sizes, the fragmented crop pattern and the high proportion of non-agricultural surfaces. Further the signal response is expected to be comparably low because of the presence of fallow and abandoned fields with complex non-agricultural vegetation cover and general water scarcity, which is supposed to flatten the NDVI profiles of crops (see section 4.3.7). This could negatively impact the use of coarser resolution imagery for crop mapping because of the higher proportion of mixed pixels when using sensors with coarser resolving power. KYZ and FER are similar in terms of crop diversity but they have different spatial patterns and field sizes. For these sites it can be expected that the pixel size for crop mapping can be coarse, in FER because of relatively large field sizes and in KYZ because of the aggregated pattern of fields with identical crops. For the experiments eight top-of-canopy (TOC) reflectance images from the RapidEye mission with a GSD of 6.5 m were available in each site (see section 3.1), nine images were available for the KKP landscape. They are well distributed along the season, approximately between DOY (day-of-year) 80 and DOY 280, in order to provide the necessary phenological information for crop discrimination. RapidEye images were available in 2011 (KKP, KYZ, FER), 2009 and 2010 (KHO), and 2012 (FER), thus in KHO and FER the experiments could be repeated in two consecutive years (Figure 6-1).
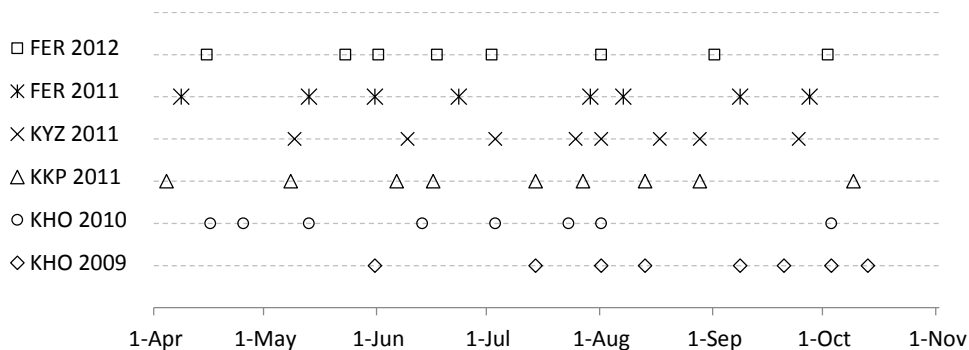
Figure 6-1: Acquisition dates of the data sets from the RapidEye instrument utilized in this study. In KKP nine acquisitions are available, in the other sites eight acquisitions.

Crop specific masks are necessary to identify the target objects (e.g. agricultural fields cultivated with a certain crop) in the landscape, and later for calculating the purity of coarser pixels with regard to specific crops. These masks take on the value "1" if a crop is present, and "0" otherwise (binary masks). Crop masks were available from previous studies for the years 2009, 2011 and 2012 (chapter 5), and for 2010 in KHO (Conrad et al., 2011b). The classification accuracies of the crop maps derived in these studies, from which the masks were generated, were more than reasonable (higher than 93 % in most cases) and assumed to have negligible error for the purpose of this study. The resulting cover proportions of crops in the landscapes, and median field sizes cultivated by certain crops, are summarized in Table 6-1.

Table 6-1: Area cover fractions (C_f) of agricultural crops in the study sites, calculated as the share of area of crops in the total cultivated area per site. MFS is the mean field size of the corresponding crops in hectares. The crop classes include cotton (C), fallow (FA), rice (R), sorghum-maize (S), alfalfa-1y (A1), alfalfa-3y (A3), melons (M), fruit trees (F), winter wheat (W), and wheat-other (WO). Below the overall classification accuracy [%] of the crop maps is given that resulted from the classification of the RapidEye time series with the CCS described in chapter 5.

KHO 2009			KHO 2010		
Crop	C_f	MFS	Crop	C_f	MFS
C	0.34	5.05±1.83	C	0.35	5.03±1.93
WO	0.27	4.49±1.95	WO	0.30	4.60±2.17
T	0.23	4.22±2.11	T	0.19	4.45±2.15
W	0.01	5.14±1.78	W	0.01	4.14±2.10
R	0.07	4.15±2.25	R	0.13	4.10±1.75
S	0.01	5.88±2.86	S	0.01	3.60±2.18
FA	0.07	2.70±1.82	FA	0.01	3.20±1.90
OA: 94.4			OA: 87.4		

KKP 2011			KYZ 2011		
Crop	C_f	MFS	Crop	C_f	MFS
C	0.12	2.50±1.68	FA	0.28	2.56±1.62
FA	0.48	2.15±1.86	R	0.52	2.36±1.36
R	0.07	2.11±1.68	A1	0.10	2.36±1.80
S	0.02	1.60±1.53	A3	0.10	2.67±1.48
W	0.22	2.39±1.83			
A1	0.07	2.29±1.67			
M	0.03	1.40±1.16			
WO	0.01	1.62±1.45			
OA: 94.6			OA: 93.9		

FER 2011			FER 2012		
Crop	C_f	MFS	Crop	C_f	MFS
WO	0.32	7.83±2.81	WO	0.27	7.54±2.75
FA	0.04	4.68±2.45	FA	0.01	4.27±2.53
W	0.05	6.49±2.84	W	0.09	8.45±2.66
C	0.42	7.72±2.71	C	0.44	7.40±2.70
T	0.19	4.70±2.42	T	0.19	4.70±2.78
OA: 96.1			OA: 95.6		

6.3. Methodology

The necessary processing steps to simulate coarser imagery and to define suitable pixel sizes for crop identification are henceforth described. The general flowchart in Figure 6-2 may guide the reader throughout the following descriptions.
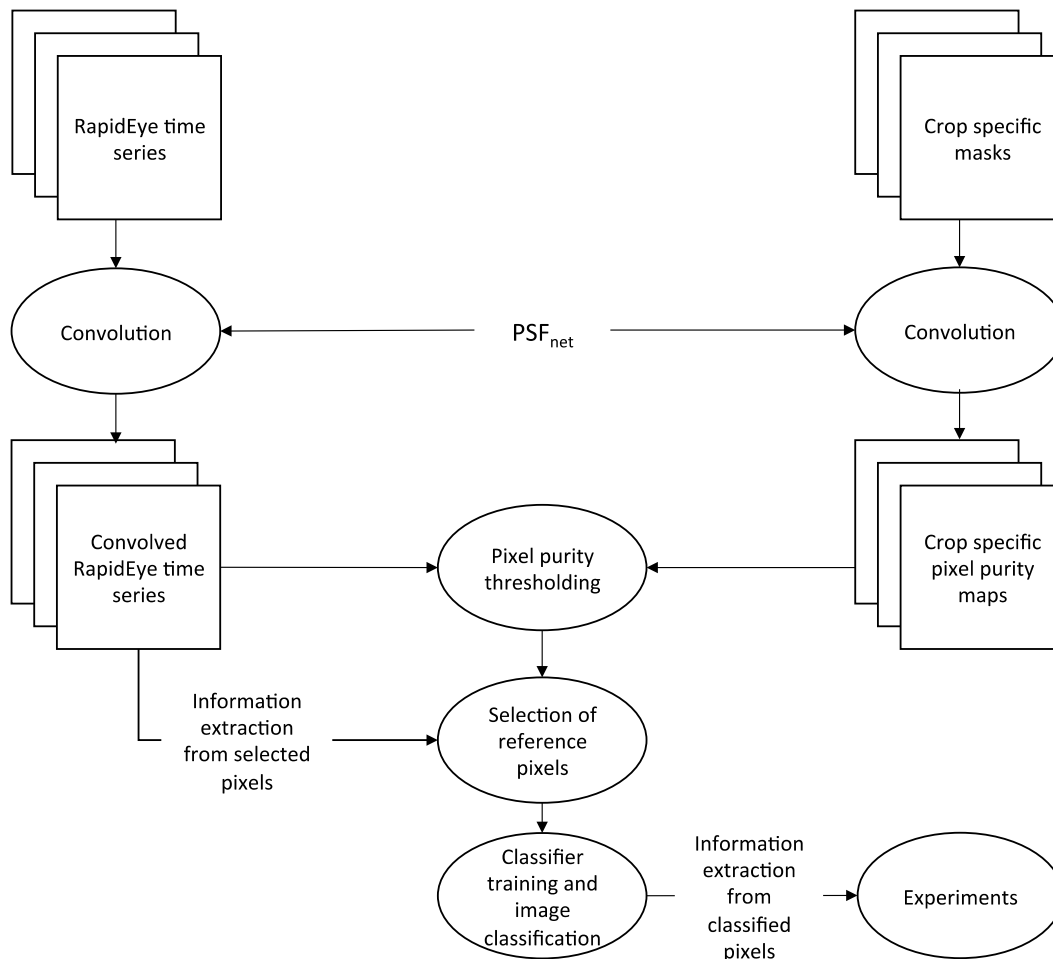
Figure 6-2: Flowchart to produce time series and pixel purity maps for different pixel sizes, and to identify pixel size requirements with different experimental setups. PSF$_{net}$ refers to the net point spread function that is used to simulate coarser pixel sizes, more details are given in section 6.3.1. "Experiments" refer to the sub-objectives in the introduction that will be reflected by the experimental setups.

6.3.1. Selecting target pixel population by aggregation and thresholding

Background

To simulate coarser pixel sizes, a spatial response model (Schowengerdt, 2007) is convolved over the original RapidEye images. This spatial response model of an imaging instrument with coarser GSD consists of a point-spread function (PSF), characterizing both optical and detector components of a generic sensor, and which can be scaled to simulate different pixel sizes. For the sake of clarity this section details the basic concepts of sensor models and sensor simulation. The interested reader is

referred to Schowengerdt (2007) for more details on this issue. The continuous radiometric signal detected by the instrument is sampled regularly. The on-ground distance that separates two such samples is called ground-sampling distance (GSD). Each sample consists of a measurement made by a detector of incoming radiance within its angular instantaneous field of view (IFOV). The detector integrates the received energy over a distinct time interval, converts it to an electric signal and assigns an integer value (the digital number, DN) to an individual square grid cell (a pixel), which constitutes the remotely sensed image. The ground sampling interval (GSI) defines the inter-pixel spacing on the ground (Figure 6-3).
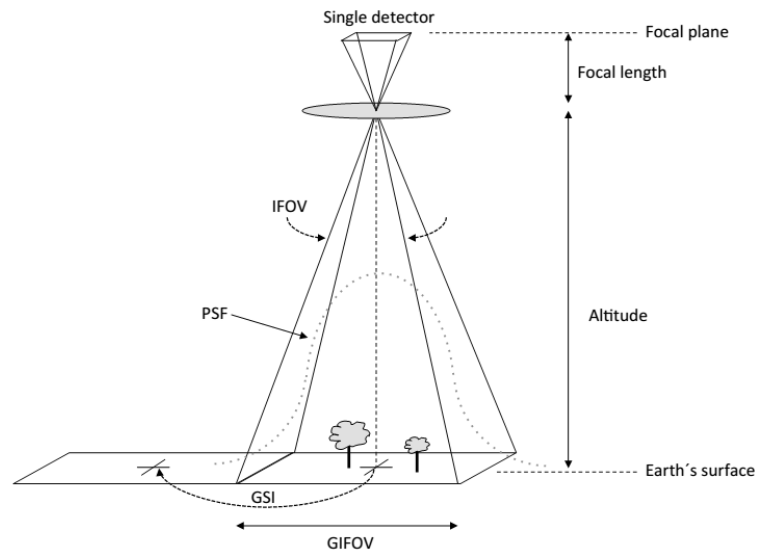
Figure 6-3: Description of a single detector element in the focal plane of an optical sensor. Here a non-overlapping GIFOV is shown, assumed in a nadir view. Adapted from Schowengerdt (2007).

Yet, a considerable fraction of the energy encoded in a pixel of a satellite image originates from its surrounding areas (Townshend, 1981), as a result of atmospheric effects and the instruments properties (Schowengerdt, 2007). This means that the geometric projection of a single detector onto the Earth's surface, which is referred to as the ground-projected IFOV or GIFOV, can be larger than the GSD in the imaging system.

This can be characterized by the sensor´s PSF (Schowengerdt, 2007), which describes the sensors response to a point signal, and how it is spread over the image plane as it is recorded by the imaging instrument and results in a blur of small objects in the image. The PSF functions like a low-pass filter to the signal retrieved by the detector. The effect is that the actual area which is integrated by a detector is larger than the squared GIFOV or GSD, and consequently the so-called actual (or effective) GIFOV is larger than the standard specified GIFOV (Schowengerdt, 2007). For instance to model a realistic spatial response of MODIS its effective GIFOV can be more than 500 m (Duveiller et al., 2011) rather than the oft-quoted 250 m (geometric GIFOV).

Several components constitute to the PSF of a sensor, which can be modelled as a spatial (two dimensional) response function that weights and integrates the signal received at a detector, and that can be used for image convolution and simulation of coarser observation supports (Schowengerdt, 2007). In this thesis the nominal (at nadir) GSD is considered, which is equivalent to pixel size (called v). Following Schowengerdt (2007) the spatial response of an imaging instrument with a GSD equal to v is given by different components:

Optical PSF

This component describes the response of the detector to a point energy source. This energy is spread over the focal plane of the sensor (Figure 6-3) and results in a blurring. The optical component can be modelled with a Gaussian function:

$$PSF_{opt}(x,y) = exp\left(-\frac{x^2+y^2}{2*(v*\sigma)^2}\right) \qquad\qquad Eq.\ 6\text{-}1$$

where x and y are the cross-track and in-track coordinates, respectively, with their origin at the centroid of the GIFOV, and σ the standard deviation of the Gaussian curve. Note that the width of the detector in both in-track and cross-track directions is assumed to be equal.

Detector PSF

The detectors in the sensor add a blurring because of their non-zero spatial area. This response can be modelled by:

$$PSF_{det}(x,y) = rect(x/v) * rect(y/v) \qquad\qquad Eq.\ 6\text{-}2$$

where $rect$ is the rectangular function, a uniform square pulse function with amplitude one and width v.

Image motion PSF

During the integration time of the signal for one pixel the detectors moves across the earth´s surface and results in a spatial smear of the signal. For whiskbroom scanners like MODIS or TM this component is modelled by:

$$PSF_{mot}(x,y) = rect(x/s) \qquad\qquad Eq.\ 6\text{-}3$$

where $rect$ is the rectangular function, and s is the spatial smear in the focal plane of the sensing instrument, which is computed by:

$$s = scan\ velocity * integration\ time. \qquad\qquad Eq.\ 6\text{-}4$$

For pushbroom scanners (e.g. SPOT), s becomes $platform\ velocity * integration\ time$.

Electronic PSF

The signal is further degraded by an electronically filtering step that is sometimes used to reduce noise. The electronic PSF (PSF_{elec}) is like a low-pass filter in some whiskbroom scanners and smooth the signal in cross-track direction.

Net PSF

In this thesis the PSF_{elec} and the PSF_{mot} are neglected for the sake of simplicity and because the analysis seeks to simulate generic sensor attributes. The PSF_{net} used in this study is not intended to mimic the exact response of a particular sensor. This does not affect the general conclusions on the impact of the PSF_{net} and the effectiveness of the proposed image convolution. Finally the PSF_{net} is computed by convolving two out of the four presented individual components:

$$PSF_{net} = PSF_{opt} * PSF_{det}.$$

<div align="right">Eq. 6-5</div>

Figure 6-4 shows the spatial integration that is involved when simulating the individual components of the sensor PSF_{net}.
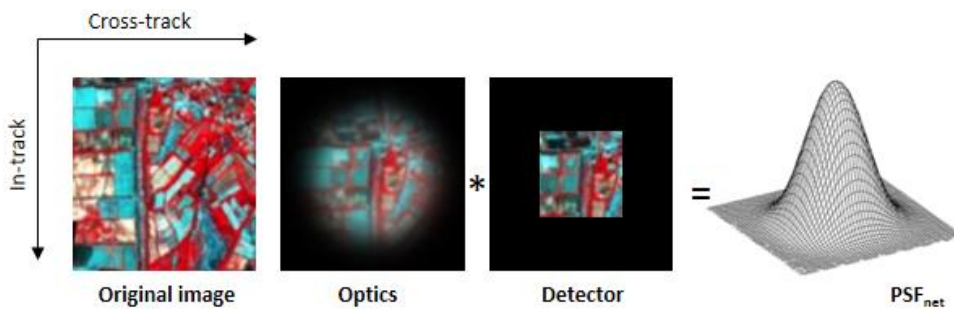
Figure 6-4: Spatial integration involved when simulating the individual components of the sensor PSF: left: original scene, middle: optical PSF component, right: detector PSF. Asterisk between optics and detector indicates convolution.

Application

As mentioned before, the methodology employed here is based on a framework presented in a previous study (Duveiller and Defourny, 2010). It relies on using high spatial resolution images and corresponding crop maps to generate various sets of pixel populations (e.g. subsets of the total image pixel population) over which a classification algorithm can be applied. The pixel populations are characterized by increasingly coarser pixel sizes and with a range of different crop specific purity thresholds. The imaging system simulation in this thesis is based on the PSF_{net} defined above. In order to simulate the spatial response of sensors at varying scales, the PSF_{net} is first discretized to the finest spatial resolution (GSD), here the original RapidEye image at 6.5 m, called v_0. Then the PSF_{net} is scaled to a range of sizes between 13 m and 747.5 m, in increments of 6.5 m, in order to simulate a continuum of coarser images. After the bi-dimensional (in x and y direction) convolution of the spatial response model, PSF_{net}, at each scale over the RapidEye time series, a subsampling is performed to result in simulated images at a given coarser pixel size. This is achieved by selecting every v/v_0 pixel from the convolved images (v being the pixel size of the coarser image). These selected pixels are assumed to be the centroids of the GIFOV of the coarser pixels, and this pixel value is assigned to the corresponding coarser pixel grid cell. This results in simulated images that mimic the spatial response of sensors with coarser pixel sizes (Figure 6-5).

The convolution of the same spatial response model over the high-resolution crop masks results in crop specific purity maps at each scale, which map the pixel purity with respect to the spatial structures represented in the high resolution crop masks (Duveiller and Defourny, 2010). This allows controlling the degree at which the footprints of coarser pixels coincide with the target structures (e.g. fields belonging to certain crops). At each spatial resolution, pixel populations can be selected based on

thresholds on the pixel purity, here denoted π[22]. A threshold can be chosen to separate the aggregated binary crop masks into two sets: target pixels and non-target pixels. The threshold can vary from "0", where all pixels in the images are selected as target, to "1", where only completely pure pixels are selected. The result is the sets of selected target pixels, or pixel populations, defined by v and by the minimum acceptable purity threshold that defines them (π).
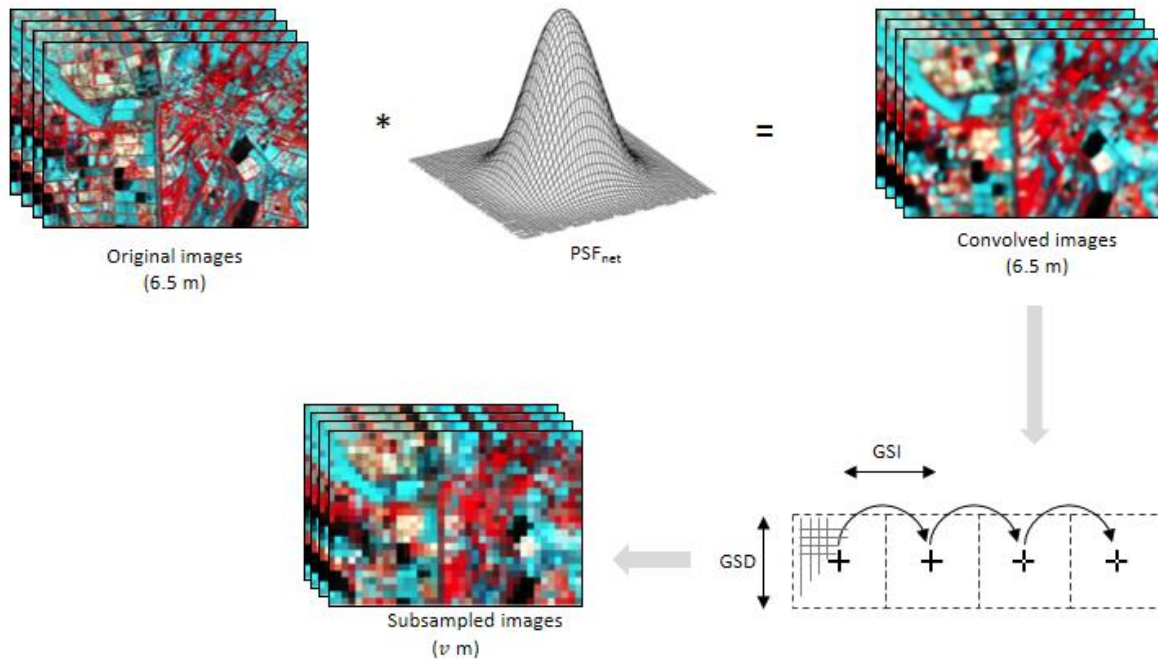
Figure 6-5: Use of *PSFnet* to convolve the RapidEye time series to simulate satellites with coarser resolving power. GSD is the ground-sampling distance, and GSI the ground-sampling interval. The upper part of the figure shows the convolution, the lower part the subsampling that results in an image with a coarser pixel size.

6.3.2. Image classification

The second step consists in applying classification procedures to each selected pixel population. Three classification algorithms were tested: two supervised classifier algorithms (SVM and RF) and one unsupervised algorithm (K-means), which are briefly described in the following. Training and classification were performed in the R programming environment (R Development Core Team, 2012). In each site the classifiers were trained and applied to classify the image stacks consisting of eight acquisition dates at each spatial scale, and all classes present in the corresponding study sites were included in the legend, expect for winter wheat in KYZ because only few fields were present (see section 4.3.6). For the supervised classifier algorithms, independent training and testing data sets (50% each) were generated from each selected pixel population, based on the purity thresholds applied to the convolved crop masks, and following an equalized random sampling (McCoy, 2005) to have the same number of reference pixels per class. The target size was initially set to 500 randomly selected pixels. For each pixel size simulated in the experiments, this initial number of pixels is set as the default, however in the case that there were not sufficient pixels to fulfil this initially targeted number (500), e.g. at coarser scales, a smaller number of pixels could be selected. Yet, the experiments were ultimately halted when for any class fewer than 20 pixels were available, which is a reasonable number for

---

[22]: For the sake of consistency the terms used in Duveiller et al. (2010) were applied, and purity is symbolized with π, and pixels size is equal to the GSD, symbolized by v.

achieving acceptable classification accuracies with SVM or tree structured classifier algorithms like RF or CARTs (Waske and Braun, 2009; Waske et al., 2009; Conrad et al., 2011a). Non-parametric classifier algorithms like SVM (Cortes and Vapnik, 1995) or RF (Breiman, 2001) have successfully been used in crop classification of diverse remote sensing data (Foody and Mathur, 2004; Loosvelt et al., 2012b; Mathur and Foody, 2008; Rodriguez-Galiano et al., 2012b; Waske and Benediktsson, 2007; Waske and Braun, 2009). Although SVM and RF completely differ in classification principle and motivation (see appendix A.2), it was shown that accuracies from RF are comparable to the SVM (Duro et al., 2012; Pal, 2005), but SVM is computationally more complex than RF.

The implementation of SVM in the package e1071 (Meyer et al., 2012) in R provides an interface to the well-known libsvm approach for SVM classification from Chang and Lin (2011). More technical details on SVM are given in appendix A.2. Training of the SVM with RBF kernel includes choosing the kernel parameter γ and the regularization parameter C. Tuning of C and γ was done using a systematic grid search in 2-D space that is spanned by γ and C. The range of γ was set to [0.00125, 4], the range of C was finally set to [1, 200]. The tuning of C and γ was performed at each spatial scale and for each pixel purity threshold, respectively.

The implementation of Breiman´s RF (Breiman, 2001) within the randomForest package (Liaw, 2013) in the R programming environment was used for classification in this study. More technical details on RF are given in appendix A.2. Two free parameters can be optimized in the RF. The number of trees was set to a relatively high value of 500 so that the OOB error can converge (see section 5.2.1). The second free parameter relevant for accurate classifications is the number of features m_{try} to split the nodes (Breiman and Cutler, 2007). As is considered adequate in literature (for categorical classification) the number of features at each node was set to the square root of the total number of input features $\sqrt{f}$, where f is the number of predictor variables within the corresponding input dataset (Gislason et al., 2006).

To evaluate the suitability of unsupervised crop identification, the K-means clustering (Tso and Mather, 2009) was tested that is implemented in the stats package (R Development Core Team, 2013) in R. A range of numbers of clusters was tested: {5,10,15,20}. The number of clusters that achieved highest accuracies was automatically selected. The maximum number of iterations was set to 50. Because random seeding is used for the initial clustering, the K-means algorithm was repeated 20 times, thereby creating different random seeds for the initial clustering. From the 20 model runs, the model with the lowest resulting sum of squared distances between the samples and their corresponding cluster centres was taken for the suitability evaluation of the unsupervised clustering. Each cluster containing at least 50 % of the samples of a class i were assigned to this class.

Input to the classification were different combinations of VIs and RapidEye bands, resulting in different feature set sizes that were tested. To enhance the reliability of the experiments the random draws of training and validation data were repeated 10 times for each classification method applied to each pixel population, and the classification performance estimates (see next section) were averaged over the 10 independent model runs. In doing so robust accuracy estimates reflective of this model evaluation approach could be obtained and possible bias in the results because of different distributional properties of the test and training sets can be reduced (Brenning, 2009). As a result of this, 24,150 classifications per study site were performed for each classifier algorithm to retrieve the parameters defined below (155 pixel sizes times 21 purity thresholds times ten model runs).

6.3.3. Characterizing classification performance

Pixel size and pixel purity can be considered as two dimensions of a $v - \pi$ space. For each selected pixel population in this $v - \pi$ space, information regarding the performance of a given application (here image classification) can be summarized as 3-D surfaces mapped along the $v - \pi$ dimensions. The standard protocol in remote sensing for evaluating the accuracy stems from quantitative metrics derived from the confusion matrix (Congalton, 1991). Yet, different metrics evaluate different components of accuracy because they are based on different statistical assumptions on the input data (Stehman, 1999) and such measures should be selected based on the requirements of the study (Stehman, 1997). Consequently, seeking to optimize or compare classifier algorithm performance (or defining suitable pixel sizes with only one metric) may lead to a non-optimal result when viewed from another point of view or quantified with a different metric that is sensitive to different features concerning accuracy (Foody, 2002; Provost and Fawcett, 1997). Next to evaluating the performance of a classifier algorithm post-classification, it must be assured that sufficient samples are drawn for a statistically significant accuracy assessment, and that the classifier algorithm can perform well on the unseen test set after the model building stage. Hence, to describe the performance of crop identification, several variables were calculated for each pixel population (e.g. each combination of π and v), whose 3-D representation is shown in Figure 6-6. These variables are later used to define the requirements for crop identification. After a short characterization of these variables, a short review on advantages and disadvantages of different types of accuracy metrics is given.

α-Quadratic entropy (AQE)

Measures of classification uncertainty like entropy assess the spatial variation of the classification quality on a per-case (e.g. per-pixel) basis, and can be used to supplement the global summary provided by standard accuracy statements like OA (Foody, 2002). It can be characterized as a quantitative measure of doubt when a classification decision is made in a hard way. Beneath the final class label, classifier algorithms such as RF or SVM generate for each classified case x (agricultural field or pixel) a soft output in form of a vector $pr(x) = (pr(x)_1, \dots, pr(x)_i, \dots, pr(x)_n)$ that contains the probabilities that a pixel is classified into a class i, n being the total number of classes. As was demonstrated in chapters 4 and 5, uncertainty measures computed from $pr(x)$ can be used to quantify classification uncertainty when a classifier makes a hard decision. From this vector, the α-quadratic entropy $H_\alpha(x)$ (Pal and Bezdek, 1994) for a given pixel (x) can be calculated as a measure of uncertainty, which is defined as:

$$AQE(x) = H_\alpha(x) = \frac{1}{n*(2-2^\alpha)} * \sum_{i=1}^{n} pr(x)_i^\alpha (1 - pr(x)_i)^\alpha \qquad \text{Eq. 6-6}$$

where $pr(x)_i$ is one element in $pr(x)$, n the number of classes, and α an exponent that determines the behaviour of $AQE(x)$, which becomes more and more selective as α increases from "0" to "1". With α close to "0", $AQE(x)$ becomes insensitive to changes in the elements in $pr(x)$, whereas for α close to "1", $AQE(x)$ is highly selective if the components in $pr(x)$ tend toward equalization. As a consequence, in this study $\alpha = 0.5$ was chosen as a good trade-off. After their calculation, the $AQE(x)$ of all pixels were scaled to a common scale [0,1]. The entropy of the total classified pixel population can be quantified with the median of all classified pixels' $AQE(x)$, called AQE. This can also be done at the per-class basis, by calculating the median entropy of all pixels classified into a class i, called AQE$_i$.

The parameter surface of AQE$_i$ (Figure 6-6 A) exhibits in general a deep furrow where the classification entropy is in its minimum. Along both sites of this furrow, toward smaller and larger pixel sizes, respectively, classification entropy increases. This can be attributed to counter effects between

increasing intra-class variability and a reduced boundary effect (due to decreasing number of mixed pixels) with smaller pixels sizes. With coarser pixel sizes the effect is vice-versa. This explains why selecting smaller pixel sizes does not necessarily lead to improved classification accuracies, and this effect is particularly strong when classes have high within-class variability (Cushnie, 1987). On the other site, when pixels become coarser and the ratio of the GSD to the field sizes increases, the intra class variability decreases. This is compensated by the increased mixed pixel error, and a phenomena called *valley-effect* occurs which means that the classification error reaches its minimum at an optimal ratio between pixel size and field width, when pure and mixed pixel errors counter-balance each other (Hsieh et al., 2001).

Overall classification accuracy (ACC)

A set of confusion matrices (Congalton, 1991) was computed on the hard result of the test sets defined along the $\pi - \nu$ dimensions. The overall accuracy parameter (ACC) is defined as the total proportion of correctly classified test pixels per total number of test pixels, and one of the most common measures of classification performance in remote sensing (Foody, 2002):

$$ACC = p_o = \frac{n_c}{n} \hspace{4cm} Eq.\ 6\text{-}7$$

where p_o is the proportion of correctly allocated test samples, n is the number of test samples, and n_c the number of correctly allocated test samples.

Class-wise classification accuracies (CA$_i$)

For each class i under investigation a class-wise measure of accuracy was employed. The general F_β-measure of Van Rijsbergen (1979) was adopted. This measure combines the precision pr_i (which gives the proportion of samples, which truly have class i among all samples that were classified as class i) and the recall tp_i (the TPR which gives the proportion of samples classified into class i among all samples which truly have class i). The former determines the error of omission (false exclusion), the latter the error of commission (false inclusion). Here as special case of the F_β-measure[23], $F_{0.5}$ was chosen that is defined as:

$$CA_i = F_{0.5} = (1 + \beta^2)\ \frac{pr_i * tp_i}{\beta^2 * pr_i + tp_i} \hspace{3cm} Eq.\ 6\text{-}8$$

where β was set to 0.5. This was done in order to put more emphasis on precision than recall, because the interest in this study lies in having highest possible precision in those pixels that were identified as the target (belonging to a class i), rather than identifying all pixels belonging to class i. CA$_i$ is computed for each class i along the $\pi - \nu$ dimensions. CA$_i$ was used to evaluate both, supervised and unsupervised classification. The difference is that in the latter case, pixels belonging to a class i are those pixels that were identified in the selected clusters (see section 6.3.2). CA$_i$ increases with increasing purities and coarser pixel sizes, respectively (Figure 6-6 B). Critically, there is a sharp drop in accuracy when purities are very low and selection of training and testing pixels includes excessive proportion of non-agricultural surfaces.

---

[23]: The traditional F_β-measure equally weights precision and recall ($\beta = 1$), and is sometimes referred to as F_1 measure.

Average area under curve (AUC)

Using receiver operating characteristic curves (ROC) and the corresponding area under the ROC (AUC) in evaluating classifier performance has been an increasingly used accuracy metric in machine learning and data mining (Fawcett, 2006). By plotting the TPR (here equivalent to recall tp_i) against the false positive rate (FPR, fraction of pixels where the crop class is incorrectly predicted as being present) for each probability threshold, ROC curves can be created. They have some advantages over the ACC, as they are especially useful when skewed class distributions or unequal classification error costs (e.g. that a false positive error is not equivalent to a false negative error, which is however assumed when using ACC, see Provost and Fawcett (1997)) have to be taken into account (Fawcett, 2006). The AUC is a measure for the probability that a classifier ranks a randomly chosen positive instance higher than a randomly chosen negative one (Fawcett, 2006). The AUC takes its values between "0" and "1", where random guessing produces a straight diagonal line between (0,0) and (1,1) and takes a value of 0.5. Here it is proposed to weight the influence of the test set size by introducing a weighting coefficient into the calculation of the AUC:

$$AUC = \sum_{i=1}^{n} AUC(i) * wg(i) \qquad\qquad Eq.\ 6\text{-}9$$

where n is the number of classes, $AUC(i)$ the area under the ROC curve for class i, and $wg(i)$ the weighting coefficient for each class i, which is computed with respect to the contribution (number of pixels) of each class to the test data set. AUC can be computed globally (e.g. for all classes) and at the per-class basis, termed AUC_i (Figure 6-6 C).

Number of reference pixels (N)

The number of available reference pixels N of a given class i, called N_i, gives the total available size of pixel populations in the $v - \pi$ dimensions that can be used as training and testing the classifier algorithm. In general N_i decreases with both higher π and v (Figure 6-6 D). The rate at which the available pixel number N_i of decreases differs among the crop types, depending on the total area of the crop in the test site, mean field sizes, and the aggregation pattern of field with the same crop. In supervised crop classification a minimum number of pixels per crop class can be desirable to assure the generalizability of the classifier model to the unseen dataset, and to reduce the influence of (random) variability in the training data on the classification result (Richards and Jia, 2005).

Sampling coverage of agricultural fields (SCF)

In crop mapping it is desirable to only include pixels that actually cover agricultural fields. This assumption might have to be relaxed when the resolving power of the instrument decreases and the amount of mixed pixels increases. The sampling coverage (SCF) describes the ratio between the sampled surface at a given scale (number of selected pixels times their surface), and the reference targeted area at the finest measured scale (here 6.5 m), and is computed as:

$$SCF = \frac{v^2 * N(\pi, v)}{v_0^2 * N_0} \qquad\qquad Eq.\ 6\text{-}10$$

where $N(\pi, v)$ is the number of agricultural pixels at a given scale and purity, v_0^2 the surface of a pixel at the finest spatial resolution v_0, and N_0 the number of pixels that cover agricultural fields at the finest investigate spatial resolution.

