Spatial-temporal Dynamic Conditional Random Fields crop type mapping using radar images

Heft 52
Schriftenreihe der Fachrichtung Geodäsie
Fachbereich Bau- und Umweltingenieurwissenschaften
Technische Universität Darmstadt
ISBN 978-3-935631-41-9

Darmstadt, September 2016

TECHNISCHE UNIVERSITÄT DARMSTADT

Heft 52

Darmstadt, September 2016

Benson Kipkemboi Kenduiywo

Spatial-temporal Dynamic Conditional Random Fields crop type mapping using radar images

Schriftenreihe
Fachrichtung Geodäsie
Fachbereich Bau- und Umweltingenieurwissenschaften
Technische Universität Darmstadt

ISBN 978-3-935631-41-9

Schriftenreihe Fachrichtung Geodäsie der Technischen Universität Darmstadt

Online unter: http://tuprints.ulb.tu-darmstadt.de/id/user/7098
Diese Arbeit ist gleichzeitig veröffentlicht in Reihe C der
Deutschen Geodätischen Kommission, München 2016

Verantwortlich für die Herausgabe der Schriftenreihe:

Institut für Geodäsie
im Fachbereich Bau- und Umweltingenieurwissenschaften
der Technischen Universität Darmstadt

"Gedruckt mit Unterstützung des Deutschen Akademischen Austauschdienstes"

Bezugsnachweis:

Technische Universität Darmstadt
Institut für Geodäsie
Franziska-Braun-Straße 7
64287 Darmstadt
ISBN: 978-3-935631-41-9

Spatial-temporal Dynamic Conditional Random Fields crop type mapping using radar images

Dem Fachbereich Bau- und Umweltingenieurwissenschaften
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Dissertation

von
M.Sc. Benson Kipkemboi Kenduiywo
aus Nakuru, Kenya

Referent:	Prof. Dr.-Ing. Matthias Becker
Korreferent:	Prof. Dr.-Ing. Uwe Soergel
Korreferent:	Prof. Dr.-Ing. Andreas Eichhorn
Tag der Einreichung:	15. Juni 2016
Tag der mündlichen Prüfung:	29. September 2016

Darmstadt, September 2016
D17

Acknowledgements

All glory to GOD, the creator of everything and beholder of all mysteries humans can not fathom, for the gift of life, care and blessing you bring my way daily.

My utmost appreciation to my supervisors, Prof. Dr.-Ing. Matthias Becker and Prof. Dr.-Ing. Uwe Sörgel for scientific advice and guidance throughout the research. The members of staff at Institute of Geodesy for various office support during research. More also, the immediate support and discussions from staff in Remote Sensing and Image Analysis research group.

I am indebted to Dr. Felx Ng'etich for providing advice on agriculture systems in Kenya. Mr. Kuria Thiong'o for collaborative field work and data sharing in Kitale, Kenya. All agriculture extension officers I worked with during field work. More also, I thank managements of Kenya Seed Company Ltd, Agricultural Development Corporation Kenya, Panocal flowers amongst other farmers for giving me access to their farms and farm information during field work. The field work and 90% of my scholarship was funded by National Commission for Science,Technology and Innovation (NACOSTI) under Kenyan Ministry of Education and the remaining 10% of the scholarship was financed by Deutscher Akademischer Austauschdienst (DAAD). TerraSAR-X data of Fuhrberg, Germany were provided by Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Sentinel 1 data by European Space Agency (ESA). Reference data campaign of Fuhrberg was funded by German Federal Ministry of Education and Research (Project 50EE1326). In Kitale, RapidEye images were provided by RapidEye Science Archive (RESA) project number 00140 and TerraSAR-X provided under project number LAN2759.

My appreciation to my parents, the late James A. Kenduiywo and Agnes A. Kenduiywo for their love and support throughout my life, my sister Nancy and brother Denis who have always encouraged me, and most importantly my lovely wife Terry, my son Ethan and our new family member Janelle to whom I dedicate this work.

Inspiration:

"*When you have eliminated the impossible, whatever remains, however improbable, must be the truth.*"
Sherlock Holmes.

Abstract

The rapid increase in population in the world has propelled pressure on arable land. Consequently, the food basket has continuously declined while the global demand for food has grown twofold. There is need for continuous monitoring and spatial information update of agriculture activities. This will support decision and policy making organs to take necessary actions towards enhancing food security. However, economic factors, farm management, natural aspects (such as weather, soils e.t.c.) and government policy for instance, influence types of crops planted in a season. Therefore, data acquisition and mapping methods need to consider these dynamics. The study adopts microwave remote sensing with synthetic aperture radar (SAR) for data acquisition. Microwave remote sensing is daylight and weather independent thus guarantees the highest temporal density of images regardless of climatic zones. This also means that images at different phenological stages can be captured by radar sensors.

Crop phenology is dynamic as it changes spatially in different times of the year. Such biophysical processes also look spectrally different to radar sensors. Some crops may depict similar spectral properties if their phenology coincide, but differ later when their phenology diverge. Thus, crop mapping methods using single-date remote sensing images can not offer optimal results in case of crops with similar phenology. In addition, methods stacking images within a cropping season for classification limits discrimination to a single high dimensional feature space vector that can suffer from overlapping classes. However, phenology can aid discrimination of crops since their backscatter varies with time. Therefore, this research seeks to fill this gap by developing a crop sequence classification method using multitemporal SAR images. The method is built to use spatial and temporal context.

The study designed first order and higher order undirected Dynamic Conditional Random Fields (DCRFs) for spatial-temporal crop classification. Basically, the DCRFs model has a repeated structure of temporally connected conditional random fields (CRFs). Each node in the sequence is connected to its temporal neighbours via conditional probability matrix. The matrix is computed using posterior class probabilities estimated by random forest classifier. We use the matrix on one hand to encode expert and image based phenological information in higher order DCRFs. On the other hand, the matrix integrates only image based phenological information in first order DCRFs. When compared to independent epoch classification, the designs improved crop discrimination at each epoch with higher order DCRFs having the highest accuracy in the sequence. However, stakeholders and policy makers need to know the quantity and spatial coverage of crops in a given season so as to ensure food security and a balanced ecosystem. Therefore, we went an extra step to develop a DCRFs ensemble classifier. The DCRFs ensemble considers a set of computed posterior crop type probabilities at each epoch in order to generate an optimal label of a node. This is done by maximizing over posterior crop type probabilities selected from the sequence based on maximum F1-score and weighted by user accuracy. Our ensemble technique is compared to standard approach of stacking all images as bands for classification using maximum likelihood classifier (MLC) and CRFs. So far it outperforms MLC and CRFs using crop type posterior probabilities estimated by both first and higher order DCRFs.

Zusammenfassung

Der rasche Anstieg der Weltbevölkerung führte zu einem erhöhten Druck auf landwirtschaftliche Flächen. Die Folge ist ein verringertes Nahrundmittelangebot bei zeitgleicher Verdopplung der Nachfrage. Dadurch entsteht ein Bedarf nach kontinuierlichem Monitoring zur Laufendhaltung der räumlichen Information über die Nutzung von Agrarflächen als eine wichtige Voraussetzung zur Unterstützung entscheidungstragenden gesellschaftlichen Organe bei der Sicherung der Versorgung mit Nahrungsmitteln. Die Wahl der angebauten Kulturen wird dabei von ökonomischen und ökologischen (z.B. Klima, Boden etc.) Aspekten, des Weiteren der Bewirtschaftungsweise und politischen Entscheidungen beeinflusst. Diese durchaus dynamischen Faktoren gilt es bei der Auswahl von Daten und Methoden zur kartografischen Erfassung von Feldfrüchten zu berücksichtigen. Die vorliegende Studie nutzt bildgebendes Radar (Synthetic Aperture Radar, SAR) zur Datenerfassung. Mikrowellenbasierte Fernerkundung ist unabhängig von Wetter- sowie Tageslichtverhältnissen und garantiert daher in allen klimatischen Zonen der Erde eine höchstmögliche zeitliche Dichte der Daten. Damit besteht die Möglichkeit, Aufnahmen aus allen phänologischen Entwicklungsstadien der einzelnen Feldfruchtarten zu liefern.

Die Phänologie der Feldfrüchte weist eine starke räumliche und zeitliche Dynamik im jährlichen Wachstumszyklus auf. Die hier auftretenden biophysikalischen Prozesse führen zu unterschiedlichen spektralen Reflektionseigenschaften im Radarbild. Die spektralen Eigenschaften verschiedener Feldfrüchte können bei gleichen Phänologien zu zeitweise identischen Messungen führen, in späteren Entwicklungsstadien jedoch variieren. Dementsprechend führen Klassifikationsverfahren mittels Bildern, die lediglich zu einem bestimmten Zeitpunkt des Pflanzenwachstums aufgenommen wurden, nicht zu optimalen Ergebnissen im Falle ähnlicher Phänologien. Zusätzlich hierzu führt ein sehr großer Merkmalsraum, in den alle Aufnahmen der Vegetationsperiode einfließen, zu einer Überlappung der Klassen und somit ebenfalls zu einer schlechten Trennbarkeit. Die Phänologie der Pflanzen kann jedoch insofern hilfreich bei der Klassentrennung sein, als dass der gemessene Rückstreukoeffizient zeitlich variiert. Das Ziel dieser Studie ist daher die Entwicklung eines Klassifikationsverfahrens, welches phänologische Sequenzen der Feldfrüchte unter Verwendung von Radarbildern berücksichtigt. Die Methode ist dabei derart konstruiert, dass sie den zeitlichen und räumlichen Kontext berücksichtigt.

Im Rahmen der Forschungsarbeit wurde ein Ansatz entwickelt, der auf Zufallsfeldern erster und höherer Ordnung (Dynamic Conditional Random Fields, DCRFs) basiert. Das DCRFs-Modell hat die Struktur sich wiederholender, zeitlich verbundener Zufallsfelder (Conditional Random Fields, CRF), die als ungerichtete Graphen definiert sind. Jeder Knoten der Sequenz ist über eine Matrix von bedingten Wahrscheinlichkeiten mit seinem jeweiligen zeitlichen Nachbarn verbunden. Die Berechnung der Matrix erfolgte über die Bestimmung der A-posteriori-Wahrscheinlichkeiten mittels des Random Forests Klassifikators. Die Matrix wird einerseits zur Implementierung expertenbasierter und phänologischer Information in den DCRFs höherer Ordnung verwendet. Zugleich ermöglicht sie die Berücksichtigung phänologischer Information ausschließlich aus den Daten bei Anwendung der DCRFs ersten Ordnung. Ein Ergebnisvergleich der Klassifikation einzelner Epochen zeigt, dass die Verwendung der DCRFs höherer Ordnung zu den höchsten Klassifikationsgenauigkeiten für alle Epochen der Sequenz führt. Beteiligte Akteure und Politiker benötigen oftmals quantitative Informationen über die vorhandenen Feldfrüchte sowie deren räumliche Verteilung zur Sicherstel-

lung der Nahrungsmittelproduktion und eines ausgeglichenen Naturhaushaltes. Aus diesem Grund wurde in einem weiteren Arbeitsschritt ein DCRFs-Ensemble-Klassifikationsverfahren entwickelt. Dieses berücksichtigt einen Satz berechneter A-posteriori-Wahrscheinlichkeiten für jede Klasse zu jeder Epoche mit dem Ziel, eine optimale Klassenentscheidung für einen Knoten zu treffen. Hierzu werden die A-posteriori-Wahrscheinlichkeiten der Sequenzen maximiert, nachdem sie basierend auf dem maximalen F1-score ausgewählt und mit den Nutzergenauigkeiten gewichtet wurden. Diese Ensemble-Technik wird anschließend mit Standardverfahren verglichen, bei denen alle Aufnahmen in einen Merkmalsraum zusammengeführt und mit dem Maximum Likelihood Klassifikator (MLC) sowie den CRFs klassifiziert werden. Es zeigt sich, dass mit dem entwickelten Verfahren bessere Ergebnisse erzielt werden als mit MLC und CRFs. Dies gilt sowohl für die Berechnung der feldfruchtspezifischen A-posteriori-Wahrscheinlichkeit mit den DCRF erster Ordnung als auch mit jenen der höheren Ordnung.

Contents

List of Figures

List of Tables

Acrynomns, abbreviations and symbols

Acronyms

Radar	Radio detection and ranging
RAR	Real Aperture Radar
SAR	Synthetic Aperture Radar

Abbreviations

A	association potential
CRF(s)	Conditional Random Field(s)
DCRF(s)	Dynamic Conditional Random Field(s)
EM	electromagnetic
GMs	Graphical Models
HMM	Hidden Markov Models
I	interaction potential
LBP	Loopy Belief Propagation
MLC	maximum likelihood classification
MRF(s)	Markov Random Field(s)
OA	Overall Accuracy
RF	Random Forests
SVM	Support Vector Machines
TP	temporal potential

Symbols		**Units**
α	aperture angle	rad
β	spatial interaction parameter	

κ	kappa statistic	
λ	wavelength of an electromagnetic wave	m
ϕ	expert based phenology weight vector	
ψ	higher order DCRFs potential weights	
ρ	Pearson correlation coefficient	
τ	pulse duration	s
θ_i	incident angle	deg
θ_l	sensor look angle	deg
B	bandwidth	Hz
D	beamwidth	m
f	frequency of an electromagnetic wave	Hz
H	sensor elevation above ground scene	m
R	slant range distance	m
r	spatial resolution at nadir	m
r_a	azimuth resolution	m
r_r	range resolution	m
v	velocity of light (in vacuum)	m/s

1 Introduction

1.1 Background

Agriculture sustains livelihoods worldwide and thus bears an economic and a social value. With the current world's population approaching six billion and expected to increase by three billion in the next five decades, the global food scenario is changing rapidly. Pressure exerted on the already limited arable land resource is greater than ever before (Seelan et al., 2003). This, coupled with challenges of balancing modern competing land-use interests, has resulted to a decline in food. Foresight by World-Bank (2011) estimates that food demand since 2010 has increased resulting into extreme poverty of about 44 million people. Currently, about a billion people face severe malnourishment (see Figure 1.1) while agricultural systems are subjected to concurrent degrading land, water, biodiversity and climate on a global scale (Foley et al., 2011). Ironically, a third of the food harvested is lost or wasted each year (FAO, 2013). This can be minimized if governments and related stakeholders are updated on expected food outputs so that they can make advanced consumption and storage plans especially for perishable crops. Therefore, mitigation measures are necessary in order to counter this trend. An alternative is to craft sound policies and risk management strategies to monitor and estimate food production.

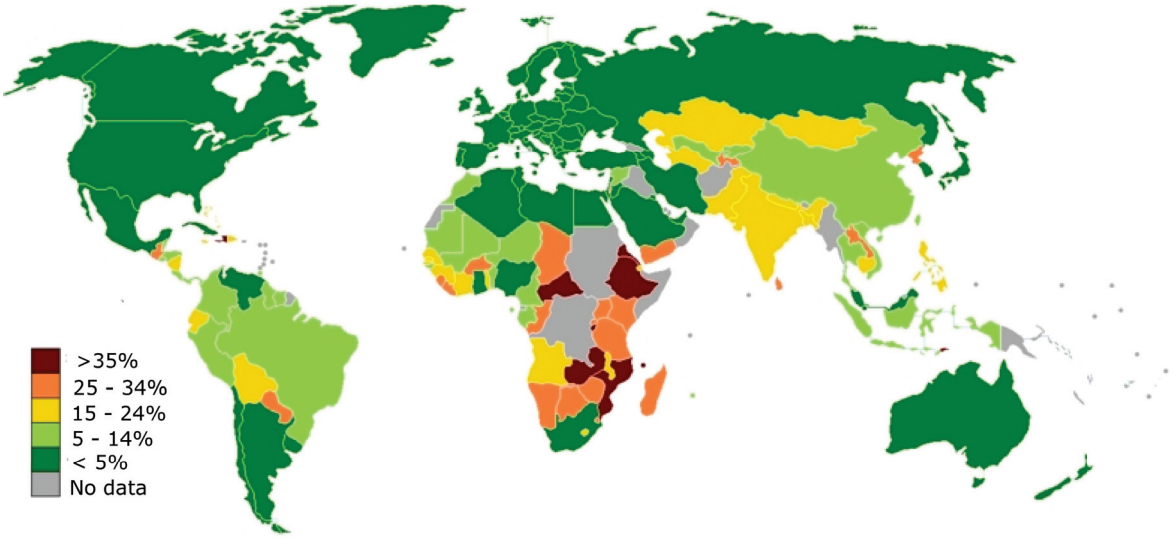


Figure 1.1.: Percentage distribution of undernourished population; source World Food Programme (2013).

To monitor and estimate food production, up-to-date precise crop type spatial information is required. Crops occupy roughly 38% of the Earth's surface which is the largest use to date (Ramankutty et al., 2008). So far only 62% of crops produce is for human consumption viz-a-vis 35% used as animal feed (which may end up as human food indirectly as meat and dairy products) and 3% for bioenergy, seed and other industrial products (Figure 1.2). Governments and related stakeholders must thus confront the challenge of feeding a growing world population while ensuring a more environmentally sustainable path is maintained despite existing stress on land resources. Such critical competing interests can be monitored spatially and temporally using remote sensing. Recent Earth observing satellites have undergone improved spatial, spectral and temporal resolutions. It is now possible to acquire frequent images within crop growing season. Optical images are able to distinguish different crop types. However, their applicability is limited to daylight and cloud free weather conditions. Therefore, crop monitoring can benefit from Synthetic Aperture Radar (SAR) images which are daylight and weather independent.

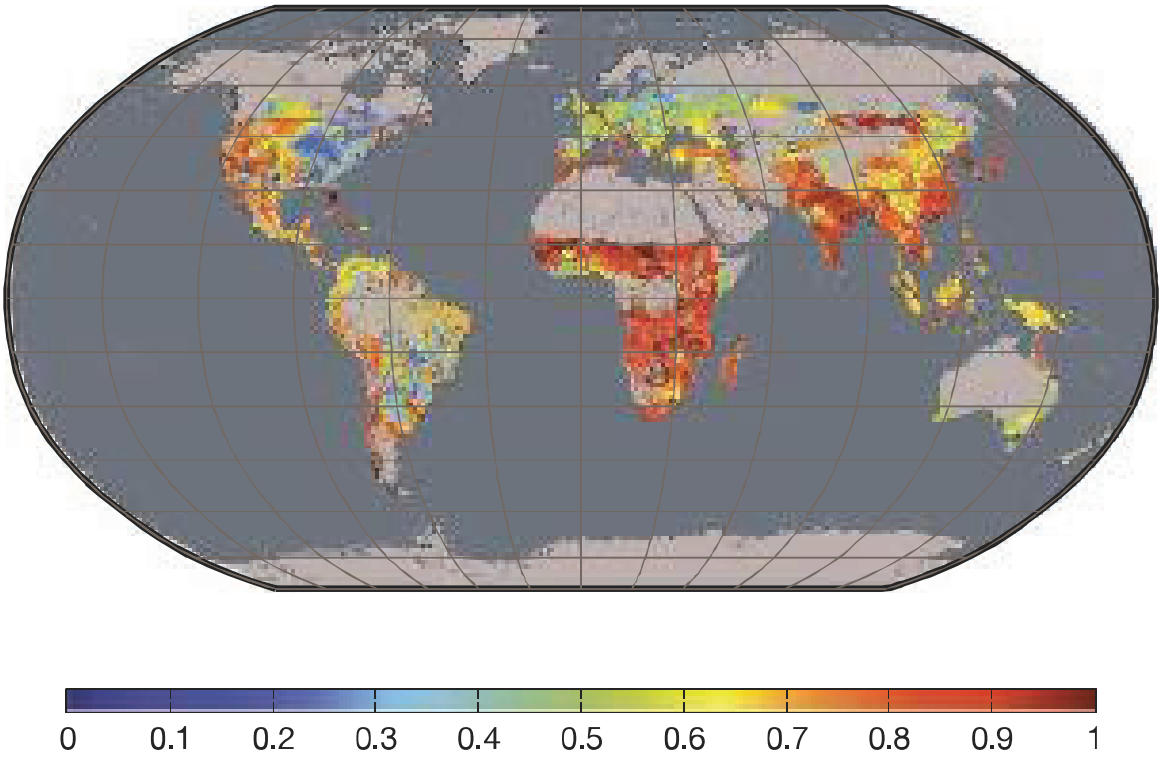


Figure 1.2.: Map depicting proportion of agricultural area adopted for food production (Foley et al., 2011).

Synthetic Aperture Radar images provide an alternative to breach the gap of crop spatial information update. The information can enhance land-use controls, development of commercial plans, regulation of agricultural products, compensation and subsidy decisions to farmers by governments and insurance companies amongst other applications. Moreover, stakeholders can monitor com-

peting demands for instance shelter and food or impact of bio-fuel policies on cropping practices. Annual updating of agricultural land-cover is necessary but requires an effective and efficient information gathering approach. So far compared to ground acquisition methods, remote sensing is a practical and cost-effective way of automatically obtaining up-to-date spatial information. Although optical satellite data are suitable for crop classification, they are weather dependent which impacts their reliability for crop monitoring. Therefore, efforts to acquire crop spatial information using optical satellites can be supplemented by relatively inexpensive SAR data recently made available for civil applications.

1.2 Motivation and problem statement

Information on agricultural activities is important to governments and related stakeholders both locally and internationally. Nevertheless, crops span over large geographic space within unique times of the year. This requires timely and effective information gathering approaches. Annual crop activity information harnessing with ground surveying is prohibitive in terms of cost and time and thus, remote sensing is a viable alternative. Remote sensing satellites capture spectral, spatial and temporal attributes of phenomena on the earth surface. This is attributed to their specific electromagnetic spectrum sensitivity, synoptic view and temporal capability. Therefore, remote sensing provides a platform where quick crop acreage information can be computed and used to estimate regional productivity. Such information can be useful for emergency applications like validating crop insurance claims.

Recent increase in imaging satellites has improved data acquisition frequency making it possible to regularly monitor phenomena on the Earth surface. Changes in an area can now be monitored regularly and also on demand. Such trend favours development of novel image classification methods which can handle temporal data (Jianya et al., 2008). The high temporal resolution also provides more chances for optical satellites to capture high quality images in temperate regions. In occasions where quality images can not be obtained, radar images are an alternative. Radar sensors overcome limitations of optical sensors since their signals can penetrate clouds and are independent of daylight (Gomez-Chova et al., 2006; Tupin, 2010). Radar signals can also penetrate vegetation canopy and dry soil thus bearing volumetric and subsurface information. These attributes render SAR a good medium to deliver a sequence of images of highest temporal density suitable for crop classification regardless of climatic zones. Though, objects on the earth surface appear differently to optical and radar sensors, they provide complementary information (Tupin, 2010). In contrast, complexity such as speckle interference, accompanying SAR data challenges conventional pixel based approaches. Object oriented classification concept minimizes this effect by classifying a block of pixels (object) instead of individual ones. Despite that, the technique does not explicitly model spatial interactions between image objects and thus speckle effect is merely averaged out within image blocks. Methods that model spatial context explicitly can be employed to overcome speckle as demonstrated in Kenduiywo et al. (2014).

Spatial interactions can improve crop classification from SAR images. However, sometimes crops in the same phenology can exhibit correlated SAR backscatter. This means spatial information alone is not enough for optimal crop discrimination. Therefore, initiatives to map agriculture activities require insights on crop dynamics such as phenology states and seasonal growth. Most crops spent a dominant part of their life cycle in vegetation period. In this period, their phenology varies strongly till harvest. Furthermore, their life cycles have different beginning and ending times such as in Figure 4.2. Dynamic changes also occur before seeding (preparation) and after harvest (post harvest). During these phases, farms are affected by deviating management, which is not necessarily unique to the types of crops to be planted or harvested. Basically, there are two types of short term changes: the crop-type specific dynamic changes occurring inside their life cycle and non crop specific changes outside the life cycle influenced by farmers management such as in Figure 1.3. These dynamics cause SAR backscatter variation in agricultural parcels as shown in Figures 2.5 and 3.3. Coincidentally, sometimes crops may at a given time be in the same phenological state, depicting similar spectral attributes, but differ significantly in another time (Siachalou et al., 2015). Therefore, according to Siachalou et al. (2015), crop mapping using single-date remote sensing images, even if acquired in critical growth stages, can not offer optimal results in case of crops with similar growing cycle. This is because it may not be possible to find one common critical growth stage in which the crops have distinguishable radar backscatter. Likewise, pixel based approaches that stack multitemporal images as bands for classification, for example Bargiel and Herrmann (2011); Forkuor et al. (2014); Sonobe et al. (2015), under exploit temporal information. This is because significant temporal information from satellite observed crop phenology is limited to a single feature space vector. Discrimination in such a high dimensional feature space can suffer from overlapping class boundaries due to large class variance from spectral/backscatter variation. Hidden Markov Models (HMM) technique has been used by studies like Leite et al. (2011); Siachalou et al. (2015) to incorporate crop temporal phenological information. The disadvantage of HMM is that it lacks a proper spatial context modelling framework. Therefore in addition to spatial context, integration of temporal context in a principled manner can help resolve classes over time. The synergy of spatial and temporal context including radar multitemporal data properties can form a robust crop classification method.