It can take values below "1" when the surface of the sampled pixels is below the reference surface. It also can take on values above "1", for instance when sampling of pixels with low purity thresholds at coarser scales results in too many pixels selected (e.g. too many pixels covering only a small proportion of agricultural field). The SCF surface is illustrated in Figure 6-6 E.
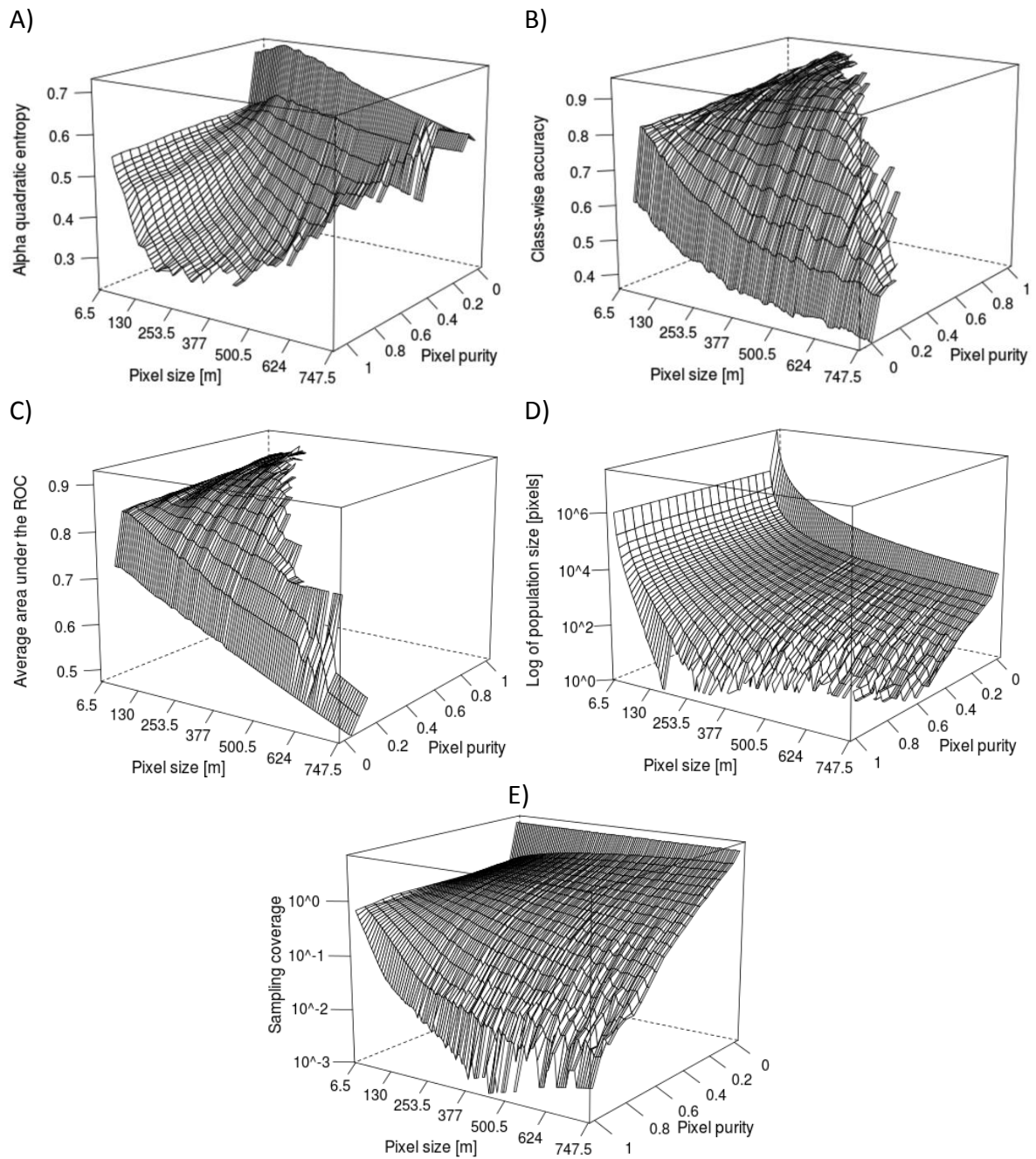
Figure 6-6: Examples of properties of pixel populations derived in KKP for cotton plotted across the pixel size-pixel purity dimensions: A) Median alpha quadratic entropy of the classified pixel populations (AQE$_i$), B) class-wise classification accuracy (CA$_i$), C) average area under the ROC (AUC$_i$), D) number of available reference pixels (N$_i$), and E) the sampling coverage of pixels related to the reference area of agricultural fields. The values shown in the surfaces are averaged over ten model runs. This parameter surface shows a pattern specific to cotton in KKP, and might deviate in the other test sites.

Choice of variables for characterizing classification performance

As already mentioned, each metric of accuracy is sensitive to different features of accuracy (Foody, 2002), consequently no single universally metric exist (Table 6-2). A straightforward way to measure classification accuracy is by calculating the proportion of correctly classified test cases from the confusion matrix (Congalton, 1991). However, the traditional accuracy or kappa k are often poor metrics for measuring performance as these assume equal error costs and are indifferent to whether a false positive or a false negative error was detected, however class distributions and misclassification costs are rarely uniform (Pontius and Millones, 2011; Provost and Fawcett, 1997). Measures like recall, precision, and F_β-measure (a combination of the former two), which are frequently being used in machine learning applications, are strong in estimating performance on different classes. CA_i, a measure that combines precision and recall and that focusses on the set of samples which truly have class i among all samples which were classified as class i is preferable, because this study focusses on correctly identifying crops in coarser pixels, irrespective of the area they cover.

However, these metrics are poor in handling negative examples (e.g. false positives), and fail to take account to chance level performance (Powers, 2011). ROC analysis are commonly used in medical sciences and are frequently being employed in the machine learning community (Fawcett, 2006). They compare true positive and false positive rates, and have advantages over the more conventional overall accuracy (Bradley, 1997) when skewed class distribution or unequal classification error costs must be taken into account. The second criterion used to measure the classification performance (AUC/AUC_i) therefore takes into account skewed class distributions or unequal classification error costs that may be expected when pixels get coarser and coarser and class distributions in the reference data sets become more unequal. On the other hand, the precision (proportion of predicted positive cases that are correctly true positives) is ignored by ROC analysis.

Even with accurate crop maps, there might be uncertainty in the final class allocation because crop identification relies on a signal that may be composed of reflected radiation coming from different adjacent land uses (mixed pixel). Entropy is a spatial measure of classification uncertainty (see section 4.2.3) and is used to assess the spatial distribution of classification uncertainty and to quantify the doubt or mistrust on a per-case basis in the final class allocation. Such measures are still not frequently being used in classifier performance assessment, although they were shown to be correlated with conventional (hard) class allocation accuracy (Giacco et al., 2010; Loosvelt et al., 2012a) and a useful supplement to the classical accuracy statement (Foody, 2002). For crop classification, the degree of tolerable entropy and, by extension, the degree of heterogeneity of the resulting pixel population can be controlled by the choice of AQE/AQE_i. If crop identification is to be used as prerequisite of crop specific monitoring in view of agricultural production estimation, then the classified pixels should ideally be composed of only one crop. Traditional accuracy metrics do not consider any sub-pixel composition. Yet, uncertainty is a function of the pixel composition (Foody, 2002) that can be used to indicate the composition of a pixel (Foody, 1996b), e.g. land cover fractions. Hence, the third criterion relates to the uncertainty on the final class allocations in the selected pixel population. It must be noted that this correlation might be susceptible to the tuning of the algorithm (Brown et al., 2009), see also discussion in chapter 5 in this thesis.

In crop classification it is desirable to assure the generalizability of the classifier model to the unseen dataset, and to reduce the influence of (random) variability in the training data on the classification result. Recent efforts showed that RF and SVM classification can achieve high accuracies even with small number of training pixels (Melgani and Bruzzone, 2004; Rodriguez-Galiano et al., 2012b; Waske and Braun, 2009), and comparatively low sensitivity to training pixel variability when using small training set sizes, e.g. using only 20 pixels per class for SVM (Shao and Lunetta, 2012). For the RF it was demonstrated that they are relatively insensitive in regard to sample size and that even small training

sets (30 or 50 pixels per class) are sufficient to achieve high quality output on complex multi-temporal data sets (Conrad et al., 2011a), but can decrease considerably when using insufficient training data, e.g. 15 pixels per class (Waske and Braun, 2009). Hence, a fourth criterion was used to assure that sufficient samples are drawn for a significant accuracy assessment, and that sufficient pixels are available for the model building stage. SCF was used in this study to interpret the effects of selecting different pixel populations on crop area coverage and to discuss possible impact on crop area estimation via pixel counting.

Table 6-2: Properties of selected classification performance metrics. "↗" indicates advantageous properties of the metric, "↘" indicates potential critical properties.

Metrics	↗	↘	Reference and application
Accuracy / Kappa	Easily interpretable / accommodate for the effects of chance agreement	Assumption of balanced class distribution, no distinction between number of correct labels of different classes	Cohen (1960), Congalton (1991), Pontius and Millones (2011), Rosenfield and Fitzpatrick-Lins (1986), and Stehman (1997)
AUC	Take into account skewed class distribution and unequal classification error costs	Incoherent in terms of misclassification costs for different classifiers	Bradley (1997), Fawcett (2006), Hand (2009), and Powers (2011)
Precision, Recall, F_β-measure	Estimate performance on different classes	Ignore performance in correctly handling negative examples, fail to take account the chance level performance	Powers (2011), Van Rijsbergen (1979), and Sokolova et al. (2006)
Entropy	Assess spatial distribution of error	As a by-product of the classifier it can be susceptible to parameter setting of the algorithm	Brown et al. (2009), Giacco et al. (2010), Loosvelt et al. (2012a), Maselli et al.(1994), and Foody (2002)

6.3.4. Definition of constraints for crop identification

The final step to determine suitable pixel sizes for crop identification is to isolate the (ν, π) combinations for which the classification performance is good enough. This is accomplished by defining acceptable thresholds for the surfaces defined above (Figure 6-6). Such thresholds will be used to slice the surfaces with a plane parallel to the $\nu - \pi$ space, thereby defining a frontier in this $\nu - \pi$ space dividing pixel populations that are above or below the acceptable threshold for a given surface. As an example, if an application requires a minimum classification accuracy of 85 %, the surface ACC is sliced by a plane passing by the value ACC=0.85. When the intersection of ACC and the plane is projected onto the 2-D space $\nu - \pi$, it separates this domain into the region where selected pixel populations have a classification accuracy higher than 85 % and the region where the accuracy of the remaining population will be lower than 85 %. The coordinates (ν, π) along the division boundary satisfy the imposed condition ACC=0.85.

Parameterization

Users will have different requirements for selecting their pixel population of interest, e.g. it might be acceptable to have crop classes identified at different levels of accuracy as long as the classes of interest are sufficiently accurately identified. Hence, a range of thresholds for the accuracy metrics was considered. For ACC/CA$_i$, and AUC/AUC$_i$, the thresholds 0.75, 0.80 and 0.85 are tested. The entropy (AQE/AQE$_i$) of the classification result must be below 0.45, 0.50, and 0.55, respectively, in order to control the degree of classification uncertainty in the results. The minimum number of reference pixels N$_i$ required for training and testing was set to 50, 75, and 100 per class. These values have been chosen so as to be realistic in an operational perspective and to fulfil commonly recommended accuracy targets

(Anderson et al., 1976), but also to allow for some flexibility. Several feature sets were tested (Table 6-3): (i) only RapidEye bands without red edge, (ii) RapidEye bands without red edge band, (iii) NDVI (Rouse et al., 1974) and EVI (Huete et al., 2002), (iv) RapidEye bands (expect for red edge band) plus NDVI and EVI, (v) red edge indices (here GRNDVI and RNDVI), (vi) red edge indices plus RapidEye bands (expect for red edge band), and (vii) all VIs and spectral bands combined.

Table 6-3: Overview of the parameterization and input data used for the calculation of suitable pixel sizes for crop identification. Note that for all parameters three thresholds were set for the experiments. For the unsupervised method classification entropy was not calculated, as no ready to use method is available to generate soft outputs for K-means.

Supervised: SVM and RF				
Data sets tested (Number of features f)	N_i	ACC/CA$_i$	AUC/AUC$_i$	AQE/AQE$_i$
i) Red or red edge band (8)				
ii) Bands without red edge (32)				
iii) NDVI and EVI (16)				
iv) Bands + NDVI + EVI (48)	> 50, 75, 100 per class	> 0.75, > 0.80, 0.85	> 0.75, > 0.80, > 0.85	< 0.55, < 0.50, < 0.45
v) REDNDVI + GRNDVI (16)				
vi) Bands + REDNDVI + GRNDVI (48)				
vii) All features combined (72)				
Unsupervised: K-means				
viii) NDVI (8)	> 50, 75, 100 per class	0.75, 0.80, 0.85	> 0.75, > 0.80, > 0.85	-

The next step is defining several thresholds for crop identification at the global (for all classes) and the per-class basis, respectively, and drawing limits on the pixel populations in the $\nu - \pi$ space to isolate the pixels that fulfil all defined requirements. By drawing limits on the different parameters, according to the thresholds defined in Table 6-3, the parameter surfaces were sliced and the intersection points of these slices in $\nu - \pi$ space were used to identify the position of the coarsest acceptable pixel sizes (ν_{max}) and the corresponding minimum acceptable pixel purities π, respectively. The intersection of the boundaries yielding the coarsest pixel size ν defines the coarsest acceptable pixel size ν_{max} (and the corresponding acceptable purity). Figure 6-7 shows examples for experimental boundaries in $\nu - \pi$ space used define ν_{max} for pixel populations that is to be used for crop identification. As is shown in this figure, adequate pixel populations are defined in this example by the ACC (Figure 6-7 A-C) or the AUC (Figure 6-7 D) constraint boundaries, which intersect the reference pixel number constraint (N_i) boundary. The point where these boundaries intersect in Figure 6-7 determines the position of the coarsest acceptable pixel size (ν_{max}) in the $\nu - \pi$ space (marked by "○"). As can be seen from this figure, a theoretical minimum pixel size (ν_{min}, marked by "■") can be derived when the application of finer pixels is restricted, e.g. due to excessive entropy (AQE) or insufficient accuracy (ACC). In these examples the use of pixels with very low purities along the pixel sizes tested, and partly the use of very small pixel sizes with medium degrees of purity is restricted by high values of classification entropy (AQE). As can be seen in Figure 6-7 D, unsupervised crop signature identification is possible but within a very limited domain in the $\nu - \pi$ space, and at the expense of relatively high pixel purity requirements. As can be seen, the adequate pixel populations for the unsupervised case are defined in by the AUC constraint boundary and the reference pixel number (N_i) constraint boundary.
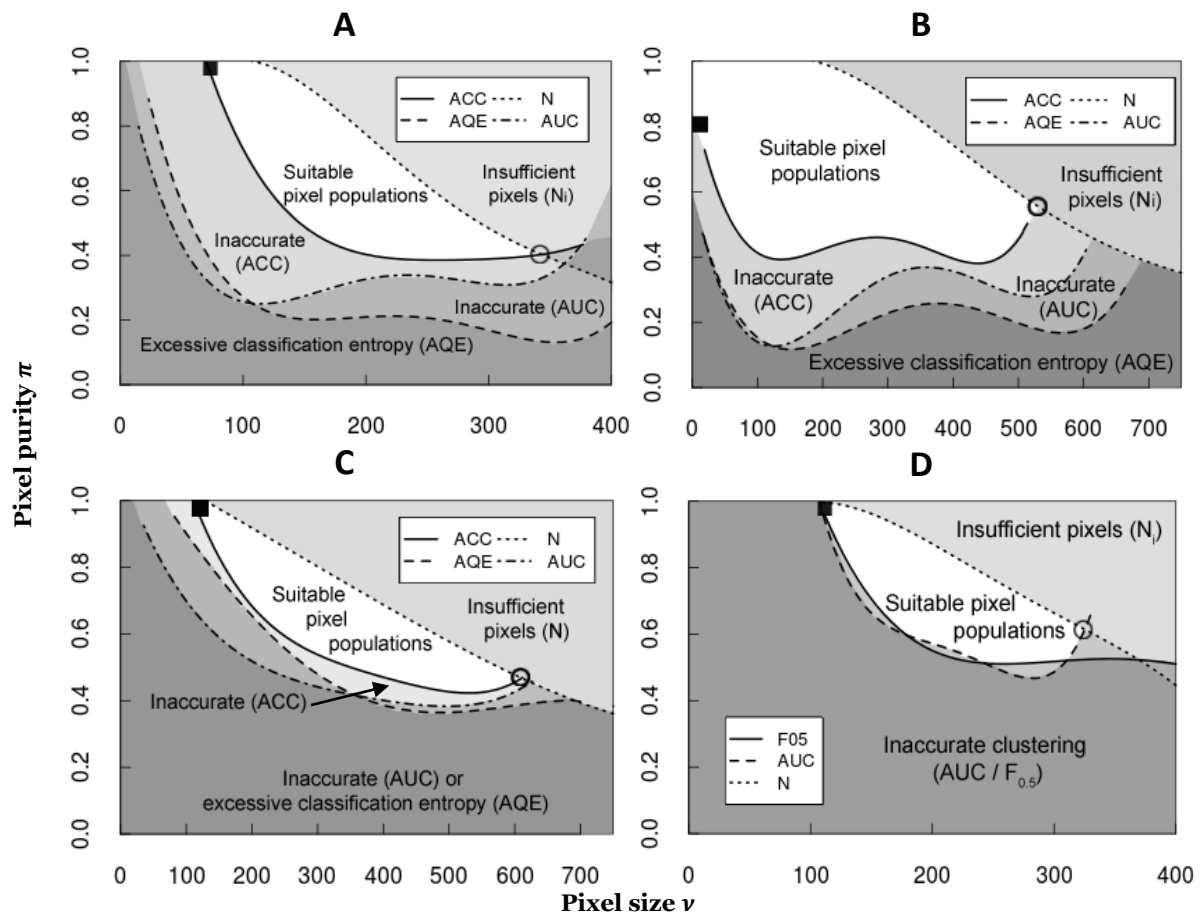
Figure 6-7: Examples for experimental boundaries in $\nu - \pi$ space used to define the coarsest acceptable pixel size (ν_{max}) for pixel populations to be used for crop identification. Examples are shown for RF classification in KKP (A), FER (B), and KYZ (C). An example for unsupervised identification of rice in KKP is given (D). Circles indicate the positions of ν_{max}, black filled squares the minimum required pixel size (ν_{min}). Boundaries were calculated using RF applied to the full dataset (bands plus indices).

Creation of suitability maps

Increasing the thresholds defined in Table 6-3 for each parameter (e.g. from 0.75 to 0.85 for ACC) to define the experimental boundaries illustrated in Figure 6-7 results in having less and less suitable pixel populations left for crop identification that can fulfil the stricter criteria. Figure 6-8 demonstrates this effect in KKP, where higher thresholds are successively selected for each parameter at the same time, according to the parameterization defined in Table 6-3. The colours in Figure 6-8 indicate how many parameters, under the corresponding parameterization, are still fulfilled. Green colours indicate that all parameters (ACC, AUC, AQE, and N$_i$) are fulfilled and the pixel population can be considers adequate ("suitable") for crop identification. Selecting higher thresholds leads to having fewer and fewer suitable pixel populations left. The combination of these three maps into one yields a "suitability" map, which shows the degree of suitability of the corresponding pixel populations for crop identification considering several thresholds at the same time. In these maps, shades of a given colour mean that a certain number of parameters is fulfilled, but at different levels: for instance, shades of green means that all four parameters, e.g. ACC, AUC, AQE, and N$_i$ are fulfilled, but not necessarily under the strictest thresholds defined in Table 6-3. Only dark green colour indicates that all parameters are fulfilled under the strictest values (e.g. ACC > 0.85). Whilst these graphs map the suitability of pixel populations with regard to identifying crops over a range of thresholds (e.g. ACC = 0.75, 0.80, and 0.85), ν_{max} and ν_{min} will only be

derived from the intersection of the experimental boundaries under the least restrictive thresholds (e.g. Figure 6-7A, B, and D) and reported in the next sections.
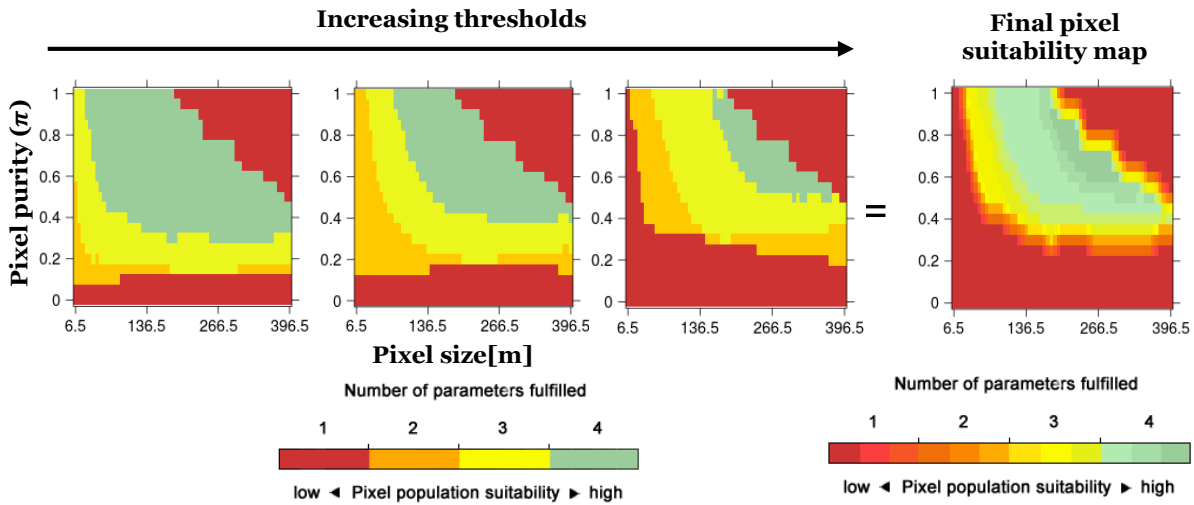
Figure 6-8: Schematic example for the evolution of the amount of suitable pixel populations (green colours) when increasing the thresholds. The first three images (from left to right) illustrate the effect of setting thresholds to 0.75, 0.80, and 0.85, respectively for ACC and AUC. N_i was set to 50, 75, and 100, and entropy values were set to 0.55, 0.50, and 0.45. The image on the right shows the final pixel suitability map, which combines the three single suitability maps. Note that pixel populations that did not fulfil any parameter are also indicated in dark red colour.

6.4. Results

The analysis of the data generated in this study focusses on reflecting the five sub-objectives that were defined in the introduction of this chapter concerning the definition of pixel size requirements for crop identification (sections 6.4.1. – 6.4.5). Then the results are discussed in section 6.5 and a conclusion and research outlook will be given in section 6.6.

6.4.1. How do pixel size and purity requirements differ per crop for each site?

In the following, the experimental results are given for the overall (considering all classes at the same time) and the per-class cases, respectively. For the experiments, the combined spectral RapidEye bands plus vegetation indices (NDVI and EVI) were used as input ($f = 48$), to allow for a better transferability of the results to sensors without red edge bands. The differences in classification performance and the resulting ranges of suitable pixel sizes, respectively from the two supervised classifier algorithms (RF and SVM) differed only marginally, yet the RF was much faster in computation. Hence, for the sake of clarity the presentation of the results focus on the RF. Finally, the spatial pattern of crops identified at different scales will be illustrated and discussed.

Pixel suitability for overall crop classification

For all study sites, the ranges of suitable pixel sizes and corresponding pixel purities for overall crop classification under the least restrictive thresholds (defined in Table 6-3), e.g. 0.75 for ACC, are reported in Table 6-4. As can be seen from this table, maximum and minimum tolerable pixel sizes, v_{max} and

v_{min}, respectively vary from site to site. In general it is possible to effectively identify crops under the given parameterization within coarser pixels while maintaining accuracies of at least 0.75 and 0.75 (ACC and AUC, respectively), which are the minimum required thresholds defined in Table 6-3. In the KKP landscape the use of coarse pixel sizes is ultimately restricted (v_{max} = 422.5 m) due to an insufficient classification accuracy that dropped below 0.75 (ACC as critical factor). In the KHO landscape the experiments were halted at 409.5 m due to an insufficient number of training pixels for sorghum/maize (that dropped below 20, see section 6.3.2), hence no value for v_{max} was defined for the overall classification. The use of very small pixel sizes (v_{min}) was restricted in KKP and KYZ by the ACC constraint, and in KHO by the AQE constraint. For the former sites this means that misclassification increased when using smaller pixel sizes, and overall classification accuracies (ACC) dropped below 0.75. For KHO this means that the classification entropy, as quantified by AQE, raise above 0.55. It must be noted that for practical reasons smaller pixels could be classified using object-based approaches, which could result in higher accuracies (Blaschke, 2010). However, the focus here is on the general relationship between pixel size and crop identification. A study on the impact of pixel size on object-based classification was provided by Turker and Ozdarici (2011). The values for v_{max} in the KYZ and FER landscapes were much coarser (728.0 m and 656.5 m, respectively) than in the KHO and KKP landscapes.

Table 6-4: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding pixel purity π for crop identification using RF in all sites in 2011 (2009 in KHO). The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75). In KHO the experiment was halted at 409.5 m because for one class the number of available pixels dropped below 20.

Site	v_{min}			Conditions for highest overall accuracy			v_{max}		
	v	π	Crit.	v	π	ACC	v	π	Crit.
KHO	110.5	1.00	AQE	331.5	0.85	0.88	409.5	0.45	*
KKP	26.0	0.95	ACC	260.0	0.65	0.89	422.5	0.35	ACC
KYZ	19.5	0.70	ACC	279.5	0.80	0.94	728.0	0.30	ACC
FER	6.5	0.15	-	260.0	0.90	0.95	656.5	0.40	ACC

The maximum achievable accuracy in $v - \pi$ space, quantified by ACC, was also investigated. In Table 6-4 it can be seen that maximum accuracies were achieved with relatively coarse pixel sizes (e.g. 260.0 m in KKP and 331.5 m in KHO). The reason why coarser pixels achieved higher accuracies than smaller pixels, which was also observed in other studies (McCloy and Bøcher, 2007) could be the interplay of increasing error-rates of smaller but purer pixels (which become more abundant when pixels become smaller), caused by increasing within-class variability (Hsieh et al., 2001) and decreasing error of mixed pixels (which become less abundant when pixels become smaller). This is illustrated for the KKP landscape in Figure 6-6 B for the CA_i parameter. Similar to this, Hsieh et al. (2001) demonstrated how classification error of pure pixels decreases with decreasing observation support (in their study quantified as ratio between pixel size and field width), and how the classification error of mixed pixels first declines with decreasing pixel size and then increases again (Figure 6-9 B). The same can be revealed for the example with the KKP landscape in Figure 6-9 A. In this figure it can be seen how classification error of purer pixels decreases with coarser observation support (e.g. coarser pixel sizes), and how classification error of mixed pixels first decreases and then increases again with very coarse pixel sizes. Further, the results presented in Table 6-4 reveal that the finer v_{min} becomes, the lower tend the pixel purity requirement for effectively identifying crops. This could be explained by the increasing proportion of pure pixels, relatively to mixed pixels, when image resolution becomes finer. In general higher pixel purities are required in KKP and KHO, compared to KYZ and FER, both for using small v_{min} and coarse v_{max} pixels.
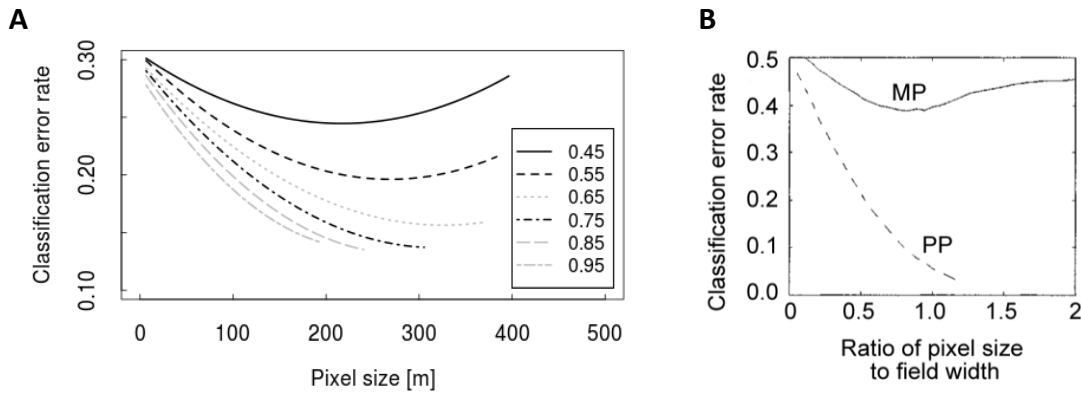
Figure 6-9: Impact of pixel size and pixel purity on error rates (1-ACC). A: With smaller pixels, the classification error of pure pixels increases, whilst the classification error of mixed pixels first declines and then increases again (experimental results are from the KKP landscape). B: experimental results from a simulated landscape showing error rates of mixed and pure pixels (MP = mixed pixels, PP = pure pixels) as a function of the ratio between pixel size to field width, figure adopted from Hsieh et al. (2001).

Pixel suitability for identification of specific crop classes

Figure 6-10 – Figure 6-13 show the suitability maps of pixel populations in $v - \pi$ space for crop identification under various constraints, and considering different thresholds at the same time (e.g. 0.75, 0.80 and 0.85 for CA_i and AUC_i, and 0.55, 0.50 and 0.45 for AQE_i). The creation of these maps was highlighted in section 6.3.4. Table 6-5 – Table 6-8 supplementary give the values for v_{min}, v_{max} and the corresponding pixel purities π for each class.

It can be seen that the crop classes displayed distinct individuality regarding the spatial resolution requirement over one particular landscape, and that the requirements for one specific crop differed over the four landscapes. To have one example for the latter case, the identification of cotton in KHO required a minimum pixel size of v_{min} = 117.0 m, whilst in FER v_{min} could be 6.5 m for the same crop. Wheat-other could be identified over a large range of pixel sizes in FER (v_{min} = 6.5 m, v_{max} = 611.0 m), whilst its identification in KKP was restricted to 32.5–91.0 m. As can be seen from the latter example, the effective identification of some classes was restricted to very limited domains within the $v - \pi$ space, e.g. melons (v_{min} = 19.5 m, v_{max} = 65.0 m) in KKP can only be identified with small pixel sizes. This is because these crops were cultivated on few and small scattered fields and together only covered a minor part of the study site (Table 6-1). Consequently the signal strength of such a target becomes too weak when using coarse pixels when other crops have already grown significantly and have comparable reflectance. This ultimately led to excessive misclassification (CA_i for wheat-other as critical factor in Table 6-5).

Other crop classes like alfalfa-1y and fallow fields in KKP required relatively coarse values for v_{min} (65.0 m and 78.0 m, respectively) and relatively high purity thresholds were required. High within-class variability and within field heterogeneity that lead to erroneous class allocation could be an explanation for this (indicated by AQE_i as critical factor for the determination of v_{min} for alfalfa-1y). Further, cotton, fruit trees, and wheat-other in KHO could not be identified adequately when using very small pixel sizes, and require relatively high purity thresholds, respectively. Again, AQE_i limited the use of small pixels to identify these crops. To have one example to explain this: fruit trees in KHO often have a second cultivar below the canopy, e.g. alfalfa, wheat, vegetables, or maize. Sometimes even cotton can be found under the fruit trees as a second cultivar, which could explain the confusion with other classes and why AQE_i limited smaller pixel sizes. Using coarser pixels is expected to reduce this variance, leading to better class discrimination. Similar to this, in KKP the irregular cutting operations that lead to indistinct NDVI profiles could explain AQE_i as critical factor for alfalfa-1y.

Whilst the aforementioned cases are examples for crops whose identification was restricted to more or less narrow pixel size ranges, there are crops that could be identified using larger ranges of pixel sizes. For instance, cotton and wheat-other in FER were identified with pixel sizes coarser than 600 m, and for a wide range of purities, respectively. This landscape is characterized by a high cover fraction of agricultural fields (0.57, see Table 2-1), and by the largest mean field size among the four sites (6.74 ha), which means that it is easier to have coarser pixels projected entirely within a field. v_{min} tended to be smaller than in KHO. Yet, fallow fields in FER only covered 4 % of the site, resulting in decreases values for v_{max} (260.0 m), compared with the other classes (Table 6-8). In general, rice fields and winter wheat fields could be identified over wide ranges of pixel sizes in all sites (e.g. with coarse pixel sizes) expect in the FER test site where it was not cultivated. In the former case this might be attested to the clear temporal signature of winter wheat fields (one vegetation peak before harvest in spring). Rice fields are leached with water before planting, in order to drain salt into the lower part of the soils. This results in a negative NDVI signal in spring, which makes the temporal signatures very distinct from other crops. This is also reflected by the comparatively low pixel purity requirements for v_{max} of rice fields (e.g. 0.25 in KKP), compared with the corresponding purities for v_{max} of other classes (e.g. 0.75 for alfalfa-1y in KKP). Rice in KYZ could be identified with pixels coarser than 700 m (v_{max}= 747.5 m, Table 6-7), which can be attested to the spatial aggregation pattern of fields in that site (see section 2.3.1). Cotton was easier to detect in KKP and FER than in KHO, where the minimum required pixel size to effectively identify cotton fields was 117.0 m. In KHO, KKP, and KYZ the use of very small pixel sizes of many classes was limited by AQE_i, which means that there was serious classification uncertainty. Again, one reason could be the increasing within-class variability and pure pixel errors, respectively.

In the KHO landscapes the values for v_{max} could not be determined for all classes (e.g. rice or wheat-other) because the experiments stopped at $v = 409.5$ m when one class (sorghum/maize) had less than 20 training pixels. This means that the boundaries in $v - \pi$ space did not intersect for these classes before the experiment stopped, and consequently v_{max} could not be determined for all classes, but can be assumed coarser than 409.5 m.

In general, the most important limiting factors that determined v_{min} and v_{max} in all test sites were either an insufficient class-wise accuracy (CA_i) or excessive classification entropy (AQE_i) (see column "Crit." in Table 6-5 – Table 6-8). Sorghum/maize could not effectively be identified in any site, neither in the KHO landscape nor in KKP, because more than two thresholds were generally exceeded (CA_i and AQE_i). Further, as was done for the overall classification (Table 6-4), the position in $v - \pi$ space where the classification performance, according to one specific metric (CA_i, AUC_i or AQE_i), is in its maximum was assessed. As can be seen from the crops´ suitability maps (Figure 6-10 – Figure 6-13), this position did not necessarily coincide with highest degree of the corresponding pixel populations´ suitability (green colours in the suitability maps). This demonstrates that taking one metric alone might be a poor indicator to define the suitability of pixel populations with regard to a specific application (see section 6.3.3). One striking characteristic that was already observed for the overall crop classification is the need for relatively coarse pixels to achieve maximum classification accuracy. The positions of maximum achievable CA_i and the corresponding minimum required purity to achieve this in $v - \pi$ space displayed individuality regarding different types of crops (see Table 6-5 – Table 6-8). As an example, best performances for melons in KKP could be expected at 39.0 m, but 182.0 m were required to achieve highest CA_i for rice fields. The required pixel purities tended to be the lower the coarser v becomes, and relatively coarse pixel sizes were required to achieve best results. Again, this might be explained with the opposite effect of spatial resolution on the classification errors associated with pure pixels and mixed pixels (see Figure 6-9).
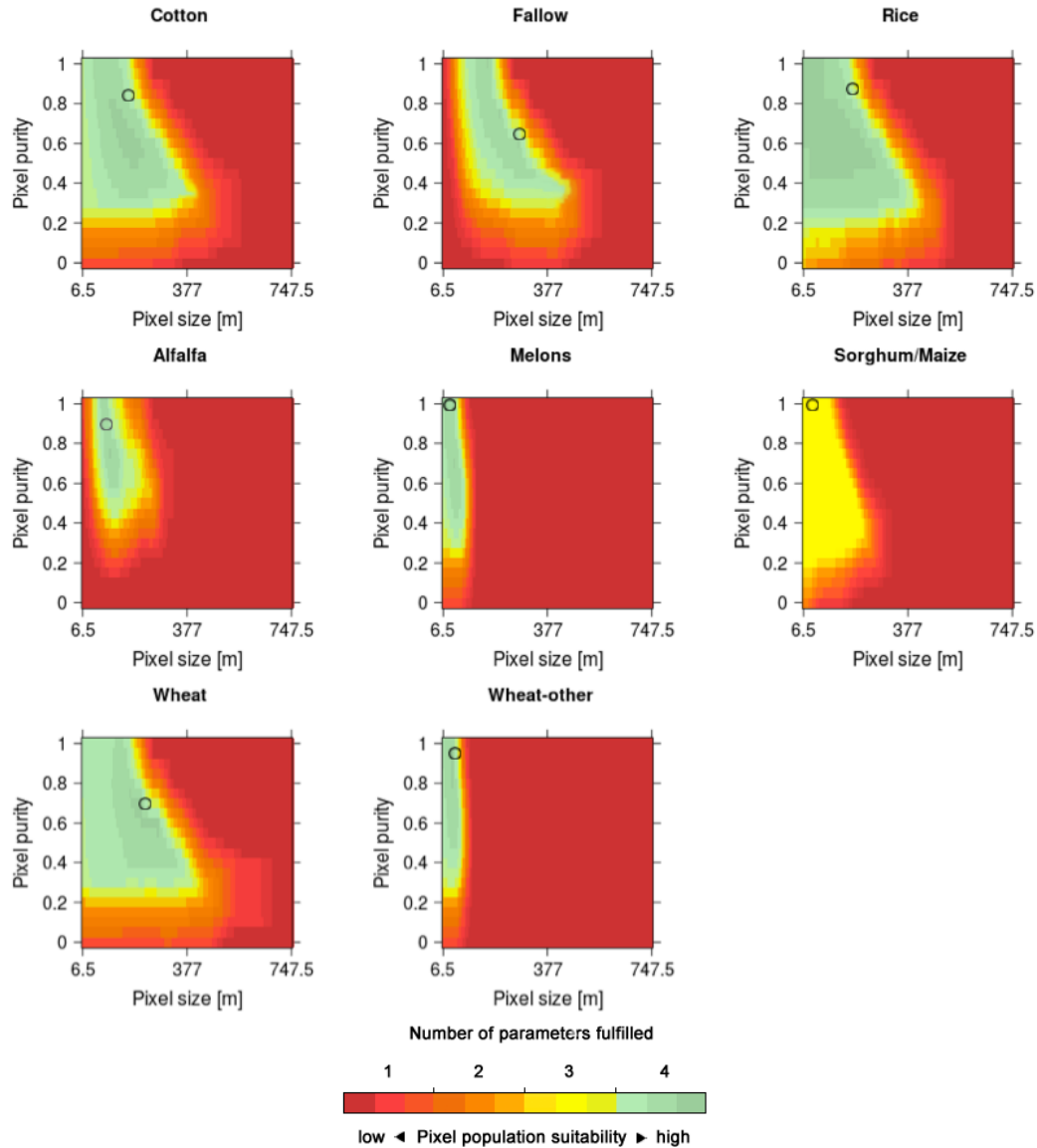
Figure 6-10: Suitable pixel populations for crop identification using RF in KKP 2011. Green colours indicate suitable populations in the pixel size-pixel purity space, where all criteria defined above are met, yellow colours indicate that one criterion is not met, orange means two criteria are not met, and finally red colours indicate that three or four criteria were not met. Circle indicates the actual position of the best values achieved in the experiments for CA$_i$.

Table 6-5: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding pixel purity π for crop identification using RF, at the per-class basis, in KKP 2011. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. Sorghum/maize in KKP could not be identified under the given parameterization. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75).

Crop	v_{min}			Conditions for highest class-wise accuracy			v_{max}		
	v	π	Crit.	v	π	CA$_i$	v	π	Crit.
Alfalfa-1y	65.0	1.00	AQE$_i$	91.0	0.90	0.81	140.5	0.75	CA$_i$
Cotton	6.5	0.80	-	182.0	0.85	0.91	429.0	0.35	AUC$_i$
Fallow	78.0	1.00	CA$_i$	279.5	0.65	0.93	487.5	0.40	CA$_i$
Rice	6.5	0.20	-	182.0	0.90	0.97	429.0	0.25	CA$_i$
Melons	19.5	0.70	AQE$_i$	39.0	1.00	0.83	65.0	0.45	N$_i$
Sorghum/maize	-	-	-	52.0	1.00	0.71	-	-	-
Winter wheat	52.0	0.60	CA$_i$	201.5	0.70	0.94	429.0	0.30	CA$_i$
Wheat-other	32.5	0.90	CA$_i$	71.5	0.95	0.81	91.0	0.50	CA$_i$
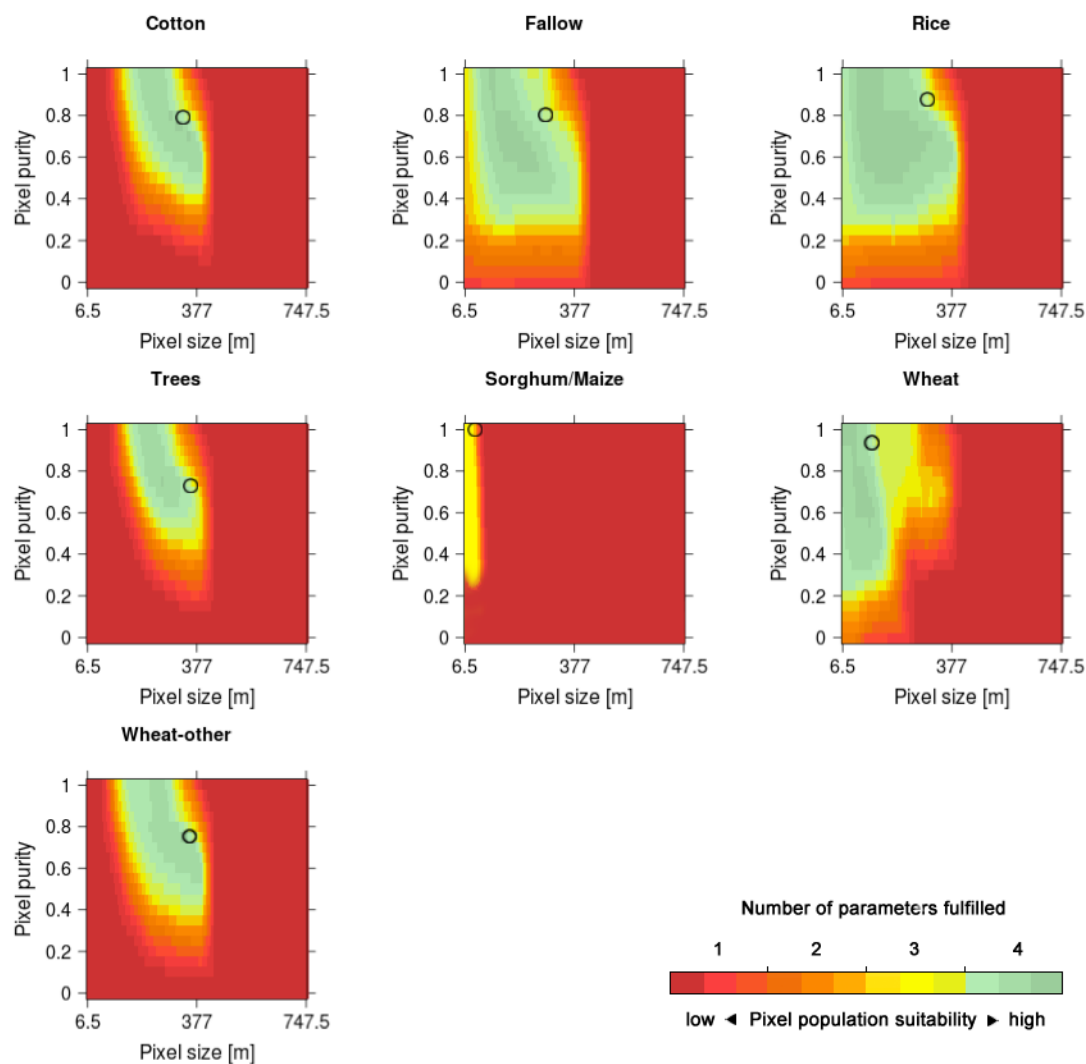
Figure 6-11: Suitable pixel populations for crop identification using RF in KHO 2009. Green colours indicate suitable populations in the pixel size-pixel purity space, where all criteria defined above are met, yellow colours indicate that one criterion is not met, orange means two criteria are not met, and finally red colours indicate that three or four criteria were not met. Circle indicates the actual position of the best values achieved in the experiments for CA_i.

Table 6-6: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding pixel purity π for crop identification using RF, at the per-class basis, in KHO 2009. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. Sorghum/maize in KHO could not be identified under the given parameterization. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75). Asterisk indicates that for this site the experiments were halted at 409.5 m because for one class the number of available training pixels dropped below 20 when using pixels coarser than 409.5 m.

Crop	v_{min}			Conditions for highest class-wise accuracy			v_{max}		
	v	π	Crit.	v	π	CA_i	v	π	Crit.
Cotton	117.0	0.95	AQE_i	344.5	0.80	0.91	403.0	0.85	CA_i
Fallow	13.0	0.95	CA_i	331.5	0.80	0.97	364.5	0.40	CA_i
Rice	13.0	0.60	CA_i	360.0	0.85	0.77	409.5	0.35	*
Fruit trees	143.0	1.00	AQE_i	377.0	0.75	0.83	390.0	0.75	CA_i
Sorghum/maize	-	-	-	71.5	1.00	0.40	-	-	-
Winter wheat	6.5	0.20	-	110.5	0.95	0.90	195.0	0.50	AUC_i
Wheat-other	117.0	1.00	AQE_i	377.0	0.75	0.91	409.5	0.60	*
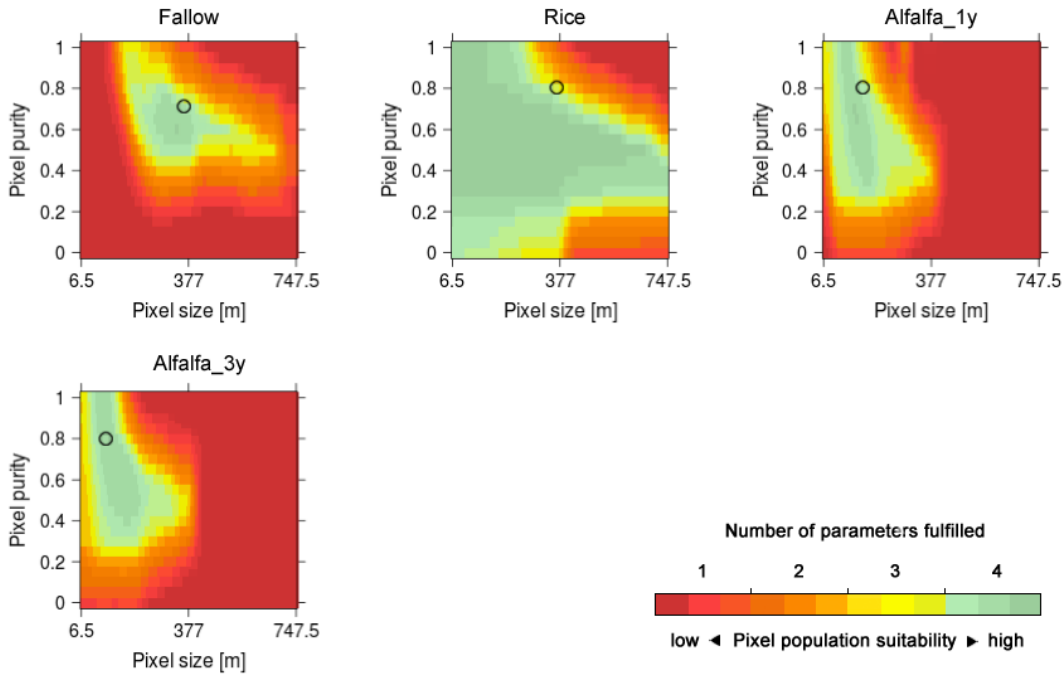
Figure 6-12: Suitable pixel populations for crop identification using RF in KYZ 2011. Green colours indicate suitable populations in the pixel size-pixel purity space, where all criteria defined above are met, yellow colours indicate that one criterion were not met, orange means two criteria are not met, and finally red colours indicate that three or four criteria are not met. Circle indicates the actual position of the best values achieved in the experiments for CA_i.

Table 6-7: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding pixel purity π for crop identification using RF, at the per-class basis, in KYZ 2011. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75).

| Crop	v_{min}			Conditions for highest class-wise accuracy			v_{max}		
	v	π	Crit.	v	π	CA_i	v	π	Crit.
Fallow	175.5	0.95	AQE_i	364.0	0.70	0.86	559.5	0.60	N_i
Rice	6.5	0.05	-	377.0	0.80	0.99	747.5	0.35	AQE_i
Alfalfa-1y	65.0	0.95	CA_i	162.5	0.80	0.89	169.0	0.40	CA_i
Alfalfa-3y	26.0	0.95	AQE_i	110.5	0.80	0.84	221.0	0.45	CA_i
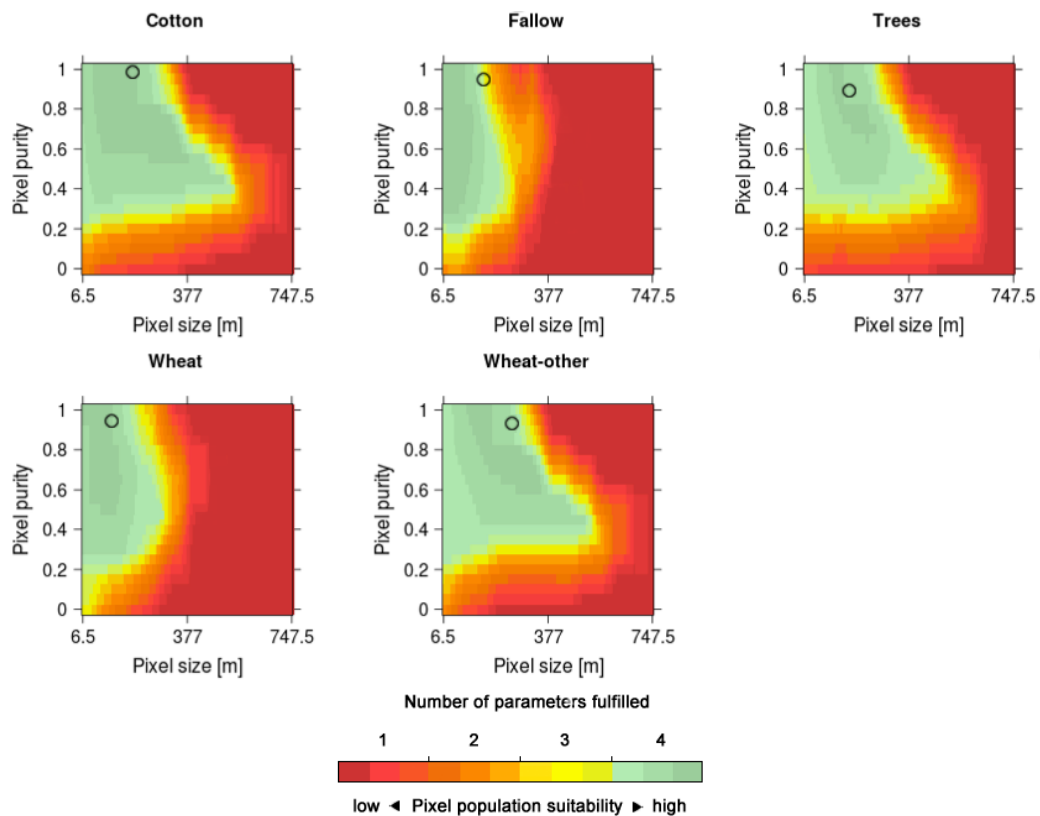
Figure 6-13: Suitable pixel populations for crop identification using RF in FER 2011. Green colours indicate suitable populations in the pixel size-pixel purity space, where all criteria defined above are met, yellow colours indicate that one criterion were not met, orange means two criteria are not met, and finally red colours indicate that three or four criteria are not met. Circle indicates the actual position of the best values achieved in the experiments for CA_i.

Table 6-8: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding pixel purity π for crop identification using RF, at the per-class basis, in FER 2011. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75).

Crop	v_{min}			Conditions for highest class-wise accuracy			v_{max}		
	v	π	Crit.	v	π	CA_i	v	π	Crit.
Fallow	13	0.15	CA_i	182.0	0.95	0.98	260.0	0.55	CA_i
Cotton	6.5	0.20	-	175.5	1.00	0.97	663.0	0.35	AUC_i
Fruit trees	19.5	0.50	AQE_i	175.5	0.85	0.90	435.5	0.25	CA_i
Winter wheat	13.0	0.20	CA_i	130.0	0.95	0.95	325.0	0.55	N_i
Wheat-other	6.5	0.20	-	266.5	0.95	0.95	611.0	0.30	CA_i

Mapping identified crops at multiple scales

When looking at the positions in $v - \pi$ space where the selected classification performance parameters CA_i was in its maximum (Figure 6-10 – Figure 6-13) it can be seen that relatively high purities were required for many classes to achieve the best values for this performance metric (e.g. highest CA_i). However, in practice excluding many (potentially mixed) pixels could result in an under-sampling of the actual agricultural (cropped) area, which means that the best classification results could be achieved with very pure pixel populations, but at the expense of excluding many pixels.

This effect is illustrated in Figure 6-14 for crop identification with 247.0 m pixel size, approximately the GSD of MODIS. In order to achieve high accuracies or to fulfil all defined criteria for crop identification, more and more (potentially mixed) pixels must be discarded. The sampling coverage SCF (see section 6.3.3) then can take on values below 1, which means that the surface sampled by the selected population of pixels is smaller than the reference target surface at the finest resolution ($v_0^2 * N_0$). The impact of selecting purer pixels on the sampling coverage (SCF) is much more pronounced over the heterogeneous landscape in KHO (Figure 6-14, top row), compared to KYZ with large blocks of more regular shaped fields (Figure 6-14, bottom row). As an example, 72 % (SCF=0.72) of the reference agricultural area ($v_0^2 * N_0$) can be identified while fulfilling all criteria defined in Table 6-3 (under the least restrictive thresholds, e.g. ACC > 0.75) in the KHO landscape, whilst in KYZ this can be achieved for 96 % (SCF=0.96) of the reference area. This difference becomes more evident when selecting only those pixels that fulfil all criteria according to the strictest thresholds (e.g. ACC > 0.85): in the KHO landscape only 16 % (SCF=0.16) of the sampled agricultural area is composed of the most suitable pixels. In the KYZ landscape 90 % (SCF=0.90) of the agricultural area crops can be identified in accordance with these strict requirements.
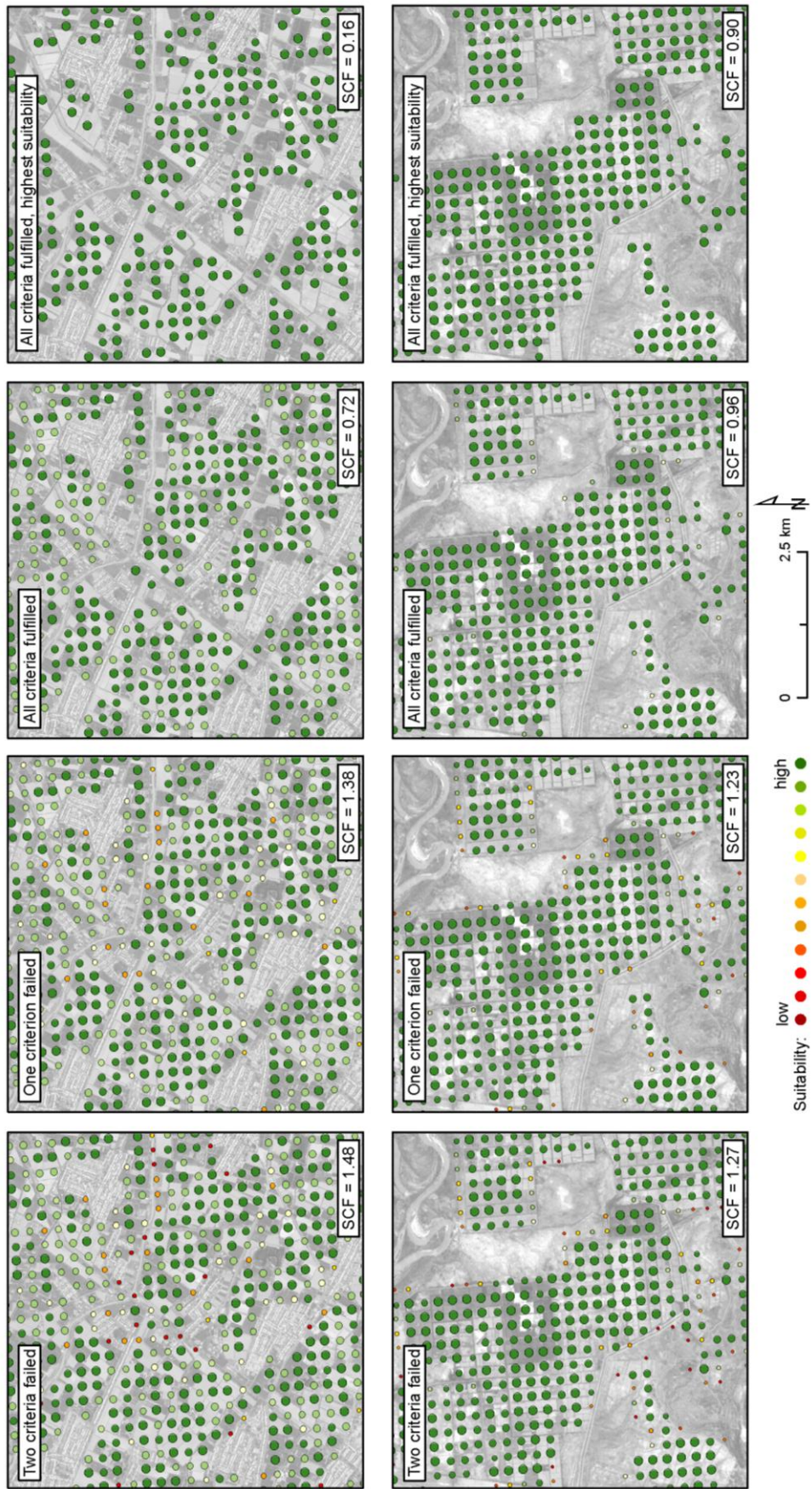
Figure 6-14: Impact of selecting purer pixels (that fulfil certain criteria for crop identification) on sampling coverage of agricultural fields (SCF) for a 6.5×6.5 km zone at 247.0 m pixel size in KHO (top) and KYZ (bottom). Dots represent the centres of coarse pixels. The size of the dots relate to the pixel purity, with larger dots being purer than smaller ones (only estimations above 0.3 are shown). SCF indicates the sampling coverage parameter, e.g. the ratio between the sampled surface at 247.0 m and the reference surface of agricultural fields at 6.5 m. The grey scale image in the background is the red channel of the corresponding RapidEye image with 6.5 m pixel size from 13-Aug 2009 (KHO) and 25-Jul 2011 (KYZ).

In a second investigation, the classified pixel populations of the two landscapes, KHO and KYZ, were analysed in more detail. Figure 6-15 and Figure 6-16 display those classified pixel populations that achieved highest values for ACC, at three different pixel sizes (97.5 m, 247.0 m, and 299.0 m). In KHO (Figure 6-15) the classification performances of RF and SVM started with values for ACC of approximately 0.81 ($v = 97.5\ m$), increased to a peak of approximately 0.91 ($v = 247.0\ m$), and then declined to about 0.86 ($v = 299.0\ m$). This figure reveals that the classification at 97.5 m was noisier then with coarser pixel sizes. At this scale there were apparent misclassifications between cotton and sorghum/maize within some of the fields (small image subsets and reference crop map in Figure 6-15). This type of class confusion has frequently been reported in the literature and might be explained with similarities in the NDVI signatures of these crops (Conrad et al., 2013). When comparing the performance of the two algorithms in KHO, RF seemed to be noisier than SVM even in homogeneous areas within larger fields (compare small image subsets of RF and SVM at 97.5 m in Figure 6-15). The most notable difference is that RF classified more pixels as sorghum/maize than the SVM. When coarser images were used for crop identification the result appears more homogeneous, but notably the borders between each field were more pixelated as compared to the highest resolution image. Classes covering smaller fractions of the landscape (e.g. sorghum/maize, winter wheat) almost diminished at coarser scales, e.g. they could not be detected anymore in the coarser pixels. The narrow range of suitable pixel sizes that can be used to detect these crops (Figure 6-11) also points to this.

In the KYZ landscape misclassifications within aggregated blocks of fields become apparent at 97.5 m (Figure 6-16). The small image subsets in Figure 6-16 illustrate the noisy classification (irrespective of the classifier algorithm used) within such a block of fields that should be classified as fallow, according to the reference crop mask. The reason for the misclassification could be the indistinct growing phenology of fields with alfalfa and fallow fields. Both kinds of fields are subjected to several, irregular cutting operations along the growing season (section 2.2), resulting in indistinct NDVI signatures. At coarser scales, this effect was markedly reduced and the classified images appear more homogeneous.

Further, selecting coarser observation supports in KYZ ($v = 247.0\ m$ and $v = 299.0\ m$) enhanced classification performances (higher ACC), and the amount of pixels classified as alfalfa-1y decreased. Critically, coarsening the observation support (pixel size) did not necessarily lead to increased classification accuracies for the alfalfa classes (alfalfa-1y and alfalfa-3y). The maximum possible value of CA_i for alfalfa-1y and alfalfa-3y (not shown in the figure) achieved at 97.5 m is 0.79 and 0.81, respectively and decreases henceforth to 0.68 and 0.76, respectively at the coarsest scale ($v = 299\ m$). This means that, despite the more homogeneous classification results and higher ACC at coarser scales (pushed by the increasing CA_i for rice and fallow fields which cover a major part of the landscape), the classes alfalfa-1y and alfalfa-3y are most effectively identified at smaller scales. This is also revealed in the pixel suitability maps in Figure 6-12.

This experiment demonstrated how classification accuracy can be affected by within field heterogeneity and by pixel size, respectively. As a consequence of using smaller pixels for classification and the resulting higher within-class variability, the entropy of the result became highest (and the ACC lowest) at the finest scale under investigation (see AQE and ACC values in Figure 6-15 and Figure 6-16). As was demonstrated by McCloy and Bøcher (2007) this effect can be reduced when selecting coarser observation supports, thereby reducing the within-class variability and achieving higher classification accuracies (Hsieh et al., 2001). This is reflected in the AQE values of the classification result, which were reduced by approximately 40 % when selecting pixel sizes coarser than 97.5 m in KHO (Figure 6-15).
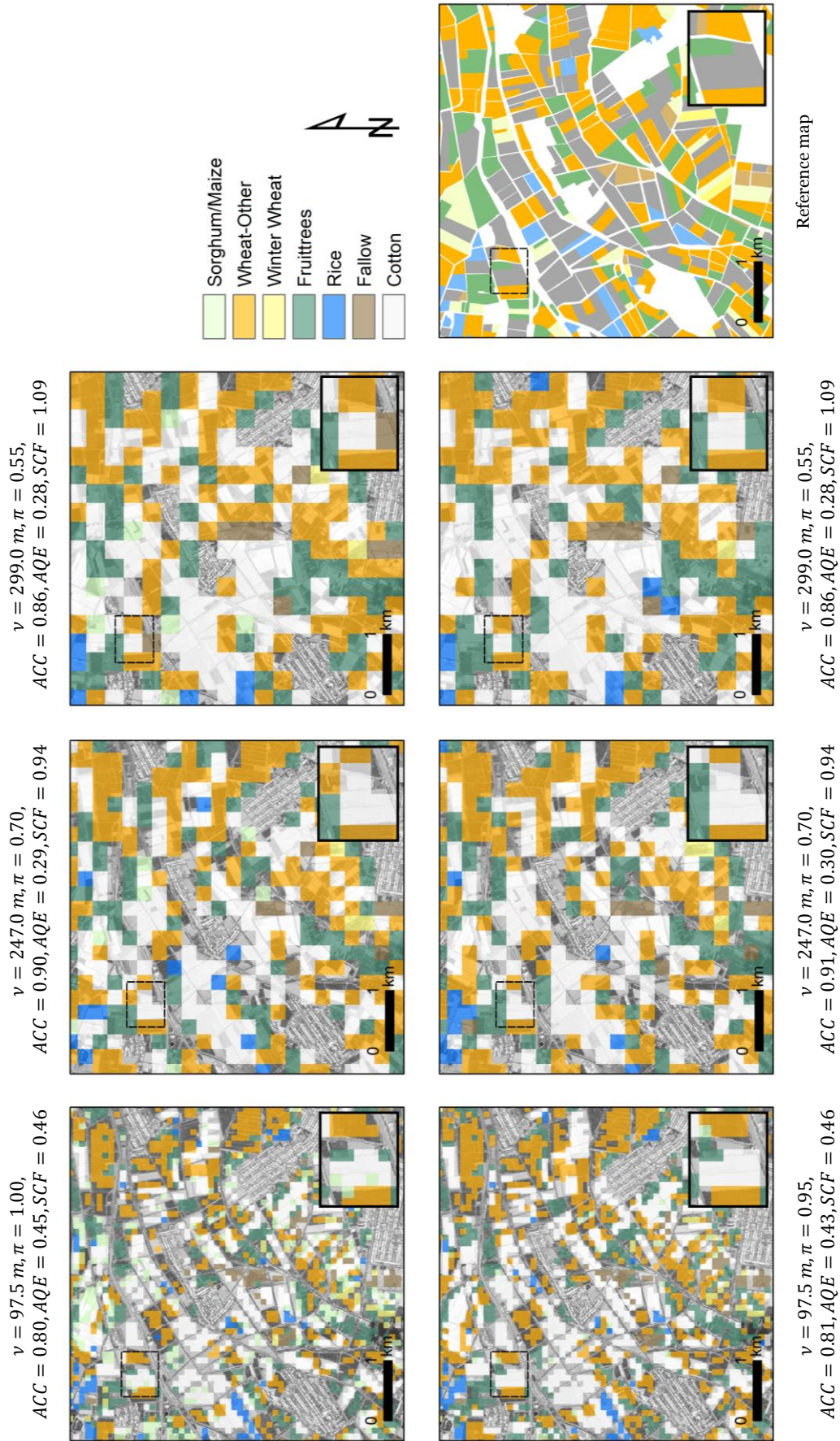
Figure 6-15: Result of image classification in KHO (in the year 2009) at three distinct scales (from left to right): 97.5 m, 247.0 m and 299.0 m. Shown are the classifications that yielded highest overall accuracies (ACC), using the RF (top row) and SVM (bottom row) as classifier. A subset of the landscape is shown in the small frame (bold contour), its footprint in the scene is indicated as a dashed square. The image backdrop displays the red band from a RapidEye image acquired on 31–May. The reference map is derived at the per-field basis with the CCS described in chapter 5.