Use of context (spatial and temporal) for image segmentation and classification has recently gained popularity. Spatial context accounts for similarities among pixels in regard to distance from each other. It determines probability of a pixel or a group of pixels occurring at a given location based on the nature of its/their neighbours in the image (Tso and Mather, 2009). Analyzing a pixel with respect to its neighbours improves classification accuracy. Additionally, use of images acquired at different times has shown significant improvement in classification for example, in classification of crops and vegetation, see a review by Lu and Weng (2007). However, complexity accompanying multitemporal data requires approaches that can efficiently exploit such data. This needs a method that can model land-cover backscatter changes with respect to time and spatial

(a) (b) (c)

Figure 1.3.: Parcels of sugar beet before seeding under different managements in (a), (b) and (c).

neighbourhood of a pixel. Furthermore, continuous increase in high temporal resolution satellites has led to a "Tsunami" of data in archives. The consequence of such phenomena is increased data dimensionality which demands novel image classification techniques. Spatial-temporal automated classification methods are therefore necessary to bridge the gap between expensive data acquisition efforts and actual beneficial data consumption.

Conventionally, Markov Random Field(s) (MRF(s))(Geman and Geman, 1984) has widely been used to integrate spatial-temporal context in image classification. Introduction of Bayesian concept by Swain (1978) to classification of multitemporal images motivated several MRFs temporal studies. Examples include: a general MRFs approach allowing unidirectional passing of temporal information from a classified image at a given date to a subsequent image of the same area at a later date proposed by Jeon and Landgrebe (1992); Solberg et al. (1996) and later extended in Melgani and Serpico (2003) to allow bidirectional exchange of temporal information. Liu et al. (2006) used temporal correlation and temporal exclusion to control certain changes in forest disease spread monitoring. In Moser and Serpico (2011) MRF is applied for multi-scale multitemporal high resolution image classification with class global transitional matrix determined using expectation maximization algorithm. These studies, except Moser and Serpico (2011), adopted a general class conditional probability matrix in MRFs determined heuristically for bidirectional temporal information transfer. While it is an easier alternative, the generalized matrix globally assumes stationarity over all pixels neglecting spatial changes that may exist in the image. This prompted Liu et al. (2008) earlier to develop a locally adjusted global matrix and a pixel-wise joint class probability matrix to represent temporal transitions in MRFs forest change detection. The pixel-wise joint class probability gave better results compared to the locally adjusted global matrix. This illustrates the significance of considering spatial heterogeneity when designing temporal models. Despite this development, MRFs' assumption of conditional independence in observed data adopted for computational tractability neglects spatial context inherent in image data (Lafferty et al., 2001; Kumar, 2006; Zhong and Wang, 2007a; Parikh and Batra, 2008; Hoberg and Rottensteiner, 2010). Remotely sensed images exhibit a coherent scene because neighbouring sites are spatially correl-

ated (Figure 2.6). This concept is modelled by Conditional Random Field(s) (CRF(s))(Lafferty et al., 2001) introduced for one-dimensional (1-D) text classification and extended to 2-D image classification by Kumar (2006). The framework provided by CRFs integrates spatial context both in class labels and data.

Conditional Random Fields ability to model context in class labels and data has triggered studies in several applications. So far CRFs has been used for: classification of settlements and urban areas (Zhong and Wang, 2007a,b; Hoberg and Rottensteiner, 2010; Niemeyer et al., 2011, 2013; Kenduiywo et al., 2014), estimation of ground heights from LiDAR data (Lu et al., 2009), building extraction (He et al., 2008; Wegner et al., 2011a,b) and interpretation of terrestrial images (He et al., 2004; Korc and Förstner, 2008; Gould et al., 2008). However, these are mono-temporal CRF studies. Despite the benefits of multitemporal images, a few temporal CRF studies exist: crop type classification using RapidEye images by Hoberg and Müller (2011), land-cover classification from IKONOS and RapidEye images (Hoberg et al., 2010) and multi-scale multitemporal study using IKONOS, RapidEye and Landsat images (Hoberg et al., 2011, 2015). The studies incorporate temporal context by passing temporal information through empirically determined global transition matrix of class conditional probability. This is an easier way of determining class transitions. However, the transition probability matrix does not optimally capture local class transitions between epochs[1]. Hoberg et al. (2015) noted that incorrect determination of transition matrix leads to erroneous transfer of information into other epochs subsequently reducing classification accuracy. Developing an approach of determining transition probabilities locally can minimize such errors. More also, according to our knowledge no studies have used multitemporal radar data for CRF classification of crops. This study develops a spatial-temporal dynamic classification framework by introducing Dynamic CRFs (DCRFs) — proposed by Sutton et al. (2007) for 1-D text sequence classification — to standard CRF as a temporal sequence classifier template. The approach is built to exploit pixel-wise crop evolution in multitemporal SAR images. We also extend the framework of Sutton et al. (2007) to accommodate higher order temporal crop phenological information exchange. The design establishes a robust spatial-temporal sequence classifier termed as DCRF because:

1. of a changing probabilistic relational model between nodes in the sequence,

2. the model captures time-changing phenomena, encodes complex interactions over the set of all possible classes and data and uncertainty in a principled manner,

3. of scalability, that is, it allows higher order temporal node interactions, and

4. the model is a conditional distribution that factorizes according to an undirected graphical model whose structure and parameters are repeated over a sequence (Sutton et al., 2007).

[1] An epoch is an image date within a sequence of acquired images.

In summary, we anticipate that harnessing of multitemporal SAR images in a given season will improve crop type classification. Crops show varied backscatter radar signals in time (at different phenological instances). We intend to maximize feature separation by exploiting this temporal phenomena. For instance, features that may not be resolved in one epoch spectrally, can be resolved in another epoch. Notably crops undergo phenological changes between epochs but spectral properties of a crop type remain similar. Integrating spatial-temporal information from different image epochs by considering pixel-wise spectral transitions, using DCRFs, is the contribution of this thesis.

1.3 Research identification

The main aim of this research is to develop a DCRF based approach for crop type classification exploiting spatial-temporal information between epochs and crop phenology. This is achievable through the following specific objectives:

1.3.1 Research objectives

1. Design a CRF spatial interaction term that considers all adjacent label scenarios given data evidence for crop classification.
2. To design a DCRFs template into standard CRFs to incorporate pixel-wise temporal phenological information exchange inherent in SAR image sequence.
3. Extend the DCRFs template to accommodate higher order pixel-wise image based temporal phenological information exchange.
4. Incorporate expert based crop phenology knowledge in multitemporal crop classification.
5. To propose an optimal seasonal crop type spatial-temporal classification approach.
6. Evaluate performance of designed approach(es).

1.3.2 Research questions

The following questions are formulated with respect to aforementioned objectives:

1. How can spatial context considering all label scenarios and data evidence be introduced to CRFs for crop classification?
2. Can site-/pixel-wise crop evolutions due to phenology be determined from SAR images using DCRFs?
3. How can the pixel-wise DCRFs model be extended to incorporate higher order temporal interactions?

4. Is it possible to introduce expert phenology knowledge to multitemporal crop classification?

5. Which is the best way to obtain one optimal seasonal crop land-cover map?

6. What is the performance of designed approach(es)?

1.3.3 State-of-the-art

This study uses SAR images and DCRFs for crop classification. This section reviews related state-of-the-art studies and reinstates the research novelty.

Crop classification from radar

Radar has been used to map and monitor crops. For instance Kuenzer and Knauer (2013) carried out a comprehensive review of studies using radar and optical images in rice mapping and monitoring. From the review most studies found maximum likelihood classification (MLC) unsuitable for mapping rice from radar and thus applied state-of-the-art approaches like neural networks and Support Vector Machines (SVM). Another review by Mosleh et al. (2015) illustrates examples of studies used to map rice areas from C- and L-band microwave images using MLC, SVM and neural networks. Random Forests (RF) state-of-art classifier is also used for crop type mapping (Sonobe et al., 2015, 2014). Sonobe et al. (2014) found that RF performed well compared to classification and regression trees (CART) in classifying five crop types from TerraSAR-X strip map dual polarimetric data. The authors later in Sonobe et al. (2015) compared the two approaches to SVM which gave the best performance. McNairn et al. (2009, 2014) used decision trees for crop classification from ALOS PALSAR and TerraSAR-X with RADARSAT-2 images respectively. Generally, in most of these studies SAR images are filtered to minimize the effect of speckle followed by pixel based classification. Filtering reduces speckle but comes with a cost of data modification and subsequent loss of some details. In addition, the studies stack several images as input for classification thereby limiting class discrimination to one feature space vector. Another disadvantage is that discrimination capability at each image acquisition is not perceived.

Object-oriented classification has been adopted as a shift from pixel based classification. The approach partitions pixels into segments. Each segment bears quantitative attributes, for example mean, area or shape, which are subsequently used for classification. The shift from pixels to segments is advocated in Blaschke and Strobl (2001). This technique reduces speckle effect when used in radar leading to improved crop classification results as shown by Jiao et al. (2014). However, a multitemporal study by Long et al. (2013) using optical images established that RF performed better than object-oriented classification in the presence of data gaps in Landsat images. The low accuracy might be driven by the fact that object-oriented approaches do not model explicitly temporal interactions but instead rely on features computed from the multitemporal images. To incorporate spatial interactions a CRF model can be applied to initial object-oriented segmentation

results like in Wang and Yung (2015). A recent in-depth review of object-oriented classification is covered by Blaschke (2010); Blaschke et al. (2014). In this thesis a probabilistic approach that models context in form of pixel's spatial and temporal neighbours is used. The approach provides a unified framework for modelling spatial-temporal context in a statistical principled manner. We review in the next section related spatial-temporal classification studies.

Spatial-temporal context classification

Spatial-temporal land-cover mapping approaches integrate spatial and temporal information during classification. So far, MRFs and CRFs have been widely used for this purpose. Exceptions include HMM which directly model temporal interactions but has been used in some studies to include spatial context indirectly via image segments. For instance, Leite et al. (2011) used HMM and Landsat optical data for multitemporal crop classification using image segments as opposed to pixels. However, this is not a principled approach of modelling spatial interactions since spectral variability is minimized within a segment hence yielding a smoother appearance perceived as context. The study includes temporal interactions between nodes via a global state transition conditional probability matrix. Another study by Siachalou et al. (2015) used HMM and RapidEye and Landsat data for crop classification. The main limitation of HMM is that it ignores spatial interactions. Structurally, MRFs and CRFs are built to consider spatial interactions. Efforts have been made to include temporal data interactions as earlier illustrated.

Our study is inspired by the works of Liu et al. (2008) and Hoberg and Müller (2011). Liu et al. (2008) used MRFs framework to compare global, locally adjusted and pixel-wise temporal interaction models in forest change detection from two Landsat image dates. The study was motivated by the limitation of the global temporal interaction model to represent local spatial heterogeneity. Generally, the global class conditional probability matrix represents temporal information exchange between all nodes in two epochs. Therefore, the authors developed pixel-wise interactions using joint probability of classes. Class joint probabilities were estimated by Gaussian models of the two images via maximum likelihood estimation based on training data. The joint probability estimates were then marginalized to conditional probabilities which formed elements of a matrix representing temporal transitions for each node. Best results were obtained with pixel-wise temporal interaction. This emphasized the significance of considering spatial heterogeneity when designing temporal models. In this study, we exploit temporal class conditional independence assumption – since evolution of a crop is temporally unique to radar – to design a pixel-wise conditional temporal interaction matrix. This assumption enables us to estimate initial class probabilities in different epochs independently. We then subsequently use the probabilities to compute conditional probability matrix between a node and each of its defined temporal neighbours via Bayesian formula. This is beneficial in terms of discrimination compared to joint classification which limits discrimination of considered epoch data to one feature space vector.

In Hoberg and Müller (2011), CRFs is used for spatial-temporal crop classification. Here the temporal model is designed using a site-wise feature difference of an image site in two epochs modified to consider mean of features of each class in those epochs. The modification was done in order to consider prevailing atmospheric and lighting condition during acquisition of optical images. Thus, without considering mean of features in each class the function would be similar to absolute difference function used in interaction potential of CRFs. Our study uses class probability which is a soft decision function that does not directly suffer from prevailing conditions during image acquisition. Moreover, radar sensors are not affected by weather and lighting conditions. A later CRFs study by Hoberg et al. (2015) defines temporal interactions via a global matrix using expert knowledge. As already mentioned, representing temporal interactions with a single matrix globally assumes stationarity in the image, which is not the case because images represent heterogenous scenes.

Dynamic conditional random fields introduced by Sutton et al. (2007) for 1-D text classification, also operate within similar concepts to model temporal interactions. The DCRFs adopt first order Markov assumption in a sequence of nodes to model interdependencies between the node's word and the previous one. So far DCRFs has been used for spatial-temporal event detection from a sequence of sensor measurements in Yin et al. (2009). In the study, temporal interactions were determined using site-wise Euclidean distance of features based on an indicator function (Kronecker delta). The indicator function considers a set of conditions which if met, Euclidean distance of a pair sites in different epochs is computed from their features as a temporal potential. This set-up is similar to the Euclidean distance model commonly used in CRFs for spatial interaction. A similar design is used by Wang et al. (2006); Yang et al. (2008) for foreground video scenes segmentation and by Zhang et al. (2014) with an additional Gaussian kernel in temporal potential to track deformable parts from video sequences. Another study by Zhang and Qiang (2008) used site-wise absolute difference of features to model temporal interactions in segmentation of video sequences. Most studies have used DCRFs for object detection from video frame sequences and event detection from sensor networks. For the first time this study introduces the design of DCRFs for crop classification using SAR multitemporal image sequences. In addition, we introduce a novel pixel-wise temporal interaction model based on conditional probability concept.

1.3.4 Research novelty

Research objectives, corresponding questions and state-of-the-art studies in crop mapping and CRFs classification have been described. Therefore, this section reinstates the main research contribution and hence novelty.

The research aims at developing a new spatial-temporal crop type classification from SAR image time series using DCRFs technique. Spatial-temporal DCRFs classification involves spatial and temporal context modelling. The study develops two suitable spatial data interaction models based on modified contrast sensitive model and Pearson correlation. Here the Pearson correlation is in-

troduced for data dependent interactions in CRFs classification for the first time. On the other hand, the new modified contrast sensitive model is expanded from the existing contrast sensitive model as suggested in Shotton et al. (2009). As for temporal interactions, we design a new approach for bidirectional information exchange between defined temporal neighbouring nodes using conditional probability. Here, conditional probability is computed using initial class probability estimates in each temporal node. A classifier is used to estimate the initial probabilities from SAR images using training data.

The spatial-temporal model is adopted for the first time to classify crops from SAR images (TerraSAR-X and Sentinel 1). In addition, a further effort is made to design a technique that generates an optimal seasonal crop map from a sequence of multitemporal images. A weighting strategy using class probabilities with maximum F1-score from a classification sequence is designed. This set-up establishes a configuration where epoch-wise DCRFs node probabilities are considered as part of the final class decision. In this case, the epoch-wise DCRFs decisions can be regarded as hidden hence, a hidden DCRFs model. It is a new method that benefits from temporal class conditional independence assumption.

1.4 Thesis outline

This thesis is organized into seven chapters. Chapter 2 discusses the basic theoretical concepts of SAR data acquisition and probabilistic reasoning via context based graphical models. Radar is selected as a data acquisition platform in this study, hence, its basic image acquisition concepts is described. Different mechanism of radar that influence backscatter of agricultural landscapes are discussed including target properties like phenology. Crop phenology is then linked to context to illustrate a modelling aspect using graphical models. Types of graphical models are discussed with emphasis on undirected graphs mainly adopted in this study. Chapter 3 explains developed methodology. This includes parameter estimation by training and final inference including validation of results. Experiments and data used to implement and test the developed methods for crop mapping are described in Chapter 4. Results from the implemented methods are presented in chapter 5 leading to a discussion in chapter 6 and conclusions of the study in chapter 7. Future prospects are also noted in chapter 7.

2 Theoretical concepts

Agriculture is both a spatial and temporal phenomenon as it spans over geographical space within some period. To map crops, a data collection medium and a subsequent map production method, from collected data, is necessary. In this study, remote sensing by radar and a probabilistic classification technique are chosen for data collection and mapping respectively. Thus, this chapter provides basic concepts of the adopted approaches. Fundamental principles of image acquisition by microwave remote sensing and crop backscatter mechanism are discussed with reference to Henderson and Lewis (1998); Woodhouse (2006); Richards (2009); Lillesand et al. (2015). Theoretical discussions on probabilistic reasoning using graphs is done with reference to Bishop (2006) and Barber (2012) and other sources which shall be cited.

2.1 Microwave remote sensing

Microwave sensing encompasses active and passive sensors mounted on distant elevated platforms – for instance, on ground, aircraft or satellite – used to derive information from the earth. Such sensors utilize microwave portion of the electromagnetic (EM) spectrum, hence the term *microwave remote sensing*. The EM spectrum represents a range of all possible wavelength λ and corresponding frequency f of EM radiation. For instance, microwaves have wavelengths in the range of approximately 1 cm – 1 m; examples of past and current microwave remote sensing programs with operating λ and f is shown in Figure 2.1. Hence compared to optical sensors, which operate in visible and infrared portion of EM spectrum, microwaves wavelengths are longer to the order of 10^5. This difference makes microwave remote sensing a unique medium. On this account, objects on the earth surface would appear differently to optical and microwave sensors. However, their information complement each other.

Active microwave imaging sensors generate their own energy to illuminate a target in an area of interest. The sensor then measures intensity of energy reflected back, known as backscatter. Consequently, active sensors are daylight independent. Their microwave wavelengths can also be chosen such that weather effects such as haze or clouds, are negligible on backscattered signal. Thus, SAR satellites guarantee the highest possible temporal density of images regardless of climatic zones. This particularly benefits crop monitoring and seasonal mapping. Wavelengths in microwave region can also penetrate vegetation and even dry soils compared to those in visible portion of EM spectrum. So in addition, volumetric and sub-surface information is conveyed. This essentially means different crop canopies and their underlying soil properties can be discriminated from microwave signals. Another benefit of active sensors is that properties of incident energy (e.g., λ and f) can be controlled. As a result, specific application data can be acquired. However,

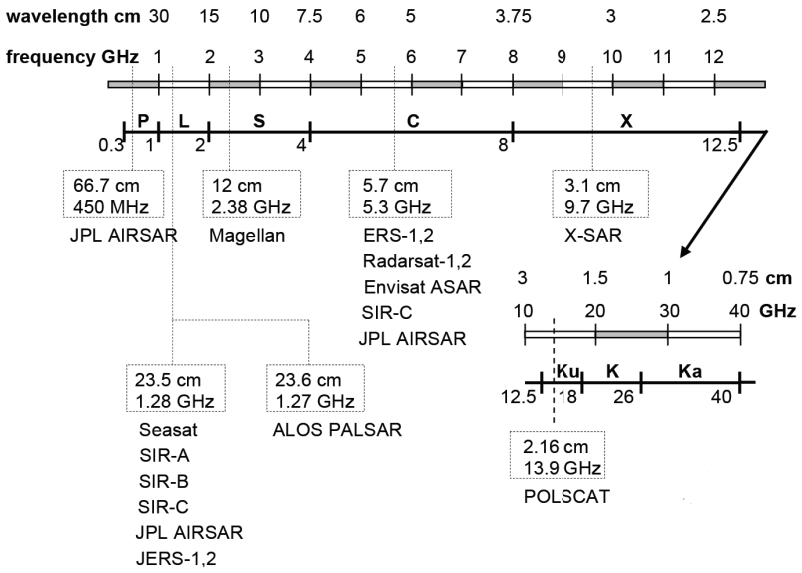


Figure 2.1.: Past and current operating microwave remote sensing satellites with their frequencies and corresponding wavelengths; modified from (Richards, 2009).

sensors operating on long wavelengths require very long antenna in order to attain sufficient spatial resolution even in regional studies. Side looking geometry and synthetic antenna, as discussed in Section 2.1.1, are used to overcome such shortcomings.

2.1.1 Imaging with synthetic aperture radar

Radio detection and ranging (radar) is a concept introduced by military and civil air traffic controllers to detect presence of objects in terms of their distance and sometimes direction. As an active sensor, it transmits short bursts (pulses) of microwave energy in direction of interest and subsequently receive "echoes" or "reflection" from objects within field of view. The system consist of a transmitter and a receiver antenna which can be located separately (bistatic) or co-located (monostatic). The later is a common mode for satellite sensors except in Shuttle Radar Topography Mission (SRTM) and TanDEM-X missions. Radar sensors transmit and/or receive microwave signals in different polarization modes. Polarization refers to the orientation of electric field in an EM wave. In remote sensing, radar signals are either transmitted with electric field plane parallel (horizontal polarization) or perpendicular (vertical polarization) to the Earth surface. Likewise, a receiver can be configured to receive particular polarized signals. For instance, in TerraSAR-X Stripmap Mode (SM), the antenna can transmit and receive in either horizontal (H) or vertical (V) single polarizations (HH[2] or VV) or cross-polarization (HV or VH). In addition, it is able to transmit dual polarized signals, that is almost simultaneous horizontal and vertical transmit and receive

[2] The first letter indicate transmit and second receive.

respectively (HH/VV) or with cross-polarization (HH/HV or VV/VH). Such polarization techniques result in different image types.

An image can simply be acquired by scanning a narrow beam of EM radiation across the earth's surface normal to spacecraft motion. However, due to wave diffraction, spatial resolution r of any imaging sensor (optical or Real Aperture Radar (RAR)) is limited by angular field of view α of antenna's aperture or beamwidth D, i.e. $\alpha = \frac{\lambda}{D}$. The resolution linearly decreases with increase in sensor elevation H above the scene:

$$r \approx \frac{\lambda}{D} \cdot H \quad \text{m.} \tag{2.1}$$

This means sufficient spatial resolution can only be attained at low altitude platforms when imaging with long wavelengths at nadir.