$v = 97.5\,m, \pi = 1.00,$
$ACC = 0.80, AQE = 0.33, SCF = 0.92$

$v = 247.0\,m, \pi = 0.85,$
$ACC = 0.90, AQE = 0.28, SCF = 0.87$

$v = 299.0\,m, \pi = 0.85,$
$ACC = 0.91, AQE = 0.28, SCF = 0.87$

$v = 97.5\,m, \pi = 1.00,$
$ACC = 0.81, AQE = 0.31, SCF = 0.92$

$v = 247.0\,m, \pi = 0.85,$
$ACC = 0.91, AQE = 0.30, SCF = 0.87$

$v = 299.0\,m, \pi = 0.80,$
$ACC = 0.92, AQE = 0.23, SCF = 0.87$

Reference map
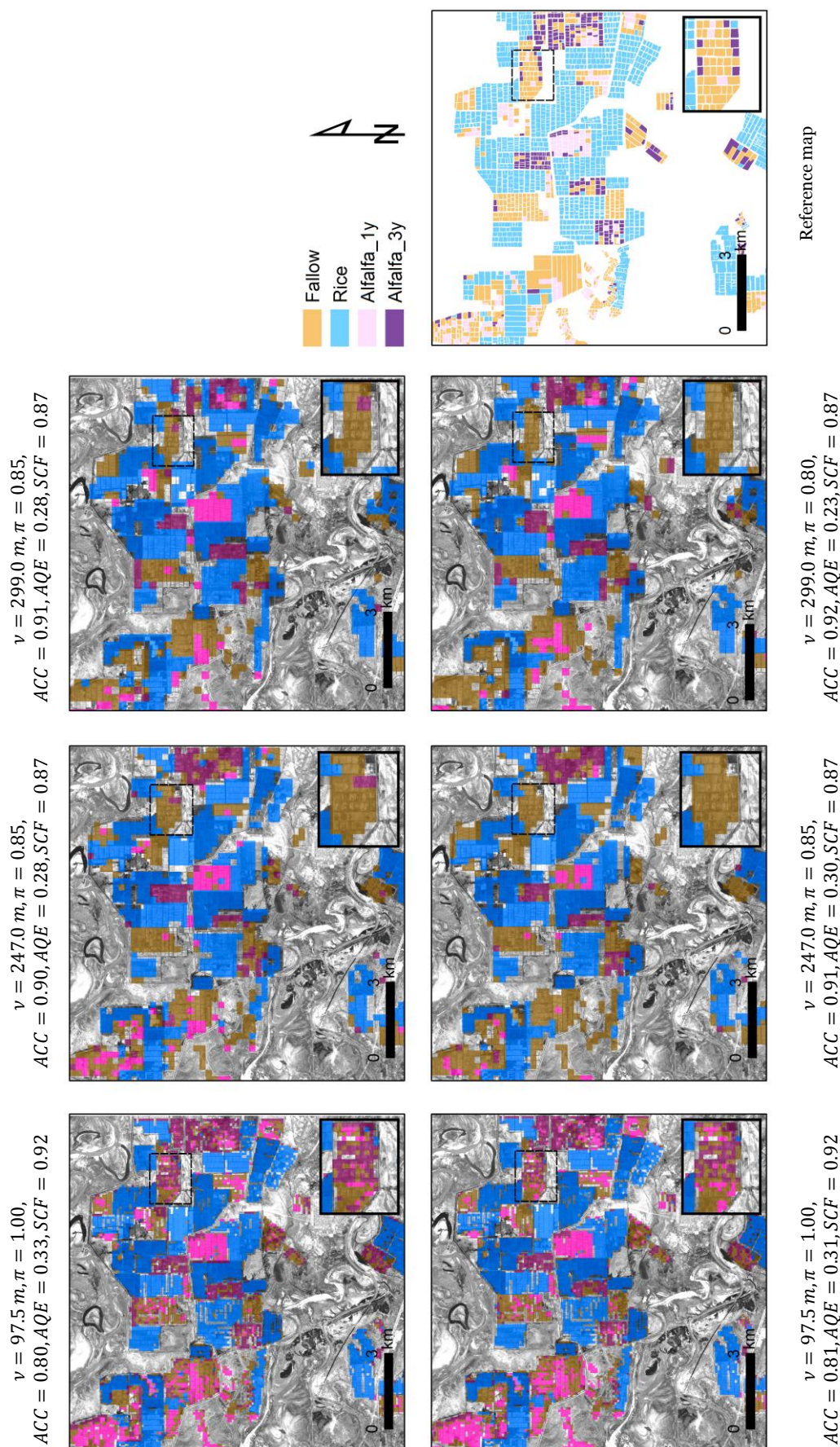
Figure 6-16: Result of image classification in KYZ (in the year 2011) at three distinct scales (from left to right): 97.5 m, 247.0 m and 299.0 m. Shown are the classifications that yielded highest overall accuracies (ACC), using the RF (top row) and SVM (bottom row) as classifier. A subset of the landscape is shown in the small frame (bold contour), its footprint in the scene is indicated as a dashed square. The image backdrop displays the red band from a RapidEye image acquired on 09–May. The reference map is derived at the per-field basis with the CCS described in chapter 5.

6.4.2. How does an unsupervised classifier algorithm influence the pixel population suitability?

Most studies on crop classification employ supervised classifier algorithms for crop mapping. However, reference samples may not be available in many regions around the world, not be of sufficient quality, or not be available for every crop type in question. The purpose of using the unsupervised K-means (Tso and Mather, 2009) here is to evaluate to what extent crop specific signals can be detected in NDVI signatures at different scales, and for different purities, respectively. In this regard, the unsupervised technique extracts temporal classes defined by their characteristics in the time series data (Ozdogan, 2010). Unsupervised classification uses clustering to identify natural groupings of pixels with similar NDVI properties, corresponding to key phenological stages (green-up, peak, senescence) in the NDVI time series (Chen et al., 1999). The purpose here is to examine which cluster(s) from this unsupervised classification reflect a specific crop type, and how accurate. The relationship of these clusters to the available reference data (crop masks) can be described by the parameters defined in section 6.3.4, and then to identify the cluster(s) that best represent a single crop type. Entropy (AQE) could not be used because this unsupervised algorithm provides no "ready-to-use" soft output from which entropy can be calculated. Having subjected the NDVI time series (eight acquisitions) to K-means clustering in each site, each cluster that contains at least 50 % pixels that belong to one specific class i was assigned to that class i.

Figure 6-17 illustrates the ranges of pixel sizes that are suitable for unsupervised crop signature identification. The most notable difference to the supervised approach (compare Table 6-5 – Table 6-8) is that fewer classes could effectively be identified, and that coarser values were required for v_{min} in KKP and KHO. Following the argumentation in the previous sections, coarser pixels are supposed to reduce within class variance, which facilitates the unsupervised crop identification. In KYZ only rice fields could be identified, most probably because of the indistinct NDVI profiles of alfalfa and fallow fields, which are characterized by heterogeneous patterns due to several irregularly scheduled cuttings throughout the season (section 2.2). In contrast, all crops (expect for wheat fields) present in the FER landscape could be identified with highest degree of suitability, e.g. all criteria defined in Table 6-3 could be fulfilled with accuracies of higher than 0.85 (CA$_i$), albeit the range of suitable pixel sizes differed from crop to crop. For instance, whilst winter wheat fields could only be detected using comparatively small pixels, cotton and wheat-other could be detected with pixel sizes of approximately 600 m. Again, relatively large field sizes and distinct NDVI signatures of relatively few crops most probably contribute to this.
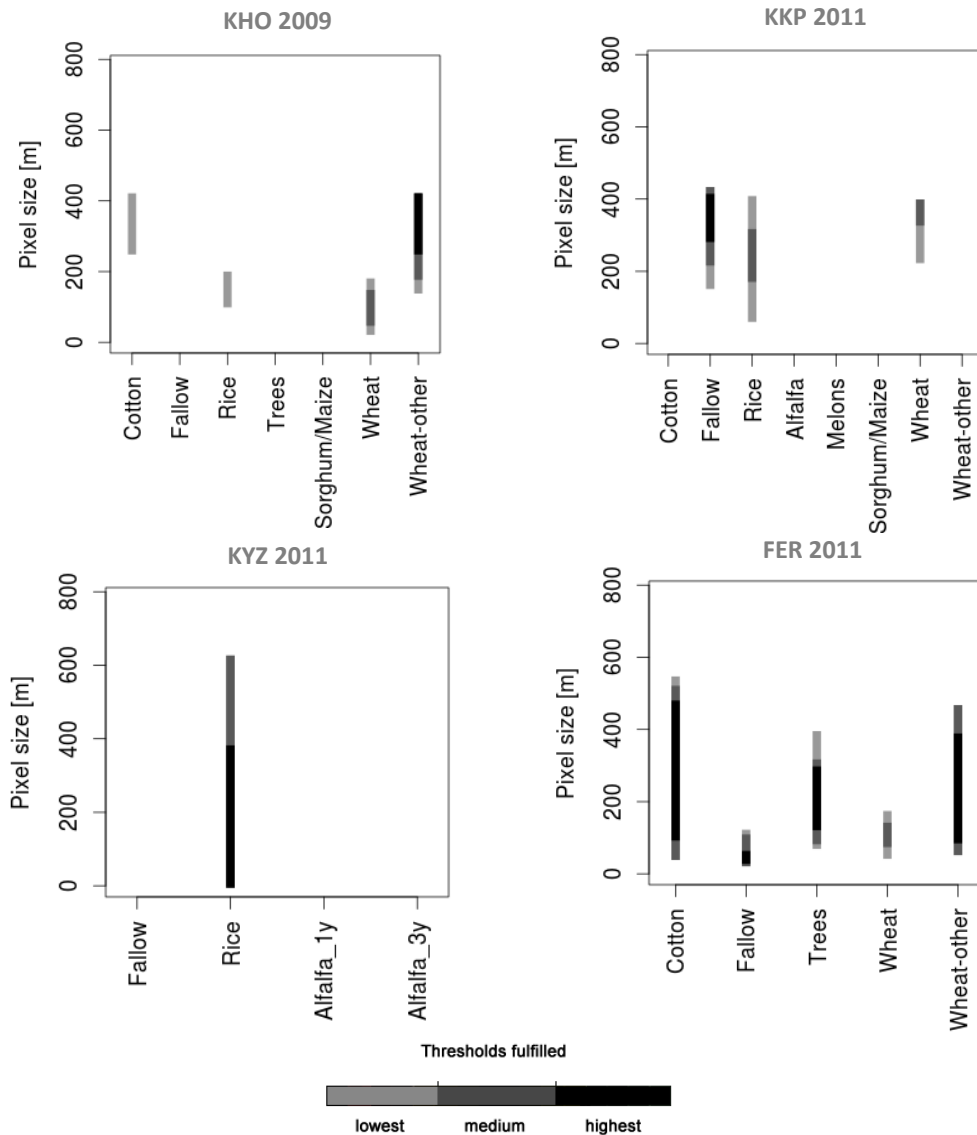
Figure 6-17: Ranges of suitable pixel sizes for different crop types using unsupervised K-means clustering. The length of the bars correspond to the range of suitable pixel sizes, shades of black indicate different levels of suitability, e.g. black means that the most stringent criteria defined in Table 6-3 are fulfilled, light grey that the least stringent criteria are fulfilled.

6.4.3. How do different input data sets influence pixel size and purity requirements?

Beside the application of different classifier algorithms, the input data was altered to assess its impact on the definition of pixel population suitability. This analysis can be reasoned with evidence from the literature that different land cover types may not only have large ranges of object sizes, but also that there may be significant differences in the local spatial variation observed at different wavelengths, which can affect the definition of suitable pixel sizes for specific applications like crop monitoring (Duveiller and Defourny, 2010). Hence, the $v - \pi$ surfaces were calculated with the same parameterization, but different input data sets (Table 6-3). These VIs were selected since they are frequently used to monitor vegetation seasonal behaviour. The choice of input features was further oriented toward transferability to sensors that (i) are equipped with similar spectral bandwidths and (ii) that sample the earth´s surface with GSD that are frequently being used in crop monitoring (Table 6-9). Critically, the bandwidths are not identical to RapidEye, but this should not affect the general

conclusions drawn from this framework, as it does not aim to simulate exactly the physical spatial response of a particular sensor. Since only few remote sensors have red edge bands, the experiments were repeated including and excluding red edge information. The features were calculated for eight acquisition dates, so the input data set size ranges from 8 to 72 features for the image classification (see Table 6-3), which is a reasonable number to achieve sufficient accuracies in multi-temporal crop classification (Carrão et al., 2008; Peña-Barragán et al., 2011; Rodriguez-Galiano et al., 2012a).

Table 6-9: Spectral bands and pixel sizes of optical sensors typically used in crop mapping. Spectral resolution is in [nm]. Only bands from other sensors that approximately correspond to the RapidEye bands are listed.

Bands	Landsat TM (30 m)	ASTER (15 m)	SPOT-5 (20 m)	PROBA-V (100 m)[24]	MODIS (250/500 m)[25]	Sentinel -3 (300 m)[26]	NPP-VIIRS (750 m)[27]
Blue	450–520	-	-	438–486	459–479	400–442	478–488
Green	520–600	520–600	500–590	-	545–565	490–560	545–565
Red	630–690	630–690	610–680	615–696	620–670	620–681	662–682
Red edge	-	-	-	-	-	681–708	-
NIR	760–900	760–860	780–890	772–914	841–876	761–900	739–885

Table 6-10 and Table 6-11, respectively reveal that the ranges of suitable pixel sizes for the RF as classifier algorithm varied among the different input data sets tested. As can be expected, the multivariate approach with more bands provided more confidence in classifying coarser pixels (higher accuracies) than selecting only one single band. Compared to the use of single bands (red or red edge), having extra spectral bands or VIs did improve the crop identification, both in terms of maximum achievable ACC and an enhanced range of suitable pixel sizes, respectively. The combined indices plus spectral bands resulted in smaller pixel sizes (v_{min}) compared to single spectral bands or indices alone. In both sites red edge indices enhanced the performance (higher accuracies) compared to feature sets without red edge information.

In KKP the use of red edge information generally resulted in smaller values for v_{min} and higher classification performance (ACC), which means that crops could be identified over a larger range of pixel sizes. This demonstrates that the additional features improve crop identification in smaller pixels, as a result of shifting the ACC bound towards smaller pixel sizes. In FER the inclusion or red edge features did slightly provide more confidence in classifying coarser pixels (higher accuracies), but there was no significant enhancement in the range of suitable pixel sizes (e.g. the value for v_{min} for most feature sets tested was 6.5 m). For pragmatic reasons (e.g. computational time) small input data sets that contain two vegetation indices can be preferred, as they suffice to identify crops within a wide range of pixel sizes and at a comparatively high level of accuracy, respectively. Yet, in KKP a larger feature set containing the combined bands plus VIs would be recommended.

When comparing both sites, the choice of input data had a much more pronounced impact on the definition of v_{min} and v_{max} in KKP than in FER. For instance, selecting all features ($f = 72$) instead of the red band ($f = 8$) resulted in a difference of 110.5 m for v_{max} in KKP, whilst this difference in FER was only 84.5 m.

---

[24]: Source: http://www.esa.int/Our_Activities/Technology/Proba_Missions/Overview2 (last accessed: 08-Sep 2013)

[25]: Source: http://modis.gsfc.nasa.gov/about/ (last accessed: 08-Sep 2013)

[26]: Source: http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-3 (last accessed: 08-Sep 2013)

[27]: Source: http://rammb.cira.colostate.edu/projects/npp/ (last accessed: 08-Sep 2013)

Table 6-10: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding pixel purity π for crop identification using RF, for different input data sets, in KKP for the year 2011. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. Maximum achievable ACC for three selected pixel sizes are given. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75). "Spectral bands" means all bands expect for red edge.

Input data sets tested	f	v_{min}			Maximum overall accuracy (ACC / π)			v_{max}		
		v	π	Crit.	97.5 m	247.5 m	299.0 m	v	π	Crit.
Red	8	182.0	0.85	ACC	0.70 / 1.00	0.73 / 0.75	0.76 / 0.65	312.0	0.55	ACC
Red edge	8	123.5	0.95	ACC	0.73 / 1.00	0.78 / 0.75	0.78 / 0.65	377.0	0.40	ACC
Spectral bands	32	39.0	0.95	ACC	0.78 / 1.00	0.86 / 0.75	0.84 / 0.65	377.0	0.40	ACC
Indices	16	71.5	1.00	AQE	0.76 / 1.00	0.82 / 0.75	0.82 / 0.65	396.5	0.35	ACC
Bands + Indices	48	32.5	0.95	ACC	0.81 / 1.00	0.83 / 0.75	0.83 / 0.65	422.5	0.35	ACC
RE Indices	16	65.0	0.95	AQE	0.78 / 1.00	0.84 / 0.75	0.84 / 0.65	409.5	0.35	ACC
Bands + RE Indices	48	26.0	0.95	ACC	0.81 / 1.00	0.84 / 0.75	0.84 / 0.65	422.5	0.35	ACC
All combined	72	19.5	0.95	ACC	0.82 / 1.00	0.89 / 0.75	0.89 / 0.65	422.5	0.35	ACC

Table 6-11: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding pixel purity π for crop identification using RF, for different input data sets, in FER for the year 2011. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. Maximum achievable ACC for three selected pixel sizes are given. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75). "Spectral bands" means all bands expect for red edge.

Input data sets tested	f	v_{min}			Maximum overall accuracy (ACC / π)			v_{max}		
		v	π	Crit.	97.5 m	247.5 m	299.0 m	v	π	Crit.
Red	8	6.5	0.15	-	0.86 / 1.00	0.91 / 1.00	0.92 / 0.90	572.0	0.40	ACC
Red edge	8	6.5	0.10	-	0.88 / 1.00	0.92 / 1.00	0.93 / 1.00	591.5	0.40	ACC
Spectral bands	32	6.5	0.15	-	0.89 / 1.00	0.92 / 1.00	0.91 / 0.80	631.5	0.40	ACC
Indices	16	32.5	0.90	ACC	0.87 / 1.00	0.92 / 1.00	0.92 / 0.85	598.5	0.40	ACC
Bands + Indices	48	6.5	0.15	-	0.88 / 1.00	0.92 / 1.00	0.95 / 1.00	604.5	0.35	ACC
RE Indices	16	6.5	0.15	-	0.87 / 1.00	0.92 / 1.00	0.93 / 0.90	650.0	0.40	ACC
Bands + RE Indices	48	6.5	0.15	-	0.88 / 0.90	0.93 / 1.00	0.94 / 0.95	656.5	0.40	ACC
All combined	72	6.5	0.15	-	0.89 / 1.00	0.93 / 0.95	0.94 / 1.00	656.5	0.40	ACC

6.4.4. How does pixel population suitability evolve along the season?

In the experiments described above the pixel size requirements were analysed by considering eight images distributed over the vegetation period within one year. However, some applications like early crop condition and warning alerts, or early crop area estimation might require to correctly identifying crops as early as possible in the season. This in turn requires accurate crop identification, which ideally should be available 1–2 months before harvest for early estimates, and 4–5 months before harvest for forecasting (Gallego et al., 2008). Given that most crops in Middle Asia are harvested between August and September (see section 2.2), the precondition for early estimation, according to Gallego et al. (2008), would be fulfilled if early crop identification can be provided as of June–July.

In order to test if the suitability of pixel populations changes along the season and to what extent, the experiments were repeated, based on the combined features plus bands, which yielded the best identification performance (see Table 6-10 and Table 6-11), but the observation length (number of images in the time series) was increased by incrementally adding images along the season, one-by-one. Then, for each incremental step, the pixel suitability for individual crops was calculated. The focus here is on two classes that can be found in KKP and FER, namely cotton and winter wheat. These two sites were selected for this experiment, because RapidEye images are available earliest in the season

(beginning of April, see Figure 6-1), which allows for a finer assessment of early estimation in the early phase of the growing season.

Figure 6-18 demonstrates for these two classes how adding images enhances the suitability of the pixel populations in $\nu - \pi$ space for crop identification. In KKP the identification of cotton was not possible till 07–Jun (DOY: 167). As of 07–Jun cotton could be identified but this was restricted to a rather small range of pixel sizes ($\nu_{min} = 162.5$, $\nu_{max} = 266.5$ m). Adding images till 14–Jul (DOY: 195) enabled the use of pixels with a wider range of resolutions ($\nu_{min} = 45.5$, $\nu_{max} = 383.5$ m) and purities, respectively. Adding an image after 14–Jul had no significant effect on the suitability of the pixel populations, albeit using very small (ν_{min}= 6.5 m) and coarse ($\nu_{max} = 429.0$ m) pixel sizes became possible. Winter wheat fields in KKP could be identified as of 09–May (DOY: 94), starting with pixel sizes ranging from ν_{min}= 117 m to $\nu_{max} = 247.0$ m, and adding more and more images improved the values for ν_{min}, which were shifted towards 13.0 m, whilst ν_{max} was further shifted towards 429.0 m.

In the FER landscape winter wheat fields could be identified with a wider range of pixel sizes early in the season, as compared with KKP. One reason could be the higher contrast between the target crop and its surroundings on the FER landscape in April (when summer crops were not yet sown but winter wheat stems were already elongated and fully covered the fields) than on the KKP landscape (when winter wheat had not already grown significantly and bare soil that covered the latent summer seeds had comparable reflectance to winter wheat fields). This difference can be explained by the earlier water availability and onset of the vegetation period, respectively in FER than in the downstream regions of the Amu Darya where the KKP site is located (Bohovic et al., 2011). Further, spectral discrimination might be hampered by the fact that signal response to vegetation was very low due to the little amount of biomass on the fields in KKP at the time when winter wheat was already fully grown in FER.

Although crop identification is in the fore of this study, an example for the evolution of pixel suitability for the classification of all crops is also given in Figure 6-19 and Figure 6-20. Again, overall crop identification was possible earlier in the season in FER. Already at the start of the season crops could be identified with relatively low error rates and entropy values (both smaller than 0.25, see Figure 6-20). Crops could be effectively identified as of April, most probable because at that time winter wheat fields were in the phase of stem elongation, and could be distinguished from fallow fields and fruit trees, respectively. As of 16–Jun (DOY: 188), suitable pixel populations for crop identification were available in KKP, with error rates below 0.25. Again, the reason for this difference in timing could be the earlier start of the growing period in FER (Bohovic et al., 2011). Along the season, the suitability did not evolve significantly henceforth.

In both sites it was observed that ν_{min} became the smaller the more images were added: from 78 m (at DOY 188) to 26.0 m (DOY 282) in KKP, and from 182.5 m (at DOY 98) to 6.5 m (DOY 270) in FER (Figure 6-19). Likewise the pixel size ν that is needed to achieve maximum possible ACC decreases, from 318.5 m to 260.0 m in KKP, and from 351.0 to 260.0 m in FER (Figure 6-20). As soon as the contrast between summer crops becomes less intense in the summer period, when crops have already grown significantly and have comparable reflectance, pixel sizes must be smaller to identify all crops because of the heterogeneous landscape pattern and to achieve a decent standard of classification accuracy.

In both figures (Figure 6-18 and Figure 6-19) the space between the upper and lower dotted lines demarcate those pixel populations whose SCF did not deviate more than ±10 % from the reference agricultural (cropped) area at the finest scale ($\nu_0^2 * N_0$, see section 6.3.3). As can be seen for winter-wheat in KKP (Figure 6-18), suitable pixel populations (green colours) that were selected before DOY 129 under-sample the true agricultural (cropped) area because they are located above the upper dashed line, which means that too many pixels were excluded due to the high pixel purities required, hence the sampled surface of the selected pixel populations did not fully cover the reference agricultural area.
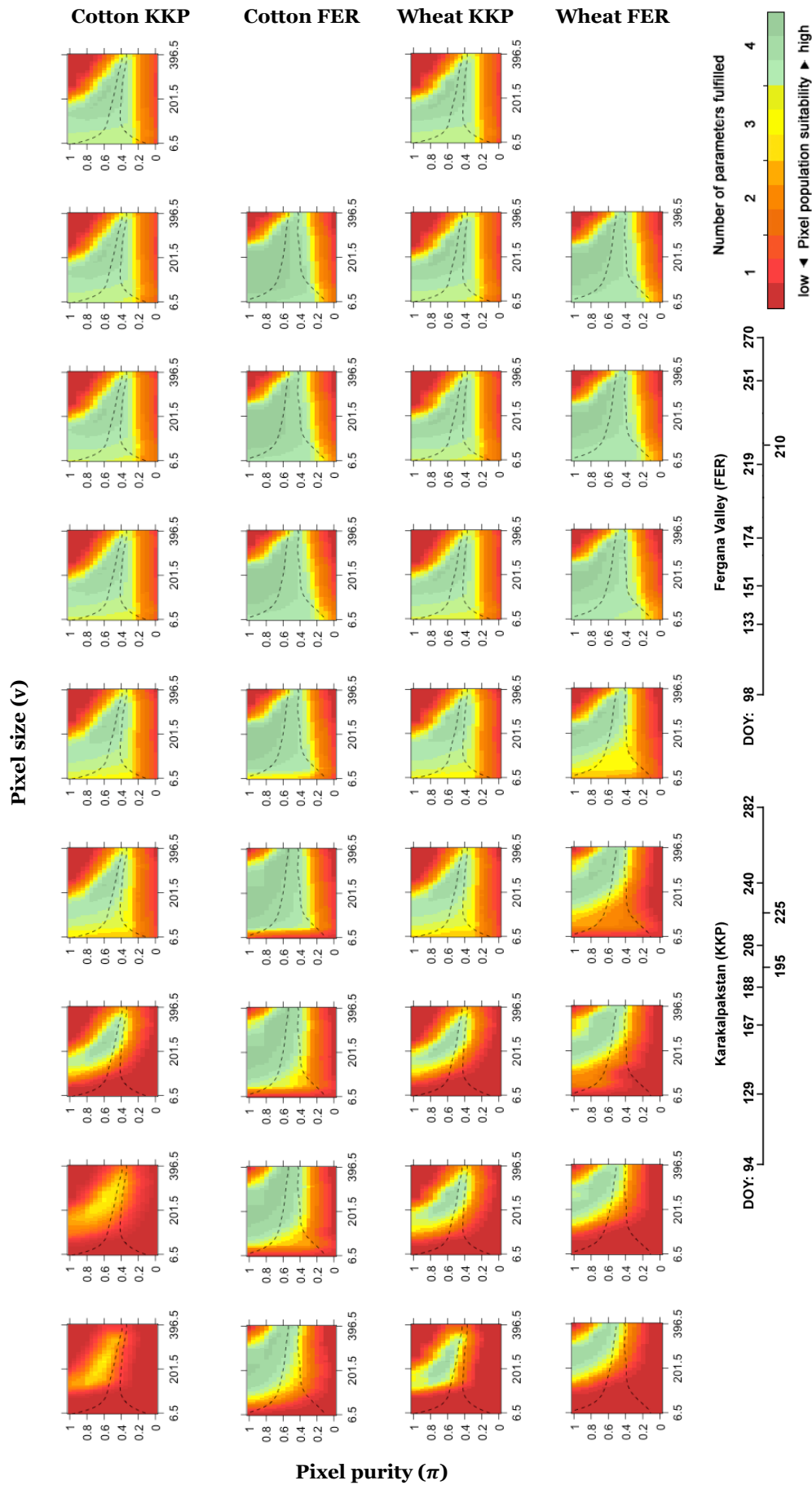
Figure 6-18: Evolution of suitable pixel sizes for cotton and winter wheat in KKP in 2011 and FER in 2011, respectively using RF as classifier algorithm. Images were incrementally added, one by one, along the season. Dashed lines indicate where over- and under sampling (lower and upper line, respectively) exceed 10 %. For better readability the suitability maps show the results for pixel sizes between 6.5 m and 396.5 m.
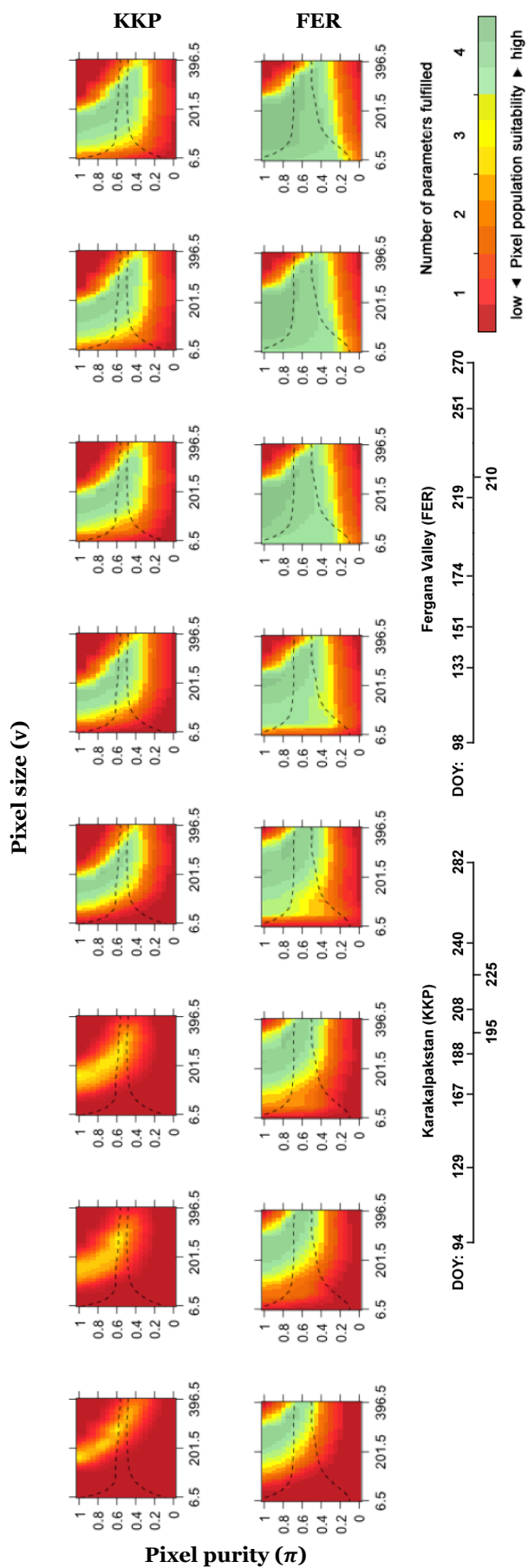
Figure 6-19: Evolution of suitable pixel sizes for overall crop classification in KKP in 2011 and FER in 2011, respectively using RF as classifier algorithm. Images were incrementally added, one by one, along the season. Dashed lines indicate where over- and under sampling (lower and upper line, respectively) exceed 10 %. For better readability the suitability maps show the results for pixel sizes between 6.5 m and 396.5 m.
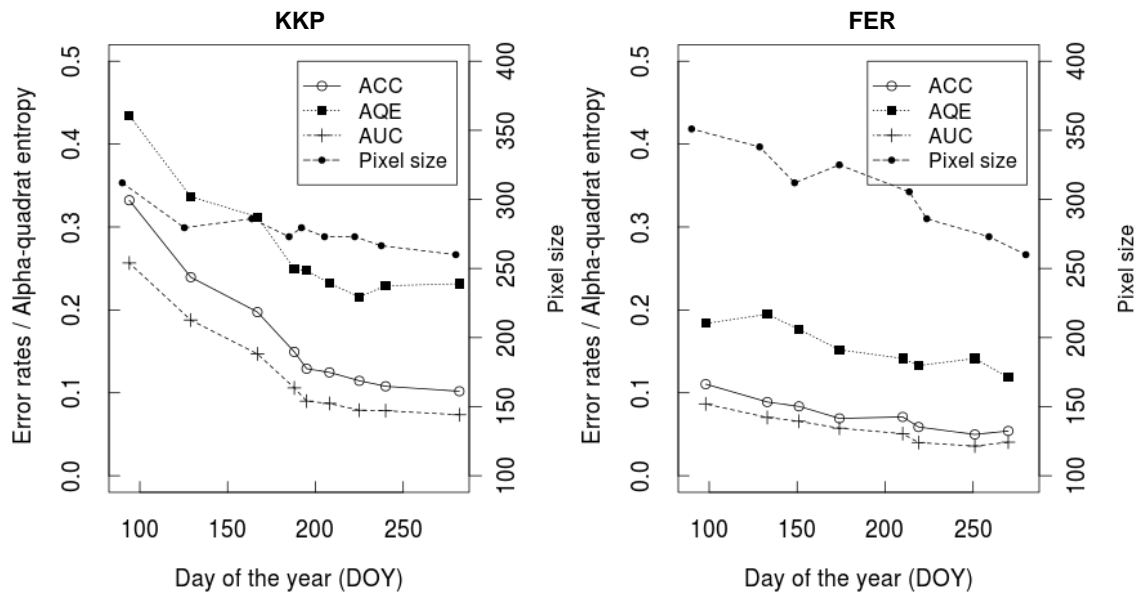
Figure 6-20: Evolution of crop identification performance for all crops in KKP and FER, respectively using RF as classifier. The maximum achievable accuracies, here expressed as 1-(ACC/100) and 1-(AUC/100), respectively for better comparison with classification entropies (AQE) are given. "Pixel size" refers to the ground sampling distance at which maximum overall accuracy (ACC) was achieved.

6.4.5. Can the defined pixel size requirements be transferred to another year?

For operational crop mapping applications, it is of interest if the requirements for crop identification are comparable from year to year. Hence, the experiments were repeated on data sets from another year in two sites, KHO (2010) and FER (2012)[28], in order to validate the consistency and transferability of the pixel size requirements among the two years. Each experiment was performed using the corresponding parameterization and following the same method as described above, using RF as classifier algorithm.

The resulting pixel size requirements for crop identification are summarized in Table 6-12 and Table 6-13. Figure 6-21 shows the ranges of suitable pixel sizes for each crop in KHO and FER, respectively in two consecutive years. In general, the requirements for crop classification in FER did not significantly change (Table 6-12 and Figure 6-21), specifically v_{min} was 6.5 m for almost all classes over the two years. However, v_{max} was coarser in 2011 than in 2012 for most classes. The identification of fallow fields in 2012 was limited to a comparatively small range of pixel sizes (v_{min} = 6.5 m, v_{max} = 202.0 m), compared to 2011 (v_{min} = 6.5 m, v_{max} = 260.5 m). One explanation could be the differences in the cover fraction C_f of fallow fields, which was almost four times higher in 2011 (Table 6-1) and found on larger fields (on average). This means that it is easier to have coarser pixels fall within target fields and thus conferring higher acceptable pixel sizes for the crop identification. The same could be found for winter wheat (v_{max} in 2011 was 331.5 m, v_{max} in 2012 was 364.0 m), which covered a larger fraction of the landscape in 2012 (C_f = 0.09) than in 2011 (C_f = 0.05). The cover fractions (C_f) of wheat-other fields decreased in 2012 (C_f = 0.27) compared to 2011 (C_f = 0.32), which was reflected in a change of v_{max} for that crop type from 656.5 m to 539.5 m. In general higher classification performances could be achieved over a wider range of pixel sizes in 2011, indicated by the length of black bars in Figure 6-21. In summary these

---

[28]: RapidEye images were made available through the Khorezm and CAWA projects in 2010 (KHO) and 2012 (FER). No RapidEye data was available for another year in KKP or KYZ.

findings indicate that in FER some sensors, like MODIS or Sentinel-3, could be suitable for crop identification in an operational context. Few classes like winter wheat or fallow fields would require finer resolutions sensor, e.g. with 100 m pixel size, and their identification was affected by varying v_{max} over the two years.

Table 6-12: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding purities π for crop identification using RF, at the per-class basis, in FER for the year 2012 and 2011. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75).

Year:	2012						2011					
Site	v_{min}			v_{max}			v_{min}			v_{max}		
	v	π	Crit.	v	π	Crit.	v	π	Crit.	v	π	Crit.
Cotton	6.5	0.15	-	591.5	0.25	CA$_i$	6.5	0.15	-	663.0	0.30	AUC$_i$
Fallow	6.5	0.15	-	202.0	0.75	CA$_i$	6.5	0.15	-	260.5	0.70	AUC$_i$
Fruit trees	6.5	0.35	-	468.0	0.85	CA$_i$	6.5	0.15	-	494.0	0.85	CA$_i$
Winter wheat	6.5	0.20	-	364.0	0.45	AQE$_i$	6.5	0.20	-	331.5	0.45	AQE$_i$
Wheat-other	19.5	0.90	CA$_i$	539.5	0.40	AUC$_i$	6.5	0.15	-	656.5	0.45	AUC$_i$

Table 6-13: Ranges of acceptable pixel sizes (limited by v_{min} and v_{max}) and corresponding purities for crop identification using RF, at the per-class basis, in KHO for the year 2010 and 2009. The column labelled "Crit." indicates which parameter constrained the spatial sampling unit. v_{min} and v_{max} are defined by the intersections of the experimental boundaries of the parameters in $v - \pi$ space, under the least restrictive thresholds (e.g. ACC = 0.75). Asterisk indicates that for this site the experiments were halted at 442.0 m (2010) and 409.5 m (2009) because for one class the number of available training pixels dropped below 20, the minimum requirement to perform the experiments (section 6.3.2).

Year:	2010						2009					
Site	v_{min}			v_{max}			v_{min}			v_{max}		
	v	π	Crit.	v	π	Crit.	v	π	Crit.	v	π	Crit.
Cotton	26.0	0.85	CA$_i$	442.0	0.45	*	104.0	0.75	AQE$_i$	409.5	0.55	*
Fallow	6.5	0.95	-	338.0	0.60	AUC$_i$	13.0	0.90	AUC$_i$	364.5	0.40	CA$_i$
Rice	6.5	0.15	-	442.0	0.45	*	6.5	0.20	-	409.5	0.55	*
Fruit trees	78.0	1.00	CA$_i$	390.0	0.65	CA$_i$	110.0	1.00	AQE$_i$	409.5	0.65	*
Sorghum/maize	26.0	1.00	AQE$_i$	84.5	0.70	CA$_i$	-	-	-	-	-	-
Winter wheat	6.5	0.15	-	123.5	0.75	CA$_i$	6.5	0.15	-	227.5	0.80	CA$_i$
Wheat-other	39.0	0.90	CA$_i$	442.0	0.45	*	117.0	0.75	AQE$_i$	409.5	0.45	*

In KHO the situation was different. The requirements for identifying some of the classes changed considerably: As an example, v_{max} for winter wheat was 123.5 m in 2010 (compared to 227.5 m in 2009) and the values of v_{min} for cotton decreased from 104 m to 26 m. Overall, there is a tendency that v_{min} of most crops is coarser in 2009, compared to 2010, whilst v_{max} was not defined in the experiments for some crops, e.g. cotton and wheat-other, because the experiments were halted in both years due to an insufficient number of pixels for sorghum/maize (< 20). Interestingly, the cover fraction of wheat-other in 2009 (C_f = 0.27) was lower compared to 2010 (C_f = 0.30), but there was no such obvious difference in the values for v_{max} of this class, as far as this can be judged because the experiments were halted. These findings indicate that crop identification in an operational (multi-year) context could be limited for some crops: e.g. winter wheat or sorghum/maize would require finer resolution sensors, e.g. with 100 m pixel size, and are also stronger affected by the varying v_{max} over the years.
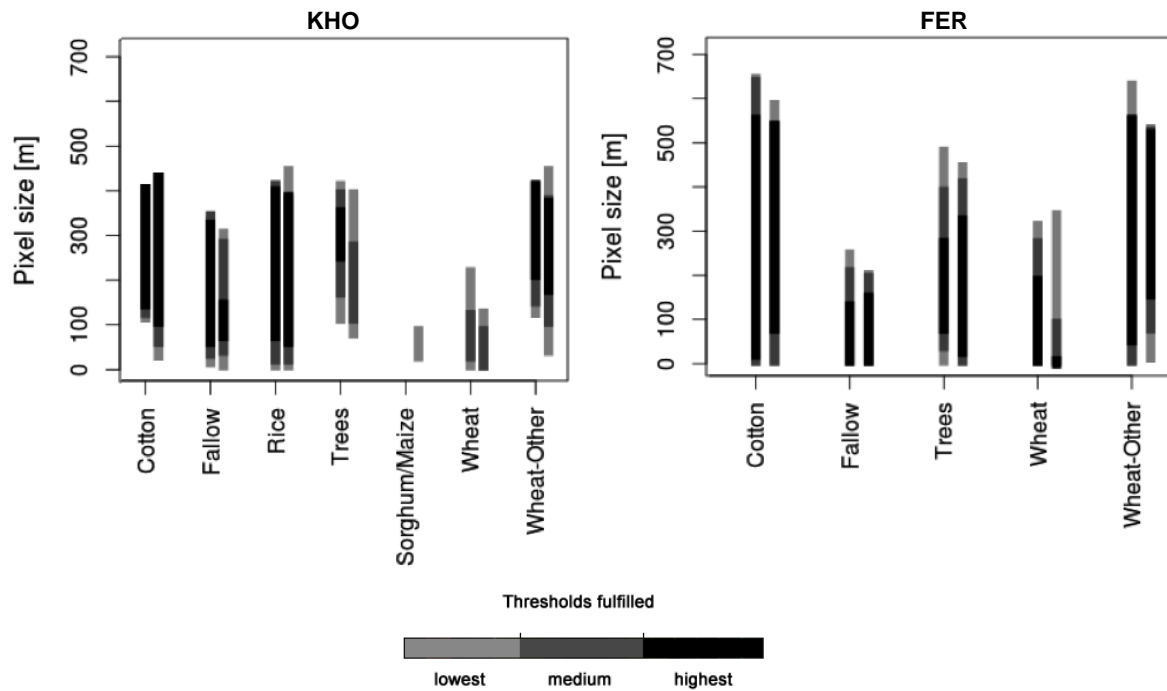
Figure 6-21: Ranges of suitable pixel sizes for different crop types in KHO (left side bars: 2009, right side bars: 2010) and FER (left side bars: 2011, right side bars: 2012). The length of the bars correspond to the range of suitable pixel sizes, shades of black indicate different levels of suitability, e.g. black means that the most stringent criteria defined in Table 6-3 are fulfilled, light grey that the least stringent criteria are fulfilled.

Another possible explanation for the pronounced difference in pixel size requirements in KHO could be the sharply reduced irrigation water supply in 2009, compared to 2010. To illustrate this: the average water intake from the Amu Darya into the KHO irrigation system (from the Tuyamujun reservoir) is 3,859 m³ (averaged over the period 2000–2011)[29]. However, in 2009 water intake was reduced to 3,660 m³. In 2010 the water intake was above the 11-year average (4,902 m³). It was illustrated in section 4.3.7 how in water sparse areas in Middle Asia a flattening of NDVI profiles, caused by reduced biomass production, can led to increasing classification entropy. It can be seen that in 2009, compared to 2010, v_{min} of many crop classes is dictated by the AQE$_i$ parameter (Table 6-13). This means that classification uncertainty due to increasing within pixel heterogeneity predominates, which ultimately leads to increased classification entropy. Reduced water supply could be causative for the occurrence of bare or salty patches within agricultural fields, which would enhance class confusion when smaller pixels fall within such patches. In this regard using coarser pixel sizes would reduce some of the variance within the pixels. As can be seen, v_{min} was coarser in 2009 compared to 2010 (coarser than 100 m for half of the classes). In this situation it is better to have coarser pixels, thereby reducing this variance and counterbalancing the effect of pure-pixel heterogeneity within smaller pixels (Hsieh et al., 2001). In the FER landscape no such pronounced differenced were observed between the two years. Water intake from the Toktogul reservoir into Fergana Valley was slightly above the 11-year average (3,940 Mio m³) in 2011 and 2012, respectively, but the difference between the two years was negligible: 4,216 and 4,476 Mio m³.

---

[29]: Water intake data from www.cawater-info.net, Russian version (last access: 11-Aug 2013).

6.4.6. Impact of reference sample data characteristics

The characteristics of the reference sample data are crucial for the accuracy statistics. By design, the sampling scheme used in this study (hereafter called mixed pixel training strategy) considers the effect of pixel purity in all stages of classification (training and testing). In general, remote sensing studies preferably employ designs that avoid undesirable (potentially mixed) pixels in the training and testing stage, respectively, e.g. prefer pixels that fall completely within a field or are homogeneous with respect to the underlying land cover (Conrad et al., 2011a). Yet, this can result in skewed accuracy estimations and the accommodation of mixed pixels seems worthwhile because it enables a more appropriate representation of the classes as well as a more meaningful evaluation of classification performance (Foody, 1996c; Shao and Lunetta, 2012). To illustrate this, an additional experiment was performed in the KKP landscape by only selecting relatively pure training pixels (e.g. $\pi \geq 0.60$), but still assessing mixed test pixels, hereafter called the pure pixel training strategy. Then the difference between the ACC of the pure and the mixed pixel training strategies, respectively was plotted in $\nu - \pi$ space.

Figure 6-22 illustrates that the difference in ACC between these two strategies is most pronounced when testing mixed pixels (reddish colours in Figure 6-22). This means that when classifying mixed pixels, an increase of ACC could be realized when mixed pixels were included in the training, as compared with the pure pixel training strategy. This means that when mixed pixels are abundant, their accommodation in the training stage enables a more appropriate representation of the classes and a fuller use of the data, as was already noted by Foody (1996c). Yet, a potential disadvantage when selecting only pure pixels for training was that for pixel sizes coarser than $286.0\,\text{m}$ not enough pixels were found for the classifier algorithm training, and the experiments were halted. When classifying purer pixels, then the ACC of both strategies seemed to be more equalized (green colours in Figure 6-22). This figure also reveals that the quality of the reference pixels and the selection of the sampling design (here: pure vs. mixed) affect the classification results more significantly than the choice of the classifier algorithm (section 6.4.1).
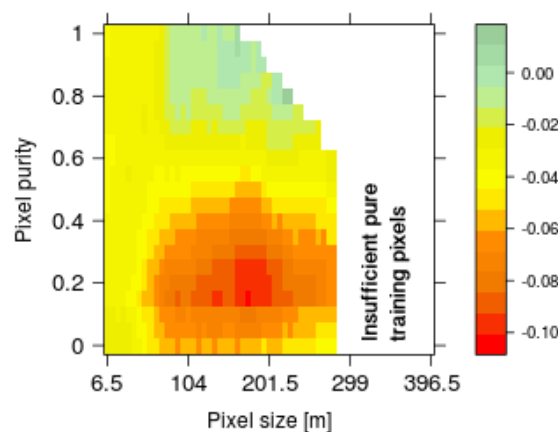

Figure 6-22: Difference of error rates ($1 - ACC/100$) between RF models based on pure pixel training strategy (e.g. only sleeting pixels for training with purities of $\pi \geq 0.60$) and mixed pixel training strategy. Example is shown for the KKP landscape. Orange and yellow colours indicate, that the pure pixel training strategy is less accurate than the mixed pixel training strategy.

6.5. Discussion

Landscape and neighbourhood heterogeneity, and the size of surface features are important factors determining classification accuracy (van Oort et al., 2004; Smith et al., 2003) and the pixel size requirements in agricultural monitoring (Duveiller and Defourny, 2010). As could be expected, when crops are grown on larger and more regular fields (such as FER and KHO), or when the cover fraction is high (such as cotton or wheat-other fields in KHO), pixel sizes could be coarser for crop identification. However, the crop classes display marked individuality in the specific nature of these relationships, and the landscape heterogeneity with respect to the spatial pattern also influences the choice of the pixel sizes. For instance, the field sizes in KYZ and KKP are comparable (Table 6-1), but the former's fields are more regular in shape, less variable in size, and the same crops are found on blocks of fields that together can aggregate to more than 100 ha in size (see section 2.2). Over such a landscape, it is easier to have coarser pixels fall within target fields and thus conferring higher acceptable pixel sizes (v_{max}) for crop identification, resulting in coarser values for v_{max} in KYZ than in FER. Crops covering small parts of the landscape like sorghum/maize or winter wheat in KHO (C_f = 0.01) cannot be detected within coarse pixels because the target signal strength is low.

Some classes demonstrated greater sensitivity to landscape heterogeneity (as quantified by cover fraction C_f), others to field size, with still others strongly affected by both variables. For instance, winter wheat in KHO, albeit grown on comparatively large fields (> 5 ha), could not be identified using pixel sizes coarser than 195 m (see section 6.4.1). Winter wheat fields covered only a small part of the scene (C_f = 0.01) and $\hat{\gamma}_r$ of winter wheat fields was very low, which indicates that these fields were dispersed over the landscape in 2009. The method reacted to the specific landscape pattern situations in the four study sites, which were characterized by field sizes and cover fractions (Table 6-1):

> KHO and FER: Similar landscapes but different number of crops with relatively large field sizes (medians: 3.21 ha and 5.47 ha, respectively). The SV of the field masks was shown to have a similar behaviour (section 2.3), but the trajectories of $\hat{\gamma}_r$ and the absolute number of different crop types differed. Crop mapping using coarser resolution imagery was easier, because many pixels fall within the relatively large fields. However, higher pixel purities are required in KHO, most probably because of the smaller field sizes and higher number of crop types in KHO compared to FER (Table 6-1), which results in a more heterogeneous landscape. Fallow fields in FER, albeit found on relatively large fields (4.68 ha), covered only 4 % of the landscape, resulting in smaller values for v_{max} (260.0 m) in that landscape, compared to the other crops.

> KKP: Complex landscape with relatively small field sizes (median: 1.73 ha). Crop identification using coarse resolution imagery was hampered in KKP for some crops because of the relatively small field sizes and C_f values (e.g. melons, alfalfa-1y, wheat-other) and the fragmented landscape pattern. Compared to KHO the landscape in KKP is more complex, with a higher proportion of non-agricultural land use (C_f in KKP: 0.32, C_f in KHO: 0.59). Still, v_{max} for some crops was relatively coarse (e.g. rice, winter wheat, cotton). The reason is probably due to the higher contrast between the target crop and its surroundings on the KKP landscape: the signal response is expected to be comparably high because of the presence of bare and sparsely vegetated land uses (e.g. open shrublands fields) in the surrounding. The net effect is that for one crop, the pixel purity requirement at a given pixel size tended to be lower in KKP than in KHO for some classes (when looking at the year 2010 in KHO).

> KYZ and FER: Similar crop diversity, but different landscape patterns. For these sites the pixel size for most crops could be coarse, in FER because of the relatively large field sizes (median: 5.47 ha). Although field sizes in FER are much larger than in KYZ (median: 2.14 ha), the spatial clustering of fields with the same crop type in KYZ allows for using coarser pixel sizes. To quantify the spatial clustering of fields, Moran´s I index (Moran, 1950) was calculated[30] for the reference crop maps, resulting in values of 0.248 (KYZ), which indicates that fields with the same crop tended to be more clustered, and 0.042 (FER), which indicates that the spatial pattern of fields with the same crop was more dispersed in FER. Fields in KYZ were characterized by a much higher degree of spatial clustering than in FER.

Over the KHO landscape it was illustrated how the definitions of suitable pixel sizes changed considerably over the years. There is evidence that in this landscape the framework reacted to environmental factors like differing magnitudes of irrigation water availability, in particular the definition of v_{min} seemed to be affected. Further, the method reacted to changes in the cover fractions or average field sizes of certain crops over the years, which might be related to management decisions. For operational mapping efforts the similarity of classification performance using pure or mixed pixels is promising. In particular at coarser scales, training-pixels can be selected with some flexibility regarding the purity and there is even an increase in classifier performance when using coarser pixels. This allows for a relaxation of the sampling scheme, because the spectral variability of individual crop types is better reflected, e.g. when pure reference pixels are not sufficiently available, or if pixels were to be selected on-screen by visual inspection of NDVI profiles, which was also concluded by Shao and Lunetta (2012) in their study on the effect of mixed pixels in classification accuracy. In this regard the framework proposed here demonstrated a detailed evaluation of the impact of pixel heterogeneity in all stages of image classification. This allowed for a more rigorous evaluation of classification accuracy, thereby extending other studies by the inclusion of the multi-scale assessment and for a continuum of purities instead of dichotomous approaches like "pure vs. mixed" (Foody, 1996c; Hsieh et al., 2001; Shao and Lunetta, 2012).

Over the landscapes studied in this paper, the parameters were set with the same values in order to illustrate how the method responds to different spatial patterns. However, the proposed framework is flexible to be adapted to the end-user´s needs. The selected settings can be adopted separately for different agricultural sites, for example for some applications higher thresholds for CA_i might be needed or some relaxation of the thresholds might be required over more fragmented landscapes. Further, the experiments could be repeated for more years of observation, providing that suitable high-resolution imagery is available, in order to trace the resolution requirements over the longer run and to better understand the impact of environmental factors (e.g. water availability) on their definition. A more detailed exercise could also be considered by applying the methodology to a wider range of agricultural landscapes. Possible candidate sites might be found in the U.S., with relatively large field sizes (Wardlow and Egbert, 2008), or sites in China with field sizes reported smaller than 2 ha (Duveiller and Defourny, 2010). Further, a fine diagnostic tailored to a specific instrument could be envisaged if the specific sensor spatial response can be reasonably approximated (Duveiller et al., 2011). The notion of v_{min} might not be appropriate in every case, e.g. when very fine pixel sizes can be used within an object-based classification context, which will most likely outperform the pixel-based approach, e.g. by decreasing the effect of high within class variability (see section 6.3.3).

---

[30]: Moran´s I was calculated with the software ArcGIS (ArcMap™ 10.1), using the Spatial Statistics Tools from ESRI ® (ESRI, 2013). Moran's I values near +1.0 indicate spatial clustering while values near −1.0 indicate dispersion.

Finally it should be noted that, although it was demonstrated that crops can be effectively identified in pixels with low purity, some application might require higher thresholds to get purer pixels, e.g. in crop growth monitoring (Duveiller and Defourny, 2010), which would further limit the range of suitable pixel sizes. In case the classified pixels are to be used for area estimation directly via pixel counting, a fine-tuning could be putting more emphasis on the recall relative to precision in the F_β-measure, which defines the CA$_i$ parameter, by selecting higher values for the β parameter in equation 6-8.

Pixel purity considers the degree of pixel mixing with regard to the proportion of signal from a specific target structure encoded in a pixel, which is primarily defined by the optical properties of the sensing system. Yet, even if a selected pixel is pure in this sense the crop signal encoded in the pixel still might be characterized by heterogeneous growing conditions. Hence, it might be discarded if the selected pixels are to be valuable for monitoring and if the signal should be from a surface with a more homogeneous behaviour along the growing season (e.g. to have distinctive temporal NDVI signature). Hence, the behaviour of the alpha quadratic entropy $H_\propto(x)$ (which defines the AQE criterion) might be relaxed by selecting lower values for the α parameter, so that the uncertainty measure can become less sensitive to changes in the elements within $pr(x)$, thereby adjusting its sensibility to surface heterogeneity and intra-class variability in the pixel populations.

6.6. Conclusions

The overall objective of this research was to analyse the EO data requirements in terms of pixel size and pixel purity, respectively for crop identification over different agricultural landscapes. A framework was proposed that automatically defines pixel size requirements for crop identification, based on user-defined criteria, for different crop types and for different agro-ecological landscapes in Middle Asia.

The practical utility of the proposed framework in the context of agricultural crop mapping include guidance (i) for selecting appropriate imagery (e.g. suitable pixel sizes) for applications focussed on crop mapping, (ii) for the selection of thresholds for minimum pixel purity required for the effective identification of crops in coarser pixels, and (iii) for EO data requirements for early crop inventory.

Different crops show marked individuality regarding the EO data requirements (in terms of pixel size and purity, respectively), and the effective identification of one given crop may considerably differ over different landscapes. It was shown that the per crop classification accuracy is affected by the pixel size of the image data, relative to the features (fields) in the image. This is most probably due to the reduced within class variability in coarser observation supports. Hence, image classification performance (irrespective of the metric used for its evaluation) could be improved by the use of suitable (and in most cases: coarser) pixel sizes.

Many studies rely on geostatistical indicators to find the optimized spatial support for image classification, however it might not be optimal to rely on only one parameter to define this. Further, these indicators are based on the entire image, whilst crop identification only focusses on a subset of spatial structures, which is one of the peculiar advantages of the framework on which this study is based on (Duveiller and Defourny, 2010). Further, the framework proposed in this study explicitly considers the influence of the temporal dimension on the definition of suitable pixel sizes for crop classification, which was not undertaken in previous studies such as from McCloy and Bøcher (2007).

The choice of supervised classifier algorithms only had a minor impact on the definition of v_{max} or v_{min}, respectively. More pronounced was the impact of the input data, e.g. red edge features enhanced the range of suitable pixel sizes for crop identification, and in general the inclusion of red edge feature results in higher accuracies. Unsupervised crop identification is possible, but limited to rather small ranges of pixel sizes and at the expense of high pixel purity requirements, compared to the supervised

strategies. Early detection of crops was shown to depend not only on the acquisition timing of images but also on the landscape pattern, and on the irrigation timing and start of season.

This study emphasizes the general need to analyse the determinants and key factors for crop mapping in very dynamic agricultural landscapes, including the spatiotemporal heterogeneity of crop types. This can be reasoned with the finding that different crop types could be detected at quite different ranges of pixel sizes over a given landscape, and that these pixel sizes requirements can considerably differed for one type of crop over different landscapes. The framework presented in this study not only provides insight into factors that influence the definition spatial resolution requirements. Such a diagnostic tool could further be of great utility to guide users in choosing the most appropriate imagery for their application, which is not restricted to crop identification. It can be of great value to guide planners in the technical implementation of any future crop-mapping component in an operational agricultural monitoring at the regional scale in Middle Asia, e.g. by evaluating the adequacy of existing remote sensing systems with regard to their use for crop mapping.

While the examples and context used in this study are from agriculture, the findings concerning the effects of spatial resolution, pixel purity and landscape characteristics are generic and applicable to many other application domains of remote sensing, including radar or hyper spectral data.

Chapter 7

OVERALL DISCUSSION AND CONCLUSIONS

Making use of remote sensing data, this thesis was introduced within the context of crop mapping, which is one fundamental input for agricultural production monitoring. The focus was on two specific research issues related to crop mapping: (i) improving crop classification at the per-field scale and providing information on the spatial distribution of classification uncertainty, and (ii) the identification of EO data requirements for an effective crop identification over heterogeneous agro-ecological landscapes.

The first issue is challenging because of the complex configuration and spatial heterogeneity of agro-ecological landscapes, and the temporal variability of crop phenology due to variable environmental and management conditions. Next to this, new remote sensing systems like RapidEye or the upcoming Sentinel-2 undoubtedly offer great potentials to classify crops with a descent standard of accuracy, but classifier algorithms must handle this huge amount of data. Further, these concepts must be robust and transferable over different landscapes and in a multi-year perspective, respectively to provide consistently accurate information on the spatial distribution of crops to be suitable for an operational crop monitoring. When crop maps are used as input to crop-specific condition modelling, the spatial distribution of error in crop maps (e.g. the classification uncertainty) should be quantified in order to be able to refine the maps or to filter out undesirable results from later modelling.

The second issue is challenging because EO data requirements for an operational crop specific monitoring, which requires the identification of crops in pixels, can differ significantly among different landscapes. A high revisit frequency is often needed to properly characterize the temporal evolution of cropped areas and to accurately identify crops. Further, a large geographical coverage (swath) is desired to have information about crops over large areas. Yet, crop mapping at regional to global scales is confronted with inherent physical limitations of remote sensing systems, because most often the observation supports of imaging systems that fulfil these requirements (frequent revisit rates and large swaths) are coarser than it might be desired. Crop classification with such coarser pixels suffers from this compromise when applied to heterogeneous agro-ecological landscapes around the world because of the effect of mixed pixels on the classification accuracy. The challenge here becomes to define suitable pixel sizes for an accurate identification of crops in such coarser and potentially mixed pixels (e.g. MODIS with 250 m, or the upcoming Sentinel-3 with 300 m pixel size).

Consequently, the overall objective of this thesis was to develop (i) advanced concepts and techniques for remote sensing-based agricultural crop classification at the per-field scale for creating accurate crop masks and (ii) a conceptual framework to quantitatively define suitable pixel sizes for the specific task of crop identification. Applications of these methods were given for heterogeneous agro-ecological landscapes in Middle Asia. Three research questions, which structure this synthesis, were formulated in the introduction chapter and investigated to contribute to the overall objective of this thesis.

What is the influence of feature selection on the accuracy and uncertainty in object-based crop classification?

After recalling concepts to derive membership estimates (soft outputs) from different classifier algorithms for measuring spatial classification uncertainty, it was established that information provided by different types of features is of complementary importance for crop classification. The different groups of features tested were of unequal importance for the classification of crops. It was further demonstrated how combining different kinds of features into one comprehensive multi-temporal input data set was useful to achieve high classification accuracies, which was also demonstrated in other studies (Loosvelt et al., 2012a; Peña-Barragán et al., 2011; Rodriguez-Galiano et al., 2012a). However the classifier algorithm must handle this amount of data. In this regard, a potential disadvantage of such a multi-type, multi-temporal data stack is its huge dimensionality. The classification performance of the SVM, which was used in this study, was claimed to be unaffected by the number of features (Cortes and Vapnik, 1995). Yet, in this study SVM was shown to be affected by the feature space size, which was also found by Pal and Foody (2010) and Waske et al. (2010). However, after a feature selection based on the RF feature importance score was performed, the SVM could even outreach the RF that was applied to the same data sets, demonstrating the general superiority of SVM in crop classification, confirming studies by Foody and Mathur (2006), Mathur and Foody (2008), or Waske et al. (2010). The RF in turn was quite unaffected by the huge feature space size, which was also demonstrated by Rodriguez-Galiano et al. (2012a). The interpretation of the RF feature score gave insight into the importance of single features and groups of features, and could be used as an alternative means to evaluate the relative importance of acquisition dates for classifying crops. Using this information the most important acquisition dates for crop classification could be determined, which provides an alternative means to more complex and computationally demanding experiments (Conrad et al., 2013; Murakami et al., 2001; Van Niel and Mcvicar, 2004). Features based on the red edge band of the RapidEye system were shown to be of particular importance for accurate classification, which is in line with a previous study from Schuster et al. (2012).

It was emphasized that classification uncertainty is not evenly distributed in the maps, but associated with different classes and possibly influenced by environmental factors, e.g. irrigation water availability. Whilst other studies found correlations between objet size and classification uncertainty (Yu et al., 2008), no such correlation was found here, most obvious because a relatively small range of object sizes was investigated. Yet, indication for an influence of irrigation water availability on classification uncertainty was discussed, e.g. there were spatial trends of increasing classification uncertainty, but further research is needed to quantify this. In accordance with Comber et al. (2012) it is suggested that the way of using the standard protocol for reporting classification accuracy should be reconsidered.

In an operational crop mapping context it could be envisaged to add such spatial classification uncertainty information as a supplement to crop maps, e.g. like the Quality Assurance Science Data Set (QA-SDS) layers in MODIS, which could be a useful input for agricultural production monitoring. For instance, this information could be used to mask out unreliable regions from later modelling. Classification uncertainty was further shown to decrease when feature space size is reduced, but even when selecting fewer features than necessary to achieve peak classification accuracy, the uncertainty could still decrease if incorrectly classified pixels were associated with lower uncertainties, as was demonstrated in two of the four study sites. This is in line with a study on the RF algorithm (Loosvelt et al. 2012a).

The main findings can be summarized as follows:

> Uncertainty information from libsvm is a valuable supplement to tradition accuracy metrics like overall classification accuracy, and an effective predictor of classification error (classification uncertainty) in maps.

> The uncertainty of correctly classified test cases generally decreased with decreasing input space size, but incorrectly classified test cases were not necessarily associated with high levels of uncertainty.

> The classification uncertainty analysis revealed the existence of spatial uncertainty trends in some of the test sites. High classification uncertainty could be an indicator for water scarcity.

> A decline in classification uncertainty was realized when selecting a smaller number of features than required to achieve peak accuracy, but without significantly reducing the classification accuracy.

Suggestions for potential improvements to the approach:

> A fine-tuning of the methodology could be the inclusion of feature selection strategies that exploit criteria intrinsically related to the SVM classifier properties.

> A more rigorous investigation could shed light on the interplay between environmental factors (e.g. water stress) and classification uncertainty, as this is particularly interesting when maps are used in spatially explicit models (e.g. crop yield). Regression analysis could be performed to evaluate explanatory factors for different levels of uncertainty, provided that suitable spatial data is available.

How do different classifier algorithms perform in crop classification at the object-level, and can combining classifier algorithms improve the overall quality of crop masks?

Different classifier algorithms (RF, SVM, and MLP) were compared with classifier combination schemes to classify crops at the object-level. An advanced combination operator was proposed, building upon existing algebraic operators that are commonly used in classifier combination. The different algorithms performed complementary concerning the classification of different crops, so it was difficult to identify the most appropriate classifier algorithms in advance. This underlines that using only one classifier algorithm in crop mapping might not yield the best results, which was also found by Foody et al. (2007). The results from the classifier combination emphasized that this technique is worthwhile for an improved crop classification, which was also found in other studies (Doan and Foody, 2007; Licciardi et al., 2009; Waske and Benediktsson, 2007).

Yet, based solely on overall classification accuracy there appeared to be no significant advantage in the classifier combination in every test site, e.g. the results were only marginally better than the single best classifier algorithm, confirming that improvements in accuracy are not granted in classifier combination (Foody et al., 2007; Giacco et al., 2010). But there are other compelling reasons for classifier combination: the strength was its transferability and robustness when applied over different test sites with different classes. For instance, the range of classification accuracies over all test sites was smallest when applying the classifier combination. Another peculiarity of combining classifier algorithms was that the improvements in classification accuracy became the more pronounced the fewer training data were available, which was also found by Waske et al. (2010). Further, classifier

combination yielded classification uncertainty information that may be used to help post-classification refinement operations and later analyses. A much smaller proportion of cases were classified with high classification uncertainty after the classifier combination, as compared with the single classifier algorithms.

Regarding the initial research question, the findings can be summarized:

> RF, SVM, and MLP perform well but complementary on the multi-temporal image stacks.

> The proposed classifier combination scheme increased classification accuracy, consistently outperformed the single classifier algorithms, and was more stable and better transferable over all landscapes tested and also to data sets from another year.

> The proposed classifier combination scheme was most accurate when testing on very small input data sets or training sets, respectively and is rather simple in implementation, which makes this approach well suited for operational applications.

> Classifier combination resulted in reduced classification uncertainty, and levelled out the spatial differences in classification uncertainty in the crop maps, as compared with the single classifier algorithms.

There are possible perspectives for pursuing this approach, but the most promising could be:

> The combination rules could be rewritten with fuzzy data fusion rules, which might allow for a finer definition and weighting of the classifier reliability during the combination process. This could in turn further enhance classification accuracy.

> The proposed approached could be tested over a wider range of landscapes with more crop classes and in particular over several years to further evaluate its transferability and stability, respectively and to further evaluate its utility for an operational application.

What is the spatial resolution requirement for crop identification via image classification and how do these requirements change over different landscapes?

To answer this question a methodological framework was established in this thesis with the objective to quantitatively define these requirements, based on user-defined criteria. This was achieved by simulating how agricultural landscapes, and more specifically the fields covered by a crop of interest, are seen by instruments with increasingly coarser resolving power. The concept of crop specific pixel purity, defined as the degree of homogeneity of the signal encoded in a pixel with respect to the target crop type, is used to analyse how mixed the pixels can be (as they become coarser), without undermining their capacity to describe the desired surface properties, e.g. to identify crop specific signals for a proper (accurate) crop classification.

Having applied this framework to the time series data in the four test sites, it was demonstrated how different crops show marked individuality regarding the pixel size and pixel purity requirements, and that there were different ranges of pixel sizes that were suitable for crop identification. These results demonstrated that it might not be possible to define one single optimal scale (Aplin, 2006) and it is in line with studies that demonstrated how the definition of suitable pixel sizes for a given application vary over different landscapes (Duveiller and Defourny, 2010; McCloy and Bøcher, 2007; Ozdogan et al., 2006; Woodcock and Strahler, 1987). Yet, previous studies were either limited to reporting pixel size requirements with regard to one specific structure in the landscapes (e.g. only fields with one certain crop) or by using only one parameter (e.g. overall classification accuracy or image variance) to find this optimal pixel size. Moreover, the findings in these studies were based on the analysis of mono-temporal

satellite images. In this regard, the proposed framework has some advantages over these studies, because it explored the effect of the observation length (e.g. number of acquisitions) on the definition of suitable pixel sizes for crop identification. These results challenge findings of previous studies, because the definition of what is "optimal" scale, or the optimal range of suitable pixel sizes, was demonstrated to vary along the vegetation season. This can be interpreted by changes in the surface spatial heterogeneity (Garrigues et al., 2008a) and demonstrates how a synergistic view on spatial and temporal scale can result in a more comprehensive understanding of the definition of suitable pixels sizes (Aplin, 2006).

Further, there is indication that the pixel sizes that yield the best crop identification are not necessarily the ones that yield the best crop area estimation, which underlines that the definition of suitable pixel sizes can be application-dependent (Duveiller et al., 2010). Whilst the application reported here is focussed on pixel-based approaches, with the interest being the general impact of pixel size on crop identification, object-based approaches might be preferred when using very small pixels as these tend to be more accurate (Lobo et al., 1996), although this is not granted (Duro et al., 2012).

This study emphasized the general need to analyse the determinants and key factors for crop mapping in very dynamic agricultural landscapes, including the spatiotemporal heterogeneity of crop types. This can be reasoned with the finding that different crop types can be detected at quite different ranges of pixel sizes over a given landscape, and that these pixel sizes requirements can considerably differ for one type of crop over different landscapes and along the season, respectively. The framework presented in this study not only provides insight into factors that influence the spatial resolution requirements. Such a diagnostic tool could further be of great utility to guide users in choosing the most appropriate imagery for their application, which is not restricted to crop identification. It can be of great utility to guide planners in the technical implementation of any future crop-mapping component in an operational agricultural monitoring, e.g. by evaluating the adequacy of existing remote sensing systems with regard to their use for crop identification. As a final conclusion to this third research question, the following findings can be summarized:

> Different crops show marked individuality regarding the EO data requirements (in terms of pixel size and purity, respectively), and the requirements for an effective identification of one given crop type may considerably differ over different landscapes.

> The choice of classifier algorithm only had a minor impact on the definition of suitable pixel sizes. More pronounced was the impact of observation length and input data, e.g. red edge features enhanced the range of suitable pixel sizes for crop identification.

> It was discussed how the spatial pattern of agricultural fields impacted the definition of pixel size requirements for crop mapping.

> In the multi-year perspective environmental cues like irrigation water availability are hypothesized to further influence the definition of pixel suitability, e.g. by enhancing intra field heterogeneity due to irrigation water shortages.

Potential improvements and future applications of the framework are:

> An application with high-resolution time series containing more acquisition dates could be envisaged, given that sufficient images are available, to better approximate the revisit frequency of sensors that have a coarser GSD like MODIS or Sentinel-3. The framework could be extended by further evaluating region specific requirements regarding the type of data (optical, radar, or hyper-spectral) to find out which is best suited for specific landscapes.