A side looking geometry, Figure 2.2, is used by imaging radar sensors to attain better spatial resolution. In this setup, radar pulses illuminate a surface at an oblique look angle θ_l with an incident angle θ_i on earth's surface. An image is recorded parallel to sensor motion (azimuth) and also orthogonal to its motion (range). The image ground spatial resolution is controlled by pulse duration τ and D. Pulse duration determines minimum possible distance for two objects to be distinguished in range direction known as range resolution r_r:

$$r_r = \frac{v\tau}{2\sin\theta_i} \quad \text{m} \tag{2.2}$$

where v is velocity of light (in vacuum) ($\approx 3 \times 10^8\ m/s$). Basically, the shorter the pulse duration the better the range resolution. However, the pulses have to be transmitted at higher energy to attain a short duration. This requires high power. In practice it is not possible to design transmitters that can produce short rectangular pulses with high peak power. To counter this, a linear frequency modulated pulse, termed *chirp*, is used. Here, the signal is send over a small range of frequencies known as bandwidth B. As a result a collection of overlapping backscattered signals are received and resolved in time to an resolution better than pulse duration. This is because of an inverse relationship between B and τ:

$$r_r = \frac{v}{2B\sin\theta_i} \quad \text{m.} \tag{2.3}$$

In azimuth direction, resolution of a cell is constrained by beamwidth (in azimuth direction) D_a and distance from the sensor to a target (slant range) R. The width of the beam in azimuth direction is in turn limited by the length of the antenna. This is a limitation caused by diffraction

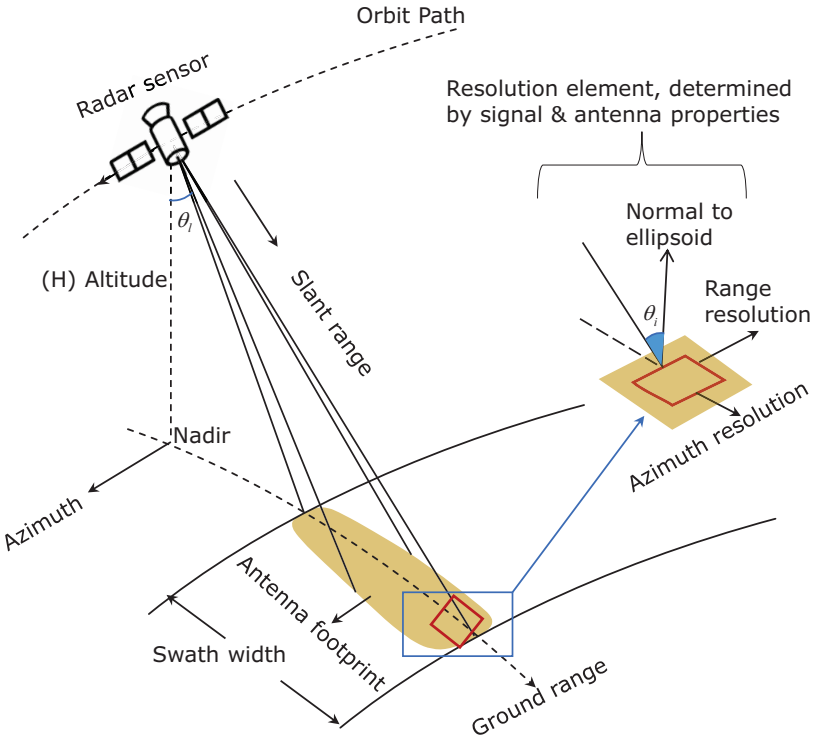


Figure 2.2.: Components of SAR imaging.

which also faces optical systems as shown in Equation (2.1). The azimuth resolution r_a deteriorates with increase in R because the antenna beam "fans out" to a target:

$$r_a = \frac{\lambda}{D_a} R \quad \text{m.} \tag{2.4}$$

Azimuth resolution in RAR can be improved by increasing length of the antenna. This option poses practical limitations in space platforms. However, using SAR principle a long antenna is synthesized to improve resolution. It is achieved using any of these two ways: 1) using Doppler frequency shift due to sensor motion and 2) by coherently combining low resolution antennas, for details see (Woodhouse, 2006, chap. 10.5). From either of the two processing concepts, a resolution half the antenna beamwidth is obtained, that is,

$$r_a = \frac{D_a}{2}. \tag{2.5}$$

So far radar principles that determine resolution of a radar image cell have been reviewed. The cell stores backscatter information corresponding to the cell area on ground scene. The backscatter is a ratio of energy received by a sensor and that incident on a target or surface. Since radar systems

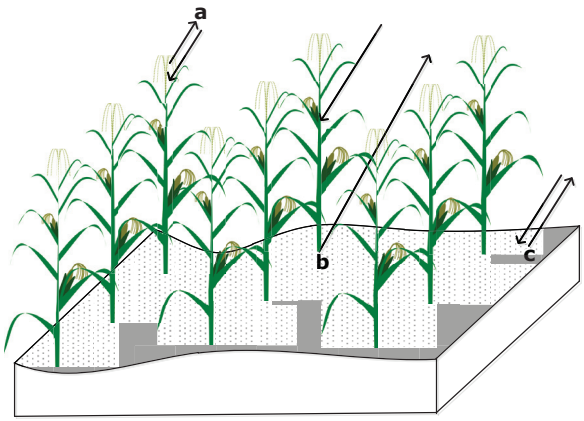


Figure 2.3: Sources of radar backscatter from crops:

a direct scattering from crop parts such as leaves, stems and heads,

b double bounce scattering from soil surface and crop canopy, and

c direct backscatter from crop undergrowth.

observe extended targets, (i.e. Earth surface), the backscattered energy comes from a portion of a distributed area (e.g., area of a pixel) as opposed to a discrete target. Thus, backscatter measured from a target area in SAR is usually normalized per unit geometric cell area known as *normalized backscatter coefficient* σ^0 (sigma nought) as shown in Equation (2.6). Magnitude of backscatter is influenced by: system parameters (frequency/wavelenght, polarization and incident angle) and target parameters (dielectric and and geometric properties of target) (Henderson and Lewis, 1998).

$$\sigma^0 = P_r \frac{(4\pi)^3 R^3}{A P_t G^2 \lambda^2} \tag{2.6}$$

From Equation (2.6) P_r refers to received energy, G is antenna gain, λ is wavelength, P_t is transmitted energy, R corresponds to the range from the sensor to the object and A is the area over which the measurement is made. This formulation expresses the so-called *SAR equation*.

Frequency (or wavelength) affects crop backscatter magnitude because of: (1) differences in dielectric constant as a function of frequency, and (2) relationship between wavelength and crop size (or leave size) and/or canopy penetration. For example, TerraSAR-X operates in high frequency (small wavelength) thus, it is dominated by canopy scattering. This is because signals with small wavelengths have less canopy penetration and vice versa. Therefore, TerraSAR-X backscatter is dominated by canopy scattering and multiple scattering from canopy and soil (i.e., case a and b in Figure 2.3). Such attributes supports its choice for crop type mapping though, this is subject to crop mix and phenology. Polarization also influences crop discrimination in SAR images. Le Toan et al. (1984); Ulaby and Wilson (1985) notes that stems of vertically oriented grain crops couple effectively with VV polarized signals resulting in increased attenuation. Such crops can be discerned in VV polarized images because their signals interact more with crop structure. In contrast, HH polarization penetrates crops and captures underlying soil roughness and moisture content. For example, potatoes are normally planted in ridges and can have significant backscatter in HH polarization especially when looking perpendicular to row direction (Ulaby and Bare, 1979; Brisco

et al., 1991; McNairn and Brisco, 2004). Cross polarized images (HV or VH) have also been found to improve crop separability (Brisco et al., 1992). In summary, knowledge of crop scattering mechanism including understanding of crop growth dynamics can enhance crop mapping. Section 2.1.2 explores how crop phenology affects backscatter changes and hence context in SAR images.

2.1.2 Phenology and Context in SAR images

Phenology studies annual rhythm of biological phenomena with respect to climate parameters. In agriculture, phenology refers to evolution of crops with reference to beginning and end of season. For instance, maize phenology (Figure 2.4) encompass: germination, leaf development, stem elongation, heading, flowering, development of fruit, ripening and senescence (plant dies) as defined by Meier (2001). In remote sensing we can consider phenology as the appearance of a crop at a particular instance in its life cycle. Such crop stages are dependent on the type of crop, climatic factors and farm management practices. Therefore, different crops may be at different phenological stages at a given instant. Economic factors also influence crop management and direct choice of certain crops. Government incentives and other forms of support drive farmers to maximize outputs within short periods. These factors introduce a complex dynamic challenge to an already evolving cropping system. Mapping such a dynamic phenomenon driven by evolving factors requires dynamic models. In contrast, these inherent factors can also aid discrimination of crops at different instances. This is possible using imaging SAR images as introduced earlier.

Phenology governs crop's water content and consequently its dielectric properties. A dielectric object/surface is "one that can withstand high electric stress without appreciable conduction" (Barton and Leonov, 1998). Materials with high dielectric constant have strong reflective surfaces. Water has a high dielectric constant and thus influences radar backscatter. During crop maturity, water content reduces proportionally decreasing backscatter contribution from plants while increasing penetration of transmitted radiation into vegetation which increases backscatter from undergrowth (Ulaby and Wilson, 1985). Generally, with support from Figure 2.3, early stages of a crop growth are dominated with backscatter from bare soil and in later stages volume scattering from plant canopy increases. For instance, broad leaved crops like corn and sugar beet return high backscatter in C- or X-band during early stages of their development with little change in the rest of the growing season (Brisco and Protz, 1980; Bouman, 1988; Tso and Mather, 1999). In contrast, grain crops like wheat and barley have temporal backscatter variation throughout growing season as portrayed by Brisco et al. (1992) results. Therefore, the correlation between plant growth stage and backscatter magnitude implies that knowledge of crop phenology is important in classification (Brisco et al., 1984).

It is evident that there are gradual or rapid backscatter changes from crop parcels. Different crops may be in the same phenology state at an instant in time and thus exhibit similar backscatter. Similarly the same crops may depict varying backscatter attributes in a different phenology state (Siachalou et al., 2015). These variations can be detected by SAR sensors because they are

(a) Germination

(b) Leaf development

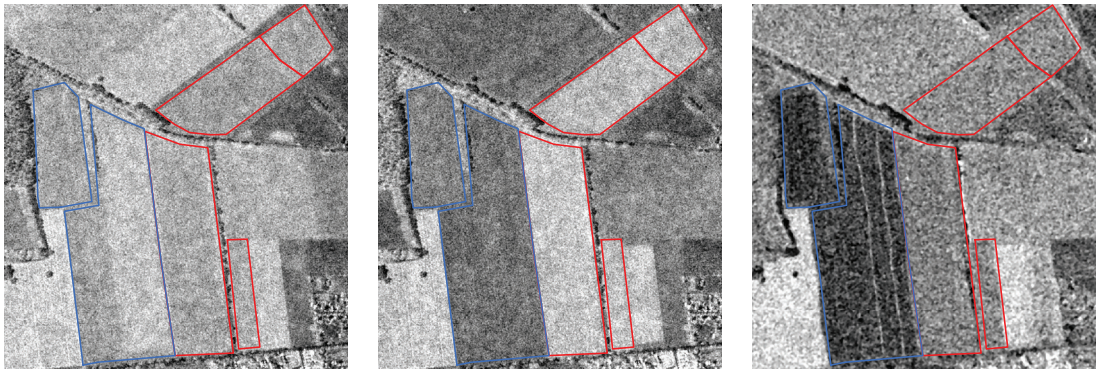
(c) Stem elongation

(d) Flowering and fruit development

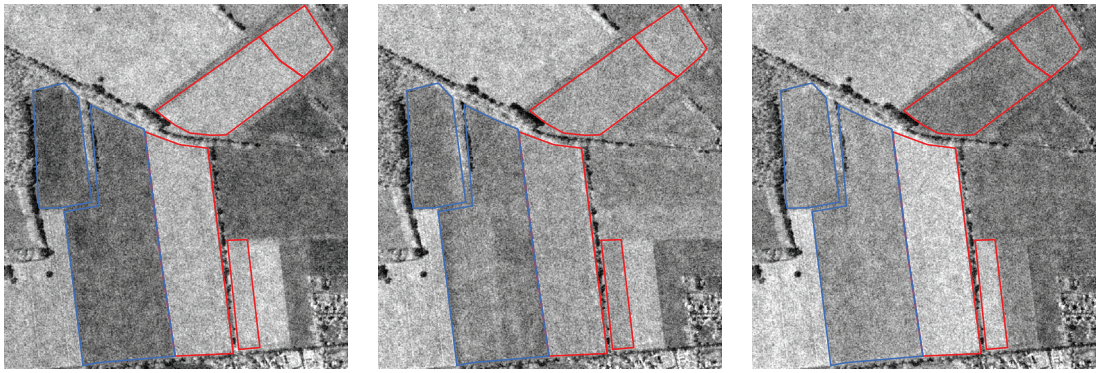
Figure 2.4.: Illustration of some phenology stages of maize.

sensitive to dielectric properties (e.g., moisture) and geometric attributes (e.g., roughness and canopy structure). Figure 2.5 illustrate variations of SAR amplitude as depicted by some parcels over time. Classification of such parcels using single date images will not be sufficient. Equally, using several images acquired at critical phenological stages as a stacked layer in per pixel classification neglects context (temporal and spatial).

Goodchild (1992) defines spatial context as "the propensity for nearby locations to influence each other and to possess similar attributes". Two types of context exist (Kittler and Föglein, 1984): class dependency context and data/pixel dependency context, all which have spatial-temporal existence. Class dependency context supports the existence of pixels of certain land-cover type(s) near each



(a) March **(b)** April **(c)** May



(d) June **(e)** July **(f)** October

Figure 2.5.: An illustration of radar amplitude variation from crop parcels in different months of 2009 as captured by TerraSAR-X in VV polarization (©DLR). Blue and red parcel line boundaries represent rye and canola crops respectively. The variations depict strong correlations within a class in time, hence temporal dependency.

other. For instance, some land-cover classes are more likely to be neighbours than others, or some classes may rarely be neighbours, for example a maize field in the middle of an ocean or forest. Pixel dependency context demonstrates underlying correlation in remote sensing data. For example, an image is made up of a coherent scene consisting of pixel regions as opposed to independent ones. Therefore, considering pixels or image segments as independent like in per pixel and segmentation approaches under-exploits context (see Figure 2.6). Additionally, in most cases land-cover classes cover a larger area than a pixel size in an image such that neighbouring pixels form a homogenous region that is more likely to be in one class. Spectrally, an area in an image varies at different times due to physical changes in phenomena (Figure 2.5). Essentially, spectral attributes at different moments in time representing the same area are correlated. This gives rise to temporal context which covers both class and pixel dependency. In this case, temporal context may be defined as class and pixel dependency in images acquired at successive times.

Spatial-temporal context supports the idea that, given a sequence of images of an area captured consecutively by a satellite, the points in space and time close to each other are likely to represent the same land-cover class and the pixels are correlated. More also, in crop classification classes undergo spectral changes as opposed to transitions to other classes thus, pixels are correlated over the entire season. Therefore, an image contains strong dependencies among its pixels, between preceding and consecutive images. Modelling this spatial-temporal dynamic context in a principled manner can help improve class discrimination.

(a) (b) (c)

Figure 2.6.: An illustration of the fact that radar remotely sensed images contain strong context – that represent spatial heterogeneity – rather than being a random collection of independent pixels or blocks/segments (Kumar, 2005). (a) Normal image scene. Image (b) is obtained by randomly combining pixel values in (a). Image (c) is obtained by randomly combining the original image blocks, ©DLR.

Probabilistic models provide a good framework of modelling spatial-temporal dynamic context. Their framework are ideal for reasoning – where many possible outcomes exist – using probability and graph theory as discussed in Section 2.2.

2.2 Reasoning with probabilistic models

Remotely sensed images represent earth phenomena at varying scales. They give and overview of relationships and pattern of objects. Earth surface phenomena form contiguous scenes with transitions between boundaries. As discussed in Section 2.1.2, such organized pattern structure support the concept of spatial-temporal context. It is the objective of image classification algorithms to capture such relationships. Probabilistic models give a framework on which such relationships can be logically build and inferred to reach sensible conclusions under uncertainty.

Reasoning under uncertainty using probabilistic models is possible with mathematical rules of probability. Specifically, Bayes' theorem with conditional probability provides a basic framework in image classification since empirical observations are available. Through them, prior information and conditions can be expressed. This allows available factors that quantify a certain belief to

be injected into a solution finding process of any given problem. For instance, crop phenology knowledge, spatial and temporal context can quantify belief of existence of a certain crop type at a given instance. These factors provide a basis for reasoning (inference) over a set of classes (crops) in order to determine an optimal class during supervised image classification.

Supervised image classification objective is to find the most probable class label[3] $\mathbf{y}$ from training set[4] given image data $\mathbf{x}$. This can be determined via Bayes' theorem:

$$P(\mathbf{y}|\mathbf{x}) = \frac{P(x|y)P(y)}{P(x)} \tag{2.7}$$

In Equation (2.7), $P(\mathbf{y}|\mathbf{x})$ is the conditional probability of $\mathbf{y}$ given $\mathbf{x}$ known as *posterior probability*, $P(x|y)$ is the likelihood function, $P(y)$ is referred to as *prior probability* and $P(x)$ is the probability of x and is referred to as a normalizing constant. Bayes' theorem plays a key role in image classification because it makes reasoning possible via probabilistic relationships. The probabilistic relationships are built from two fundamental rules:

$$\textbf{Sum rule: } P(x) = \sum_{\mathbf{y}} P(x, y) \tag{2.8}$$

where $P(x)$ is marginal probability of the joint distribution $P(x, y)$ of x and y. It is computed by summing the joint probability over all possible states of y, a process called *marginalization*.

$$\textbf{Product rule: } P(x, y) = P(y|x)P(x) \tag{2.9}$$

The product rule and symmetric property of joint probability $P(x, y) = P(y, x)$ give rise to Bayes' theorem. Substituting Equation (2.8) into Equation (2.7) gives the following relationship between conditional probabilities

$$P(\mathbf{y}|\mathbf{x}) = \frac{P(x|y)P(y)}{\sum_{\mathbf{y}} P(x, y)} \tag{2.10}$$

which is referred to as *Bayes' theorem*. Revisiting the objective of modelling time series, consider a sequence of images $\mathbf{x} = x^t$ to be modelled over corresponding labels $\mathbf{y} = y^t$ acquired at different

[3] Bold face symbols illustrate a random variable while normal font symbols are realizations of the random variable

[4] A training set is a subset of data, with a given label, used as an example to predict particular class from the entire data set.

times t where $t = 1, \ldots T$ (see Figure 2.7). Then, time series image classification challenge is to determine the optimal label y for a given image pixel from a sequence of labels given data:

$$\max_{y} \left\{ P(y^{t=1}, \ldots, y^{t=T} | x^{t=1}, \ldots, x^{t=T}) \right\} \tag{2.11}$$

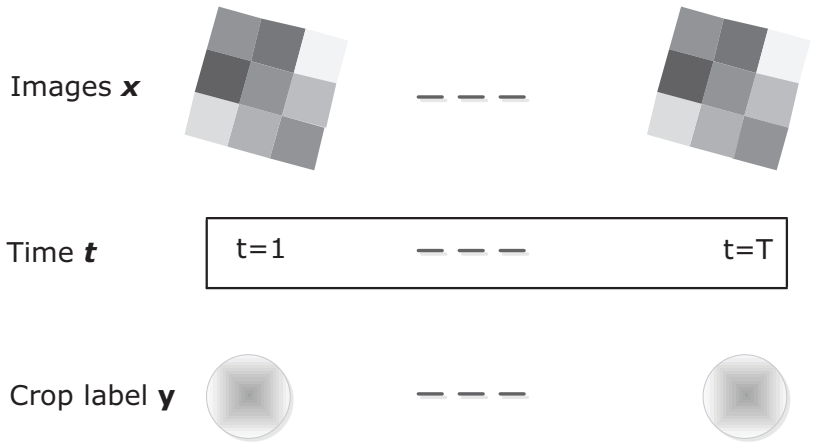


Figure 2.7.: Illustration of classification over a sequence of images.

Solution to Equation (2.11) is complicated as it involves a joint distribution over label and data sequences. Therefore, an alternative approach to modelling time series exploiting probabilistic reasoning benefits is necessary. Probabilistic reasoning framework introduces valuable tools for simplification and for minimizing of computational complexity. For instance, independency assumptions can be incorporated leading to compromise solutions. This eliminates the trade-off between computational efficiency and models that are still rich enough to provide the required discriminating power (Hornegger et al., 2000). Other benefits that justify choice of probabilistic reasoning according to Hornegger et al. (2000) include:

1. Sensor signals and associated features show a probabilistic behavior due to sensor noise, varying illumination conditions or segmentation errors.

2. Image classification tasks should use all available sources of information including prior knowledge and empirical data. A unified mathematical formulation incorporating all modules is given by probabilistic models.

3. Decision theory guarantees the optimality of Bayesian classifiers, which maximize posterior probabilities, if a correct model is chosen for the underlying probability distribution function.

4. The design of learning algorithms can utilize comprehensive results in statistics and statistical learning theory.

5. The success of probabilistic models in different areas of applied pattern recognition (e.g., speech recognition and handwritten character recognition) motivate the use of statistical methods.

The discussed fundamentals on probabilistic reasoning will be used throughout this text. It is now clear that probabilistic reasoning can be expressed using two basic equations of probability: Equations (2.8) and (2.9). More details on probabilistic reasoning can be found in (Bishop, 2006, chap. 1.2) and (Barber, 2012, chap. 1). Machine learning and inference tasks in image classification are an application of the two equations. Probabilistic reasoning as regards image classification problems can be expressed better using *graphical models*. Graphical models augment solution finding process using diagrammatic representations of probability. Their framework offer suitable properties for spatial-temporal crop classification with phenology as discussed in Section 2.2.1.

2.2.1 Graphical models

Before reasoning precedes any real-world problem, assumptions about the structure of the problem are necessary. Such structural abstractions encompasses relevant variables and their interrelationships. Graphical models provides a formal representation of these assumptions. A Graphical Model (GM) is a family of probability distributions that depicts their independence/dependence relationships diagrammatically using graphs. Graphical models (GMs) are a union of probability and graph theory. Their structure is suitable for modelling with several useful properties:

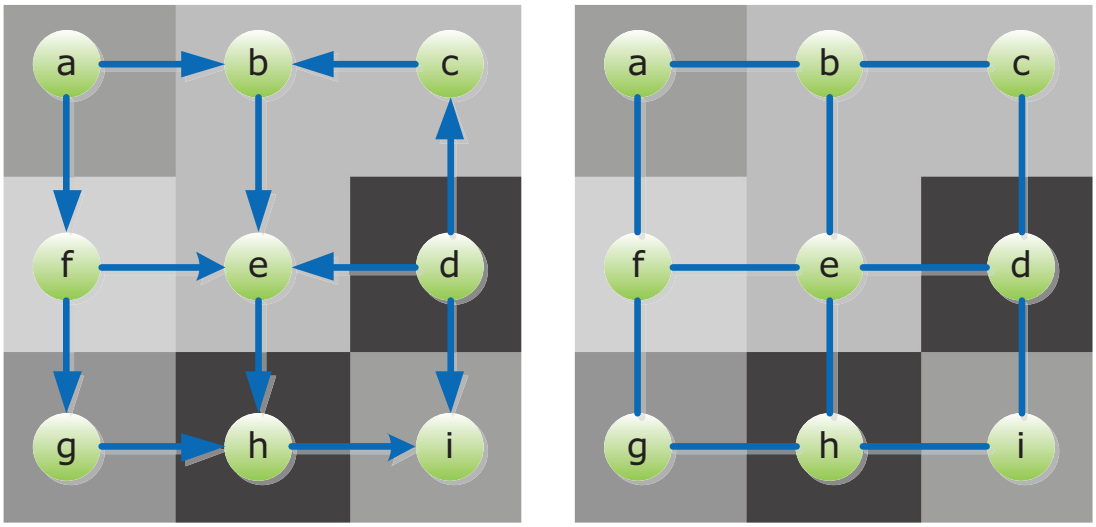
1. They provide a visual representation of probabilistic models useful in design and motivation of new models.
2. Graphs helps deduce properties of a model being designed, e.g. independence/dependence assumptions.
3. Provide a framework for studying probabilistic models.
4. They provide a unified framework in which learning and inference algorithms in different disciplines can be related.

Types of GMs include and not limited to: Bayesian networks, Markov networks (Markov random fields and conditional random fields), factor graphs amongst others see (Barber, 2012, chap. 4) and (Sutton and McCallum, 2011). Since GMs rely on graphs to illustrate probabilistic reasoning, we review basics of Graphs.

Graphs

A graph **G** consists of a set of nodes **S** and edges **E** connecting the nodes. An edge may be directed, Figure 2.8a, which is illustrated with an arrow in a single direction or with no particular direction, i.e. undirected, as shown in Figure 2.8b. Directed edges form directed graphical models known as

Bayesian Networks and undirected edges establishes undirected graphical models called random fields. These models encode conditional dependency/independence relationships between nodes in a graph. On the other hand, nodes represent random variables in reference to a modelling environment or problem such as an image pixel or segment in classification or image site[5]. So if two nodes are not linked then they are independent while if they are linked they have a dependency.



(a) Directed graph **(b)** Undirected graph

Figure 2.8.: Directed and undirected graphs modelled over pixels of an image. Nodes are represented by green circles while edges have blue lines. Each node has a spatial extent of corresponding pixel.

Directed and undirected graphs are used by graphical models to express probabilistic operations. For instance, probabilistic mathematical operations like sums of conditional and marginal probabilities and multiplication, see Equations (2.8) and (2.9), can be deduced from a given graph. This makes it possible to decompose the joint distribution over the entire image $\mathbf{x}$ and class labels $\mathbf{y}$ as a product of conditional and marginal probabilities of adjacent nodes within a graph. The process is referred to as factorization and are expressed by factor graphs mainly used by inference algorithms.

To illustrate further how reasoning is done using graphs we refer to a directed graphical model in Figure 2.8a. Here, the joint distribution $P(a,b,c,d,e,f,g,h,i)$ in the directed graph can be expressed as:

$$P(a,b,c,d,e,f,g,h,i) = P(a)P(b|a,c)P(c|d)P(d)P(e|b,d,f)P(f|a)P(g|f)P(h|e,g)P(i|h,d) \tag{2.12}$$

[5] The terms node, site and pixel are synonymous and will be used interchangeably.