> The experiments could be repeated in other years in order to trace the resolution requirements over the longer run and to better understand the impact of environmental cues (e.g. water availability) on their definition. An application over a wider range of different agricultural landscapes could be envisaged (e.g. in Africa, China, the U.S., or Europe), and possible spatial shifts in pixel size requirements could be analysed by applying the framework over several subsets within one landscape.

> Whilst the examples presented in this study are from agriculture, the framework can be applied to any other type of classes, e.g. natural land cover types. Finally, with very coarse pixel sizes one might try to narrow the classification problem down to a simple binary case (e.g. cropland vs. non-cropland).

Synthesis and prospect

Regarding the operationalization of EO-based techniques for agricultural monitoring and its application to a broader range of agricultural landscapes, it can be noted that, despite the high performance of existing methods (e.g. classifier algorithms), transferability and stability of such methods remain one important research issue. This means that methods developed and tested in one place might not necessarily be portable to another place or over several years, respectively. Hence, investigating concepts to employ or combine existing techniques that can adapt to specific regional or local contexts seems worthwhile. To have one example, the use of multi-temporal data sets with different kind of data has advantages over mono-temporal data sets and can enhance classification accuracy. Yet, is was demonstrated in this study that the configuration of such huge data sets (e.g. its size) had to be optimised for the classifier algorithms and classes investigated, respectively and for each test site. In this context, classifier combination is a worthwhile and effective method that should be preferred in an operational crop-mapping context, because the findings in this thesis indicate that these concepts can be more stable, e.g. they consistently provided higher classification accuracies over different landscapes and over different years than individual classifier algorithms. Specifically in Middle Asia, which was selected as study region for this thesis, classifier combination makes sense due to its easy implementation and because it enhanced classification accuracy for classes with insufficient training samples. This observation makes it interesting for operational contexts and when field reference data availability is limited. Further, an investigation of multi-sensory remote sensing data, including radar or even hyper spectral, seems worthwhile. Space-born radar missions provide spatially enhanced observations (with swaths suitable at least for local applications) with high revisit frequencies and are not dependent on the need for cloud free atmospherically conditions. The concepts and frameworks presented in this thesis are not restricted to optical data.

Recalling Figure 1-1 in the introduction chapter, it can be noted that crop identification, which is one prerequisite for crop specific monitoring, is possible over a wide range of pixel sizes and over various landscapes, but selecting an appropriate imagery is far from being trivial. Consequently, the application of existing satellite sensors might not be equally suitable for crop identification and, by extension, for crop specific monitoring in different heterogeneous agricultural landscapes. In this regard, similar to the transferability of methods, the application of only one certain kind of EO data (e.g. with one specific pixel size) over different landscapes needs to be revisited and the synergistic use of multi-scale data, e.g. combining remote sensing imagery of both fine and coarse spatial resolution, should be fostered. The necessity to predict and control the effects of spatial and temporal scale on crop classification is recognized here as a major goal to achieve in EO-based agricultural monitoring, specifically if the focus is on crop specific monitoring.

REFERENCES

Alexandratos, N. and Bruinsma, J. (2012), *World agriculture: Towards 2030 / 2050 - The 2012 revision (Report)*, p. 154.

Ali, S. and Smith-Miles, K. (2006), "Improved support vector machine generalization using normalized input space", *Advances in Artificial Intelligence*, Vol. 4304, pp. 362–371.

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, Government Printing Office (US Geological Survey, Professional Paper 964)*, Washington, DC.

Aplin, P. (2006), "On scales and dynamics in observing the environment", *International Journal of Remote Sensing*, Vol. 27 No. 11, pp. 2123–2140.

Asian Development Bank. (2012), *Key indicators for Asia and the Pacific (Report)*, Mandaluyong City, Philippines, p. 301.

Atkinson, P.M. and Aplin, P. (2004), "Spatial variation in land cover and choice of spatial resolution for remote sensing", *International Journal of Remote Sensing*, Vol. 25 No. 18, pp. 3687–3702.

Atkinson, P.M. and Curran, P. (1997), "Choosing an appropriate spatial resolution for remote sensing investigations", *Photogrammetric Engineering & Remote Sensing*, Vol. 63 No. 12, pp. 1345–1351.

Atkinson, P.M. and Tatnall, A. (1997), "Introduction neural networks in remote sensing", *International Journal of Remote Sensing*, Vol. 18 No. 4, pp. 699–709.

Atzberger, C. (2013), "Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs", *Remote Sensing*, Vol. 5 No. 2, pp. 949–981.

Baatz, M. and Schäpe, A. (2000), "Multiresolution segmentation - an optimization approach for high quality multi-scale image segmentation", in Strobl, J., Blaschke, T. and Griesebner, G. (Eds.),*Angewandte Geographische Informationsverarbeitung XII*, Wichmann-Verlag, Heidelberg, pp. 12–23.

Bailey, J.T. and Boryan, C.G. (2010), *Remote sensing applications in agriculture at the USDA national agricultural statistics service*, Fairfax, USA, Vol. 22030, p. 14.

Balaguer, A., Ruiz, L.A., Hermosilla, T. and Recio, J.A. (2010), "Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification", *Computers & Geosciences* Vol. 36 No. 2, pp. 231–240.

Barrett, E.C. and Curtis, D.E. (1992), *Introduction to environmental remote sensing*, Springer, 3rded., p. 448.

Bastiaanssen, W.G.M. and Bos, M.G. (1999), "Irrigation performance indicators based on remotely sensed data: A review of literature", *Irrigation and Drainage Systems*, Vol. 13 No. 4, pp. 291–311.

Bastiaanssen, W.G.M. and Harshadeep, N.R. (2005), "Managing scarce water resources in Asia: The nature of the problem and can remote sensing help?", *Irrigation and Drainage Systems*, Vol. 19 No. 3-4, pp. 269–284.

Bastiaanssen, W.G.M., Molden, D.J. and Makin, I.W. (2000), "Remote sensing for irrigated agriculture: Examples from research and possible applications", *Agricultural Water Management*, Vol. 46 No. 2, pp. 137–155.

Bazi, Y. and Melgani, F. (2006), "Toward an optimal SVM classification system for hyperspectral remote sensing images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44 No. 11, pp. 3374–3385.

Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., Tucker, C., Anyamba, A., Small, J., et al. (2010), "Monitoring global croplands with coarse resolution earth observations: The Global Agriculture Monitoring (GLAM) project", *Remote Sensing*, Vol. 2 No. 6, pp. 1589–1609.

Beekhuizen, J. and Clarke, K.C. (2010), "Toward accountable land use mapping: Using geocomputation to improve classification accuracy and reveal uncertainty", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 12 No. 3, pp. 127–137.

Belousov, A.I., Verzakov, S.A. and Frese, J. Von. (2002), "A flexible classification approach with optimal generalisation performance: Support vector machines", *Chemometrics and Intelligent Laboratory Systems*, Vol. 64 No. 1, pp. 12–25.

Benediktsson, J.A., Chanussot, J. and Fauvel, M. (2007), "Multiple classifier systems in remote sensing: From basics to recent developments", in Haindl, M., Kittler, J. and Roli, F. (Eds.), *Multiple classifier systems*, Springer, Vol. 22, pp. 501–512.

Benediktsson, J.A. and Kanellopoulos, I. (1999), "Classification of multisource and hyperspectral data based on decision fusion", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37 No. 3, pp. 1367–1377.

Benediktsson, J.A. and Swain, P.H. (1992), "Consensus theoretic classification methods", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, pp. 688–704.

Benediktsson, J.A., Swain, P.H. and Erase, O.K. (1990), "Neural network approaches versus statistical methods in classification of multisource remote sensing data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 28, pp. 540–551.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M. (2004), "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 58 No. 3-4, pp. 239–258.

Béquignon, J., Caughey, J., Cramer, W., Heip, C., Justice, C.O., Key, J.R., Koike, T., et al. (2010), *GEO and science (Report)*, p. 50.

Berberoglu, S., Lloyd, C.D., Atkinson, P.M. and Curran, P.J. (2000), "The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean", *Computers & Geosciences*, Vol. 26 No. 4, pp. 385–396.

Beyer, S. (2012), *Degradationsprobleme in den Bewässerungsgebieten des Aralbeckens - Eine fernerkundungsbasierte Untersuchung der Bewässerungsgebiete im Amudarja-Delta (Diploma thesis)*, Heidelberg University.

Blaes, X., Vanhalle, L. and Defourny, P. (2005), "Efficiency of crop identification based on optical and SAR image time series", *Remote Sensing of Environment*, Vol. 96 No. 3-4, pp. 352–365.

Blaschke, T. (2010), "Object based image analysis for remote sensing", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 65 No. 1, pp. 2–16.

Blaschke, T. and Hay, G.J. (2001), "Object-oriented image analysis and scale-space: Theory and methods for modeling and evaluating multiscale landscape structure", *International Archives of Photogrammetry and Remote Sensing*, Vol. 34 No. 4, pp. 22–29.

Bloch, I. (1996), "Information combination operators for data fusion: A comparative review with classification", *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 26 No. 1, pp. 52–67.

Bloch, I. and Maître, H. (1994), "Fusion de données en traitement d'images: Modèles d'information et décisions", *TS. Traitement du signal*, Vol. 11 No. 6, pp. 435–446.

Blum, A. and Langley, P. (1997), "Selection of relevant features and examples in machine learning", *Artificial Intelligence*, Vol. 97 No. 1-2, pp. 245–271.

Bohovic, R., Klein, D. and Conrad, C. (2011), "Phenological patterns in Central Asia: Monitoring vegetation dynamics using remote sensing at different scales", *EARSeL Remote Sensing and Geoinformation not only for Scientific Cooperation*, pp. 289–297.

Boroffka, N., Oberhänsli, H., Sorrel, P., Demory, F., Reinhardt, C., Wünnemann, B., Alimov, K., et al. (2006), "Archaeology and climate: Settlement and lake-level changes at the Aral Sea", *Geoarchaeology*, Vol. 21 No. 7, pp. 721–734.

Brachet, G. (2004), "From initial ideas to a European plan: GMES as an exemplar of European space strategy", *Space Policy,*, Vol. 20 No. 1, pp. 7–15.

Bradley, A.P. (1997), "The use of the area under the ROC curve in the evaluation of machine learning algorithms", *Pattern Recognition*, Vol. 30 No. 7, pp. 1145–1159.

Bradley, A.V. and Millington, A.C. (2006), "Spatial and temporal scale issues in determining biomass burning regimes in Bolivia and Peru", *International Journal of Remote Sensing*, Vol. 27 No. 11, pp. 2221–2253.

Breiman, L. (1996), "Bagging predictors", *Machine learning*, Vol. 24, pp. 123–140.

Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5–32.

Breiman, L. and Cutler, A. (2007), "Random forests — Classification description: Random forests", *stat-www.berkeley.edu/users/breiman/RandomForests/cc_home.htm (last accessed: 01-Nov 2013)*.

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1984), *Classification and regression trees*, CRC Press, Boca Raton, p. 368.

Brenning, A. (2009), "Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection", *Remote Sensing of Environment*, Vol. 113 No. 1, pp. 239–247.

Briem, G.J., Benediktsson, J.A. and Sveinsson, J.R. (2002), "Multiple classifiers applied to multisource remote sensing data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40 No. 10, pp. 2291–2299.

Broge, N. and Leblanc, E. (2000), "Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density", *Remote Sensing of Environment*, Vol. 76 No. 2, pp. 156–172.

Brown, J.C., Kastens, J.H., Coutinho, A.C., Victoria, D.D.C. and Bishop, C.R. (2013), "Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data", *Remote Sensing of Environment*, Vol. 130, pp. 39–50.

Brown, K.M., Foody, G.M. and Atkinson, P.M. (2009), "Estimating per-pixel thematic uncertainty in remote sensing classifications", *International Journal of Remote Sensing*, Vol. 30 No. 1, pp. 209–229.

Bruinsma, J. (2003), *World agriculture: Towards 2015 / 2030 - An FAO perspective (Report)*, Food and Agriculture Organization (FAO), p. 444.

Bruzzone, L., Conese, C., Maselli, F. and Roli, F. (1997), "Multisource classification of complex rural areas by statistical and neural-network approaches", *Photogrammetric Engineering & Remote Sensing*, Vol. 63 No. 5, pp. 523–533.

Bruzzone, L., Prieto, D.F. and Serpico, S.B. (1999), "A neural-statistical approach to multitemporal and multisource remote-sensing image classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37 No. 3, pp. 1350–1359.

Bucknall, J., Klytchnikova, I., Lampietti, J., Lundell, M., Scatasta, M. and Thurman, M. (2003), *Irrigation in Central Asia. Social, economic and environmental considerations (Report)*, Washington, DC, p. 104.

Buddenbaum, H., Schlerf, M. and Hill, J. (2005), "Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods", *International Journal of Remote Sensing*, Vol. 26 No. 24, pp. 5453–5465.

Burges, C.J.C. (1998), "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, Vol. 2, pp. 121–167.

Cai, X., McKinney, D.C. and Rosegrant, M.W. (2003), "Sustainability analysis for irrigation water management in the Aral Sea region", *Agricultural Systems*, Vol. 76 No. 3, pp. 1043–1066.

Canters, F. (1997), "Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification", *Photogrammetric Engineering & Remote Sensing*, Vol. 63 No. 4, pp. 403–414.

Carfagna, E. and Gallego, J.F. (2005), "Using remote sensing for agricultural statistics", *International Statistical Review*, Vol. 73 No. 3, pp. 389–404.

Carrão, H., Gonçalves, P. and Caetano, M. (2008), "Contribution of multispectral and multitemporal information from MODIS images to land cover classification", *Remote Sensing of Environment*, Vol. 112 No. 3, pp. 986–997.

Chan, J.C.-W., Paelinckx, D., Beckers, P., Spanhove, T. and Borre, J. Vanden. (2008), "Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery", *Remote Sensing of Environment*, Vol. 112 No. 6, pp. 2999–3011.

Chang, C.C. and Lin, C.J. (2011), "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, Vol. 2 No. 3, pp. 1–39.

Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., et al. (2000), "Consequences of changing biodiversity.", *Nature*, Vol. 405 No. 6783, pp. 234–42.

Chemin, Y., Platonov, A., Ul-Hassan, M. and Abdullaev, I. (2004), "Using remote sensing data for water depletion assessment at administrative and irrigation-system levels: Case study of the Ferghana Province of Uzbekistan", *Agricultural Water Management*, Vol. 64 No. 3, pp. 183–196.

Chen, J. (1996), "Evaluation of vegetation indices and a modified simple ratio for boreal applications", *Canadian Journal of Remote Sensing*, Vol. 22 No. 3, pp. 229–242.

Chen, X., Tateishi, R. and Wang, C. (1999), "Development of a 1-km landcover dataset of China using AVHRR data", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, pp. 305–316.

Chica-Olmo, M. and Abarca-Hernandez, F. (2000), "Computing geostatistical image texture for remotely sensed data classication", *Computers & Geosciences*, Vol. 26 No. 4, pp. 373–383.

Clausi, D.A. (2002), "An analysis of co-occurrence texture statistics as a function of grey level quantization", *Canadian Journal of Remote Sensing*, Vol. 28 No. 1, pp. 45–62.

Cohen, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 20 No. 1, pp. 37–46.

Colditz, R., Schmidt, M., Conrad, C., Hansen, M.C. and Dech, S. (2011), "Land cover classification with coarse spatial resolution data to derive continuous and discrete maps for complex regions", *Remote Sensing of Environment*, Vol. 115 No. 12, pp. 3264–3275.

Collins, J.B. and Woodcock, C.E. (1999), "Modeling the distribution of cover fraction of a geophysical field", in Atkinson, P.M. and Tate, N.J. (Eds.), *Advances in remote sensing and GIS analysis*, Wiley, pp. 119–134.

Comber, A., Fisher, P., Brunsdon, C. and Khmag, A. (2012), "Spatial analysis of remote sensing image classification accuracy", *Remote Sensing of Environment*, Vol. 127, pp. 237–246.

Congalton, R.G. (1991), "A review of assessing the accuracy of classifications of remotely sensed data", *Remote Sensing of Environment*, Vol. 37 No. 1, pp. 35–46.

Congalton, R.G. and Green, K. (2009), *Assessing the accuracy of remotely sensed data: Principles and practices*, *Assessment*, CRC Press, 2nded., Vol. 48, p. 200.

Conrad, C. (2006), *Fernerkundungsbasierte Modellierung und hydrologische Messungen zur Analyse und Bewertung der landwirtschaftlichen Wassernutzung in der Region Khorezm, Usbekistan (PhD Thesis)*, Wuerzburg University.

Conrad, C., Colditz, R., Dech, S., Klein, D. and Vlek, P.L.G. (2011a), "Temporal segmentation of MODIS time series for improving crop classification in Central Asian irrigation systems", *International Journal of Remote Sensing*, Vol. 32 No. 23, pp. 8763–8778.

Conrad, C., Dech, S., Dubovyk, O., Klein, D., Schorcht, G. and Zeidler, J. (2013), "Utilizing RapidEye time series for crop monitoring and per-field classifications in heterogeneous agricultural areas of arid Central Asia", *Submitted to Computers and Electronics in Agriculture*.

Conrad, C., Dech, S., Hafeez, M., Lamers, J.P.A., Martius, C. and Strunz, G. (2007), "Mapping and assessing water use in a Central Asian irrigation system by utilizing MODIS remote sensing products", *Irrigation and Drainage Systems*, Vol. 21 No. 3-4, pp. 197–218.

Conrad, C., Fritsch, S., Lex, S., Löw, F., Rücker, G., Schorcht, G., Sultanov, M., et al. (2012), "Potenziale des 'Red Edge' Kanals von RapidEye zur Unterscheidung und zum Monitoring landwirtschaftlicher Anbaufrüchte am Beispiel des usbekischen Bewaesserungssystems Khorezm", in Borg, E., Daedelow, H. and Johnson, A.R. (Eds.), *RapidEye Science Archive (RESA) - Vom Algorithmus zum Produkt*, Springer, GITO, pp. 201–217.

Conrad, C., Fritsch, S., Zeidler, J., Rücker, G. and Dech, S. (2010), "Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data", *Remote Sensing*, Vol. 2 No. 4, pp. 1035–1056.

Conrad, C., Machwitz, M., Schorcht, G., Löw, F., Fritsch, S. and Dech, S. (2011b), "Potentials of RapidEye time series for improved classification of crop rotations in heterogeneous agricultural landscapes: Experiences from irrigation systems in Central Asia", *Proceedings of SPIE 8174*, Prague, Czech Republic.

Cortes, C. and Vapnik, V. (1995), "Support-vector networks", *Machine learning*, Vol. 20 No. 3, pp. 273–297.

Cowan, P.J. (2007), "Geographic usage of the terms Middle Asia and Central Asia", *Journal of Arid Environments*, Vol. 69 No. 2, pp. 359–363.

Cracknell, A.P. (1998), "Synergy in remote sensing—what's in a pixel?", *International Journal of Remote Sensing*, Vol. 19 No. 11, pp. 2025–2047.

Curran, P.J. (1988), "The semivariogram in remote sensing: An introduction", *Remote Sensing of Environment*, Vol. 24 No. 3, pp. 493–507.

Cushnie, J.L. (1987), "The interactive effect of spatial resolution and degree of internal variability within land-cover types on classification accuracies", *International Journal of Remote Sensing*, Vol. 8 No. 1, pp. 15–29.

Daughtry, C.S.T., Walthall, C.L., Kim, M.S., Brown de Colstoun, E. and McMurtrey III, J.E. (2000), "Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance", *Remote Sensing of Environment*, Vol. 74 No. 2, pp. 229–239.

DeFries, R., Hansen, M. and Townshend, J. (1995), "Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data", *Remote Sensing of Environment*, Vol. 54, pp. 209–222.

Van Diepen, K. and Boogaard, H. (2009), "History of CGMS in the MARS project", *Agro Informatica*, Vol. 22, pp. 11–14.

Doan, H.T.X. and Foody, G.M. (2007), "Increasing soft classification accuracy through the use of an ensemble of classifiers", *International Journal of Remote Sensing*, Vol. 28 No. 20, pp. 4609–4623.

Döll, P. (2002), "Impact of climate change and variability on irrigation requirements: A global perspective", *Climatic Change*, Vol. 54, pp. 269–293.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., et al. (2012), "The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission", *Remote Sensing of Environment*, Vol. 120, pp. 37–57.

Doraiswamy, P., Hatfield, J., Jackson, B., Akhmedov, B., Prueger, J. and Stern, A. (2004), "Crop condition and yield simulations using Landsat and MODIS", *Remote Sensing of Environment*, Vol. 92 No. 4, pp. 548–559.

Douglas, J. (1962), "The Virgin and Idle Lands Program reappraised", *Annals of the Association of American Geographers*, Vol. 52, pp. 69–79.

Dubovyk, O., Menz, G., Conrad, C., Kan, E., Machwitz, M. and Khamzina, A. (2012), "Spatio-temporal analyses of cropland degradation in the irrigated lowlands of Uzbekistan using remote-sensing and logistic regression modeling", *Environmental Monitoring and Assessment*, Vol. 185 No. 6, pp. 4775–4790.

Duro, D.C., Franklin, S.E. and Dubé, M.G. (2012), "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery", *Remote Sensing of Environment*, Vol. 118, pp. 259–272.

Duveiller, G., Baret, F. and Defourny, P. (2011), "Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy", *Remote Sensing of Environment*, Vol. 115 No. 10, pp. 2686–2701.

Duveiller, G., Baret, F. and Defourny, P. (2012), "Remotely sensed green area index for winter wheat crop monitoring: 10-Year assessment at regional scale over a fragmented landscape", *Agricultural and Forest Meteorology*, Vol. 166-167, pp. 156–168.

Duveiller, G. and Defourny, P. (2010), "A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing", *Remote Sensing of Environment*, Vol. 114 No. 11, pp. 2637–2650.

Duveiller, G., López-Lozano, R., Seguini, L., Bojanowski, J.S. and Baruth, B. (2013), "Optical remote sensing requirements for operational crop monitoring and yield forecasting in Europe", in Ouwehand, L. (Ed.), *Proc. "Sentinel-3 OLCI/SLSTR and MERIS/(A)ATSR Workshop" Vol. 21027, 15–19 October 2012, Frascati, Italy*.

Eckhardt, D.W., Verdin, J.P. and Lyford, R.G. (1990), "Automated update of an irrigated lands GIS using SPOT HRV imagery", *Photogrammetric Engineering & Remote Sensing*, Vol. 56 No. 11, pp. 1515–1522.

Ehammer, A., Fritsch, S., Conrad, C. and Lamers, J.P.A. (2010), "Statistical derivation of fPAR and LAI for irrigated cotton and rice in arid Uzbekistan by combining multi-temporal RapidEye data and ground measurements", *Proceedings of SPIE 7824*, Toulouse, France.

Eitel, J., Long, D. and Gessler, P. (2007), "Using in-situ measurements to evaluate the new RapidEye™ satellite series for prediction of wheat nitrogen status", *International Journal of Remote Sensing*, Vol. 28 No. 18, pp. 4183–4190.

El-Magd, A.I. and Tanton, T.W. (2003), "Improvements in land use mapping for irrigated agriculture from satellite sensor data using a multi-stage maximum likelihood classification", *International Journal of Remote Sensing*, Vol. 24 No. 21, pp. 4197–4206.

ERDAS. (2010), "IMAGINE AutoSync ™ User´s guide".

Erickson, J.D. (1984), "The LACIE experiment in satellite aided monitoring of global crop production", in Woodwell, G.M. (Ed.), *The role of terrestrial vegetation in the global carbon cycle: Measurement by remote sensing*, Wiley, pp. 191–217.

ESRI. (2013), "ArcGIS 10.1 (Computer program)", ESRI, available at: http://www.esri.com/ (last accessed: 15-Nov 2013).

FAO. (2006), *World agriculture: Towards 2030/2050 - Prospects for food, nutrition, agriculture and major commodity groups (Report)*, Rome, p. 78.

FAO. (2013a), "Food and Agriculture Organization of the United Nations (FAOSTAT)", *http://faostat3.fao.org/faostat-gateway/go/to/home/E*, available at: http://faostat3.fao.org/faostat-gateway/go/to/home/E (last accessed 30-Nov 2013).

FAO. (2013b), "FAO aquastat", *http://www.fao.org/nr/water/aquastat* (last accessed 13-Nov 2013).

Fassnacht, F.E., Latifi, H. and Koch, B. (2012), "An angular vegetation index for imaging spectroscopy data— Preliminary results on forest damage detection in the Bavarian National Park, Germany", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 19, pp. 308–321.

Fauvel, M., Benediktsson, J.A., Chanussot, J. and Sveinsson, J.R. (2008), "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46 No. 11, pp. 3804–3814.

Fawcett, T. (2006), "An introduction to ROC analysis", *Pattern Recognition Letters*, Vol. 27 No. 8, pp. 861–874.

Fisher, P. (1997), "The pixel: A snare and a delusion", *International Journal of Remote Sensing*, Vol. 18 No. 3, pp. 679–685.

Foley, J.A., Defries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., et al. (2005), "Global consequences of land use", *Science*, Vol. 309, pp. 570–574.

Foley, J.A., Ramankutty, N., Brauman, K. A, Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., et al. (2011), "Solutions for a cultivated planet.", *Nature*, Vol. 478 No. 7369, pp. 337–42.

Foody, G.M. (1995a), "Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 50 No. 5, pp. 2–12.

Foody, G.M. (1995b), "Land cover classification by an artificial neural network with ancillary information", *International Journal of Remote Sensing*, Vol. 9 No. 5, pp. 527–542.

Foody, G.M. (1996a), "Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data", *International Journal of Remote Sensing*, Vol. 17 No. 7, pp. 1317–1340.

Foody, G.M. (1996b), "Relating the land-cover composition of mixed pixels to artificial neural network classification output.", *Photogrammetric Engineering & Remote Sensing*, Vol. 62 No. 5, pp. 491–499.

Foody, G.M. (1996c), "Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications", *Pattern Recognition Letters*, Vol. 17 No. 13, pp. 1389–1398.

Foody, G.M. (2000), "Mapping land cover from remotely sensed data with a softened feedforward neural network classification", *Journal of Intelligent & Robotic Systems*, Vol. 29 No. 4, pp. 433–449.

Foody, G.M. (2002), "Status of land cover classification accuracy assessment", *Remote Sensing of Environment*, Vol. 80 No. 1, pp. 185–201.

Foody, G.M. (2004), "Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes", *International Journal of Remote Sensing*, Vol. 25 No. 15, pp. 3091–3104.

Foody, G.M. (2005), "Local characterization of thematic classification accuracy through spatially constrained confusion matrices", *International Journal of Remote Sensing*, Vol. 26 No. 6, pp. 1217–1228.

Foody, G.M. (2008), "RVM based multi class classification of remotely sensed data", *International Journal of Remote Sensing*, Vol. 29 No. 6, pp. 1817–1823.

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*, Vol. 18 No. 4, pp. 799–810.

Foody, G.M. and Atkinson, P.M. (2002), *Uncertainty in remote sensing and GIS*, Wiley, p. 331.

Foody, G.M., Boyd, D.S. and Sanchez-Hernandez, C. (2007), "Mapping a specific class with an ensemble of classifiers", *International Journal of Remote Sensing*, Vol. 28 No. 8, pp. 1733–1746.

Foody, G.M., Campbell, N., Trodd, N. and Wood, T. (1992), "Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification", *Photogrammetric Engineering & Remote Sensing*, Vol. 58 No. 9, pp. 1335–1341.

Foody, G.M. and Mathur, A. (2004), "Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification", *Remote Sensing of Environment*, Vol. 93 No. 1-2, pp. 107–117.

Foody, G.M. and Mathur, A. (2006), "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM", *Remote Sensing of Environment*, Vol. 103 No. 2, pp. 179–189.

Foody, G.M., Mathur, A., Sanchez-Hernandez, C. and Boyd, D.S. (2006), "Training set size requirements for the classification of a specific class", *Remote Sensing of Environment*, Vol. 104 No. 1, pp. 1–14.

Forshaw, M.R.B., Haskell, A., Miller, P.F., Stanley, D.J. and Townshend, J.R.G. (1983), "Spatial resolution of remotely sensed imagery - A review paper", *International Journal of Remote Sensing*, Vol. 4 No. 3, pp. 497–520.

Freund, Y. and Schapire, R.E. (1996), "Experiments with a new boosting algorithm", *Machine Learning: Proceedings of the Thirteenth International Conference, June3rd-6th*, Bare, Italy, p. 148–156.

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*, IEEE, Vol. 37 No. 2, pp. 969–977.

Fritsch, S. (2013), *Spatial and temporal patterns of crop yield and marginal land in the Aral Sea Basin: Derivation by combining multi-scale and multi-temporal remote sensing data with a light use efficiency model (PhD Thesis)*, University of Wuerzburg.

Fritsch, S., Machwitz, M., Ehammer, A., Conrad, C. and Dech, S. (2012), "Validation of the collection 5 MODIS FPAR product in a heterogeneous agricultural landscape in arid Uzbekistan using multitemporal RapidEye imagery", *International Journal of Remote Sensing*, Vol. 33 No. 21, pp. 6818–6837.

Fritz, S., Massart, M., Savin, I., Gallego, J.F. and Rembold, F. (2008), "The use of MODIS data to derive acreage estimations for larger fields: A case study in the south-western Rostov region of Russia", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 10 No. 4, pp. 453–466.

Gallego, J.F. (1999), "Crop area estimation in the MARS project", *Conference on ten years of the MARS Project*, Brussels, Belgium, pp. 1–11.

Gallego, J.F. (2004), "Remote sensing and land cover area estimation", *International Journal of Remote Sensing*, Vol. 25 No. 15, pp. 3019–3047.

Gallego, J.F., Craig, M., Michaelsen, J., Bossyns, B. and Fritz, S. (2008), *Best practices for crop area estimation with remote sensing (Report)*, Joint Research Center, Ispra, p. 14.

Gallego, J.F., Delince, J. and Rueda, C. (1993), "Crop area estimates through remote sensing: Stability of the regression correction", *International Journal of Remote Sensing*, Vol. 14 No. 18, pp. 3433–3445.

Garrigues, S., Allard, D. and Baret, F. (2008a), "Modeling temporal changes in surface spatial heterogeneity over an agricultural site", *Remote Sensing of Environment*, Vol. 112 No. 2, pp. 588–602.

Garrigues, S., Allard, D., Baret, F. and Morisette, J. (2008b), "Multivariate quantification of landscape spatial heterogeneity using variogram models", *Remote Sensing of Environment*, Vol. 112 No. 1, pp. 216–230.

Garrigues, S., Allard, D., Baret, F. and Weiss, M. (2006a), "Quantifying spatial heterogeneity at the landscape scale using variogram models", *Remote Sensing of Environment*, Vol. 103 No. 1, pp. 81–96.

Garrigues, S., Allard, D., Baret, F. and Weiss, M. (2006b), "Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data", *Remote Sensing of Environment*, Vol. 105, pp. 286–298.

Genovese, G., Vignolles, C., Nègre, T. and Passera, G. (2001), "A methodology for a combined use of normalised difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain", *Agronomie*, Vol. 21 No. 1, pp. 91–111.

Gessner, U., Machwitz, M., Conrad, C. and Dech, S. (2013), "Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles", *Remote Sensing of Environment*, Vol. 129, pp. 90–102.

Giacco, F., Thiel, C., Pugliese, L., Scarpetta, S. and Marinaro, M. (2010), "Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48 No. 10, pp. 3769–3779.

Giese, E. and Sehring, J. (2007), *Regionalexpertise – Destabilisierungs- und Konfliktpotenzial prognostizierter Umweltveränderung in der Region Zentralasien 2020/2050 (Report)*, WBGU, p. 48.

Gislason, P., Benediktsson, J.A. and Sveinsson, J.R. (2006), "Random Forests for land cover classification", *Pattern Recognition Letters*, Vol. 27 No. 4, pp. 294–300.

Gitelson, A.A., Kaufman, Y.J., Stark, R. and Rundquist, D. (2002), "Novel algorithms for remote estimation of vegetation fraction", *Remote Sensing of Environment*, Vol. 80 No. 1, pp. 76–87.

Gitelson, A.A., Kaufman, Y.J. and Merzlyak, M.N. (1996), "Use of a green channel in remote sensing of global vegetation from EOS-MODIS", *Remote Sensing of Environment*, Vol. 58 No. 3, pp. 289–298.

Glantz, M. (2005), "Water, climate, and development issues in the Amu Darya Basin", *Mitigation and Adaptation Strategies for Global Change*, Vol. 10, pp. 23–50.

Glantz, M.H. (2009), *Creeping environmental problems and sustainable development in the Aral Sea Basin*, Cambridge University Press, p. 304.

Gleick, P.H. (2003), "Global freshwater resources: Soft-path solutions for the 21st century", *Science*, Vol. 302 No. 5650, pp. 1524–1528.

Glenn, E.P., Huete, A.R., Nagler, P.L. and Nelson, S.G. (2008), "Relationship between remotely-sensed vegetation indices, canopy attributes, and plant physiological processes: What vegetation indices can and cannot tell us about the landscape", *Sensors*, Vol. 8, pp. 2136–2160.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., et al. (2010), "Food security: The challenge of feeding 9 billion people", *Science*, Vol. 327, pp. 812–818.

Gonçalves, L. (2011), "A hybrid approach classification of remote sensing images", *EARSeL Remote Sensing and Geoinformation not only for Scientific Cooperation*, Prague, pp. 328–335.

Gonçalves, L., Fonte, C.C., Julio, E. and Caetano, M. (2009), "A method to incorporate uncertainty in the classification of remote sensing images", *International Journal of Remote Sensing*, Vol. 30 No. 20, pp. 5489–5503.

Gong, P., Pu, R. and Chen, J. (1996), "Mapping ecological land systems and classification uncertainties from digital elevation and forest cover data using neural networks", *Photogrammetric Engineering & Remote Sensing*, Vol. 62 No. 11, pp. 1249–1260.

Guissard, V., Defourny, P. and Ledent, J.F. (2004), "Crop specific information extraction based on coarse resolution pixel sampling", *2nd International VEGETATION Users Conference*, Antwerp, Belgium.

Guyon, I. and Elisseeff, A. (2003), "An Introduction to variable and feature selection", *Journal ofMachine Learning Research*, Vol. 3, pp. 1157–1182.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002), "A gene selection method for cancer classification using support vector machines", *Machine learning*, Vol. 46, pp. 389–422.

Haack, B. and Rafter, A. (2010), "Regression estimation techniques with remote sensing: A review and case study", *Geocarto International*, Vol. 25 No. 1, pp. 71–82.