From Equation (2.12), probability of nodes a and d are independent from other nodes hence marginal probabilities $P(a)$ and $P(d)$. In this case, a is a parent of nodes b and f and similarly nodes b and f are its children. Apart from nodes a and d, probability of other nodes have a dependency as illustrated by edges, i.e. conditional probability. This structure in the graph reduces the solution to a product of factors. In this way, the graph presents a modelling and/or reasoning environment where probabilities are used to express relationships/conditions in a solution finding process. In principle, directed graphical models assume that the observed random variables, in this case nodes, are generated by a causal latent process. The graph structure indicates direct dependencies among nodes. Here the general *conditional independence assumption* is stated as; "a node is conditionally independent of all other nodes given the values of its parents, its children and its children's parents".

Undirected graphical models on the contrary, model random variables that do not have explicit causal dependence. For example an image grid structure can be intuitively illustrated as an undirected graph as depicted in Figure 2.8b. In this case, the *conditional independence assumption* encoded by the graph can be stated as; "a node is conditionally independent of all the nodes in the graph given its neighbours N". This graph structure is exploited to represent global context in an image via local dependencies. Normally, relationships in a given neighbourhood set is expressed via cliques. This helps maintain the idea of a coherent image scene (see Figure 2.6) by factorizing global pixel dependency via potential functions defined on cliques. We utilize this set up in this study to encode context (spatial and temporal) because it is suitable for images.

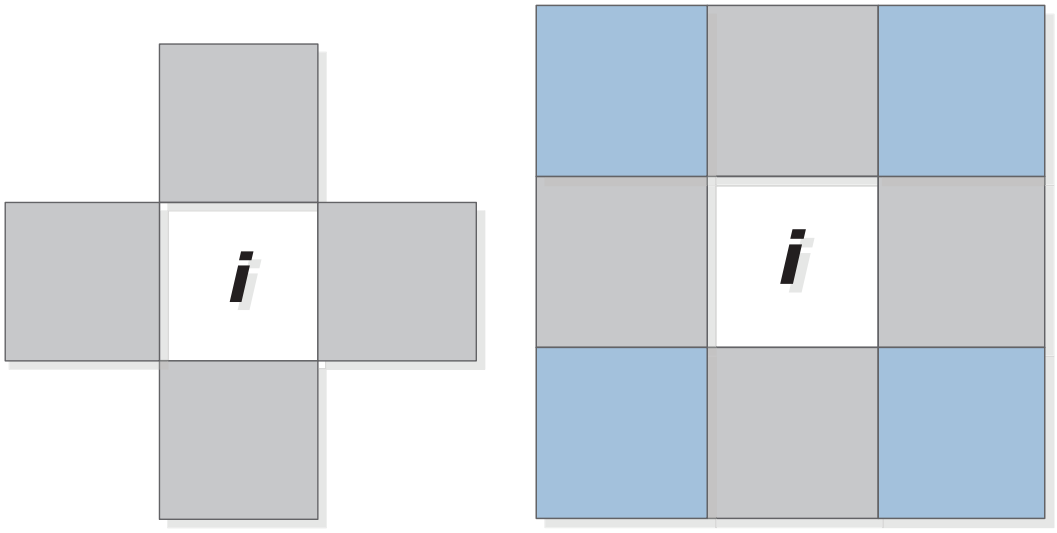
Neighbourhood structure on image graphs

Spatial dependency of a set of sites[6] $\mathbf{S}$ in an image is defined by a neighbourhood system. A guiding principle is that information contained in the local neighbourhood of a site i is sufficient to obtain a good global image representation. This is attributed to equivalence of markovianity and Gibbs Random Field (GRF) properties as per Hammersley and Clifford (1971) theorem. Markovianity indicates that a label of a site i is only dependent on its neighbours N_i, while in GRF the label given to a specific site is affected by labels given to all other sites.

Common neighbourhood structures used in image analysis are the first and second order neighbourhood systems see (Tso and Mather, 2009, chap. 8). First order neighbors of a site/pixel are the four pixels sharing a side with the given pixel, as shown in Figure 2.9a. Second order neighbours are the four pixels having corner boundaries with the pixel of interest, as depicted by Figure 2.9b. Higher order neighbours can be extended in a similar fashion. Basically, the neighbourhood systems are used to define relationship between nodes via clique potential functions.

In an undirected graph, a clique is a subset of nodes $C \in S$ which are fully connected as illustrated in Figure 2.10. All members of a clique are mutual neighbours. Cliques play a central role

[6] Site refers to a pixel at a given location in the image.



(a) First order four neighbours of pixel i **(b)** Second order eight neighbours of pixel i

Figure 2.9.: Neighbourhood systems as defined in (Tso and Mather, 2009, chap. 8). Higher-order can be extended in a similar fashion.

in graphical modelling and inference. They are the framework in which potentials are built. A potential such as $\phi(x, y)$, is a non-negative function of the variable x and y, $\phi(x, y) \geq 0$. Potentials are expressed by exponential functions because they have the property of never being zero. Thus, cliques are exploited to express spatial dependency via potentials. In addition, since cliques express relationships, they can also be adopted to model temporal dependency in a sequence of images. For instance, a site i in time t can depend on a site k in a subsequent or previous, i.e. $t \pm 1$, as shall be discussed in Section 3.1.3.

It has been demonstrated that GMs captures context (spatial and/or temporal) via graphs by representing the entire pixel dependency as a product of local ones. This reduces complexity of the model because distribution over a large number of random variables is factorized to a product of local functions that each depend on only a small number of variables. Such a robust framework supports the first law of Geography that "everything is related to everything else, but near things are more related than distant things" by Tobler (1970). In the next sections we discuss the two types of undirected GMs.

2.2.2 Markov random fields

Markov random fields are popular undirected graphical models as a result of earlier works by Geman and Geman (1984); Besag (1986). MRFs offer a generative structure where the posterior distribution over the labels $\mathbf{y} = y_{i \in S}$ given the data $\mathbf{x} = x_{i \in S}$ is modelled from prior and likelihood of Bayesian formula in Equation (2.7). For computation tractability data features/pixels in the

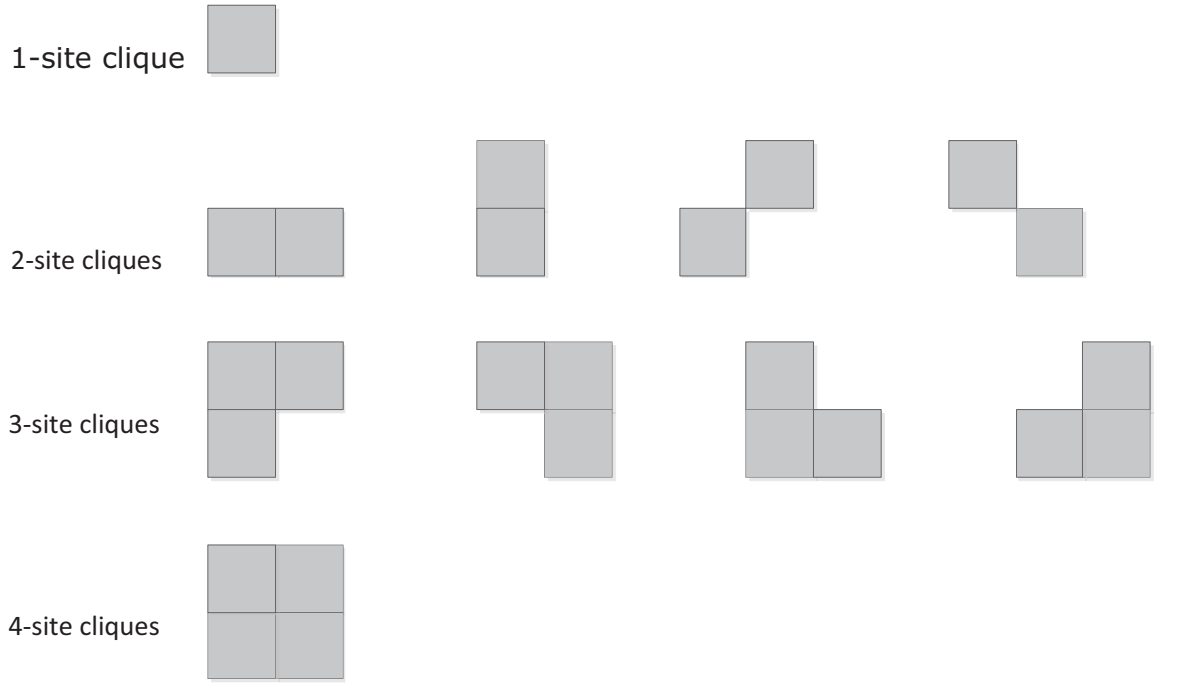


Figure 2.10.: Configuration of cliques in a second order neighbourhood system.

likelihood are assumed to be conditionally independent given the class labels (Figure 2.11a). This simplifies the model as:

$$P(\mathbf{y}|\mathbf{x}) \propto \prod_{i \in S} P(x_i|y_i)P(y). \tag{2.13}$$

The prior $P(y)$ is modelled as a MRF over labels only with the following properties:

1. Positivity: $P(y) > 0$ $\forall$ possible configurations of y,
2. Markovianity: $P(y_i|y_{\setminus i}) = P(y_i|y_{N_i})$, and
3. Homogeneity : $P(y_i|y_{N_i})$ is the same $\forall$ sites i.

where $y_{\setminus i}$ is the set of all pixels in S excluding i and N_i are the neighbours of site i. Markovianity property shows that labelling of a site i is dependent on its neighbours (local neighbourhood property). Homogeneity property defines the likelihood for a label at site i given its neighbourhood regardless of the relative position of i in S (Figure 2.9).

With reference to the three properties, MRF classification considers the prior model to be homogeneous and isotropic Potts model with only pairwise clique potentials, that is,

$$P(y) = \beta(y_i, y_j) \equiv \begin{cases} \beta & \text{if } y_i = y_j \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

where i and j are adjacent sites and β is the spatial interaction parameter that regulates smoothness of land-cover based on similarity of adjacent labels. Therefore, posterior distribution as per MRF formulation can be written as:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left\{\sum_{i\in S} \log P(x_i|y_i) + \sum_{i\in S}\sum_{j\in N_i} \beta(y_i, y_j)\right\} \tag{2.15}$$

where $Z(\mathbf{x})$ is a data normalizing constant known as the partition function computed as:

$$Z(\mathbf{x}) = \sum_{\forall \text{ forms of } \mathbf{y}} \exp\left\{\sum_{i\in S} \log P(x_i|y_i) + \sum_{i\in S}\sum_{j\in N_i} \beta(y_i, y_j)\right\}. \tag{2.16}$$

As depicted in Equation (2.13) the likelihood model of MRFs is assumed to be a fully factorized form for computational tractability. The assumption is too restrictive as it ignores spatial dependencies inherent in images when assigning labels to classes (Hoberg and Rottensteiner, 2010; Zhong and Wang, 2007b). In addition, when the observations are conditioned on the labels, they are not independent from each other as assumed by MRF (Parikh and Batra, 2008). While it is essential to have a model that attains a tractable inference, it is also desirable that it represents the data without making unwarranted independence assumption. To fulfil both requirements, the conditional distribution over the labels given the observed data is modelled by CRFs instead of the joint probability distribution over both labels and the observations (Sutton and McCallum, 2006). This avoids modelling a complicated probability distribution function over the data, $P(\mathbf{x})$, that can lead to intractable models. The difference between MRFs and CRFs is attributed to modelling approaches adopted for them, that is generative and discriminative respectively.

Generative methods learn a model of the joint probability, $P(x, y)$, of data x and label y. Posterior probability estimates, $P(y|x)$, are then made using Bayes' rules in Equation (2.10) and the most likely label y selected. This requires learning the prior and the class conditional densities using likelihood function in order to compute posterior probabilities. In contrast, discriminative approaches model the posterior probability directly, or determine a suitable direct mapping of data x to labels y. Basically, the motivation is to directly address the classification problem as opposed to an intermediate problem of solving the likelihood function. For sure it simplifies the model but, as noted in (Ulusoy and Bishop, 2005) generative approaches have some benefits:

1. They can handle missing data or partially labelled data, and can augment small quantities of expensive labelled data with large quantities of cheap unlabelled data.

2. A new class can be added incrementally by learning its likelihood function independently of all the previous classes.

3. Generative models can readily handle diverse class composition whereas standard discriminative models need to see all possible class combinations during training.

In summary, Bradski and Kaehler (2008) notes that discriminative models are good for yielding predictions given the data while generative models are good for giving more powerful representations of data or for conditionally synthesizing new data. Therefore, this means that MRF performs better when it meets new data distribution such as not represented in the training site and can allocate it to nearest class something discriminative models may be weaker in. On the contrary, discriminative approaches like CRFs are good for making predictions. Generally, discriminative classifiers are often preferred to generative ones. In Section 2.2.3 we discuss the discriminative approach adopted by CRFs.

2.2.3 Conditional random fields

Conditional random fields were introduced by Lafferty et al. (2001) for one-dimensional text classification and extended to two-dimensional image classification by Kumar (2006). They are undirected graphs that represent conditional probability distribution over a set of data/data sequence. The conditional probabilities are represented in potentials functions defined over cliques. In CRFs, posterior probability of a distribution is computed directly, unlike in MRF, as a product of potentials through inference techniques.

Definition Consider a sequence of images $\mathbf{x} = \mathbf{x}_t$ modelled over corresponding discrete labels $\mathbf{y} = \mathbf{y}_t$, from a given set of class labels $l \in m$, acquired at different times t where $t = 1, \ldots, T$ (see Figure 2.7). Let $G = \{S, E\}$ be a graph with spatial edges E defined over a pair of cliques i and j in a neighbourhood set N such that $\mathbf{y} = (y_i)_{i \in S}$ so that $\mathbf{y}$ is indexed by nodes S of the Graph G. In mono-temporal classification, the random variable $(\mathbf{y}, \mathbf{x})$ is a CRF only if, when conditioned on $\mathbf{x}$, the random variable y_i obeys the Markov property with respect to G: $P(y_i|\mathbf{x}, \mathbf{y}_{\setminus i}) = P(y_i|\mathbf{x}, \mathbf{y}_{N_i})$, where $\mathbf{y}_{\setminus i}$ is the set of all nodes in the G except node i and N_i is a set of neighbours of node i in G.

Following Hammersley and Clifford (1971) basic theorem, the joint distribution over the labels $\mathbf{y}$ given the data $\mathbf{x}$ can be written as:

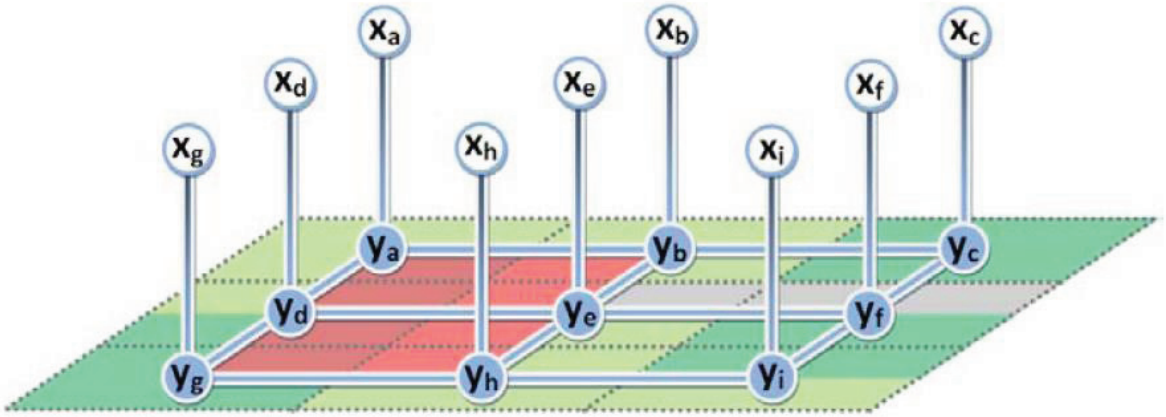
$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left\{\sum_{i \in S} A(y_i, \mathbf{x}) + \sum_{i \in S} \sum_{j \in N_i} I(y_i, y_j, \mathbf{x})\right\} \tag{2.17}$$

where A and I are known as association (unary) and interaction (pairwise) potentials respectively. In CRF, the partition function can be expressed as:

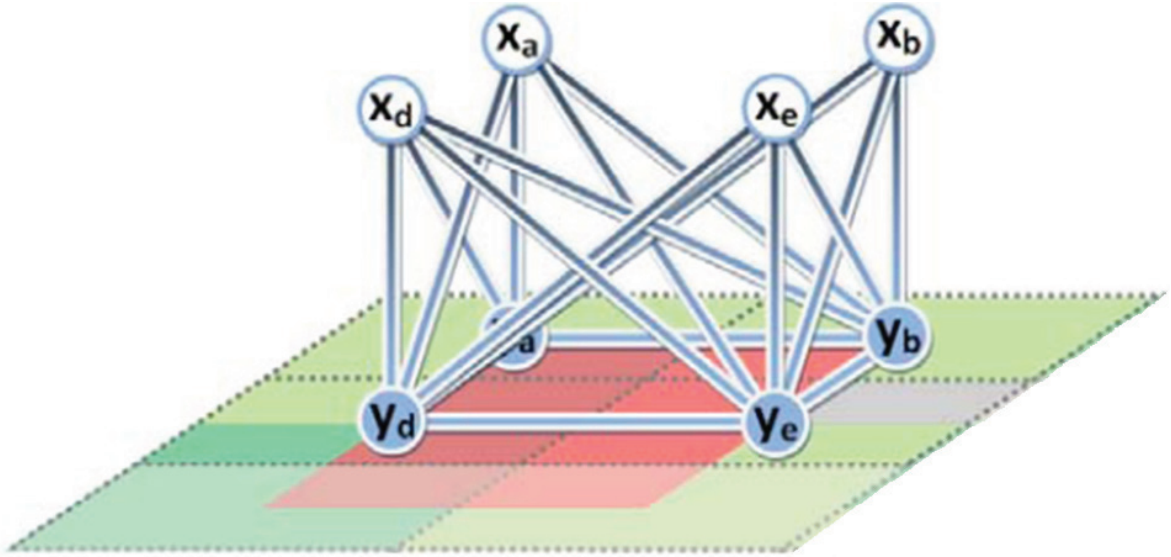
$$Z(\mathbf{x}) = \sum_{\forall \text{ forms of } \mathbf{y}} \exp\left\{\sum_{i \in S} A(y_i, \mathbf{x}) + \sum_{i \in S} \sum_{j \in N_i} I(y_i, y_j, \mathbf{x})\right\}. \tag{2.18}$$

The association and interaction potentials can be regraded as arbitrary local classifiers. This property enables use of domain specific discriminative classifiers in structured data rather than restricting the potentials to a certain form (Zhong and Wang, 2007b). The *association potential* determines how likely a site i takes a label y_i given the observed data $\mathbf{x}$. It can be determined in several ways, for instance Kumar and Hebert (2003); Zhong and Wang (2007b,a); Korc and Förstner (2008) use a general logistic function as a local discriminative classifier and in (Hoberg et al., 2015) a multivariate Gaussian model is used. On the other hand, *interaction potential* quantifies the influence of data and neighbouring labels on site i. It ensures that site i, as initially determined by A, is labelled to its corresponding "true class" given data evidence $\mathbf{x}$ and neighbourhood dependency N where $j \in N_i$. By so doing, it imposes spatial interaction and can be seen as a label and data dependent smoothing function. The homogenous and isotropic Potts model, $I = \beta y_i y_j$, in MRF framework does not permit data dependent interaction due to the assumption of conditional independence in observed data. This ignores existing spatial dependencies in data. The MRF interaction model penalizes each dissimilar pair of labels by a constant smoothness parameter (β). Such a model gives preference to piecewise constant smoothing without explicitly accounting for discontinuities in the data (Kumar and Hebert, 2003). In contrast, I in CRF is a function of all the observed data $\mathbf{x}$ see Figure 2.11b. Different models of data dependent spatial interactions are used to design I in CRF, e.g. concatenation of features (Zhong and Wang, 2007a,b; Li et al., 2009), absolute difference of features (He et al., 2008; Wegner et al., 2011a), contrast sensitive model (Shotton et al., 2009; Schindler, 2012) and inverse/transformed euclidean distance (Kenduiywo et al., 2014).

The CRF model is normally used for classification after determining interaction and association potentials. This requires determination of posterior probability $P(\mathbf{y}|\mathbf{x})$ in Equation (2.17) and the optimal label $\hat{y}$ of each node i. Computation of posterior probability involves evaluation of partition function $Z(\mathbf{x})$. Solution of Equation (2.18) is computationally intractable because it involves determination of marginal probability over m label configurations by a sum-product operation over the set of nodes S, which is exponential in the number of nodes, i.e. m^S. Thus, sampling techniques or approximations such as pseudo-likelihood, mean field or Loopy Belief Propagation (LBP) are normally used to infer a solution. This study adopts sum-product LBP which is a standard approximate inference algorithm in undirected graphs (Murphy et al., 1999). Loopy belief propagation applies belief propagation, an inference technique originally developed for trees in Pearl (1988) but recently modified for "loopy" (cyclic) graphs such as MRFs or CRFs. It works by passing messages (called marginals) from variable nodes to neighbouring nodes via edges using factor graphs. This is a specific form of sum-product algorithm (Bishop, 2006, chap. 8.4.4) used for exact inference on trees. However, for cyclic graphs LBP is not guaranteed to converge, that is, it may end up in a local extremum. Nonetheless, experimental results have demonstrated that LBP gives good estimates of marginal probability (Murphy et al., 1999). For detail description of LBP inference the reader is referred to Bishop (2006) and Barber (2012).



(a) Labelling of nodes in MRF



(b) Labelling of nodes in CRF

Figure 2.11.: Comparison of how spatial context is incorporated to labelling of nodes in; (a) MRF and (b) CRF (Wegner, 2011).

After posterior probabilities of nodes are computed through inference, image labelling is performed by determining optimal class label $\hat{y}$. This is determined by Maximum A Posterior (MAP) estimate:

$$\hat{y} = \underset{l \in m}{\arg\max} [P(y_l|x, i)] \tag{2.19}$$

where *arg max* denotes maximum value of the argument.

In summary, it is important to note the following distinct properties of CRF:

1. It is a discriminative model,

2. the conditional independence of data adopted in MRF (Equation (2.13)) for tractability is relaxed,

3. global data interactions may be incorporated in the association potential,

4. the interaction potential models interactions in both labels and observations unlike in MRF where only labels are considered as shown in Figure 2.11,

5. through markov and Gibbs equivalence, labels and observations in the interaction potential are not limited to the local neighbours, but factorizes globally.

These properties guided the choice of CRF to model temporal interactions and phenology on crops. On the other hand, radar sensors are daylight and weather independent. These renders them a good medium to deliver a sequence of images of highest temporal density suitable for crop classification regardless of climatic zones. The next chapter describes study area, data and methodology used for crop classification.

3 Methodology

Techniques implemented in this research including new contributions are described in this chapter. Dynamic conditional random fields classification framework is designed in order to model spatial and temporal dependencies. The DCRFs framework was designed for supervised spatial-temporal classification of a sequence of images; where class labels were predicted from temporal data using a sample of pre-labelled data instances. This involved design of association, interaction and temporal potentials that determined spectral, spatial and temporal information respectively. First order and higher temporal interactions were modelled.

3.1 Dynamic conditional random fields

Sequence classification requires determination of posterior probability $P(\mathbf{y}_{1,\ldots,T}|\mathbf{x}_{1,\ldots,T})$. The computation is intractable and exponential in time as it involves computation of S^T functions in a 2-D space. Since satellite observation of crops is unique and discrete in each instance, we assumed that their evolution is independent. In this way, the conventional class conditional independence (Swain, 1978) was attained: $P(\mathbf{y}_{1,\ldots,T}|\mathbf{x}_{1,\ldots,T}) = P(\mathbf{y}_1|\mathbf{x}_1), \ldots, P(\mathbf{y}_T|\mathbf{x}_T)$. This simplified the classification problem to an independent estimation of class posterior probabilities $P(y_i)$ for each node i in t. Spatial interactions are also considered at each epoch by I in Equation (2.17).

To exploit crop phenology information, we extended DCRFs proposed by Sutton et al. (2007) for 1-D text sequence classification to 3-D image sequence classification. A DCRF is a conditionally trained undirected graphical model whose structure and parameters are repeated over a sequence. The DCRF can be designed using a template that defines graphical structure, features, and weights for any two epochs. The same set of features and weights are used at each epoch so that the parameters are tied across the network. Therefore, the study developed an undirected DCRF graph template that factorizes according to first order Markov assumption as depicted in Figures 3.1 and 3.2. In the design, each node i at time t can depend on node data from the previous (if $t \neq 0$) and subsequent (if $t \neq T$) epochs. The objective is to connect a set of all possible temporal cliques C of nodes S using a conditional probability matrix distribution $P(\mathbf{y}|\mathbf{y}_{t-1}, \mathbf{x}, \mathbf{x}_{t-1})$. This set-up gives a DCRF sequence template model, considering crop phenological information captured by radar, such that a node can have at least one or two temporal neighbours (Figure 3.2).

Definition Let c, $c \in C$, be a temporal clique index of a node k in epoch $\Delta t = t \pm 1$ of a label vector $\mathbf{y}_{\Delta t}$ which corresponds to another node i in label vector $\mathbf{y}_t$ at time t such that $c = \{k, \Delta t\}$ as illustrated by Figure 3.2. In this case, a set of random variables $\mathbf{y}_{i,t,c} \equiv \{y_{i,t}|y_{(k,\Delta t)\in c}\}$ is the set of variables of the evolving clique index c at time t in the sequence of T images. Then, our spatial-temporal DCRF template can be expressed as illustrated by Equation (3.1).

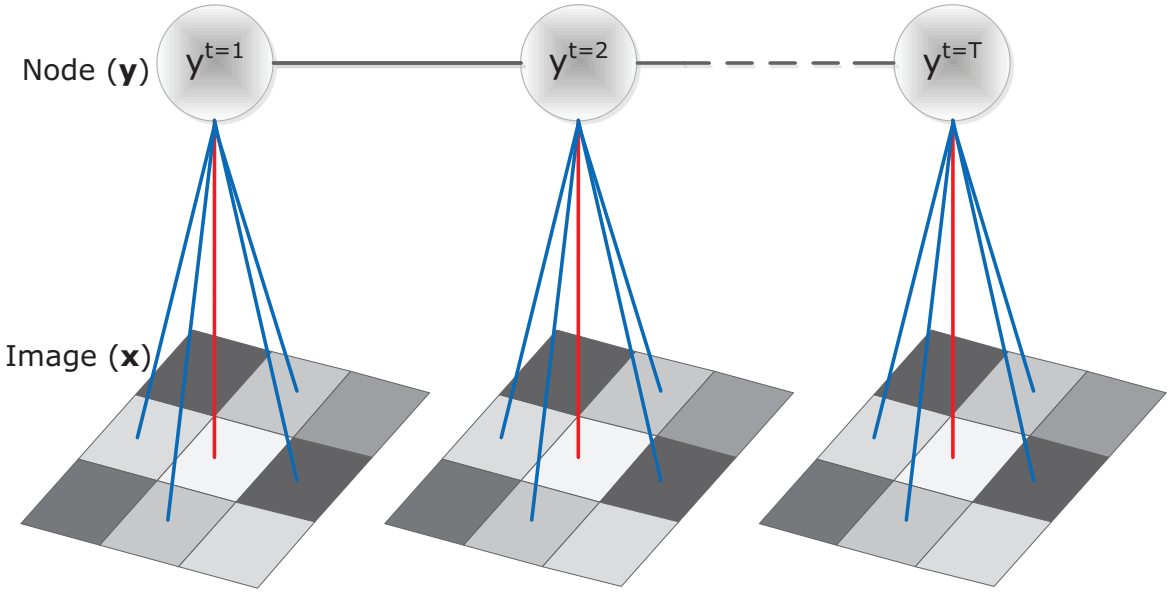


Figure 3.1.: First order DCRF sequence graph showing a subset of temporal nodes and image data in a sequence. First order spatial (N=4) and temporal interactions are illustrated by blue and black lines respectively while red lines depict node potentials.

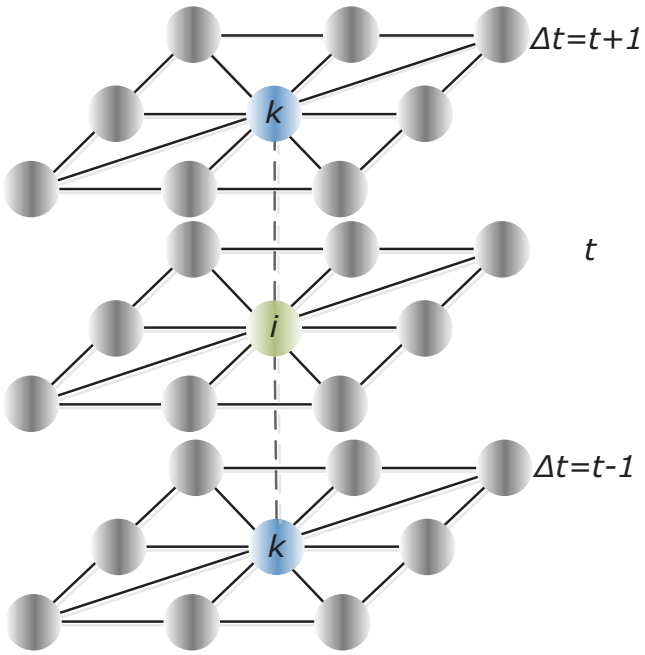


Figure 3.2.: First order temporal (k) and second order spatial (N=8) neighbours of node i in a sequence of T images (see Figure 3.1). Spatial and temporal edges are indicated by solid and dashed lines respectively.

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left\{\sum_{i\in S} A(y_i, \mathbf{x}) + \sum_{i\in S}\sum_{j\in N_i} I(y_i, y_j, \mathbf{x}) + \sum_{t\in T}\sum_{i\in S}\sum_{c\in C} TP(y_{i,t,c}, \mathbf{x}, \mathbf{x}_{\Delta t})\right\} \tag{3.1}$$

In Equation (3.1), *TP* denotes temporal potential. The association, interaction and temporal potentials can be considered as local domain specific classifiers which were implemented as described in Sections 3.1.1 to 3.1.3.

3.1.1 Association Potential

The association potential determines how likely an image site i takes a label y_i given the data $\mathbf{x}$:

$$A(y_i, \mathbf{x}) = P(y_i|\mathbf{f}_i(\mathbf{x})) \tag{3.2}$$

where $\mathbf{f}_i(\mathbf{x})$ is a site-wise feature vector (Kumar, 2006). We used random forests (Breiman, 2001) to determine A by independent classification of different epochs assuming that evolution of a crop is unique over the sequence. RF consists of several decision tree classifiers D_T where each tree is generated using a random vector sampled independently from the training set of input vectors. Each tree then casts a placement vote for the most popular class given an input vector $\mathbf{x}$. For instance, if the number of votes cast for a given class label y by RF is V_y, then our A at site i is:

$$P(y_i = y|\mathbf{f}_i(\mathbf{x})) = \left(\frac{V_y}{D_T}\right). \tag{3.3}$$

We set $D_T = 350$ because over 200 trees RF stabilizes (Hastie et al., 2011, chap. 15) and defined tree depth as 25.

The decision trees presents a hierarchical categorization of data by splitting (e.g. using Gini impurity index). Basically, RF is an ensemble technique that applies a combination of bagging and random subspace methods (Tso and Mather, 2009). The principle of bagging (Breiman, 1996) is to average many noisy but approximately unbiased classifiers, hence reduce the variance. This improves classification accuracy and avoids overfitting. Trees are ideal candidates for bagging, because they can capture complex relationships in data, and if grown sufficiently deep, have relatively low bias. Since trees are noisy, they benefit greatly from the averaging (Hastie et al., 2011, chap. 15) which reduces variance. For more details on RF the reader is referred to (Breiman, 2001; Tso and Mather, 2009; Hastie et al., 2011). In summary, Breiman (2001) observes the following benefits that made us select RF for determining A:

1. its accuracy is as good as Adaboost and sometimes better,

2. it is relatively robust to outliers and noise,

3. it is faster than bagging or boosting, and

4. it shows useful internal error estimates and variable importance.

3.1.2 Interaction Potential

Remote sensing images of the physical world portray a coherent and spatially smooth scene because of strong spatial context among neighbouring pixels (Figure 2.6). Moreover, the first law of Geography "everything is related to everything else, but near things are more related than distant things" by Tobler (1970), supports the premise. This elementary piece of prior knowledge is modelled by the interaction potential. It measures the influence of data and neighbouring labels on site i. It ensures that site i, as initially determined by association potential, is labelled to its corresponding "true class" given data evidence $\mathbf{x}$ and neighbourhood dependency N where $j \in N_i$. We set $N = 8$, second order neighbourhood structure, as shown in Figure 2.9b, in order to include more spatial interactions to node i. In principle the I imposes context using pairwise node cliques to penalize local changes in class labels if data (image features) are similar and also penalize identical class labels when the features are different. So far the proposed contrast sensitive model (Shotton et al., 2009; Schindler, 2012) prefers to enforce a higher smoothness penalty to a pair of similar labels than to dissimilar ones as shown in Equation (3.4).

$$I(y_i, y_j, \mathbf{x}) = \begin{cases} \beta \cdot \exp\left(-\eta \cdot d_{ij}\right) & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases} \tag{3.4}$$

In Equation (3.4), β is a spatial interaction parameter that regulates smoothness and parameter η controls inclusion ($\eta = 1$) or exclusion ($\eta = 0$) of data interactions d_{ij} of adjacent node features $\mathbf{f}_i$ and $\mathbf{f}_j$, that is:

$$d_{ij} = \frac{\sqrt{\sum_{i=1}^{R} \mid \mathbf{f}_i(\mathbf{x}) - \mathbf{f}_j(\mathbf{x}) \mid^2}}{R} \tag{3.5}$$

where R is the number of features/elements in vectors $\mathbf{f}_i$ and $\mathbf{f}_j$. Division by R ensures identical influence of I in all images (Hoberg et al., 2015) because it converts the sum of Euclidean distance of features from adjacent sites to mean which is a better estimate of central tendency.

This study introduced two designs of I that enforce context by considering scenarios when adjacent labels are either similar or dissimilar given data evidence. The two models are based on a new version of contrast sensitive Potts model and Pearson correlation.

New version of contrast sensitive Potts model

This model of I is modified from Equation (3.4) as:

$$I(y_i, y_j, \mathbf{x}) = \begin{cases} \beta \cdot \exp\left(-\eta \cdot d_{ij}\right) & \text{if } y_i = y_j \\ \beta \cdot \left(\text{Max}[1 - \exp(-\eta \cdot d_{ij}), \epsilon]\right)^{-1} & \text{if } y_i \neq y_j \end{cases} \tag{3.6}$$

where Max is a function that returns a maximum between two values and ϵ is a value close to zero (it prevents division by zero). Therefore, the model is different from contrast sensitive Potts model because transitions of adjacent labels are now moderated based on data evidence both when initial labels from A are similar or dissimilar. In this manner, the model regulates smoothing while preserving edges.

Pearson correlation coefficient model

In statistics, Pearson product-moment correlation coefficient ρ or simply the correlation coefficient, is a measure of the linear correlation between two variables, in this case node $\mathbf{x_i}$ and $\mathbf{x_j}$ (Webster and Oliver, 2007). It is a dimensionless value between +1 and -1, where 1 refers to total positive correlation, 0 is no correlation, and -1 is total negative correlation. Correlation quantifies interdependence of variables hence, the reason we used it to quantify context in adjacent nodes i and j as:

$$I(y_i, y_j, \mathbf{x}) = \begin{cases} \beta \cdot \exp\left(\eta \cdot \rho_{ij}\right) & \text{if } y_i = y_j \\ \beta \cdot \left(\exp(-\eta \cdot \rho_{ij})\right)^{-1} & \text{if } y_i \neq y_j \end{cases} \tag{3.7}$$

where ρ_{ij} is the Pearson correlation coefficient of two adjacent nodes i and j. Pearson correlation coefficient is computed as:

$$\rho_{ij} = \frac{\sum_{q=1}^{q=R} \left\{(x_i - \bar{x}_i)(x_j - \bar{x}_j)\right\}}{\sqrt{\sum_{q=1}^{q=R} \{(x_i - \bar{x}_i)^2} \sqrt{\sum_{q=1}^{q=R} (x_j - \bar{x}_j)^2}} \tag{3.8}$$

where R, as earlier mentioned, is the number of elements in vectors $\mathbf{x}_i$ and $\mathbf{x}_j$.

3.1.3 Temporal Potential

Multitemporal mapping techniques exploit temporal context to improve classification. Temporal context describes inter-class and/or intra-class transitions in time. Changes during crop growth are largely due to phenology and other external dynamics such as farm management, artificial or

natural activities. We modelled these aspects using the temporal potential. It captures interactions between nodes in the sequence of images T. This potential can be considered as a classifier that ensures mutual information exchange between nodes i and k in epochs t and Δt. Since the posterior probabilities $P(y_i = l)$ and $P(y_k = l)$ are determined by A, our TP can be expressed as:

$$TP(y_{i,t,c}, \mathbf{x}, \mathbf{x}_{\Delta t}) = P(y_i = l | y_k = l). \tag{3.9}$$

To solve Equation (3.9), Bayesian formula was used to estimate the probability of a node i being assigned a class label l given a corresponding temporal node k in the sequence:

$$P(y_i = l | y_k = l) = \frac{P(y_i = l, y_k = l)}{P(y_k = l)} \equiv \frac{P(y_i = l, y_k = l)}{\sum_{y_i} P(y_i = l, y_k = l)}. \tag{3.10}$$

Equation (3.10) determines the probability of a crop label l, where $l = \{1, \ldots, m\}$, being assigned to a node i given data from the two epochs t and Δt. It is used to compute a pixel-wise $m \times m$ conditional probability matrix representing temporal interactions in m classes between a pair of epochs as illustrated by Table 3.1 and earlier proposed in Kenduiywo et al. (2015).

Table 3.1.: Determination of pixel-wise conditional probability matrix between two sites i and k in two different epochs.

$y_i \Rightarrow y_k$ \ $y_k \Rightarrow y_i$	$y_k = 1$	$y_k = 2$	$\ldots$	$y_k = m$	**Sum**
$y_i = 1$	$P(y_i = 1, y_k = 1)$	$P(y_i = 1, y_k = 2)$		$P(y_i = 1, y_k = m)$	$P(y_i = 1)$
$y_i = 2$	$P(y_i = 2, y_k = 1)$	$P(y_i = 2, y_k = 2)$		$P(y_i = 2, y_k = m)$	$P(y_i = 2)$
$\vdots$					
$y_i = m$	$P(y_i = m, y_k = 1)$	$P(y_i = m, y_k = 2)$		$P(y_i = m, y_k = m)$	$P(y_i = m)$
Sum	$P(y_k = 1)$	$P(y_k = 2)$		$P(y_k = m)$	**1**

3.2 Higher order DCRFs with expert phenology

The standard DCRFs model introduced in Section 3.1 considers first order temporal interactions. This is based on the belief that neighbouring nodes in a time sequence are more dependent than nodes further away. However, phenological information varies across the life cycle of a crop and this is captured by radar images as depicted by Figure 3.3. Hence, it is necessary to include temporal dependencies from the entire sequence of images to a node in a given epoch. Nonetheless, there exist different instances in the life cycle where backscatter information of a crop is high as highlighted in Section 2.1.2. The backscatter may also be low or completely from undergrowth or crops left over after harvesting. In each case, the backscatter information is useful for discriminating that particular crop parcel. This means that temporal nodes have varying magnitude of

phenological information. We thus used expert phenology knowledge to assign weights to each crop type in each epoch in the sequence. The weights are estimated based on anticipated proportion of crop backscatter given its phenology knowledge at the instance of image acquisition. Therefore, the temporal interaction model in Section 3.1.3 was redesigned to accommodate expert weighted temporal phenological information in the entire sequence as depicted in Figure 3.4. In this configuration, each node has temporal interactions up to n^{th} order (where $n = T - 1$) in the sequence of T images. Therefore, higher order DCRFs can also be termed as n^{th} order DCRFs.

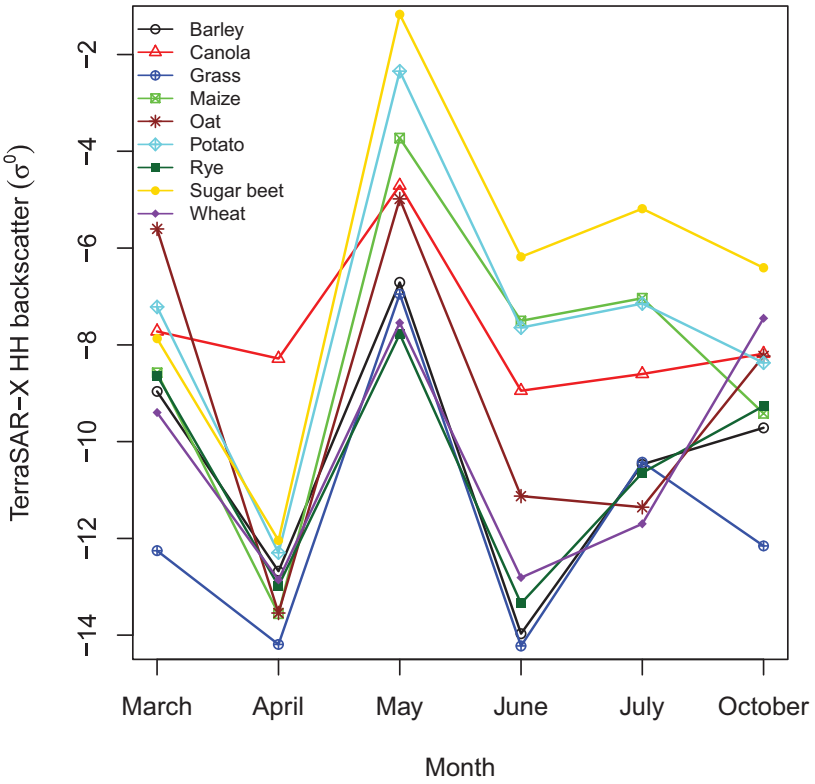


Figure 3.3.: Mean backscatter variation of crops as captured in different TerraSAR-X HH polarized image acquisitions in the year 2009.

Now, let k be an index of a set of temporal nodes K in all other epochs $1 : T \setminus t$ (the set of all other epochs except t) that corresponds to node i in epoch t. Then, the temporal potential can be redefined as:

$$TP(y_i, y_{k \in K}, \mathbf{x}_{1:T}, t) = P(y_i | y_{k \in K}) \tag{3.11}$$

where $1 : T$ refers to the sequence of epochs from 1 to T.

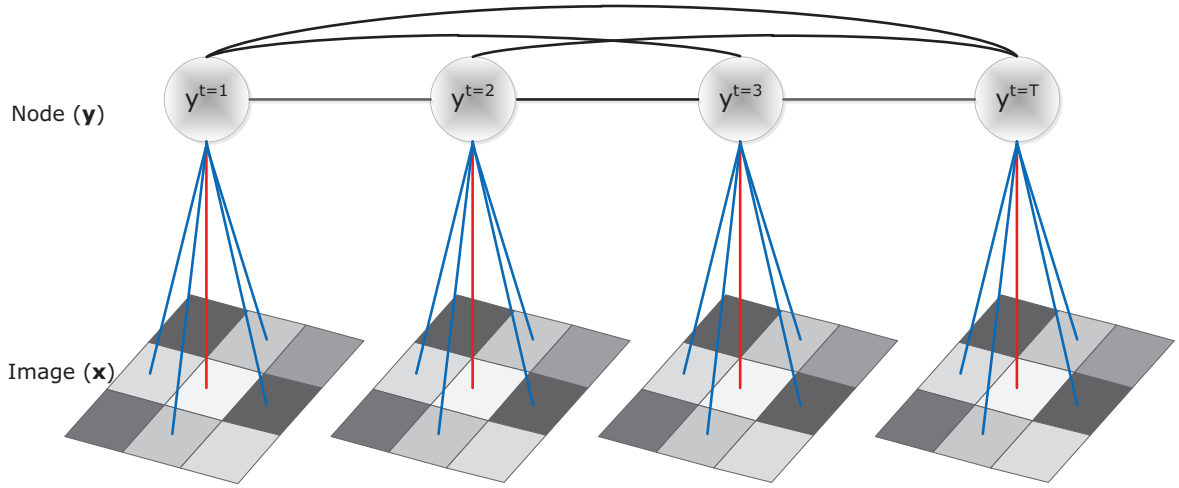


Figure 3.4.: Higher order DCRFs sequence graph showing a subset of nodes and image data in a sequence. First order spatial and higher (3^{rd}) order temporal interactions are illustrated by blue and black lines respectively while red lines are node potentials.

Since class conditional independence is already adopted (see Section 3.1) then Equation (3.11) can be simplified as:

$$P(y_i|y_{k\in K}) = P(y_i|y_{k=1})\,P(y_i|y_{k=2})\ldots P(y_i|y_{k=K}) \tag{3.12}$$

A weights vector ϕ representing prior expert phenological information of each crop type was then introduced to a vector of probabilities determined in each epoch by A. This means phenological information exchange between all temporal nodes now include expert knowledge. Therefore, Equation (3.12) can further be simplified and expert phenology integrated by:

$$P(y_i|y_{k\in K}) = \prod_{k=1}^{K} P(\phi_i \cdot y_i|\phi_k \cdot y_k). \tag{3.13}$$

where ϕ_i and ϕ_k are vectors of expert phenology weights corresponding to each crop type l at nodes i and k in different epochs respectively. Each of the conditional probability terms in Equation (3.13) were used to compute the conditional probability matrix between pairwise temporal node cliques i and k as earlier demonstrated in Equation (3.10) and Table 3.1 but, now the probabilities are weighted. These matrices capture inherent image and expert phenological information

which enhance bidirectional temporal information exchange between node i and all other K temporal nodes. To sum up, the designed higher order DCRFs model can now be expressed as:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left\{ \sum_{i \in S} A(y_i, \mathbf{x})^{\psi_1} + \sum_{i \in S} \sum_{j \in N_i} I(y_i, y_j, \mathbf{x})^{\psi_2} + \sum_{t \in T} \sum_{i \in S} \sum_{k \in K} TP(y_i, y_k, \mathbf{x}_{1:T}, t)^{\psi_3} \right\} \tag{3.14}$$

where $\psi_1, \ldots, \psi_3$ are weights corresponding to A, I and TP which are association, interaction and temporal potentials respectively applied at each node. We set $N = 4$ as defined in Figure 2.9a in order to reduce computational complexity. The spatial interaction potential was given a weight $\psi_2 = 2$ which is double that of temporal and association potentials. This acted as a compensation that ensured spatial interactions are not dominated by temporal ones.

3.3 Optimal crop mapping with DCRF

Farmers, governments, insurers, agricultural stock markets and traders and other stakeholders are mostly interested in the quantity of a certain crop in a given season. The proposed DCRFs approaches incorporates temporal information exchange between epochs. However, at each epoch we obtain posterior class probabilities incorporating phenology information inherent in images and expert knowledge in the case of higher order DCRFs. In order to generate an optimal seasonal crop map, an ensemble classifier was necessary. For this reason, an ensemble classifier using maximum F1-score to select the best posterior probability of each class from a sequence of epochs was developed. Selected posterior probabilities were then weighted using user accuracy measure. To generate an optimal seasonal crop map the posteriori probabilities were maximized.

Consider our classification problem where each node $i \in S$ is to be assigned a discrete class label y from $l \in m$ possible classes $(y_1, \ldots, y_m)$ in each epoch t from the sequence of T images. Now given that posterior probabilities $P(y_l|x, t)$ have been determined, then for each class l we selected a probability with maximum F1-score from the sequence and weighted them with user accuracy. A discrete class for node i was determined by maximizing over probabilities selected from the sequence. Since the same training sites are used throughout the sequence (crop season), prior probability in each epoch was assumed equal:

$$\hat{y}_i = \arg\max_{l=1}^{m} \left\{ \max_{t=1}^{T} (\text{F1}\,[P(y_l|\mathbf{x}), t]) \cdot \text{User accuracy} \right\} \tag{3.15}$$

where F1$[P(y_l|\mathbf{x}), t]$ is a probability corresponding to class l with maximum F1-score (F1) at time t and $\hat{y}_i$ is the estimated class label for node i. Sokolova et al. (2006) defines F1-score as:

$$\text{F1-score} = \frac{2(\text{Producer accuracy} \cdot \text{User accuracy})}{\text{Producer accuracy} + \text{User accuracy}}. \tag{3.16}$$

Figure 3.5 illustrates the approach used to design ensemble classifiers for optimal labelling. For instance, Figure 3.5a depicts optimal labelling using first order DCRFs while labelling based on higher order (2^{nd} order) DCRFs is shown in Figure 3.5b.