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P. and Dextraze, L. (2002), "Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture", *Remote Sensing of Environment*, Vol. 81 No. 2-3, pp. 416–426.

Hakimov, N., Lines, A., Elmuratov, P. and Hakimov, R. (2007), "Climate change and water resource alteration in Central Asia: The case of Uzbekistan", in Lal, R., Suleimenov, M., Stewart, B.A., Hansen, D.O. and Doraiswamy, P. (Eds.), *Climate change and terrestrial carbon sequestration on Central Asia*, Taylor & Francis, pp. 75–82.

Hand, D.J. (2009), "Measuring classifier performance: A coherent alternative to the area under the ROC curve", *Machine Learning*, Vol. 77 No. 1, pp. 103–123.

Haralick, R.M., Shanmugam, K. and Dinstein, I. (1973), "Textural features for image classification", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3 No. 6, pp. 610–621.

Harris, D.R. (2010), *Origins of agriculture in western Central Asia. An environmental-archaeological study*, University of Pennsylvania Press, p. 304.

Hastie, T. and Tibshirani, R. (1998), "Classification by pairwise coupling", *The Annals of Statistics*, Vol. 26, pp. 451–471.

Hastie, T., Tibshirani, R. and Friedman, J. (2011), *The elements of statistical learning: Data mining, inference, and prediction*, Springer, 2nded., p. 745.

Hornik, K., Buchta, C., Hothorn, T., Karatzoglou, A., Meyer, D. and Zeileis, A. (2013), "R package 'RWeka'".

Howell, T. (2001), "Enhancing water use efficiency in irrigated agriculture", *Agronomy Journal*, Vol. 93, pp. 281–289.

Hsieh, P., Lee, L.C. and Chen, N. (2001), "Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39 No. 12, pp. 2657–2663.

Hsu, C.W. and Lin, C.J. (2002), "A comparison of methods for multi-class support vector machines", *IEEE Transactions on Neural Networks*, Vol. 13 No. 4, pp. 415–425.

Hu, X. and Weng, Q. (2009), "Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks", *Remote Sensing of Environment*, Vol. 113 No. 10, pp. 2089–2102.

Huang, C., Davis, L.S. and Townshend, J.R.G. (2002a), "An assessment of support vector machines for land cover classification", *International Journal of Remote Sensing*, Vol. 23 No. 4, pp. 725–749.

Huang, C., Townshend, J.R.G., Liang, S., Kalluri, S.N. V. and Defries, R.S. (2002b), "Impact of sensor's point spread function on land cover characterization: Assessment and deconvolution", *Remote Sensing of Environment*, Vol. 80 No. 2, pp. 203–212.

Huete, A., 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*, Vol. 83, pp. 195–213.

Huete, A.R. (1988), "A soil-adjusted vegetation index (SAVI)", *Remote Sensing of Environment*, Vol. 25 No. 3, pp. 295–309.

Hughes, G. (1968), "On the mean accuracy of statistical pattern recognizers", *IEEE Transactions on Information Theory*, Vol. 14 No. 1, pp. 55–63.

Husson, F., Josse, J., Le, S. and Mazet, J. (2013), "R package 'FactoMineR' - Multivariate Exploratory data analysis and data mining with R".

Hüttich, C., Gessner, U., Herold, M., Strohbach, B.J., Schmidt, M., Keil, M. and Dech, S. (2009), "On the suitability of MODIS time series metrics to map vegetation types in dry savanna ecosystems: A case study in the Kalahari of NE Namibia", *Remote Sensing*, Vol. 1 No. 4, pp. 620–643.

Ibrakhimov, M., Khamzina, A., Forkutsa, I., Paluasheva, G., Lamers, J.P.A., Tischbein, B., Vlek, P.L.G., et al. (2007), "Groundwater table and salinity: Spatial and temporal distribution and influence on soil salinization in Khorezm region (Uzbekistan, Aral Sea Basin)", *Irrigation and Drainage Systems*, Vol. 21 No. 3-4, pp. 219–236.

Jain, A., Duin, R. and Mao, J. (2000), "Statistical pattern recognition: A review", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 22 No. 1, pp. 4–37.

Jeon, B. and Landgrebe, D.A. (1999), "Decision fusion approach for multitemporal classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37 No. 3, pp. 1227–1233.

JICA, Oriental Consultants, Sanyu Consultants and International NTC. (2010), *The study on regional development in Karakalpakstan in the Republic of Uzbekistan (Progress report)*, *Perspectives in public health*, Vol. 130, p. 238.

Johnson, D.M. and Mueller, R. (2010), "The 2009 Cropland Data Layer", *Photogrammetric Engineering & Remote Sensing*, Vol. 76 No. 11, pp. 1201–1205.

Jones, H.G. and Vaughan, R.A. (2010), *Remote sensing of vegetation: Principles, techniques, and applications.*, Oxford University Press, New York, p. 380.

Jordan, C.F. (1969), "Derivation of leaf-area index from quality of light on the forest floor", *Ecology*, Vol. 50 No. 4, pp. 663–666.

Ju, J., Gopal, S. and Kolaczyk, E.D. (2005), "On the choice of spatial and categorical scale in remote sensing land cover classification", *Remote Sensing of Environment*, Vol. 96 No. 1, pp. 62–77.

Justice, C. and Becker-Reshef, I. (2007), "Report from the workshop on developing a strategy for global agricultural monitoring in the framework of Group on Earth Observations (GEO), 16-18 July 2007, FAO, Rome", University of Maryland: College Park, MD, USA, p. 67.

Kaiser, G. and Schneider, W. (2008), "Estimation of sensor point spread function by spatial subpixel analysis", *International Journal of Remote Sensing*, Vol. 29 No. 7, pp. 2137–2155.

Karatzoglou, A., Meyer, D. and Hornik, K. (2006), "Support vector machines in R", *Journal of Statistical Software*, Vol. 15 No. 9, pp. 1–28.

Kariyeva, J. and van Leeuwen, W.J.D. (2012), "Phenological dynamics of irrigated and natural drylands in Central Asia before and after the USSR collapse", *Agriculture, Ecosystems & Environment*, Vol. 162, pp. 77–89.

Kastens, J.H., Kastens, T.L., Kastens, D.L.A., Price, K.P., Martinko, E.A. and Lee, R.-Y. (2005), "Image masking for crop yield forecasting using AVHRR NDVI time series imagery", *Remote Sensing of Environment*, Vol. 99 No. 3, pp. 341–356.

Kaufman, Y.J., Tanré, D., Holben, B.N., Markham, B. and Gitelson, A.A. (1992), "Atmospheric effects on the NDVI - Strategies for its removal", *Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, pp. 1238–1241.

Kavzoglu, T. (2001), *An investigation of the design and use of feed-forward artificial neural networks in the classification of remotely sensed images (PhD Thesis)*, The University of Nottingham.

Kavzoglu, T. (2009), "Increasing the accuracy of neural network classification using refined training data", *Environmental Modelling & Software*, Vol. 24 No. 7, pp. 850–858.

Kavzoglu, T. and Mather, P.M. (2003), "The use of backpropagating artificial neural networks in land cover classification", *International Journal of Remote Sensing*, Vol. 24 No. 23, pp. 4907–4938.

Kittler, J. and Alkoot, F.M. (2003), "Sum versus vote fusion in multiple classifier systems", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 25 No. 1, pp. 110–115.

Kittler, J., Hatef, M., Duin, R.P.W. and Matas, J. (1998), "On combining clasifiers", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 20 No. 3, pp. 226–239.

Kotte, K., Löw, F., Huber, S.G., Krause, T., Mulder, I. and Schöler, H.F. (2012), "Organohalogen emissions from saline environments – spatial extrapolation using remote sensing as most promising tool", *Biogeosciences*, Vol. 9 No. 3, pp. 1225–1235.

Kramer, H.J. (2001), *Observation of the earth and its environment: Survey of missions and sensors*, Springer, 4thed., p. 1510.

Kramer, H.J. and Cracknell, A.P. (2008), "An overview of small satellites in remote sensing", *International Journal of Remote Sensing*, Vol. 29 No. 15, pp. 4285–4337.

Kuncheva, L.I. (2002), "A theoretical study on six classifier fusion strategies", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24 No. 2, pp. 281–286.

Kuncheva, L.I. (2004), *Combining pattern classifiers: Methods and algorithms*. Wiley, p. 360.

Lal, R., Suleimenov, M., Steward, B.A., Hansen, D.O. and Doraiswamy, P. (2007), *Climate change and terrestrial carbon sequestration in Central Asia*, Taylor & Francis, p. 493.

Létolle, R. and Mainguet, M. (1996), *Der Aralsee: Eine ökologische Katastrophe*, Springer, p. 517.

Levin, S.A. (1992), "The problem of pattern and scale in ecology", *Ecology*, Vol. 73 No. 6, pp. 1943–1967.

Liaw, A. (2013), "R package 'randomForest', Breiman and Cutler's random forests for classification and regression".

Licciardi, G., Pacifici, F., Tuia, D., Prasad, S., West, T., Giacco, F., Thiel, C., et al. (2009), "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47 No. 11, pp. 3857–3865.

Lin, H.T., Lin, C.-J. and Weng, R.C. (2007), "A note on Platt's probabilistic outputs for support vector machines", *Machine Learning*, Vol. 68 No. 3, pp. 267–276.

Linderman, M., Liu, J., Qi, J., An, L., Ouyang, Z., Yang, J. and Tan, Y. (2004), "Using artificial neural networks to map the spatial distribution of understorey bamboo from remote sensing data", *International Journal of Remote Sensing*, Vol. 25 No. 9, pp. 1685–1700.

Lioubimtseva, E. and Henebry, G.M. (2009), "Climate and environmental change in arid Central Asia: Impacts, vulnerability, and adaptations", *Journal of Arid Environments*, Vol. 73 No. 11, pp. 963–977.

Liu, W., Gopal, S. and Woodcock, C.E. (2004), "Uncertainty and confidence in land cover classification using a hybrid classifier approach", *Photogrammetric Engineering & Remote Sensing*, Vol. 70 No. 8, pp. 963–972.

Lloyd, C.D., Berberoglu, S., Curran, P.J. and Atkinson, P.M. (2004), "A comparison of texture measures for the per-field classification of Mediterranean land cover", *International Journal of Remote Sensing*, Vol. 25 No. 19, pp. 3943–3965.

Lobell, D.B., Schlenker, W. and Costa-Roberts, J. (2011), "Climate trends and global crop production since 1980", *Science*, Vol. 333, pp. 616–20.

Lobo, A., Chic, O. and Casterad, A. (1996), "Classification of Mediterranean crops with multisensor data: Per-pixel versus per-object statistics and image segmentation", *International Journal of Remote Sensing*, Vol. 17 No. 12, pp. 2385–2400.

Loosvelt, L., Peters, J. and Skriver, H. (2012a), "Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50 No. 10, pp. 4185 – 4200.

Loosvelt, L., Peters, J., Skriver, H., Lievens, H., Van Coillie, F.M.B., De Baets, B. and Verhoest, N.E.C. (2012b), "Random forests as a tool for estimating uncertainty at pixel-level in SAR image classification", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 19, pp. 173–184.

Löw, F., Michel, U., Dech, S. and Conrad, C. (2013a), "Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 85, pp. 102–119.

Löw, F., Navratil, P. and Bubenzer, O. (2012), "Landscape dynamics in the southern Aralkum: Using MODIS time series for land cover change analysis", in Breckle, S.-W., Wucherer, W., Dimeyeva, L.A. and Ogar, N.P. (Eds.), *Ecological Studies 218*, Springer, pp. 83–96.

Löw, F., Navratil, P., Kotte, K., Schöler, H.F. and Bubenzer, O. (2013b), "Remote sensing based analysis of landscape change in the desiccated seabed of the Aral Sea - A potential tool for assessing the hazard degree of dust and salt storms", *Environmental Monitoring and Assessment*, Vol. 185 No. 10, pp. 8303–8319.

Löw, F., Schorcht, G., Michel, U., Dech, S. and Conrad, C. (2012), "Per-field crop classification in irrigated agricultural regions in middle Asia using random forest and support vector machine ensemble", *Proceedings of SPIE 8538,* Edinburgh, UK.

Lu, D. and Weng, Q. (2007), "A survey of image classification methods and techniques for improving classification performance", *International Journal of Remote Sensing*, Vol. 28 No. 5, pp. 823–870.

Lu, S., Oki, K., Shimizu, Y. and Omasa, K. (2007), "Comparison between several feature extraction/classification methods for mapping complicated agricultural land use patches using airborne hyperspectral data", *International Journal of Remote Sensing*, Vol. 28 No. 5, pp. 963–984.

Machwitz, M., Bloethe, J., Klein, D., Conrad, C. and Dech, S. (2010), "Mapping of large irrigated areas in Central Asia using MODIS time series", *Proceedings of SPIE 7824*, Toulouse, France.

Mannig, B., Müller, M., Starke, E., Merkenschlager, C., Mao, W., Zhi, X., Podzun, R., et al. (2013), "Dynamical downscaling of climate change in Central Asia", *Global and Planetary Change*, Vol. 110, pp. 26–39.

Marceau, D.J., Gratton, D., Fournier, R. and Fortin, J. (1994a), "Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution", *Remote Sensing of Environment*, Vol. 49 No. 2, pp. 105–117.

Marceau, D.J., Howarth, P.J. and Gratton, D.J. (1994b), "Remote sensing and the measurement of geographical entities in a forested environment. 1: The scale and spatial aggregation problem", *Remote Sensing of Environment*, Vol. 49, pp. 93–104.

Martius, C., Rudenko, I., Lamers, J.P.A. and Vlek, P.L.G. (2012), *Cotton, water, salts and soums. Economic and ecological restructering in Khorezm, Uzbekistan*, Springer, p. 360.

Maselli, F., Conese, C. and Petkov, L. (1994), "Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 49 No. 2, pp. 13–20.

Mathur, A. and Foody, G.M. (2008), "Crop classification by support vector machine with intelligently selected training data for an operational application", *International Journal of Remote Sensing*, Vol. 29 No. 8, pp. 2227–2240.

McCabe, M.F. and Wood, E.F. (2006), "Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors", *Remote Sensing of Environment*, Vol. 105 No. 4, pp. 271–285.

McCloy, K.R. and Bøcher, P.K. (2007), "Optimizing image resolution to maximize the accuracy of hard classification", *Photogrammetric Engineering & Remote Sensing*, Vol. 73 No. 8, pp. 893–903.

McCoy, R.M. (2005), *Field methods in remote sensing*, The Guilford Press, p. 177.

McIver, D. and Friedl, M. (2001), "Estimating pixel-scale land cover classification confidence using nonparametric machine learning methods", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39 No. 9, pp. 1959–1968.

McKinney, D.C. (2004), "Cooperative management of transboundary water resources in Central Asia", in Burghart, D.L. and Sabonis-Helf, T. (Eds.),*In the tracks of Tamerlane: Central Asia's path to the 21st century*, Center for Technology and national Security Policy, Washington, DC, pp. 187–219.

McNairn, H., Champagne, C., Shang, J., Holmstrom, D. and Reichert, G. (2009), "Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 64 No. 5, pp. 434–449.

McNairn, H., Ellis, J., Van Der Sanden, J.. J., Hirose, T. and Brown, R.J. (2002), "Providing crop information using RADARSAT-1 and satellite optical imagery", *International Journal of Remote Sensing*, Vol. 23 No. 5, pp. 851–870.

Melgani, F. and Bruzzone, L. (2004), "Classification of hyperspectral remote sensing images with support vector machines", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42 No. 8, pp. 1778–1790.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. and Leisch, F. (2012), "R package 'e1071': Misc functions of the Department of Statistics (e1071), TU Wien".

Micklin, P. (2002), "Water in the Aral Sea Basin of Central Asia: Cause of conflict or cooperation?", *Eurasian Geography and Economics*, Vol. 43 No. 7, pp. 505–528.

Micklin, P. (2007), "The Aral Sea disaster", *Annual Review of Earth and Planetary Sciences*, Annual Reviews, Vol. 35, pp. 47–72.

Micklin, P. (2010), "The past, present, and future Aral Sea", *Lakes & Reservoirs: Research and Management*, Vol. 15 No. 3, pp. 193–213.

Micklin, P.P. (1988), "Desiccation of the Aral Sea: A Water management disaster in the Soviet Union", *Science*, Vol. 241, pp. 1170–1176.

Monfreda, C., Ramankutty, N. and Foley, J.A. (2008), "Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000", *Global Biogeochemical Cycles*, Vol. 22 No. 1, pp. 1–19.

Moran, P.A.P. (1950), "Notes on continuous stochastic phenomena", *Biometrika*, Vol. 37 No. 1/2, pp. 17–23.

Mountrakis, G., Im, J. and Ogole, C. (2011), "Support vector machines in remote sensing: A review", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 66 No. 3, pp. 247–259.

Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N. and Foley, J.A. (2012), "Closing yield gaps through nutrient and water management.", *Nature*, Vol. 490 No. 7419, pp. 254–257.

Munoz, G. and Grieser, J. (2006), "CLIMWAT 2.0 for CROPWAT (Computer program)", Food & Agriculture Organization of the United Nations (FAO), Rome.

Murakami, T., Ogawa, S., Ishitsuka, N., Kumagai, K. and Saito, G. (2001), "Crop discrimination with multitemporal SPOT/HRV data in the Saga Plains, Japan", *International Journal of Remote Sensing*, Vol. 22 No. 7, pp. 1335–1348.

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*, Vol. 24 No. 23, pp. 4871–4890.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S. and Weng, Q. (2011), "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery", *Remote Sensing of Environment*, Vol. 115 No. 5, pp. 1145–1161.

Nentwig, W. (2005), *Humanökologie*, Springer, 2nded., p. 474.

Neubert, M., Herold, H. and Meinel, G. (2006), "Evaluation of remote sensing image segmentation quality– further results and concepts", *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36.4/C42.*

Neubert, M., Herold, H. and Meinel, G. (2008), "Assessing image segmentation quality - Concepts, methods and application", in Blaschke, T., Lang, S. and Hay, G.J. (Eds.), *Object-based image analysis - Spatial concepts for knowledge-driven remote sensing applications*, Springer, pp. 769–784.

Van Niel, T.G. and Mcvicar, T.R. (2004), "Determining temporal windows for crop discrimination with remote sensing: A case study in south-eastern Australia", *Computers and Electronics in Agriculture*, Vol. 45 No. 1-3, pp. 91–108.

Nijland, W., Addink, E.A., De Jong, S.M. and Van der Meer, F.D. (2009), "Optimizing spatial image support for quantitative mapping of natural vegetation", *Remote Sensing of Environment*, Vol. 113 No. 4, pp. 771–780.

Oberhänsli, H., Boroffka, N., Sorrel, P. and Krivonogov, S. (2007), "Climate variability during the past 2,000 years and past economic and irrigation activities in the Aral Sea basin", *Irrigation and Drainage Systems*, Vol. 21 No. 3-4, pp. 167–183.

Van Oort, P. a. J., Bregt, a. K., de Bruin, S., de Wit, a. J.W. and Stein, A. (2004), "Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database", *International Journal of Geographical Information Science*, Vol. 18 No. 6, pp. 611–626.

Ortiz, A. and Oliver, G. (2006), "On the use of the overlapping area matrix for image segmentation evaluation: A survey and new performance measures", *Pattern Recognition Letters*, Vol. 27 No. 16, pp. 1916–1926.

Ozdogan, M. (2010), "The spatial distribution of crop types from MODIS data: Temporal unmixing using independent component analysis", *Remote Sensing of Environment*, Vol. 114 No. 6, pp. 1190–1204.

Ozdogan, M. and Woodcock, C.E. (2006), "Resolution dependent errors in remote sensing of cultivated areas", *Remote Sensing of Environment*, Vol. 103 No. 2, pp. 203–217.

Ozdogan, M., Yang, Y., Allez, G. and Cervantes, C. (2010), "Remote sensing of irrigated agriculture: Opportunities and challenges", *Remote Sensing*, Vol. 2 No. 9, pp. 2274–2304.

Pachauri, R.K. and Reisinger, A. (2008), *Climate change 2007: Synthesis report. Contribution of working groups I, II and III to the fourth assessment report of the Intergovernmental Panel on Climate Change*, Geneva, Switzerland, p. 104.

Pal, M. (2005), "Random forest classifer for remote sensing classification", *International Journal of Remote Sensing*, Vol. 26 No. 1, pp. 217–222.

Pal, M. and Foody, G.M. (2010), "Feature selection for classification of hyperspectral data by SVM", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48 No. 5, pp. 2297–2307.

Pal, M. and Mather, P.M. (2005), "Support vector machines for classification in remote sensing", *International Journal of Remote Sensing*, Vol. 26 No. 5, pp. 1007–1011.

Pal, N.R. and Bezdek, J.C. (1994), "Measuring fuzzy uncertainty", *IEEE Transactions on Fuzzy Systems*, Vol. 2 No. 2, pp. 107 – 118.

Pax-Lenney, M. and Woodcock, C.E. (1997), "The effect of spatial resolution on the ability to monitor the status of agricultural lands", *Remote Sensing of Environment*, Vol. 61 No. 2, pp. 210–220.

Pebesma, E. and Graeler, B. (2013), "R package 'gstat': Spatial and spatio-temporal geostatistical modelling, prediction and simulation".

Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E. and Six, J. (2011), "Object-based crop identification using multiple vegetation indices, textural features and crop phenology", *Remote Sensing of Environment*, Vol. 115 No. 6, pp. 1301–1316.

Peter, N. (2004), "The use of remote sensing to support the application of multilateral environmental agreements", *Space Policy*, Vol. 20 No. 3, pp. 189–195.

Platt, J. (2000), "Probabilities for SV Machines", in Smola, A.J., Bartlett, P., Schölkopf, B. and Schuurmans, D. (Eds.), *Advances in large margin classifiers*, MIT Press, pp. 61–74.

Plourde, L. and Congalton, R. (2003), "Sampling method and sample placement: How do they affect the accuracy of remotely sensed maps?", *Photogrammetric Engineering & Remote Sensing*, Vol. 69 No. 3, pp. 289–297.

Pohl, C. and van Genderen, J.L. (1998), "Multisensor image fusion in remote sensing: Concepts, methods and applications", *International Journal of Remote Sensing*, Vol. 19 No. 5, pp. 823–854.

Policar, R. (2006), "Ensemble based systems in decision making", *IEEE Circuits and Systems Magazin*, Vol. 6 No. 3, pp. 21–45.

Pontius, R.G. and Millones, M. (2011), "Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment", *International Journal of Remote Sensing*, Vol. 32 No. 15, pp. 4407–4429.

Powers, D.M.W. (2011), "Evaluation: From precision, recall and f-mesure to ROC, informedness, markedness & correlation", *Journal of Machine Learning Technologies*, Vol. 2 No. 1, pp. 37–63.

Prasad, A.M., Iverson, L.R. and Liaw, A. (2006), "Newer classification and regression tree techniques: Bagging and random forests for ecological prediction", *Ecosystems*, Vol. 9 No. 2, pp. 181–199.

Provost, F. and Fawcett, T. (1997), "Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions", *Proceedings of the third international conference on knowledge discovery and data mining*, AAAI Press, pp. 43–48.

Pudil, P., Novovičová, J. and Kittler, J. (1994), "Floating search methods in feature selection", *Pattern recognition letters*, Vol. 15, pp. 1119–1125.

Pyrcz, M. and Deutsch, C. (2003), "The whole story on the hole effect", *Geostatistical Association of Australasia Newsletter 18*, West Perth, Australia, Vol. 18, p. 18.

Qadir, M., Noble, A.D., Qureshi, A.S., Gupta, R.K., Yuldashev, T. and Karimov, A. (2009), "Salt induced land and water degradation in the Aral Sea basin: A challenge to sustainable agriculture in Central Asia", *Natural Resources Forum*, Vol. 33 No. 2, pp. 134–149.

R Development Core Team. (2012), "A language and environment for statistical computing", Vienna, Austria, available at: http://www.r-project.org/(last accessed 16-Nov 2012).

R Development Core Team. (2013), "The R Stats Package".

Ramankutty, N., Evan, A.T., Monfreda, C. and Foley, J.A. (2008), "Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000", *Global Biogeochemical Cycles*, Vol. 22 No. 1, doi:10.1029/2007GB002952.

RapidEye AG. (2012), "Satellite imagery product specifications, v. 4.1 (www.rapideye.com, last accessed 13-Nov 2012)", RapidEye AG, Brandenburg an der Havel.

Read, J.J., Tarpley, L., McKinion, J.M. and Reddy, K.R. (2002), "Narrow-waveband reflectance ratios for remote estimation of nitrogen status in cotton", *Journal of Environmental Quality*, Vol. 31 No. 5, pp. 1442–1452.

Richards, J.A. and Jia, X. (2005), *Remote Sensing Digital Image Analysis*, *New York*, Springer, 5thed., p. 464.

Richardson, A.J. and Everitt, J.H. (1992), "Using spectral vegetation indices to estimate rangeland productivity", *Geocarto International*, Vol. 7 No. 1, pp. 63–69.

Richter, R. (2011), "Atmospheric/Topographic Correction for Satellite Imagery. ATCOR-2/3 User Guide 7.1".

Van Rijsbergen, C.J. (1979), *Information Retrieval, Butterworths*, Butterworth & Co Publishers Ltd, 2nded., p. 224.

Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M. and Jeganathan, C. (2012a), "Random forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture", *Remote Sensing of Environment*, Vol. 121, pp. 93–107.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. (2012b), "An assessment of the effectiveness of a random forest classifier for land-cover classification", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 67, pp. 93–104.

Rosenfield, G.H. and Fitzpatrick-Lins, K. (1986), "A coefficient of agreement as a measure of thematic classification accuracy", *Photogrammetric Engineering & Remote Sensing*, Vol. 52 No. 2, pp. 223–227.

Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M. and Dobson, C. (2003), "A review of remote sensing technology in support of the Kyoto Protocol", *Environmental Science & Policy*, Vol. 6 No. 5, pp. 441–455.

Roujean, J. and Breon, F. (1995), "Estimating PAR absorbed by vegetation from bidirectional reflectance measurements", *Remote Sensing of Environment*, Vol. 51 No. 3, pp. 375–384.

Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W. (1974), "Monitoring vegetation systems in the Great Plains with ERTS", in Freden, S.C., Mercanti, E.P. and Becker, M.A. (Eds.), *Proceedings of the Earth Resources Technology Satellite Symposium NASA SP-351*, Washington, DC, p. 309–317.

Ruiz, L.A., Recio, J.A., Fernández-Sarría, A. and Hermosilla, T. (2011), "A feature extraction software tool for agricultural object-based image analysis", *Computers and Electronics in Agriculture Electronics in Agriculture*, Vol. 76 No. 2, pp. 284–296.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986), "Learning representations by back-propagating errors", *Nature*, Vol. 323, pp. 533–536.

Saiko, T.A. and Zonn, I.S. (2000), "Irrigation expansion and dynamics of desertification in the Circum-Aral region of Central Asia", *Applied Geography*, Vol. 20 No. 4, pp. 349–367.

Sakamoto, T., Wardlow, B.D., Gitelson, A. A., Verma, S.B., Suyker, A.E. and Arkebauer, T.J. (2010), "A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data", *Remote Sensing of Environment*, Vol. 114 No. 10, pp. 2146–2159.

Sapper, M., Weichsel, V. and Huterer, A. (2007), *Machtmosaik Zentralasien. Traditionen, Restriktionen, Aspirationen*, BWV - Berliner Wissenschafts-Verlag, p. 648.

Sauer, T., Havlík, P., Schneider, U. A., Schmid, E., Kindermann, G. and Obersteiner, M. (2010), "Agriculture and resource availability in a changing world: The role of irrigation", *Water Resources Research*, Vol. 46 No. 6, pp. 1–12.

Schorcht, G., Löw, F., Fritsch, S. and Conrad, C. (2012), "Crop classification at subfield level using RapidEye time series and graph theory algorithms", *Proceedings of SPIE 8531*, Edinburgh, UK.

Schowengerdt, R.A. (2007), *Remote sensing: Models and methods for image processing*, Academic Press, 3rded., p. 560.

Schuster, C., Förster, M. and Kleinschmit, B. (2012), "Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data", *International Journal of Remote Sensing*, Vol. 33 No. 17, pp. 5583–5599.

Seelan, S.K., Laguette, S., Casady, G.M. and Seielstad, G.A. (2003), "Remote sensing applications for precision agriculture: A learning community approach", *Remote Sensing of Environment*, Vol. 88 No. 1-2, pp. 157–169.

Sepulcre-Cantó, G., Gellens-Meulenberghs, F., Arboleda, A., Duveiller, G., Piccard, I., de Wit, A., Tychon, B., et al. (2010), "Combining remote sensing imagery of both fine and coarse spatial resolution to estimate crop evapotranspiration and quantifying its influence on crop growth monitoring.", *Geophysical Research Abstracts*, Vol. 12.

Serpico, S.B. and Bruzzone, L. (2001), "A new search algorithm for feature selection in hyperspectral remote sensing images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, pp. 1360–1367.

Serra, P. and Pons, X. (2008), "Monitoring farmers' decisions on Mediterranean irrigated crops using satellite image time series", *International Journal of Remote Sensing*, Vol. 29 No. 8, pp. 2293–2316.

Shannon, C.E. (1948), "The mathematical theory of communication.", *Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656.

Shao, Y. and Lunetta, R.S. (2012), "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 70, pp. 78–87.

Siegfried, T., Bernauer, T., Guiennet, R., Sellars, S., Robertson, A.W., Mankin, J., Bauer-Gottwein, P., et al. (2012), "Will climate change exacerbate water stress in Central Asia?", *Climatic Change*, Vol. 112 No. 3-4, pp. 881–899.

Simonneaux, V., Duchemin, B., Helson, D., Er-Raki, S., Olioso, A. and Chehbouni, A.G. (2008), "The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco", *International Journal of Remote Sensing*, Vol. 29 No. 1, pp. 95–116.

Smith, J.H., Stehman, S. V., Wickham, J.D. and Yang, L. (2003), "Effects of landscape characteristics on land-cover class accuracy", *Remote Sensing of Environment*, Vol. 84 No. 3, pp. 342–349.

Sokolova, M., Japkowicz, N. and Szpakowicz, S. (2006), "Beyond accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation", in Sattar, A. and Kang, B. (Eds.), *Advances in Artificial Intelligence (Lecture Notes in Computer Science) Vol. 4304*, Springer, pp. 1015–1021.

Solberg, A.H.S., Jain, A.K. and Taxt, T. (1994), "Multisource classification of remotely sensed data: Fusion of Landsat TM and SAR images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32 No. 4, pp. 768–778.

Song, C., Woodcock, C.E., Seto, K.C., Pax-Lenney, M. and Macomber, S.A. (2001), "Classification and change detection using Landsat TM data: When and how to correct atmospheric effects?", *Remote Sensing of Environment*, Vol. 75 No. 2, pp. 230–244.

SPOT IMAGE. (2012), "Spot Image Homepage", http://spot-image.fr (last accessed 18-Nov 2012).

Steele, B. (2000), "Combining multiple classifiers: An application using spatial and remotely sensed information for land cover type mapping", *Remote Sensing of Environment*, Vol. 47 No. 3, pp. 545–556.

Steele, B.M., Winne, J.C. and Redmond, R.L. (1998), "Estimation and mapping of misclassification probabilities for thematic land cover maps", *Remote Sensing of Environment*, Vol. 66 No. 2, pp. 192–202.

Stehman, S. V. (1997), "Selecting and interpreting measures of thematic classification accuracy", *Remote sensing of Environment*, Vol. 62 No. 1, pp. 77–89.

Stehman, S. V. (1999), "Comparing thematic maps based on map value", *International Journal of Remote Sensing*, Vol. 20 No. 12, pp. 2347–2366.

Stehman, S. V. and Milliken, J. (2007), "Estimating the effect of crop classification error on evapotranspiration derived from remote sensing in the lower Colorado River basin, USA", *Remote Sensing of Environment*, Vol. 106 No. 2, pp. 217–227.

Stehman, S. V. and Wickham, J.D. (2011), "Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment", *Remote Sensing of Environment*, Vol. 115 No. 12, pp. 3044–3055.

Stellmes, M., Udelhoven, T., Röder, a., Sonnenschein, R. and Hill, J. (2010), "Dryland observation at local and regional scale — Comparison of Landsat TM/ETM+ and NOAA AVHRR time series", *Remote Sensing of Environment*, Vol. 114 No. 10, pp. 2111–2125.

Stuart, N., Jaas, T., Zisopoulos, I. and Viergever, K. (2008), "Methods for estimating the accuracy of per-pixel , per-parcel and expert visual classification of high resolution optical satellite imagery", *Proceedings of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences, June 25-27*, Shanghai, P.R. China, pp. 163–170.

Tadjudin, S. and Landgrebe, D.A. (1999), "Covariance estimation with limited training samples", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37 No. 4, pp. 2113–2118.

Tan, C.-P., Koay, J.-Y., Lim, K.-S., Ewe, H.-T. and Chuah, H.-T. (2007), "Classification of multi-temporal SAR images for rice crops using combined entropy decomposition and support vector machine technique", *Progress in Electromagnetics Research*, Vol. 71, pp. 19–39.

Tarnavsky, E., Garrigues, S. and Brown, M.E. (2008), "Multiscale geostatistical analysis of AVHRR, SPOT-VGT, and MODIS global NDVI products", *Remote Sensing of Environment*, Vol. 112 No. 2, pp. 535–549.

Townshend, J.R.G. (1981), "The spatial resolving power of earth resources satellites", *Progress in Physical Geography*, Vol. 5, pp. 32–55.

Trimble Germany GmbH. (2011), *eCognition Developer 8.7 Reference Book*.

Tso, B. and Mather, P.M. (2009), *Classification methods for remotely sensed data*, CRC Press Taylor & Francis Group, 2nded.

Turker, M. and Arikan, M. (2005), "Sequential masking classification of multi-temporal Landsat7 ETM+ images for field-based crop mapping in Karacabey, Turkey", *International Journal of Remote Sensing*, Vol. 26 No. 17, pp. 3813–3830.

Turker, M. and Ozdarici, A. (2011), "Field-based crop classification using SPOT4 , SPOT5 , IKONOS and QuickBird imagery for agricultural areas: A comparison study", *International Journal of Remote Sensing*, Vol. 32 No. 24, pp. 37–41.

Tyc, G., Tulip, J., Schulten, D., Krischke, M. and Oxford, M. (2005), "The RapidEye mission design", *Acta Astronautica*, Vol. 56 No. 1-2, pp. 213–219.

Udelhoven, T., van der Linden, S., Waske, B., Stellmes, M. and Hoffmann, L. (2009), "Hypertemporal classification of large areas using decision fusion", *IEEE Geoscience and Remote Sensing Letters*, Vol. 6 No. 3, pp. 592–596.