The approach in Equation (3.15) was compared to the following classifier combinations rules in (Kittler et al., 1998):

$$\textbf{Max rule: } \hat{y}_i = \arg\max_{l=1}^{m} \left\{ \max_{t=1}^{T} P(y_l|\mathbf{x}, t) \right\} \tag{3.17}$$

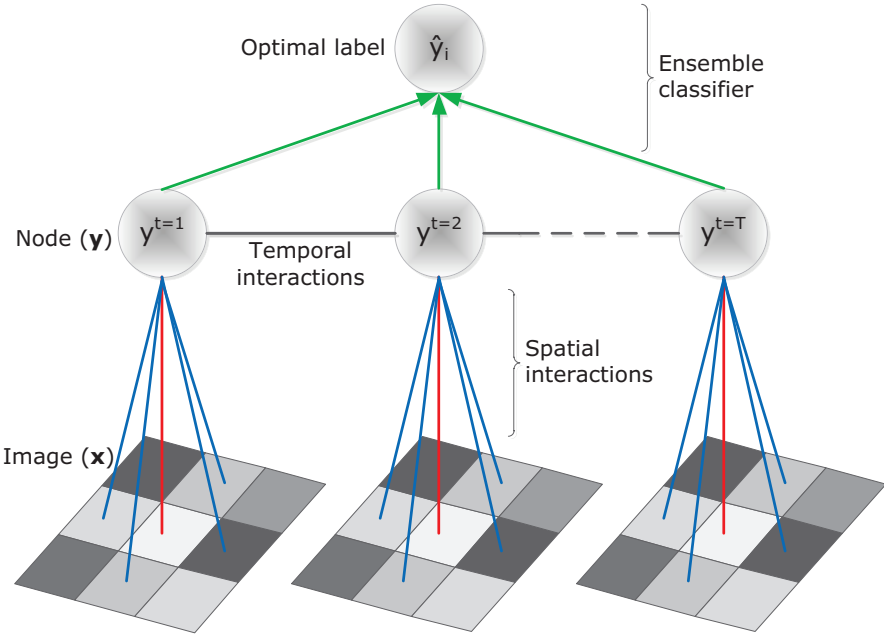
$$\textbf{Majority vote: } \hat{y}_i = \arg\max_{l=1}^{m} \left\{ \sum_{t=1}^{T} \Delta_{lt} \right\} \tag{3.18}$$

$$\textbf{Median rule: } \hat{y}_i = \arg\max_{l=1}^{m} \left\{ \text{med}_{t=1}^{T} P(y_l|\mathbf{x}, t) \right\} \tag{3.19}$$

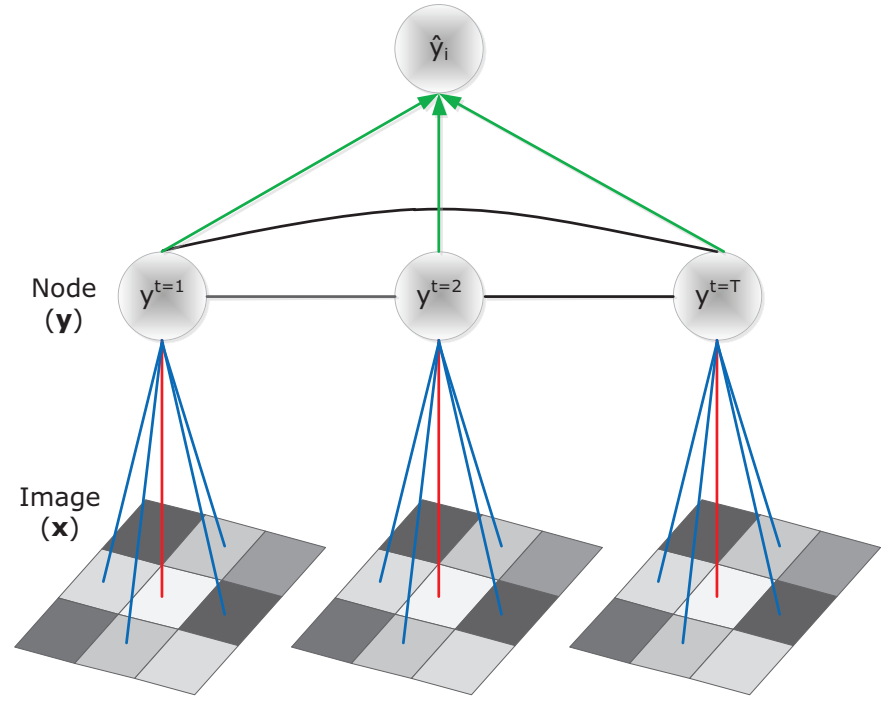
$$\textbf{Product rule: } \hat{y}_i = \arg\max_{l=1}^{m} \left\{ \prod_{t=1}^{T} P(y_l|\mathbf{x}, t) \right\} \tag{3.20}$$

where med is the median estimate and Δ is the frequency of occurrence of a class label l in the sequence.

Equation (3.17) finds the maximum probability of a class from the sequence. The selected posterior probabilities are then maximized to get the final label of a node. Majority vote simply counts the votes received for a class in the sequence. The class which receives the largest number of votes is chosen as the consensus (majority) label. Median rule in Equation (3.19) simply finds the mid probability of class from the sequence. It is preferred to sum rule because if any of the epoch classifiers outputs a posteriori probability for some class which is an outlier, it will affect the average which could lead to an incorrect labelling. Moreover, it is established that a robust estimate of the mean is the median. Therefore, it is more appropriate to base the combined decision on the median of the a posteriori probabilities. On the other hand, the design in Equation (3.20) quantifies the likelihood of a label by combining the a posteriori probabilities generated by the individual classifiers using the product rule. It can be a severe rule of fusing classifier outputs as it is sufficient for a single epoch to assign a close to zero probability to a class. In such a case, the combined output will also be a value close to zero due to reduction effect of multiplication. This effectively undermines the decision of any correctly predicted class in the sequence and denies it an opportunity for its identity to be finally accepted or rejected.



(a) Optimal labelling in first order DCRFs.



(b) Optimal labelling in 2^{nd} order DCRFs.

Figure 3.5.: Illustration of optimal node labelling using class posterior probabilities estimated by first and higher order DCRFs.

3.4 Training

Solutions to Equations (3.1) and (3.14) are obtained by maximizing probabilities, spectral (A), spatial (I) and temporal (TP) using Bayes' MAP estimate as earlier mentioned in Section 2.2.3. This requires an inference algorithm to determine posterior probabilities $P(\mathbf{y}|\mathbf{x})$ and a maximization algorithm to estimate optimum labels $\hat{y}$. We apply sum-product LBP (Murphy et al., 1999), a standard inference algorithm in graphs with cycles. To estimate class labels, we design a maximization algorithm. The association potential probabilities used in both I and TP are estimated from training data using RF implemented in OpenCV (OpenCV, 2014).

3.5 Accuracy assessment

Users of classification maps derived from remote sensing data need to know credibility of the results. Accuracy assessment is thus necessary in order to ascribe confidence limits to the derived classification products. In addition, quality assessment is also key in evaluating the performance of a classification approach. It places a degree of confidence to results and implicitly the classification approach. In remote sensing, confusion matrix/error matrix, is mostly used for this purpose. It shows proportions of correctly classified and misclassified sites in a table. From the table more accuracy measures can derived (see ITC, 2010, chap. 6.2.4). The most direct measure is the *Overall Accuracy* (OA). It is represents the proportion of correctly classified pixels in a classified map with respect to ground reference data used for verification.

Other measures derived from the error matrix such as error of omission and commission are computed per class. *Error of omission* also known as Type I error or false positive, refers to those sample pixels that are omitted in the interpretation result. In contrast, the *error of commission* also called Type II error or false negative refers to incorrectly classified pixels. Omission error is the corollary of producer accuracy, while user accuracy is the corollary of commission error. The user accuracy determines the probability that a pixel labelled as a certain land-cover class in the map indeed corresponds to the initial defined class in training data. Similarly, producer accuracy measure is the probability that a sampled pixel on the map is indeed that particular class on the ground. An additional accuracy measure is the kappa statistic κ. This measure estimates the coefficient of agreement κ by taking into account omission and commission errors in addition to the values of the main diagonal in the confusion matrix. Kappa statistic is a measure of agreement after chance agreement is eliminated. So it can be used to determine whether a classified map is significantly better than chance or random. It has value between 0 and 1. If $\kappa = 0$ then there is no agreement between the classified output and reference data while $\kappa = 1$ indicates a perfect match. Based on kappa, one can test if two classification results, have different levels of accuracy. This type of testing is used to evaluate different techniques of remote sensing classification (ITC, 2010).

4 Experiments

This chapter describes data used and experiments conducted using the designed techniques in Chapter 3 for crop type classification. The tests were conducted in three sites Hannover (Section 4.2) and Fuhrberg (Section 4.1.1) in Germany and Kitale (Section 4.1.2) in Kenya. These sites are in different continents, that is Germany in Europe and Kenya in Africa. TerraSAR-X data acquisition were used in Fuhrberg and Kitale while Sentinel 1 was adopted for Hannover region.

4.1 Crop type mapping using TerraSAR-X images

Fuhrberg and Kitale areas were selected to conduct crop classification experiments from a sequence of TerraSAR-X images.

4.1.1 Fuhrberg

Study site location and data

This experiment site is located in Northern Germany approximately 52.56°N, 9.84°E as illustrated in Figure 4.1. The average annual precipitation and temperature are 656 mm and 8.9°C respectively (Deutscher Wetterdienst, 2012). The region is characterized by intensive agriculture with large farms. Area under active farming is approximately 50%.

Crops grown in Fuhrberg region include: 1) barley, 2) canola, 3) grassland, 4) maize, 5) oat, 6) potato, 7) rye 8) sugar beet and, 9) wheat. These crops go through different phenological stages within a season, a fact that can enhance discrimination. Five general phenology phases from farm activity point of view — preparation, seeding, growing, harvesting and post harvest — were defined (Figure 4.2). Preparation phase involves ploughing and soil grooming processes before seeding. In seeding phase, crop seeds are placed in the soil. Growing phase includes the period between crop germination to ripening. After ripening, harvesting starts by gathering mature crops from the fields. The last stage is post harvest phase, where the field could be fallow or with some remaining ripe crops.

Data acquisitions from TerraSAR-X satellite were used. The temporal sequence consists of six dual polarized (HH and VV) TerraSAR-X High Resolution Spotlight images (see Table 4.1). These images were delivered as ground range products with equidistant pixel spacing. They are radiometrically calibrated to σ^0 according to Equation (2.6) in Envi SARscape 5.2. All images are co-registered to an extent of (7.1 × 11.8) km^2 using WGS 1984 datum on UTM zone 32N coordinate projection system. Selected experiment site covers an extent of (5.4 × 5.4) km^2 as highlighted in Figure 4.1.

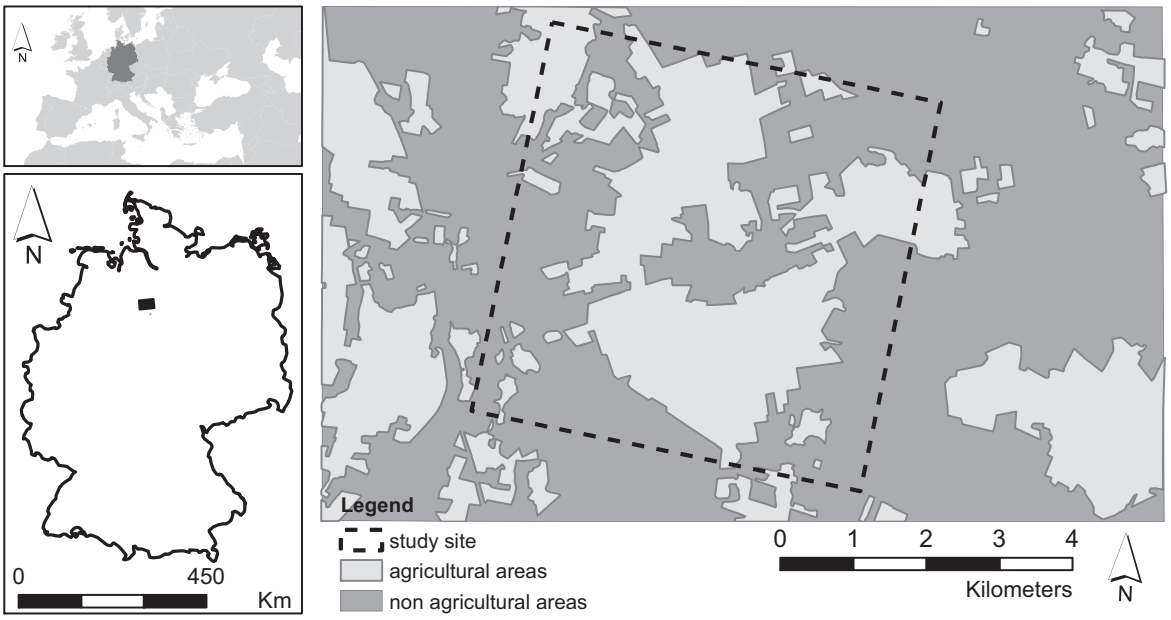


Figure 4.1.: Fuhrberg study area located in Northern Germany.

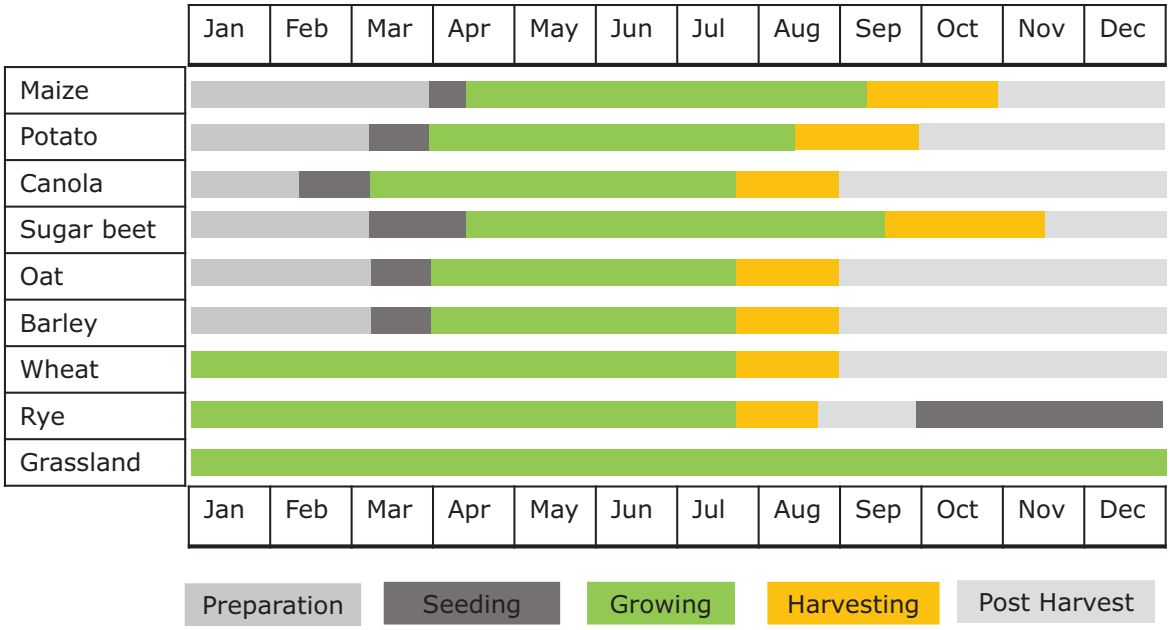


Figure 4.2.: Phenology stages of crops considered for classification in Fuhrberg.

Reference data campaign was conducted during the same year of image acquisition. Table 4.2 illustrates distribution of training and validation data for various crops in the study area. All parcel boundaries were buffered inward by 10 m. The buffers were used to avoid selecting training pixels from the edge of a field, which would create a mixed signal and affect classification accuracy.

Acquisition date	Incidence angle	Resolution (m)	
		Range	Azimuth
11^{th} March 2009	34.75°	2.1	2.4
13^{th} April 2009	34.75°	2.1	2.4
22^{nd} May 2009	43.65°	3.4	2.9
18^{th} June 2009	34.75°	2.1	2.4
10^{th} July 2009	34.75°	2.1	2.4
17^{th} October 2009	34.75°	2.1	2.4

Table 4.1.: TerraSAR-X image acquisitions captured in ascending mode over the study area.

The parcels were then separated into approximately 50% training and validation sets, for each crop type, using stratified random sampling design tool in ArcGIS 10.0 (Buja and Menza, 2013). A minimum spatial threshold of 100 m between parcels was used. The spatial threshold and stratified random sampling ensured that samples were spread over the study area and have some form of randomness with representation of unpopular classes as advocated by Stehman (2009).

Crop	**Training set (ha)**	**Validation set (ha)**
Barley	38.54	41.30
Canola	38.60	40.87
Grassland	69.97	55.10
Maize	27.63	33.10
Oat	10.11	17.39
Potato	55.76	66.45
Rye	97.23	79.04
Sugar beet	52.49	47.51
Wheat	34.88	34.84

Table 4.2.: Distribution of training and validation data.

Feature selection

Image features were computed from TerraSAR-X data of Furhberg in order to mine more information as only fewer images were available in each epoch. The features are image input variables that are pre-processed into some new space of variables where it is intended to make image classification easier (see Bishop, 2006, p. 2). Therefore, Gray Level Co-occurrence Measures (GLCM) were computed using a 3×3 matrix. Eight features — mean, variance, correlation, homogeneity, contrast, dissimilarity, entropy and 2nd moment — were computed in directions 0°, 45°, 90°, and 135° giving rise to a total of 32 features in each polarization. RF variance of importance was used to select 4 significant features from the 8 GLCM features in each direction and polarization (a total of 32 features per epoch). Important GLCM features as per RF include: correlation, homogeneity, variance and mean. For each selected feature, a super pixel/block was generated from a mean of

3×3 pixels. All block features were normalized between 0 and 1 to minimize undue influence by features with high values during classification. Block size selection was done in consideration to the minimum mappable unit. The shift from pixels to block segments classification is advocated in (Blaschke and Strobl, 2001).

Spatial interaction parameters and model selection

The DCRFs techniques proposed in Equations (3.1) and (3.14) have a spatial interaction model. Normally a suitable model should be selected or designed and its parameters determined in a given classification task. Two spatial interaction models in Equations (3.6) and (3.7) were designed. These models were compared to the commonly used contrast sensitive model in Equation (3.4) over a series of classifications tests. During the tests, a range of possible β and η parameter values in 2-D logarithmic scale were utilized. In each test, average overall classification accuracy and kappa of two epochs in growing season (June and July) were computed as an evaluation for each set of parameters. These epochs were chosen because within the period, returned radar backscatter are dominantly from crops. Finally, a suitable model of I including β and η parameters were selected using initial 2-D parameter search results.

DCRFs crop type mapping

After determining suitable I model and corresponding parameters, the technique in Equation (3.1) was adopted for sequence crop type classification. The approach classifies each epoch integrating first order temporal information and spatial information from 8 neighbouring nodes. We compare this approach to MLC and mono-temporal CRF, Equation (2.17), in single epoch classifications using 3×3 amplitude pixel blocks. This was done in order to test the impact of temporal interactions in DCRFs.

Posterior probabilities from each epoch were combined to generate an optimal map. Different methods introduced in Section 3.3 were tested and compared to our new approach in Equation (3.15). Moreover, studies like Bargiel and Herrmann (2011); Forkuor et al. (2014); Sonobe et al. (2015) stacked temporal images/features as bands for multitemporal classification. Therefore, we also stacked a sequence of amplitude images for classification using MLC (MLC-stack) and CRF (CRF-stack). This technique of stacking images was compared to optimal sequence classification method introduced in Equation (3.15). The approach introduced is referred to as DCRF max F1-score. This approach is compared to higher order DCRFs also developed in this study as discussed in the next section.

Crop type mapping with higher (n^{th}) order DCRFs

The capability of higher order temporal phenological information exchange with expert knowledge in Section 3.2 was tested for crop classification compared to first order DCRFs in Section 3.1.

Here, image based and expert temporal phenological information from the entire sequence was adopted for classification. Expert weights ϕ in Equation (3.13) were determined using crop phenology calendar (Figure 4.2) as shown in Table 4.3. The weights range between $1 \leq \phi \leq 10$. A weight of 5 was chosen if the anticipated backscatter is mainly contributed by the crop and 2.5 otherwise. This form of weighting emphasizes both unique crop and corresponding parcel characteristics respectively in all epochs.

Table 4.3.: Expert weights in different epochs determined using crop phenology.

Crop \ Epoch	March	April	May	June	July	October
Barley	2.5	2.5	2.5	5.0	5.0	2.5
Canola	2.5	2.5	5.0	5.0	5.0	2.5
Grassland	5.0	5.0	5.0	5.0	5.0	5.0
Maize	2.5	2.5	2.5	5.0	5.0	2.5
Oat	2.5	2.5	2.5	5.0	5.0	2.5
Potato	2.5	2.5	2.5	5.0	5.0	2.5
Rye	2.5	2.5	2.5	5.0	5.0	2.5
Sugar beet	2.5	2.5	2.5	5.0	5.0	2.5
Wheat	2.5	2.5	2.5	5.0	5.0	2.5

4.1.2 Kitale

Study site location and data

This experiment site is located west of Kenya approximately 1.01°N, 34.87°E as illustrated in Figure 4.3. Average annual precipitation and temperature in the area are 1600–1800 mm and 18.0°C, for more information see Jaetzold and Schmidt (1982); Kenya Meteorological Department (2015). The region is characterized by intensive small scale and large agriculture activities. Several crops are grown in Kitale. This study focused on crops grown in large scale within the area of interest. They include: 1) coffee, 2) grass, 3) maize, 4) rose flower, 5) sugarcane, and, 6) wheat. Coffee, rose flower and sugarcane are perennial crops while maize is annual. Rose flower is mainly grown in greenhouse controlled environment. The greenhouse structures are perceivable in satellite images. All these crops go through different phenological stages within a season as defined in Figure 4.4. Similar phenology phases defined in Section 4.1.1 are also used here.

Radar data from TerraSAR-X strip map were acquired for crop classification. The temporal sequence consists of 18 dual polarized images with range and azimuth resolutions of 2.6 m and 6.6 m in ascending and 2.1 m and 6.6 m in descending mode respectively (see Table 4.4). All dates have VH and VV polarized images except 6^{th} May which has HH and VV. The images were acquired at average incidence angles of 43° and 35° in ascending and descending modes respectively. These images were delivered as Single Look Slant Range Complex (SSC) with equidistant pixel spacing.

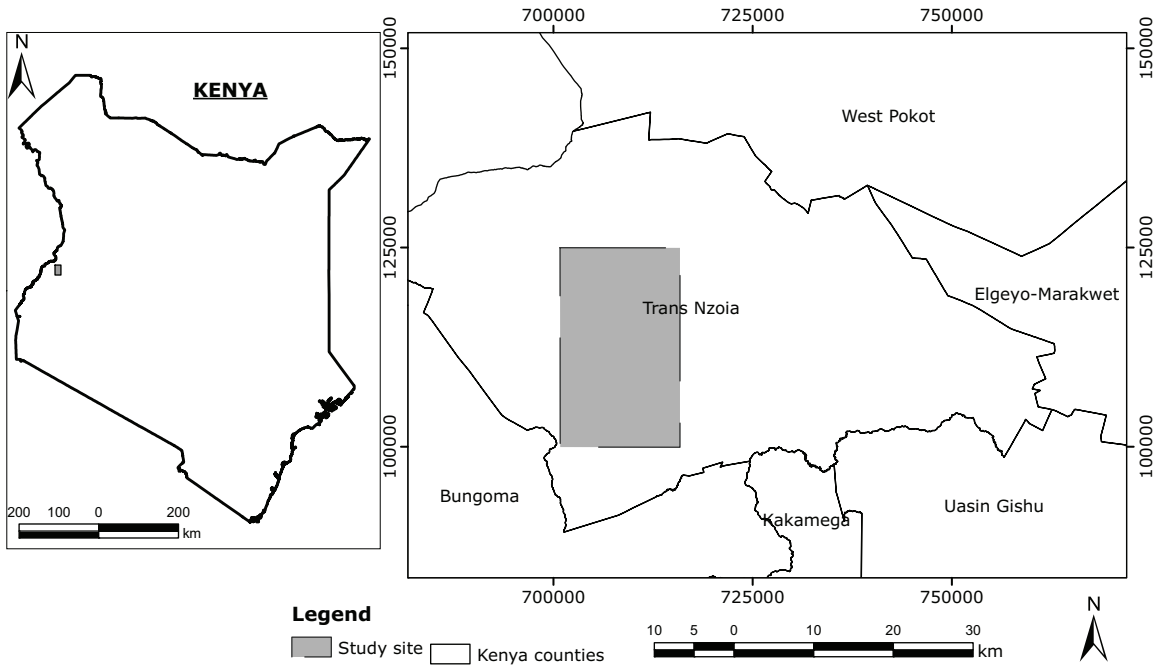


Figure 4.3.: Study area in Kitale, Kenya.