UN. (2003), *World population prospects: The 2002 revision (Report)*, New York, p. 22.

UNDP. (2005), *Central Asia human development report (Report)*, Bratislava, p. 268.

UNDP. (2008), *Development & transition (Report)*, Bratislava, p. 24.

Vapnik, V. (1998), *Statistical learning theory*, Wiley, p. 768.

Verbeiren, S., Eerens, H., Piccard, I., Bauwens, I. and Van Orshoven, J. (2008), "Sub-pixel classification of SPOT-VEGETATION time series for the assessment of regional crop areas in Belgium", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 10 No. 4, pp. 486–497.

Vina, A. and Gitelson, A.A. (2005), "New developments in the remote estimation of the fraction of absorbed photosynthetically active radiation in crops", *Geophysical Research Letters*, Vol. 32 No. 17, doi:10.1029/2005GL023647.

Vintrou, E., Soumare, M., Bernand, S., Begue, A., Baron, C. and Lo Seen, D. (2012), "Mapping fragmented agricultural systems in the Sudano-Sahelian environments of Africa using random forest and ensemble metrics of coarse resolution MODIS imagery", *Photogrammetric Engineering & Remote Sensing*, Vol. 78 No. 8, pp. 839–848.

Wardlow, B.D. and Egbert, S. L. (2008), "Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains", *Remote Sensing of Environment*, Vol. 112 No. 3, pp. 1096–1116.

Wardlow, B.D., Egbert, S. L. and Kastens, J.H. (2007), "Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains", *Remote Sensing of Environment*, Vol. 108 No. 3, pp. 290–310.

Waske, B. and Benediktsson, J.A. (2007), "Fusion of support vector machines for classification of multisensor data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45 No. 12, pp. 3858–3866.

Waske, B., Benediktsson, J.A. and Sveinsson, J.R. (2009), "Classifying remote sensing data with support vector machines and imbalanced training data", in Benediktsson, J.A., Kittler, J. and Roli, F. (Eds.),*Multiple Classifier Systems, Lecture Notes in Computer Science Volume 5519*, Springer, pp. 375–384.

Waske, B. and Braun, M. (2009), "Classifier ensembles for land cover mapping using multitemporal SAR imagery", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 64 No. 5, pp. 450–457.

Waske, B. and van der Linden, S. (2008), "Classifying multilevel imagery from SAR and optical sensors by decision fusion", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46 No. 5, pp. 1457–1466.

Waske, B., van der Linden, S., Benediktsson, J.A., Rabe, A. and Hostert, P. (2010), "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48 No. 7, pp. 2880–2889.

Watts, J.D., Lawrence, R.L., Miller, P.R. and Montagne, C. (2009), "Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing", *Remote Sensing of Environment*, Vol. 113 No. 9, pp. 1843–1852.

Watts, J.D., Powell, S.L., Lawrence, R.L. and Hilker, T. (2011), "Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery", *Remote Sensing of Environment*, Vol. 115 No. 1, pp. 66–75.

Wesseling, J.G. and Feddes, R.A. (2006), "Assessing crop water productivity from field to regional scale", *Agricultural Water Management*, Vol. 86 No. 1-2, pp. 30–39.

Wichelns, D. and Oster, J.D. (2006), "Sustainable irrigation is necessary and achievable, but direct costs and environmental impacts can be substantial", *Agricultural Water Management*, Vol. 86 No. 1-2, pp. 114–127.

Wiesmann, D. and Quinn, D. (2013), "R package 'rasclass': Supervised raster image classification".

De Wit, A. and Clevers, J.G.P.W. (2004), "Efficiency and accuracy of per-field classification for operational crop mapping", *International Journal of Remote Sensing*, Vol. 25 No. 20, pp. 4091–4112.

Witten, I. and Frank, E. (2005), *Data Mining. Practical Machine Learning Tools and Techniques*, Elsevier, 2nded., p. 528.

Wolpert, D.H. (1992), "Stacked generalization", *Neural Networks*, Vol. 5 No. 2, pp. 241–259.

Woodcock, C.E. and Harward, V. (1992), "Nested-hierarchical scene models and image segmentation", *International Journal of Remote Sensing*, Taylor & Francis, Vol. 13 No. 16, pp. 3167–3187.

Woodcock, C.E. and Strahler, A.H. (1987), "The factor of scale in remote sensing", *Remote Sensing of Environment*, Vol. 21 No. 3, pp. 311–332.

Wu, H. and Li, Z.-L. (2009), "Scale issues in remote sensing: A review on analysis, processing and modeling", *Sensors*, Vol. 9 No. 3, pp. 1768–1793.

Wu, T.-F., Lin, C.-J. and Weng, R.C. (2004), "Probability estimates for multi-class classification by pairwise coupling", *Journal of Machine Learning Research*, Vol. 5, pp. 975–1005.

Yang, C., Everitt, J.H., Fletcher, R.S. and Murden, D. (2007), "Using high resolution QuickBird imagery for crop identification and area estimation", *Geocarto International*, Vol. 22 No. 3, pp. 219–233.

Yu, Q., Gong, P., Tian, Y., Pu, R. and Yang, J. (2008), "Factors affecting spatial variation of classification uncertainty in an image object-based vegetation mapping", *Photogrammetric Engineering & Remote Sensing*, Vol. 74 No. 8, pp. 1007–1018.

Zhang, J. (2010), "Multi-source remote sensing data fusion: Status and trends", *International Journal of Image and Data Fusion*, Vol. 1 No. 1, pp. 5–24.

Zhang, Y.J. (1996), "A survey on evaluation methods for image segmentation", *Pattern Recognition*, Vol. 29 No. 8, pp. 1335–1346.

Zhu, A.-X. (1997), "Measuring uncertainty in class assignment for natural resource maps under fuzzy logic", *Photogrammetric Engineering & Remote Sensing*, Vol. 63 No. 10, pp. 1195–1202.

APPENDIX

A.1: Imaging instruments used in crop mapping and monitoring

Extensive information about instruments including hyper-spectral, microwave and radar systems, Lidar, and airborne systems can be found in Kramer and Cracknell (2008) and Kramer (2001). One way to categorize satellite systems is by the scope of the study area, e.g. local, regional, and global (Ozdogan et al., 2010), but more commonly by resolution or pixel size, which is their ability to resolve high spatial frequencies (Schowengerdt, 2007). NASA´s Landsat fleet with its Multi-Spectral Scanner (MSS), TM and Enhanced Thematic Mapper (ETM+) instruments on board of the Landsat 1–7 platforms succeeds in observing the earth since the launch of the first Landsat satellite on 23-Jul 1972. In 2013 the Landsat 8 mission was launched to continue the legacy of the Landsat series. The ASTER sensor on board the Terra platform complements the fleet with high-resolution observation capabilities. These instruments provide information on pixel grids in the order of 60 m (e.g. MSS) to 15 m (e.g. ASTER, ETM+). The French SPOT satellites that include the HRV, HRVIR and HRG instruments with pixel grids between 20 m–2.5 m complement this. In 2008 the German RapidEye mission complemented these well-established passive optical systems. RapidEye is a constellation of five satellites that provide five band multi-spectral images with a ground sampling distance (GSD) of 6.5 m at nadir (Tyc et al., 2005). One of its features is its red edge canal that measures the spectral region between the red absorbance and the near infrared (NIR) reflection (690–730 nm) for vegetation characterization (Eitel et al., 2007; Read et al., 2002). Its high frequent revisit times, spatial resolution, and the red edge band predestine it for crop mapping (Conrad et al., 2011a; Löw et al., 2012; Schuster et al., 2012) and allows for assessing within field heterogeneity (Ehammer et al., 2010).

Table A-1: Overview of some commonly used sensor systems in crop mapping applications, sorted by their nominal ground sampling distance (at nadir). Asterisk indicates that the scale of application is given according to Ozdogan et al. (2010). Examples for applications of these systems in agricultural studies are given.

System	GSD [m]	Swath [km]	Revisit times [days]	Scale of application	Free of charge	Application
Quickbird	2.4	16.5	Irregular, on demand	Local	No	Myint et al. (2011), Turker and Ozdarici (2011), Yang et al. (2007)
Ikonos	4	11	Irregular, on demand	Local	No	Turker and Ozdarici (2011)
RapidEye	6.5	77	3–5, on demand	Local, regional	No	Conrad et al. (2011b), Löw et al. (2012)
SPOT-5 HRG	20	60	Irregular, on demand	Local*	No	Foody (2005), Murakami et al. (2001)
ASTER	15	60	Irregular, on demand	Local*	No	Conrad et al. (2010), Peña-Barragán et al. (2011)
Landsat-5 TM / 7 ETM+	30	180	16	Local*, regional*, continental*	Yes	El-Magd and Tanton (2003), Turker and Arikan (2005)
MODIS	250	2,330	1–3	Regional*, continental*, global*	Yes	Conrad et al. (2011a), Wardlow and Egbert (2008), Wardlow et al. (2007)

A.2: Classifier algorithms

The classification algorithms employed in this thesis belong to the group of supervised classifiers algorithms (Tso and Mather, 2009), which can further be distinguished in parametric and non-parametric. Supervised algorithms rely on the pre-existence of reference data (sometimes referred to as a priori knowledge), which contains class-specific information and data structure. As is beyond the scope of this thesis to evaluate or improve a broad range of classifier algorithms (let alone unsupervised approaches) the interested reader is referred to Richards and Jia (2005) and Tso and Mather (2009) for further details and additional references. This thesis employs three recent non-parametric algorithms, the RF, SVM and MLP. Further, one traditional probabilistic approach, the MLC, is introduced and will be used in the experiments for comparison. Each classifier algorithm used in this thesis is implemented in a "package" within the freely available programming environment R (R Development Core Team, 2012):

> Random forest: `randomForest` package (Liaw, 2013).

> Support vector machine: `e1071` package (Meyer et al., 2012).

> Multilayer perceptron: `MultilayerPerceptron` function in `RWeka` package (Hornik et al., 2013).

> Maximum likelihood classifier: `rasclass` package (Wiesmann and Quinn, 2013).

Random forest

The RF of Breiman (2001) belongs to the non-parametric machine learning algorithms. It consists of an ensemble of classification and regression tree (CART)-like structured classifiers (Breiman et al., 1984), based on bootstrapped samples of the training data (Figure A-1) (Breiman, 2001). It applies the ideas of bagging and random-subspace strategies to tree structured classifiers. Each tree within the forest is trained independently ("weak learners") from the others based on a subset of the original training samples, which is called the "in-the-bag" sample in bagging[31] (Breiman, 1996). The split rule at each split in the trees is determined using this randomly selected subset (with replacement) of the input data. The remaining, so-called "bootstrap sample" is put down the tree to generate a test classification, and to calculate the out-of-bag error (OOB error). The OOB prediction error can provide a more conservative classification error assessment than using a separate test set for the accuracy assessment (Gislason et al., 2006). OOB error rates tend to stabilize with an increasing number of trees in the ensemble (Hastie et al., 2011; Rodriguez-Galiano et al., 2012b).

Each tree in the ensemble is grown to its maximum depth without pruning using the Gini index of node impurity as a splitting criterion (Breiman et al., 1984), calculated as:

$$G = \sum_{i=1}^{n} p_i(1 - p_i) \hspace{4cm} Eq. \; i$$

where p_i is the probability or relative frequency of class i at a node, computed as:

$$p_i = \frac{n_i}{n} \hspace{5cm} Eq. \; ii$$

where n_i is the total number of samples belonging to i, and n the number of samples within one node. Gini ranges from [0,1] and reaches zero at a maximum level of homogeneity of a resulting split at a given

---

[31]: An acronym for bootstrap aggregation.

node. The Gini index determines for each node the best binary split. This is done by comparing the total impurity of child nodes that result from a possible split at a parent node to the parent Gini value, and the final split is successful when the Gini of the children is less than that of a parent (e.g. the split with maximum reduction in impurity). The tree reaches its maximum depth when the splitting reaches a Gini of zero, which means only one class is present at each terminal node.

Through the strategy of random permutation of training samples and features the computational complexity of the trees within the ensemble is kept simple and the correlation between the trees is decreased. This in turn enables the RF to handle multi-modal distributions, to deal with huge input feature spaces, and antagonizes data over-fitting that is associated with other tree bases classifier algorithms (Breiman, 2001). From a computational perspective RF is relatively effective when using huge feature sets, because the trees in the RF are only based on subsets of the input data (Gislason et al., 2006). The generation of n trees results in n class predictions C_n. The final class assignment to a case x (pixel or object, depending on the application) is based on fusion via majority voting:

$$C_{fin}(x) = majority\ vote\{C_n(x)\} \qquad\qquad\qquad Eq.\ iii$$
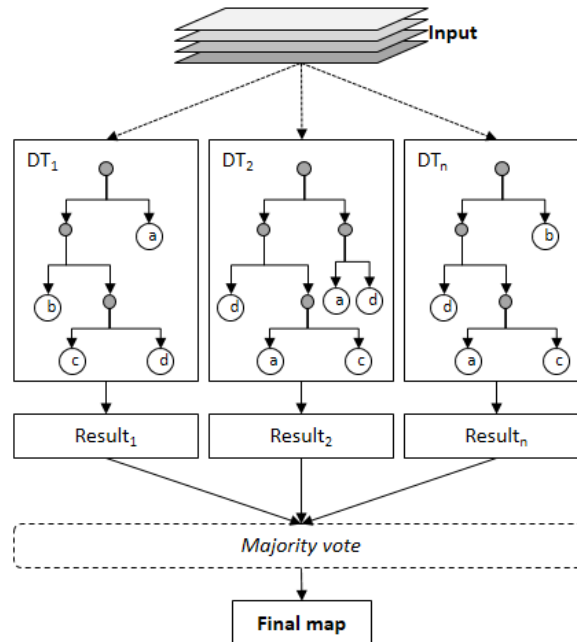
where $C_n(x)$ is the prediction of the nth tree.

Figure A-1: Schematic overview of a RF, separating four classes (a-d).

Studies have shown the good performance of RF in land cover and crop mapping applications (Brown et al., 2013; Chan et al., 2008; Pal, 2005; Rodriguez-Galiano et al., 2012a, 2012b; Vintrou et al., 2012; Waske and Braun, 2009; Watts et al., 2011; Conrad et al., 2013; Löw et al., 2012; Watts et al., 2009). RF were shown to be robust against huge number of input features, and can be used for feature selection (Pal and Foody, 2010). The application of RF requires two user-defined parameters to be optimized: number n of trees in the ensemble, and the number of features used for slitting each node in the trees. The number of trees should be high enough to assure convergence of the OOB error estimates, whilst the number of features per split is usually set to the square root of all features (Gislason et al., 2006). In general 500 trees per ensemble give reasonable results (Gislason et al., 2006). In remote sensing-based applications the influence of these parameter settings on the output (e.g. classification accuracy) was demonstrated in Chan et al. (2008), Duro et al. (2012), and Rodriguez-Galiano et al. (2012b) and should not be neglected.

Support vector machines

In remote sensing applications SVM became very popular because of their ability to handle small training sets and high-dimensional data, and many applications can be found with hyper-spectral data sets (Giacco et al., 2010; Melgani and Bruzzone, 2004; Mountrakis et al., 2011). They were shown to perform competitive or even more accurate than other classifier algorithms (Huang et al., 2002a; Melgani and Bruzzone, 2004; Shao and Lunetta, 2012; Waske and Benediktsson, 2007). Emerging applications for multi-spectral and multi-temporal data sets are given in Huang et al. (2002a) and Shao and Lunetta (2012) and for hyper-spectral data in Melgani and Bruzzone (2004). More recent applications of SVM in crop classification can be found in Foody et al. (2006), Löw et al. (2012, 2013a), Mathur and Foody (2008), Waske and Benediktsson (2007), and Waske et al. (2009). Burges (1998) provides a detailed introduction of SVM to the remote sensing community and Mountrakis et al. (2011) summarized empirical results from over 100 articles using the SVM.

SVM is a non-parametric classification approach that can separate multi-modal class distributions in high-dimensional feature spaces by using kernel functions, where the optimization problem is solved based on structural risk minimization (SRM) (Cortes and Vapnik, 1995). SVM is based on the notion of fitting an optimal separating hyper-plane (OSH) to the training data of two classes within the multi-dimensional feature space (Foody and Mathur, 2004), and to maximize the margins between the OSH and the closest training samples (the so-called support vectors). SVM focus only the training samples that are closest to the edge of the class distributions in the feature space (Mathur and Foody, 2008). Based on the mapping of the data into a higher dimensional Hilbert feature space ("kernel-trick") the OSH can be fit to a more complex class distribution, which might not be separable in the original input space. In this regard one of the most appealing properties of SVMs is their high capacity for generalization with relatively small training data set sizes (Mathur and Foody, 2008).

Two broad strategies for SVM exist: OAO and OAA. By undertaking a set of $n(n-1)/2$ classifications however (where n = number of classes) a multi-class classification can be derived, the so-called OAO approach, and finally the outputs from the individually trained binary classifiers are combined by a majority voting strategy (Hsu and Lin, 2002). The other approach is OAA where the classification problem is split into n classifications, where in each binary classification one class is separated from all the others. Although the OAO approach suggests a higher number of SVMs to be trained the computation time is usually less compared to OAA, because the binary problems are smaller and the optimization problem scales super-linearly (Karatzoglou et al., 2006). Further, the OAO approach was shown to perform very accurate when solving classification problems with many classes (Hsu and Lin, 2002).

A more rigorous and detailed description of SVM is given in Burges (1998). In the following the basis of SVM classification is illustrated by the example of a binary classification problem for linear separable cases. Let for the binary classification problem with two classes in a n-dimensional feature space $\Re^n$, $x_i \in \Re^n$, $i = 1,2,\dots M$ be a training set of M samples with class labels $y_i \in \{1,-1\}$. A hyperplane separates these two classes and is defined by the decision function

$$f(x) = w * x + b \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{Eq. iv}$$

where x is a point on the hyperplane, w is a normal vector to the hyperplane, b the bias, and $|b|/\|w\|$ is the distance between the OSH and the origin.

The support vectors lie on two hyperplanes H_1 and H_2, which are in turn parallel to the OSH, and are defined by

$$wx_i + b = \pm 1. \hspace{4cm} Eq.\ v$$

The resulting margin between these two hyper-planes is $\frac{2}{\|w\|}$. The aim in SVM classification here becomes to maximize the margin, based on the following optimization problem

$$min\left\{\frac{\|w\|^2}{2} + C\sum_{i=1}^{M}\xi_i\right\} \hspace{3cm} Eq.\ vi$$

where ξ_i is the slack variable that indicate the distance the sample is from the hyperplane H_1 or H_2 that passes through the support vectors of the class to which the sample belongs and C the regularization parameter that is added as a penalty for misclassified cases that lie on the wrong site of the OSH. The extension of this basic SVM approach to non-linear decision surfaces (see Figure A-2) can be realized by so-called kernel methods. This is based on a non-linear transformation of the input data to a higher dimensional Hilbert feature space. Through this "kernel-trick" the distribution of the data points is spread in a way that allows for fitting of a linear hyper-plane to a complex class distribution (Foody and Mathur, 2004), which is generally not separable in the original data space. This transformation process is done using kernels, which must meet Mercers conditions (Cortes and Vapnik, 1995). The final hyper-plane decision function then becomes

$$f(x) = sgn\left(\sum_{i=1}^{M}\alpha_i y_i\, k(x_i x_j) + b\right) \hspace{2cm} Eq.\ vii$$

where α_i denote Lagrange multipliers. Figure A-2 illustrates the concept of a SVM for a linearly non-separable case. Kernel functions that, are commonly used in remote sensing applications are RBF, polynomial, and linear kernels:

Gaussian radial basis function kernel:

$$k(x_i, x_j) = exp\left(-\gamma\|x_i - x_j\|^2\right) \hspace{2.5cm} Eq.\ viii$$

where γ controls the width of the Gaussian kernel.

Polynomial kernel:

$$k(x_i, x_j)' = \left(scale * \langle x_i, x_j\rangle' + offset\right)^{degree}. \hspace{1.5cm} Eq.\ ix$$

Linear kernel:

$$k(x_i, x_j)' = \langle x_i, x_j\rangle'. \hspace{3.5cm} Eq.\ x$$

The original output of a SVM is the distances of each case x to the OSH. Each SVM will, for a given sample x, compute a distance $d_i(x), i = 1,2,\dots,L$ that x has to the hyperplane. The distances can be used to determine the final class label, depending on the multi-class strategy chosen: in the OAO approach a set of $n(n-1)/2$ distances is created and can be transferred to class labels (Chang and Lin, 2011).
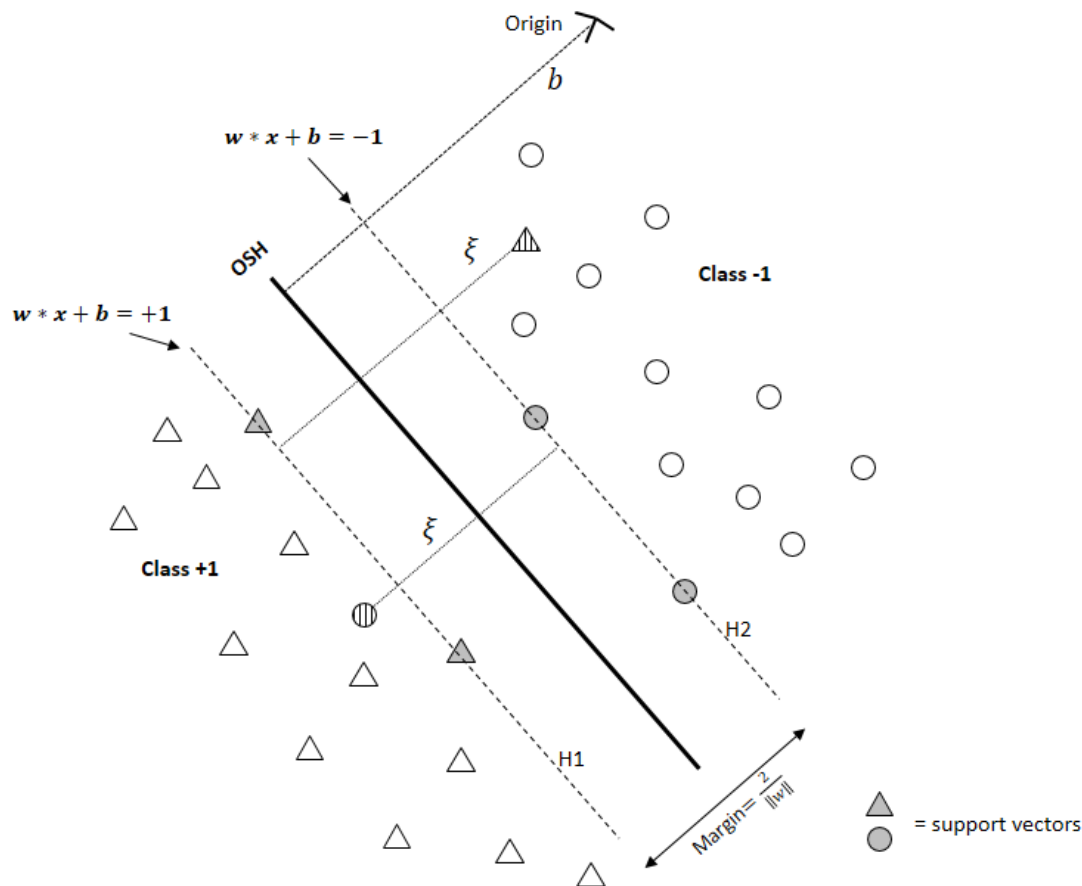
Figure A-2: Concept of a SVM for a linearly non-separable case of classes x_i and x_j. The circles and triangles indicate the samples of class x_i and x_j, respectively. The grey samples refer to the support vectors, which lie on the two separating hyper-planes H_1 and H_2. The OSH is between these two functions (hyper-planes).

Multilayer perceptron neural networks

ANNs are a group of non-parametric classifier algorithms that can solve complex non-linear multi-class problems (Richards and Jia, 2005). ANN are composed of an interconnected group of nodes, which use mathematical methods to process information (Hu and Weng, 2009). There exist a variety of ANN models (or "architectures"), e.g. self-organizing feature maps (SOM), counter-propagation networks, and Hopfield networks. However the feed-forward MLP with back-error propagation (BP) is one of the most widely used neural network architectures in remote sensing applications (Benediktsson et al., 1990; Foody, 1995b; Kavzoglu and Mather, 2003).

The MLP is characterized by a typical layer structure, where each layer has several nodes ("artificial neurons"), the basic units of the MLP. The MLP is a self-adaptive system whose nodes were designed to mimic some of the functions of human brains, by summing the input information and performing a transformation with a mathematical function on the summed input. The widely used three-layer structure of the MLP (Figure A-3) has one input layer (with as many nodes as input features), one or more hidden layer(s) (for the actual data processing, e.g. classification), and one output layer (with as many nodes as classes). The nodes within each layer are connected with each other, but without interconnections between the neurons within the same layer. Information passing the layers is modified by numerical weights that are assigned to the node-interconnections, the so-called neurons. The outputs at each output layer node are called "activities" (Tso and Mather, 2009). During an iterative training process the training samples (vectors with the feature values from the training sample) are input into

the MLP model and the activities of the neurons are updated layer-by-layer (from input to output layer). The most commonly used algorithm for updating the neuronal activities and weights is the BP algorithm (Rumelhart et al., 1986; Tso and Mather, 2009). During the forward pass the input values are multiplied by weights associated with the nodes and summed, resulting in the net input x_j received by a single neuron j, characterized with:

$$x_j = \sum_i a_i w_{ji} \qquad\qquad\qquad Eq.\ xi$$

where a_i is the output (activity) from a single node j, and w_{ji} the weight of an interconnection between the jth and ith nodes. Adding a bias to the network affects each neuron using different values and changes to

$$x_j = \sum_i a_i w_{ji} - \theta_j \qquad\qquad\qquad Eq.\ xii$$

where θ_j is a bias for each neuron j. A non-linear sigmoid function is commonly used to transform x_j to an output value or "activity", and is defined as

$$a_j = \frac{1}{1+\frac{1}{exp\left(\frac{x_j}{p}\right)}} \qquad\qquad\qquad Eq.\ xiii$$

where p is a parameter that influences the shape of the sigmoid curve (Tso and Mather, 2009). After the forward pass of the input data the activities of the output layer neurons are compared with the input (e.g. their expected activities). The difference of the calculated output from the expected output is calculated with an error function, and is called "network-error". If the difference is larger than the initial threshold value, the mapping algorithm updates the weights in order to minimize the difference between the actual and the desired output (the network error). A backward pass of the information on the network-error from the output layer, thereby updating the weights, accomplishes this and the whole process is repeated until certain criterions are fulfilled, e.g. a predefined level of accuracy or the maximum number of iterations is achieved.
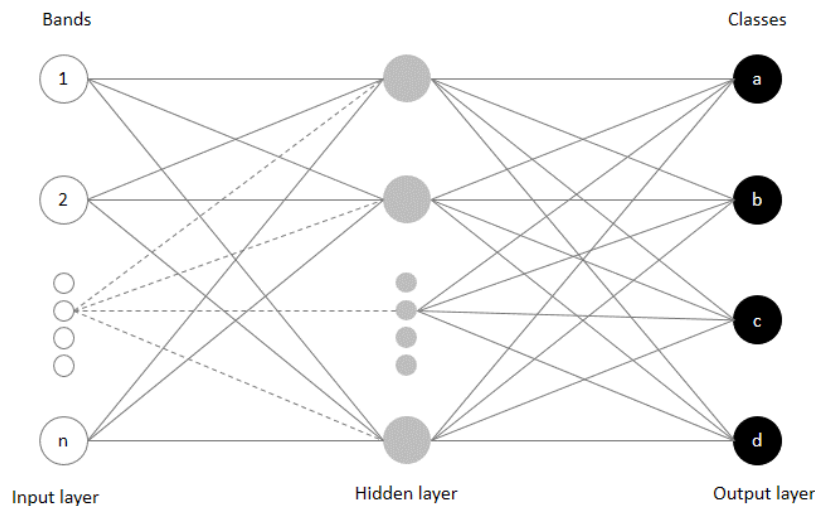

Figure A-3: Exemplary structure of a MLP neural network with the typical three layer structure: one input layer with n nodes (corresponding to the number of input features), one hidden layer with four nodes for classification, and one output layer with four nodes (corresponding to four classes a−d).

The performance of MLP depends on the way it is trained, because the number of free parameters that have to be defined by the user in the BP algorithm is larger than in SVM. They can have a significant impact on the performance of the model, and more precisely affect the ability of the networks to generalize and maximize classification accuracy (Tso and Mather, 2009). Important parameters to define are the initial weights, number of hidden layer nodes, momentum factor, and the learning rate. Further a potential effect of the feature space on the performance of the neural network has to be considered (Foody and Arora, 1997). Too few hidden layers can cause under-fitting of the network, caused by the inability of small networks to fully explore the data structure present in the training data, whilst in too large networks the decision boundaries in feature space may be too influenced by the specific properties of the training data and loose its generalizability. Tso and Mather (2009) suggest that one hidden layer can suffice to solve most classification problems with relatively few classes. A too high learning rate will cause the model to be instable and prevent it from converging, and with a too low value the model might end up finding a local minimum (Kavzoglu and Mather, 2003). Kavzoglu (2001) proposed leaning rates in the range [0.1, 0.2], momentum factors in the range [0.5, 0.6], and the initial values of the weights in the range [-0.25, 0.25]. Further in BP the number of iterations must be chosen to prevent over-fitting, which can result in a loss of generalizability of the model. In general a higher number of iterations lead to higher accuracies but can result in over-fitting to the training data structure (Atkinson and Tatnall, 1997). The use of neural network classifiers have frequently been reported in remote sensing classification applications (Atkinson and Tatnall, 1997; Berberoglu et al., 2000; Bruzzone et al., 1997, 1999; Foody, 2004; Kavzoglu and Mather, 2003; Kavzoglu, 2009; Linderman et al., 2004).

Maximum likelihood classification

The MLC is a well-known statistical (parametric) supervised classification method. MLC is a particular case of Bayes' decision rule when classes are assumed to have equal priorities. It assumes that the probability density function of each class is multivariate and has a Gaussian distribution. A more detailed and rigorous description of MLC and its underlying statistics is given in Richards and Jia (2005), and in the following an overview is given.

MLC assumes that each class has a class-specific probability density function. Let assume there are $\omega_i, i = 1, \dots n$ classes, where n is the total number of classes. The membership of a sample vector x (e.g. a pixel) to a particular class ω_i in a d-dimensional feature space $\mathfrak{K}^d, x = \{x_1, x_2, \dots x_d\}$ is defined by the conditional probability $p(\omega_i, x), i = 1, \dots n$. This gives the likelihood or posterior probability that x belongs to a class ω_i. The decision rule for classification then is

$$x \in \omega_i, \quad if \quad p(\omega_i, x) > p(\omega_j, x). \qquad\qquad\qquad Eq.\ xiv$$

That means x belongs to class ω_i if the conditional probability $p(\omega_i, x)$ of that class is maximum, yet these class-conditional densities $p(\omega, x)$ are usually not known a priori for the whole image and estimated from reference samples of each class (Richards and Jia, 2005). The class-specific information and data structure are estimated from reference samples, which are used to train the classifier. The resulting probability density functions (PDF) $p(x, \omega_i)$ are estimated for each class (resulting in as many functions as classes). For each x (e.g. pixel vector) there exists a set of n PDFs, each giving the relative likelihood that this case belongs to each available class. The relationship between the known $p(x, \omega_i)$ and the required $p(\omega_i, x)$ is modelled as follows (Richards and Jia 2005):

$$p(\omega_i, x) = p(x, \omega_j)p(\omega_i)/p(x)$$ *Eq. xv*

where $p(\omega_i)$ is the so called a priori probability that a class ω_i occurs in the imagery, and is estimated from the training data before classification.

The final class assignment then is calculated as:

$$\omega \in x_i, \quad if \quad p(x, \omega_i)p(\omega_i) > p(x, \omega_j)p(\omega_j),$$ *Eq. xvi*

where $p(x, \omega_i)$ and $p(\omega_i)$ are estimated from the training data. Although MLC is an easy to use method, its application in many image classification tasks is hampered by its lacking ability to handle multi-source or multi-temporal data. This is because the PDFs are assumed to have Gaussian distributions when estimating $p(x, \omega_i)$, which is done by the class specific mean vector and the covariance matrix. However multi-modal input data sets cannot be modelled by means of convenient statistical models, and MLC requires a complete description of each class from the feature space, which requires large training data sets (Richards and Jia, 2005).

Despite its limitations MLC have frequently been used for crop mapping, for instance in mono-temporal applications (El-Magd and Tanton, 2003; Lobo et al., 1996). Multi-temporal crop classification with high-resolution time series is demonstrated in Fritz et al. (2008) and Murakami et al. (2001). Blaes et al. (2005) combined optical and SAR time-series in a nested classification approach using MLC. More recent studies however demonstrate that non-parametric machine learners generally outperform the MLC in crop mapping, irrespective of the application design or input data used (Löw et al., 2012; Murthy et al., 2003; Waske and Benediktsson, 2007). Although the MLC assumes a Gaussian distribution and is supposed not to be ideal for classifying multi-temporal datasets, it is used in this thesis as a benchmark because MLC still is a very common supervised classifier implemented in many commercial remote sensing software packages, and because of its simplicity and lack of exhaustive parameter tuning.

CURRICULUM VITAE

Persönliche Daten

Fabian Löw, geboren am 23. Juli 1983 in Heidelberg

Studium und Berufserfahrung

Seit 10/2013	*Lehrstuhl für Fernerkundung, Universität Würzburg* Wissenschaftlicher Angestellter im Projekt CAWA (Central Asian Water)
10/2010 – 09/2013	*Lehrstuhl für Fernerkundung, Universität Würzburg* Promotionsstipendium der Studienstiftung des deutschen Volkes
12/2009 – 09/2010	*Spang. Fischer. Natzschka GmbH, Walldorf* Werksstudent / Wissenschaftlicher Mitarbeiter
09/2004 – 12/2009	*Ruprecht-Karls-Universität Heidelberg* Studium der Geographie
08/2003 – 06/2004	Zivildienst
07/2003	Abitur in Schriesheim

Erklärung zur selbstständigen Verfassung der vorliegenden Arbeit

Hiermit erkläre ich, Fabian Löw, geboren am 23.07.1983 in Heidelberg, dass ich die vorliegende Arbeit mit dem Titel "*Agricultural crop mapping from multi-scale remote sensing data - Concepts and applications in heterogeneous Middle Asian agricultural landscapes*" selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Würzburg, 26. November 2013   Fabian Löw