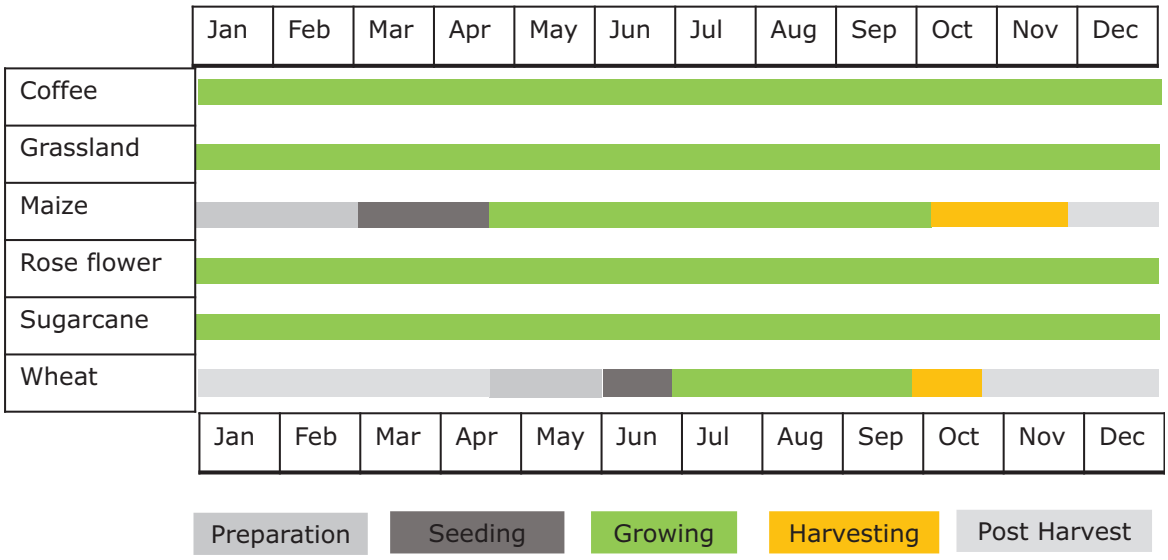


Figure 4.4.: Phenology stages of crops considered for classification in Kitale.

They were radiometrically calibrated to σ^0 according to Equation (2.6) and co-registered to an extent of (31 × 66) km^2 using WGS 1984 datum on UTM zone 36N coordinate projection system. Selected experiment site covers an extent of (15 × 25) km^2 as highlighted in Figure 4.3. Super pixels/blocks were computed for all images in the site using a mean of 2 × 2 pixels. Reference

Table 4.4.: TerraSAR-X image acquisitions in Kitale.

Year	Month	Date
2015	May	6, 8, 17, 28 & 30
	June	8, 19, 21 & 30
	July	11, 13 & 22
	August	2, 4, 13, 24 & 26
	November	9

data used for classification is illustrated in Table 4.5. The data was separated into training and validation sets using the procedure described in Section 4.1.1.

Table 4.5.: Distribution of training and validation data in hectares (ha) used in Kitale.

Crop type	Training set (ha)	Validation set (ha)
Coffee	93.17	78.11
Grassland	147.78	203.04
Maize	2038.28	1993.26
Rose flower	16.03	19.21
Sugarcane	113.94	138.56
Wheat	133.76	123.76

DCRFs crop type mapping

A predefined image sequence is required for DCRFs classification. Therefore, the images in Table 4.4 were divided into a sequence of 12 epochs. The sequence include the following image dates: 1) 6^{th} – 8^{th} May, 2) 17^{th} May, 3) 28^{th} – 30^{th} May, 4) 8^{th} June, 5) 19^{th} – 21^{st} June, 6) 30^{th} June, 7) 11^{th} – 13^{th} July, 8) 22^{nd} July, 9) 2^{nd} – 4^{th} August, 10) 13^{th} August, 11) 24^{th} – 26^{th}, and 12) 9^{th} November. Crops in this sequence were classified using the method in Equation (3.1). Parameters, $\beta = 10$ and $\eta = 1.0$, were used following prior knowledge from experiments in Section 4.1.1. Similarly, posterior probabilities from each epoch were combined to generate an optimal map. Different methods introduced in Section 3.3 were again tested and compared to our new approach in Equation (3.15). Likewise, we also stacked a sequence of amplitude images for classification using MLC, RF and CRFs as done by Bargiel and Herrmann (2011); Forkuor et al. (2014); Sonobe et al. (2014, 2015). This technique is compared to optimal sequence classification method introduced in Equation (3.15).

4.2 Crop type mapping from Sentinel 1 images

A study area covering Hannover region was selected to conduct crop classification experiments using Sentinel 1 images.

4.2.1 Study site and data

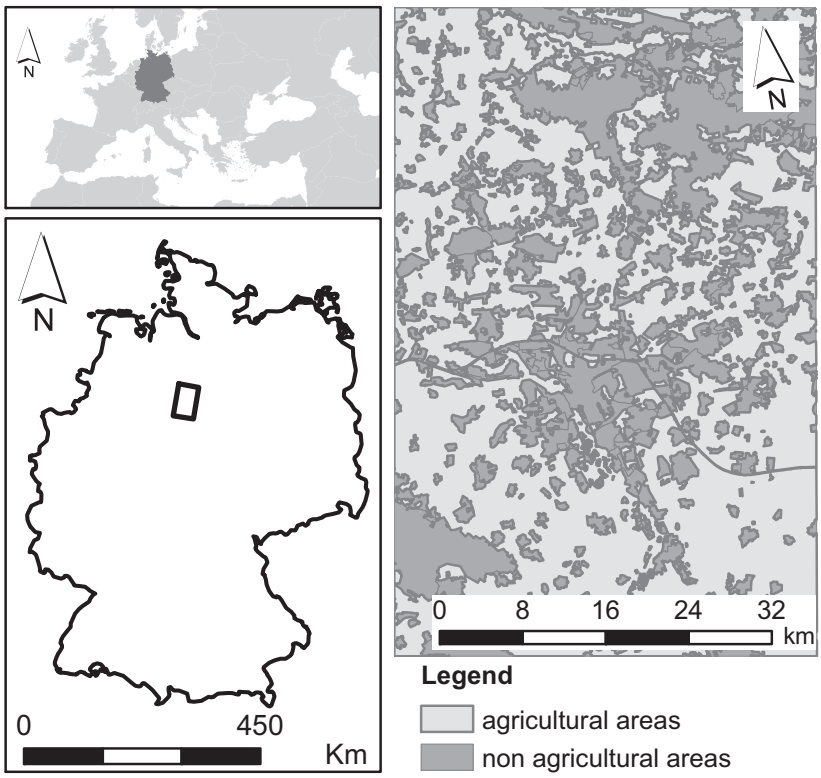


Figure 4.5.: Hannover study area.

The study area covers Hannover region and is located in Northern Germany (52.26°N, 9.84°E) (see Figure 4.5). The average annual precipitation and temperature are 656 mm and 8.9°C respectively (Deutscher Wetterdienst, 2012). The region is characterized by intensive agriculture with large farms. Crops in the area include: 1) summer barley, 2) winter barley, 3) canola, 4) grassland, 5) maize, 6) potato, 7) rye and, 8) sugar beet. These crops go through different phenological stages within a season, a fact that can enhance discrimination. Five phenology phases described in Section 4.1.1 are considered as shown in Figure 4.6.

Data used consists of a sequence of 45 dual polarized (VH and VV) Sentinel 1 C-band images (Table 4.6) acquired using Interferometric Wide (IW) swath mode at incidence angles between 31° to 46° with range and azimuth resolutions of 5 m and 20 m respectively (Torres et al., 2012). The images were delivered as Ground Range Detected (GRD) products. GRD products consist of SAR data which has been detected, multi-looked and projected to ground range using an ellipsoid such as WGS84 with equidistant pixel spacing. We radiometrically calibrated and co-registered the images to σ^0 and WGS 1984 datum on UTM Zone 32N coordinate projection system respectively

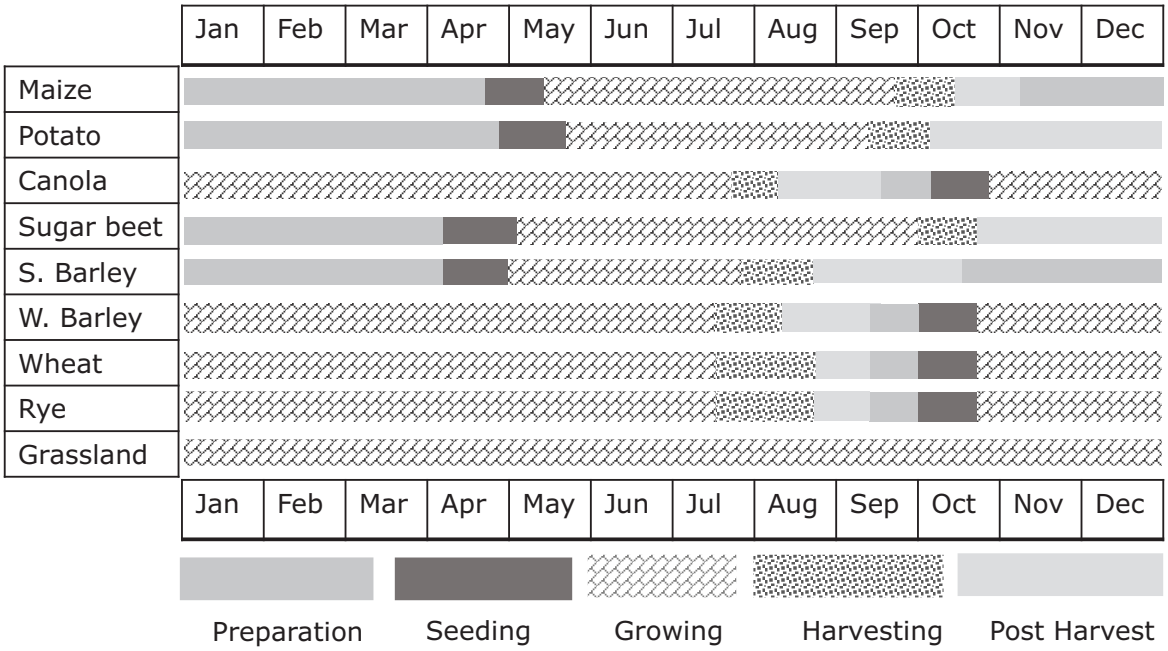


Figure 4.6.: Crop phenology stages. Seeding and growing phases are within plant life cycle and the rest are outside. S. Barley and W. Barley correspond to summer and winter barley respectively.

in Envi SARscape 5.2. Our experiment site covers an extent of (36 × 48) km^2 as highlighted in Figure 4.5.

Similar to Fuhrberg, reference data campaign was conducted during the same year the images were acquired. Table 4.7 illustrates distribution of training and validation parcel samples for various crops in the study area. The samples were prepared in the same procedure described in Section 4.1.1. However, to test the demand for larger training samples by classification algorithms when subjected to high dimensional data sets, the samples were divided into two categories of training and validation sets using stratified random sampling design tool in ArcGIS 10.0 (Buja and Menza, 2013). First category had approximately 50% training and 50% validation parcels per crop type while the second contained 20% and 80% respectively.

Spatial interaction parameters and model selection

A similar procedure of parameter estimation used for TerraSAR-X images in Section 4.1.1 was adopted. However, since Sentinel 1 images have a lower spatial resolution than TerraSAR-X, the parameter search range was restricted to between 0.1 and 25 for both β and η. This was done because intra- and inter-class backscatter variability is less in lower resolution images. For this reason, parameters with lower smoothness magnitude are anticipated to give better classification accuracy. Category I training and validation data in table 4.7 in conjunction with epochs in growing

Table 4.6.: Sentinel one data acquisitions over the study area.

Year	Month	Date
2014	October	13 & 22
	November	15 & 27
	December	09 & 21
	January	02, 14, 29
	February	07, 10, 19, & 22
	March	03, 06, 15, & 27
2015	April	08, 11, 20, & 23
	May	02, 05, 14, 17, & 26
	June	07, 10, 19, & 22
	July	01, 04, 13, 16, & 25
	August	06, 09, 18, 21, & 30
	September	02, 11, 14, 23, & 26

	Training set				Validation set			
	Category I		Category II		Category I		Category II	
	Count	ha	Count	ha	Count	ha	Count	ha
Summer Barley	3	9.97	2	4.91	3	14.64	4	19.70
Winter Barley	3	20.38	2	14.04	3	24.83	4	31.17
Canola	7	35.01	2	10.43	8	51.23	13	75.81
Grassland	15	41.96	6	19.42	15	43.84	24	66.38
Maize	14	67.07	6	26.71	16	78.51	24	118.87
Potato	6	36.37	3	19.78	6	34.43	9	51.02
Rye	14	81.72	5	23.62	14	80.47	23	138.56
Sugar beet	21	124.13	8	44.76	21	113.1	34	192.47
Wheat	19	113.88	7	44.76	19	123.88	31	193.00

Table 4.7.: Distribution of training and validation data in terms of size in hectares (ha) and number of parcels (count) per crop. Proportions of reference data (training set/validation set) is: category I (50%/50%) and category II (20%/80%).

season (June and July) were used for parameter search experiments. These epochs were selected because within the period, returned radar backscatter are dominantly from crops.

4.2.2 DCRFs crop type mapping

The images in Table 4.6 were divided into 7 epoch categories, that is, epoch 0 consist of winter images between October to March, while epoch 1–6 correspond to images in April–September respectively. We adopt the technique in Equation (3.1) for pixel-wise crop classification in the epoch sequence. The approach classifies each epoch integrating first order temporal information and spatial information from 8 neighbouring nodes. We compare this approach to MLC, RF and

mono-temporal CRF, Equation (2.17), in single epoch classification. These comparison was done using category I training and validation data in Table 4.7.

All classifications were done using a core i7 3.6 GHz computer with 64 gigabytes memory. For DCRFs, a tile based classification with an overlap of 1 pixel around a tile was implemented to minimize memory demands. The overlaps were reflected (copied) to form a complete tile in the outermost borders of an image. Thus, the algorithm is scalable, that is, it can be used in any computer only that the processing time will vary. For instance, DCRFs classifications over the study area in all the 7 epochs using 12 tiles took approximately 27 hours.

Similar to experiment in Section 4.1.1, posterior probabilities from each epoch were combined to generate an optimal map. Different methods introduced in Section 3.3 were again tested and compared to our new approach in Equation (3.15). Likewise, we also stacked a sequence of amplitude images for classification using MLC, RF and CRF as done by Bargiel and Herrmann (2011); Forkuor et al. (2014); Sonobe et al. (2014, 2015). This technique is compared to optimal sequence classification method introduced in Equation (3.15).

Finally, we tested the robustness of DCRF max F1-score and MLC-stack techniques by subjecting them to high dimensional data and less training data in category II of Table 4.7. An ideal remote sensing land-cover mapping method should be able to use less training data even in high dimensional feature space and still deliver high classification accuracy thus, the motivation to conduct the experiment.

5 Results

Results from experiments conducted in three different regions using first order and higher order DCRFs compared to MLC, RF and standard CRFs are presented in this chapter.

5.1 Results of crop type mapping using TerraSAR-X images

Results of crop mapping using TerraSAR-X data are presented in this section part of which appear in Kenduiywo et al. (2016). The results are based on experiments conducted from two different regions, Fuhrberg and Kitale, in Europe and Africa respectively.

5.1.1 Fuhrberg

Feature selection results

Results of GLCM features selection are described here. Four features were selected using RF variable importance criteria. Figure 5.1 shows RF importance computed from an average of four directions, 0°, 45°, 90°, and 135°, for each feature and subsequently their average over all the six epochs. Generally with the exception of HH-polarized correlation features, features computed from VV-polarization backscatter have a higher importance compared to features computed from HH-polarization.

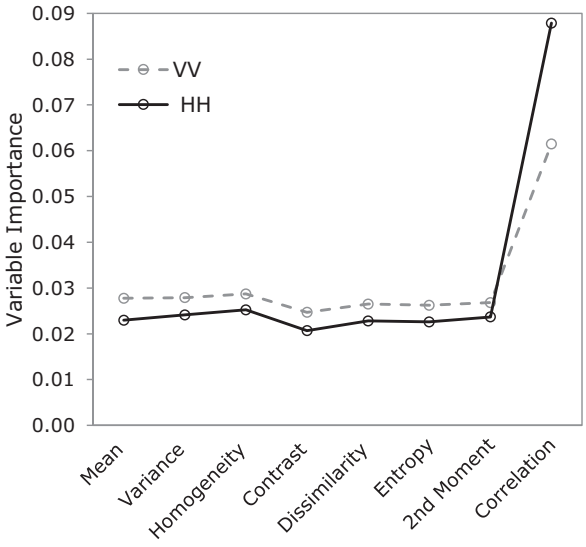


Figure 5.1.: Random forest variable importance of different GLCM features based on a mean of four directions, 0°, 45°, 90°, and 135°, and subsequently their average over all epochs.

Choice of a data interaction model for I and corresponding parameters was guided by results in Figures 5.2 to 5.4. The new version of contrast sensitive model, Equation (3.6), outperforms standard contrast sensitive, Equation (3.4), in OA and κ over most parameters. The new model returns high OA for $10^1 \leq \beta \leq 10^3$ and $10^{-2} \leq \eta \leq 10^0$ (Figure 5.2a). A high κ is observed over the same range of parameter values indicating that the classifications are not by chance (Figure 5.2b). In contrast, OA and κ for contrast sensitive Potts model reduces if $\beta > 10^{-1}$ in combination with any η parameter values as depicted by Figures 5.3a and 5.3b respectively. Another novel model

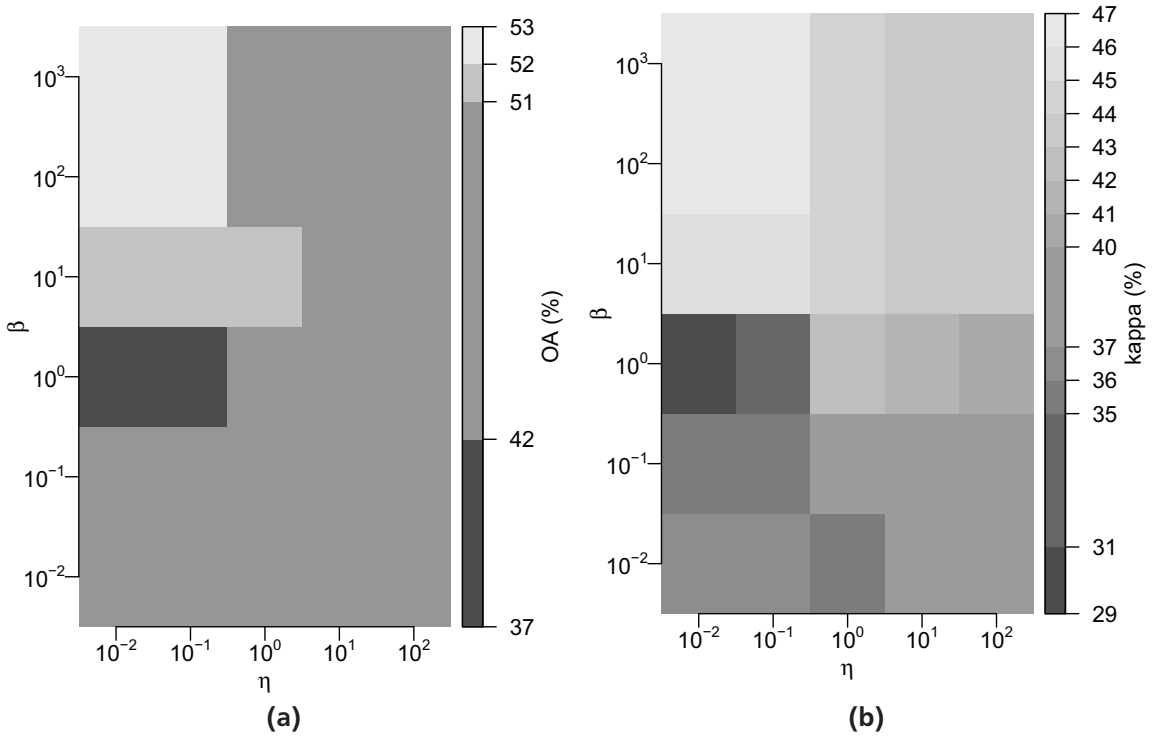


Figure 5.2.: Classification accuracy of the new expanded version of contrast sensitive model; (a) is overall accuracy and (b) is kappa coefficient in percentage.

designed and tested in this study is the Pearson correlation coefficient in Equation (3.8). This model outperforms standard contrast sensitive model in OA and κ $10^1 \leq \beta \leq 10^3$ and $10^{-2} \leq \eta \leq 10^0$ (Figure 5.4). However, for $\eta > 10^2$ parameter values the model over-smoothens the entire image allocating all pixels to one class as indicated by κ in Figure 5.4b. Therefore, this data interaction function requires a careful parameter search because it is very sensitive to high values. Nonetheless, the model performs competitively and comparable to the expanded version of contrast sensitive model as shown by both results in Figures 5.2 and 5.4 respectively.

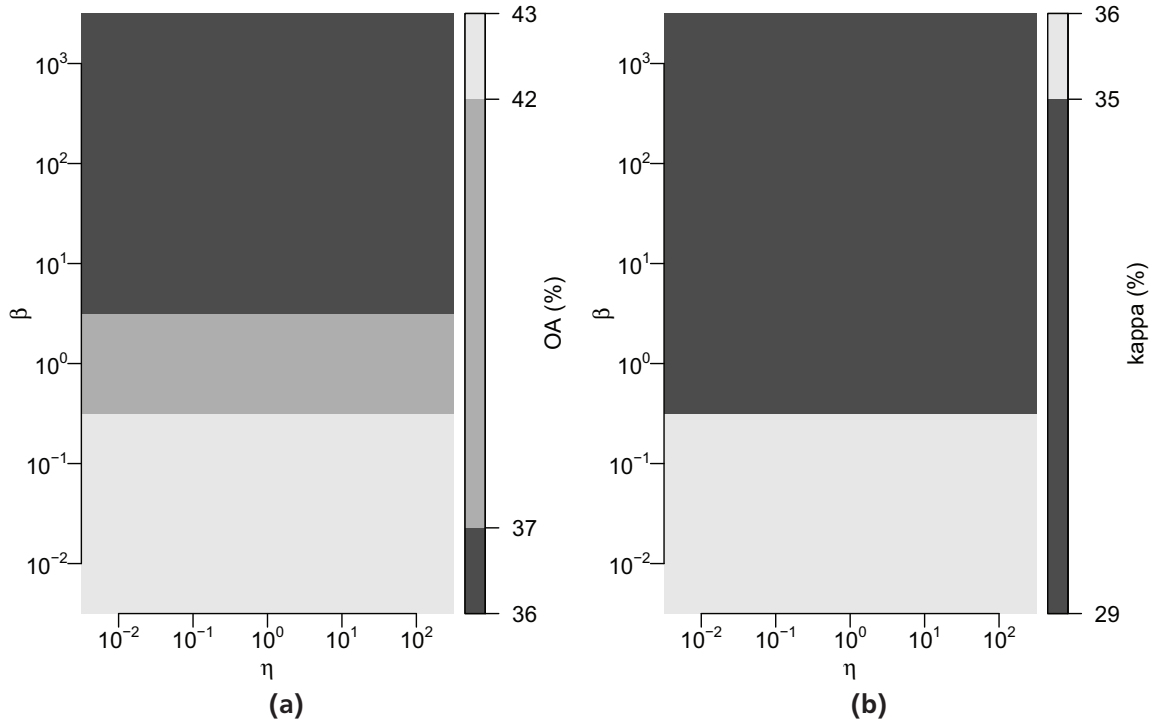


Figure 5.3.: Standard contrast sensitive model classification accuracy; (a) is overall accuracy and (b) is kappa coefficient in percentage.

In summary, the new contrast sensitive model in Equation (3.6) was selected since it is robust over a range of parameter values. Parameter search results from the model, with a compromise between high accuracy and over-smoothing, informed the choice of $\beta = 10$ and $\eta = 1$ for classification. These parameter values are used across the sequence for comparability.

Classification results

Results from 5^{th} order DCRFs and standard first order DCRFs epoch classification compared to other approaches are illustrated in Figure 5.5. The results show that DCRFs approach outperforms CRF, RF and MLC in both OA and κ. In all epochs MLC has the least accuracy followed by RF and CRF respectively. The addition of temporal information also improved classification accuracy in each epoch since DCRFs outperforms CRF which considers only spatial information. Spatial information also improved classification as demonstrated by CRF performance compared to RF and MLC which have lower accuracy. Moreover, 5^{th} order DCRFs including expert knowledge gave even higher classification accuracy in each epoch compared to standard DCRFs. This proves that image based and expert temporal phenological information from the entire sequence of images are significant in classification.

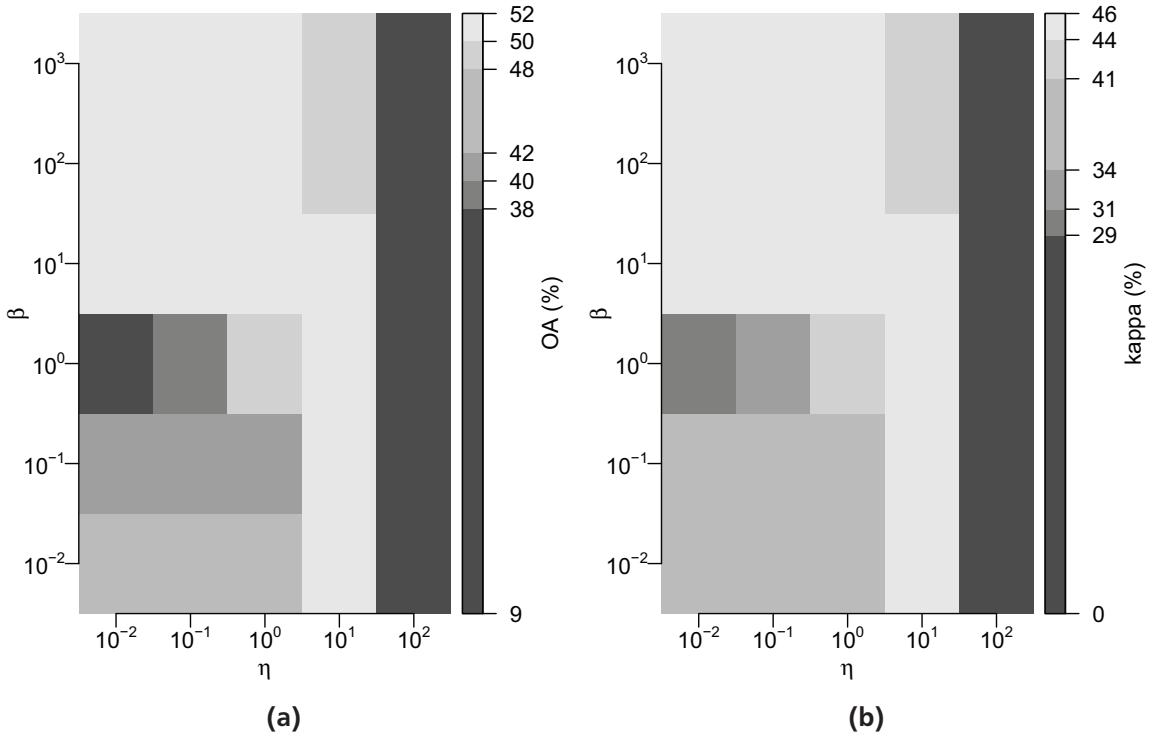


Figure 5.4.: Proposed Pearson correlation interaction data dependent function classification accuracy; (a) is overall accuracy and (b) is kappa coefficient in percentage.

An optimal crop map is generated from epoch-wise DCRFs posterior probabilities using different classifier combination strategies as depicted by results in Figure 5.6a. The first technique we introduce, DCRF max F1-score, surpasses max rule, majority vote, product rule and median rule by 33.54%/0.37, 9.89%/0.11, 7.76%/0.09 and 7.71%/0.09 in OA/κ respectively. Thus, max rule has the least accuracy followed by majority vote, product rule and median rule respectively. On the other hand, when higher order expert weighted temporal interactions are considered, classification accuracy of all ensemble approaches improves except product rule. From Figure 5.6b OA/κ of max rule, majority vote, median rule and max F1-score improve by 18.51%/0.20, 2.39%/0.04, 0.34%/0.1, and 1.31%/0.02 respectively while product rule drops by 5.75%/0.06. As observed the product rule is unstable and severe classifier combination rule as mentioned in Section 3.3. These outcomes justified the use of max F1-score for determination of an optimal seasonal crop map in both DCRFs and 5^{th} order DCRFs.

The developed max F1-score classifier ensemble based on 5^{th} and first order DCRFs is compared to the conventional approach of stacking multitemporal images as input bands to a classifier. Table 5.1 illustrates that first order DCRF max F1-score outperforms MLC and CRF methods adopting stacked images by OA/κ of 6.46%/0.08 and 5.90%/0.07 respectively. Moreover, higher order

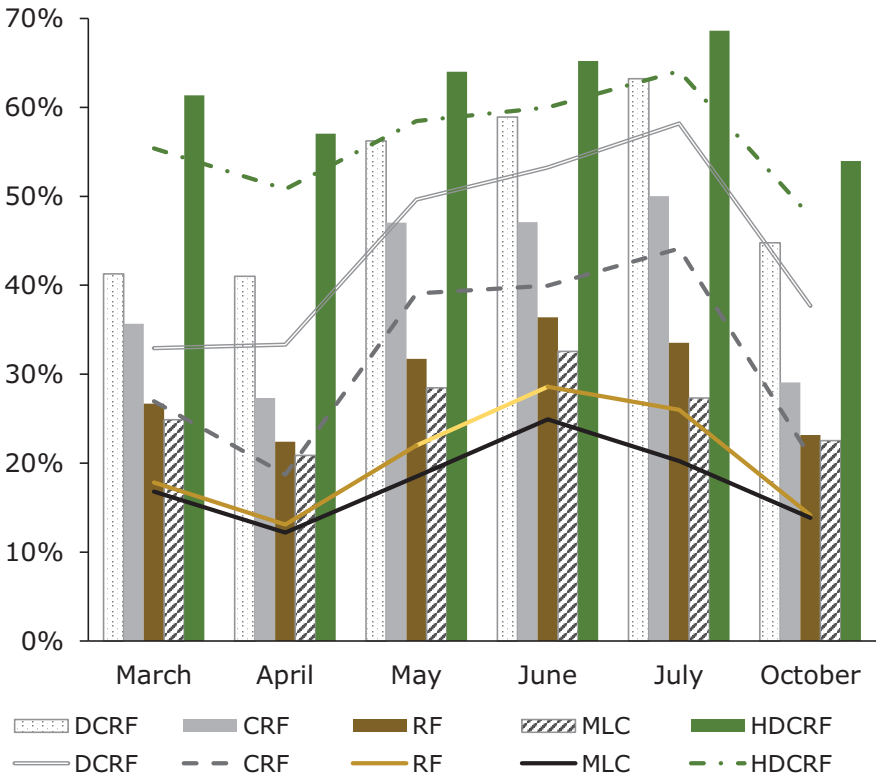
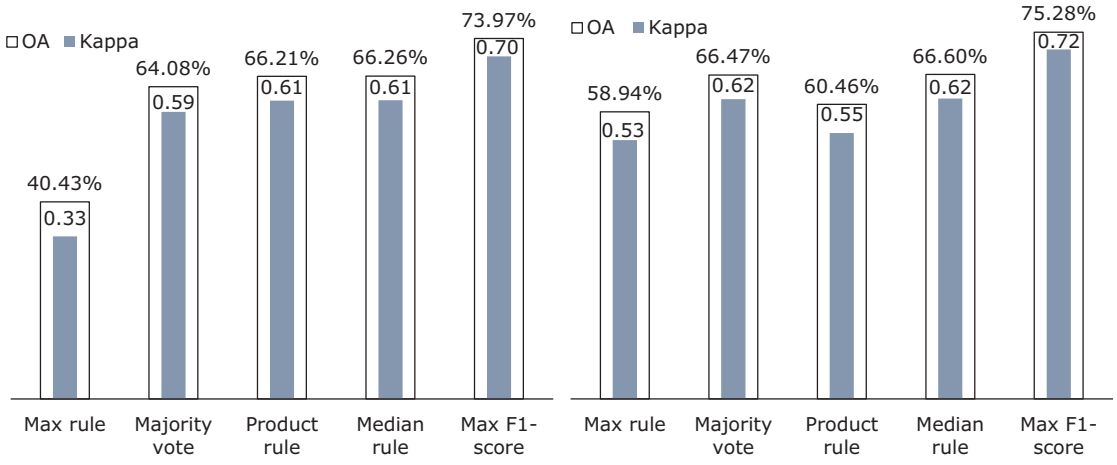


Figure 5.5.: Fuhrberg epoch-wise classification results from different approaches; HDCRF refers to 5th order DCRF. Overall accuracy and kappa coefficient are shown by bars and lines in percentage respectively.



(a) First order DCRFs. **(b)** 5^{th} order DCRFs.

Figure 5.6.: Comparison of different strategies of integrating DCRFs posterior probabilities to produce an optimal seasonal crop map. Methods in (a) and (b) are based on estimated class probabilities from standard and higher order DCRFs respectively.

DCRFs surpasses DCRFs by 1.31%/0.02 in OA/κ. Therefore, it can be observed that classifying stacked images even using spatial context methods does not lead to higher accuracy.

Table 5.1.: Results from max F1-score ensemble methods versus stacking multitemporal images together as input bands for classification.

Method	OA	κ
MLC-stack	67.51%	0.62%
CRF-stack	68.07%	0.63%
DCRFs max F1-score	73.97%	0.70%
Higher order DCRFs max F1-score	75.28%	0.72%

Examination of standard first order DCRFs based on max F1-score and MLC-stack 3×3 pixel blocks classifications using producer and user accuracy of each class is made in Table 5.2. This exposes in detail how each crop is recognized by the classifiers as opposed to a limited view on overall accuracy in Table 5.1. It is only grassland that has a lower classification accuracy in our approach (user accuracy -11.60% / producer accuracy -2.3%) compared to MLC-stack approach. All other classes were classified better or comparable to MLC-stack approach. This is especially true for barley (+3.71% / +14.21%), maize (+13.04% / +24.56%), sugar beet (+15.99% / +46.68%) and wheat (+0.99% / +7.90%). The producer accuracy of rye and potato decreased by 6.67% and 15.52% while their user accuracy increased by 12.08% and 21.03% respectively. In contrast, user accuracy of canola and oat dropped by 4.60% and 25.92% while producer accuracy increased by 0.19% and 47.78% respectively.

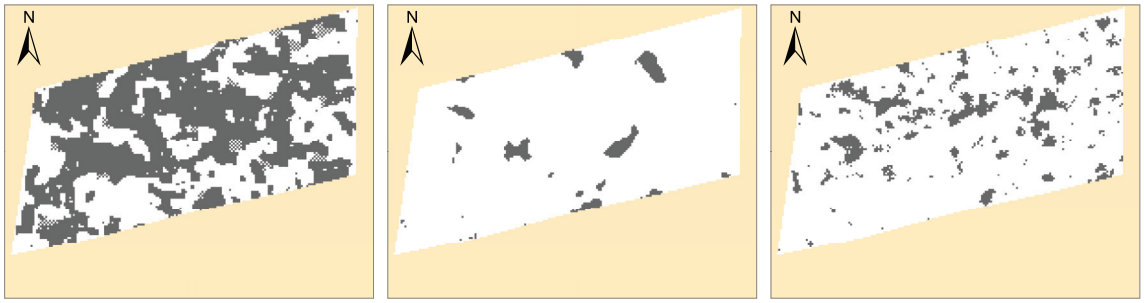
Introduction of higher order temporal interactions increased crop discrimination accuracy. Discrimination accuracy of potato and barley slightly decreased in 5^{th} order DCRFs compared to DCRFs. However, recognition of wheat, oat and maize significantly improved. For instance, Figure 5.7 shows a maize parcel well detected by 5^{th} order DCRFs compared to DCRFs and MLC-stack. Most maize pixels in the parcel are misclassified to potato and sugar beet which are all broad leaved crops. Oats are well classified by 5^{th} order DCRFs and DCRFs max F1-score techniques as opposed to MLC-stack which completely misclassified a parcel as rye (see Figure 5.8). Oat and rye belong to the same family (cereals) thus, MLC fails to effectively discriminate them. To sum up, the producer accuracy of rye and sugar beet decreased while user accuracy decreased in grassland and maize. For example, Figure 5.9 illustrates a sugar beet parcel which is well recognized by DCRFs max-F1 in contrast to 5^{th} order DCRFs max F1-score and MLC-stack approaches. The MLC-stack performs the poorest by misclassifying a bigger region of the parcel as potato and partly as maize. Sugar beet, maize and potato are broad leaved crops. For an in-depth view, confusion matrices of MLC-stack, 5^{th} order DCRFs and DCRFs based on max F1-score can be referred to in Appendix A.1 Tables A.1, A.2 and A.4 respectively.

Contribution of spatial interactions in classification of stacked images can also be observed in Table 5.2. This is illustrated by CRFs classification where, the producer accuracy of canola, grass-

Table 5.2.: Crop producer and user accuracy from first and higher order DCRFs max F1-score including CRF and MLC using stacked image features.

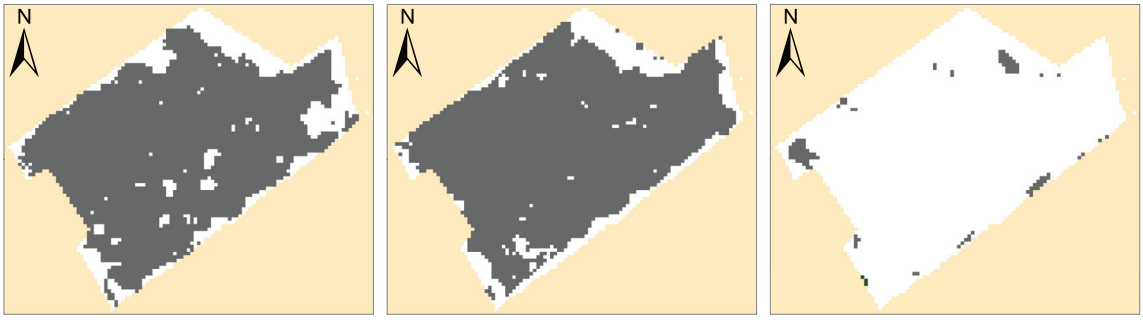
Crop type	Producer accuracy				User accuracy			
	DCRFs	5th order DCRFs	CRF stack	MLC stack	DCRFs	5th order DCRFs	CRF stack	MLC stack
Barley	69.66%	63.57%	52.07%	55.45%	73.32%	72.71%	76.69%	69.61%
Canola	97.33%	97.45%	97.97%	97.14%	90.23%	88.89%	95.17%	94.83%
Grassland	85.93%	92.71%	97.91%	94.62%	77.30%	75.88%	87.18%	88.90%
Maize	49.81%	60.74%	17.28%	25.25%	47.30%	53.06%	22.01%	34.26%
Oat	88.93%	84.30%	94.29%	41.15%	59.06%	90.14%	98.97%	84.98%
Potato	63.19%	73.17%	71.08%	78.71%	74.15%	67.41%	51.08%	53.12%
Rye	64.38%	62.92%	67.34%	71.05%	74.93%	75.77%	67.69%	62.85%
Sugar beet	89.35%	77.17%	42.03%	42.67%	86.59%	93.51%	71.30%	70.60%
Wheat	69.66%	74.46%	71.20%	61.76%	66.60%	71.53%	59.14%	65.61%

land, oat and wheat increased by 0.83%, 3.28%, 53.15% and 9.45% respectively compared to classification of pixel blocks by MLC. However, it is interesting that the producer accuracy of barley, maize, potato, rye and sugar beet dropped by -3.38%, -7.97%, -7.63%, -3.71% and -0.64% respectively compared to MLC. Thus, for these crops spatial interactions did not improve their discrimination compared to pixel block classification by MLC. This is because the CRFs interaction potential averaged the temporal features hence suppressed changes that might have occurred. In general, CRFs performed poorly in mapping barley, maize and sugar beet compared to all other approaches (MLC, first and higher order DCRFs). Since most crop mapping studies use pixel based classification methods based on stacked images, maps of MLC and the novel DCRFs are analyzed in the next paragraph. The confusion matrix of CRFs is available in Table A.3 for an in-dept view.



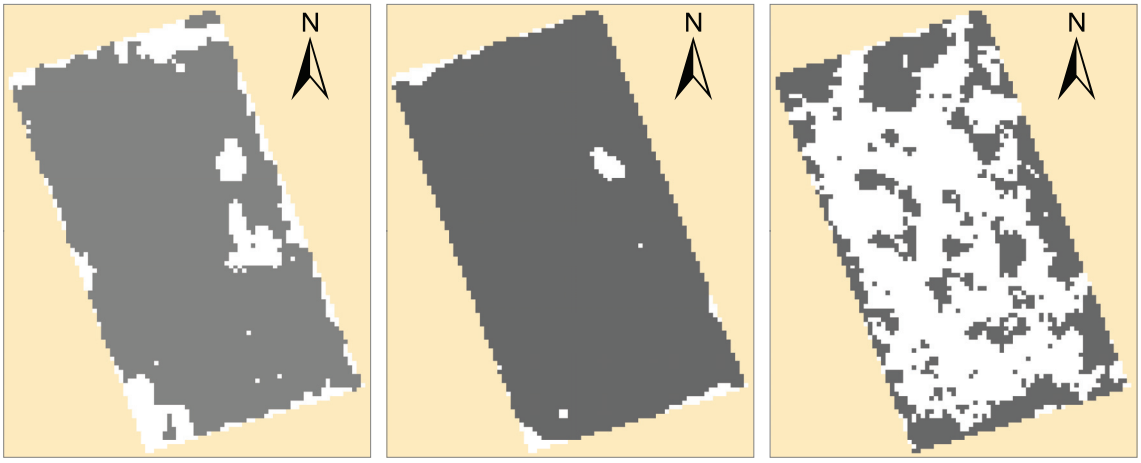
(a) 5^{th} order DCRFs. **(b)** DCRFs. **(c)** MLC-stack.

Figure 5.7.: A Maize parcel well discriminated by 5^{th} order DCRFs max F1-score compared to MLC-stack and first order DCRFs max F1-score methods. False positives (white areas) are potato and sugar beet crops.



(a) 5^{th} order DCRFs. **(b)** DCRFs. **(c)** MLC-stack.

Figure 5.8.: An oat parcel well detected by 5^{th} and first order DCRFs max F1-score methods but dominantly misclassified as rye by MLC-stack. In all maps false positives are white regions.

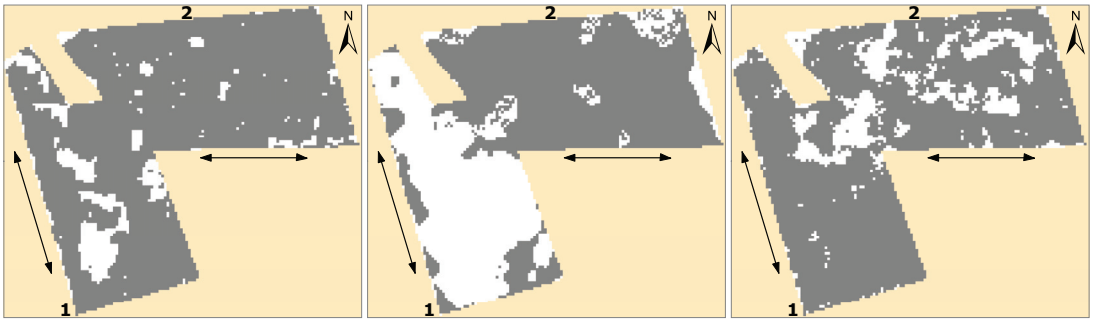


(a) 5^{th} order DCRFs. **(b)** DCRFs. **(c)** MLC-stack.

Figure 5.9.: A sugar beet parcel well detected by 5^{th} and first order DCRFs max F1-score methods but dominantly misclassified as potato and partly as maize by MLC-stack. In all maps false positives are white regions.

An analysis of crop maps highlights challenges faced by MLC-stack, 5^{th} order DCRFs and DCRFs classification methods. For instance, potato producer accuracy decreased in both DCRFs approaches compared to MLC-stack mainly due to false positives which were probably caused by the influence of row direction on SAR backscatter. However, DCRFs experiences lower discrimination in comparison to higher order DCRFs as seen in Figures 5.10a and 5.10b. It can also be observed from Figure 5.10, that the potato parcel with rows in the direction South East to North West had the lowest accuracy. This is the approximate direction of TerraSAR-X satellite in ascending mode. Rye is also a challenge to all the three methods as illustrated in Figure 5.11. This is because rye had earlier been planted (Figure 5.11d) and after harvesting (Figure 5.11e), straw-

berry was planted (Figure 5.11f). This dynamic change was not incorporated into validation data set. Thus, the false positives are actually true. Nonetheless, MLC-stack detected these changes the least as shown in Figure 5.11c.

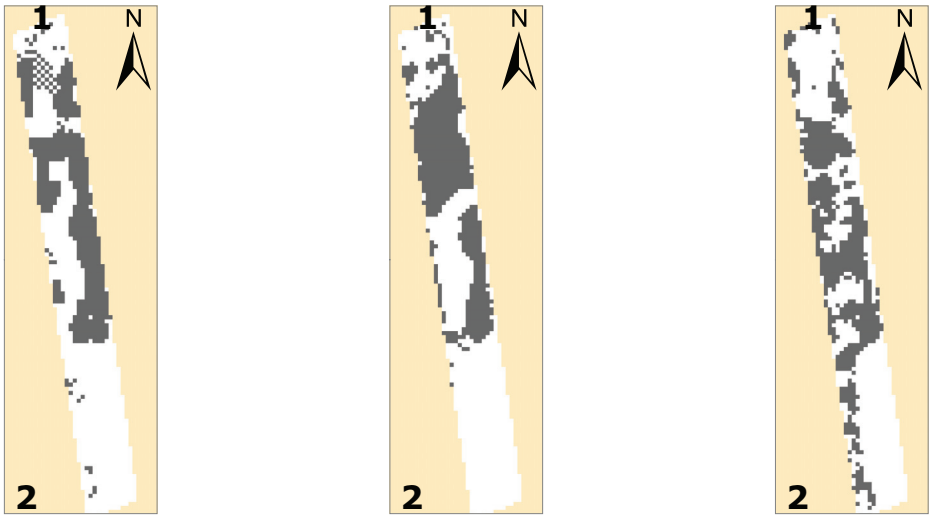


(a) 5^{th} order DCRFs. **(b)** DCRFs. **(c)** MLC-stack.

(d) Photo taken at position 1 in (a)–(c). **(e)** Photo taken at position 2 in (a)–(c).

Figure 5.10.: A potato parcel as mapped in MLC-stack and 5^{th} and first order DCRFs max F1-score techniques. The arrows indicate row directions which possibly influenced radar signal and hence false positives (white areas) in the maps.

A further analysis of standard first order DCRFs is made to establish why it has difficulty in discriminating grassland. Figure 5.12 depicts scatterplot of first order DCRFs max F1-score against MLC-stack producer accuracy computed from each grassland validation parcel. Parcel numbers 8, 43, 49, and 51 have higher accuracy in MLC-stack compared to DCRF max F1-score. However, observations from ground referencing photos in Figure 5.13 demonstrate that errors encountered by DCRF max F1-score in those parcels are true positives. For instance, parcel number 49 was used as storage which exposed the top soil. Likewise parcel number 8 has inhomogeneous grass because of weeds infestation. Hence, inhomogeneous grassland maps from DCRF max F1-score technique reflect true changes on ground not anticipated in ground reference data (see Figure 5.13). Remaining parcels have comparable producer accuracy in both methods.



(a) 5^{th} order DCRFs. **(b)** DCRFs. **(c)** MLC-stack.

(d) Photo taken in April at position 1 of figures (a)–(c). **(e)** Photo taken in August at position 2 of figures (a)–(c). **(f)** Photo taken in October at position 1 of figures (a)–(c).

Figure 5.11.: A rye parcel as mapped in MLC-stack and 5^{th} and first order DCRFs max F1-score techniques. False positives (white areas) caused by harvested rye in photo (e) and planting of strawberry in photo (f).

Despite the challenges faced by DCRFs it still delivers results better than MLC-stack and comparable to 5^{th} order DCRFs. This is demonstrated in Table 5.2 and by maps in Figures 5.8 and 5.9 as classified by both MLC and DCRF max F1-score. It can be seen from the maps that DCRF max F1-score technique produces homogenous parcels compared to MLC-stack. This emphasizes the contribution of temporal phenological information inherent in images and spatial context in crop classification. The final map generated from 5^{th} order DCRFs and DCRFs max F1-score ensemble classifications including MLC-stack are shown in Appendix A.1 Figures A.1 to A.3.

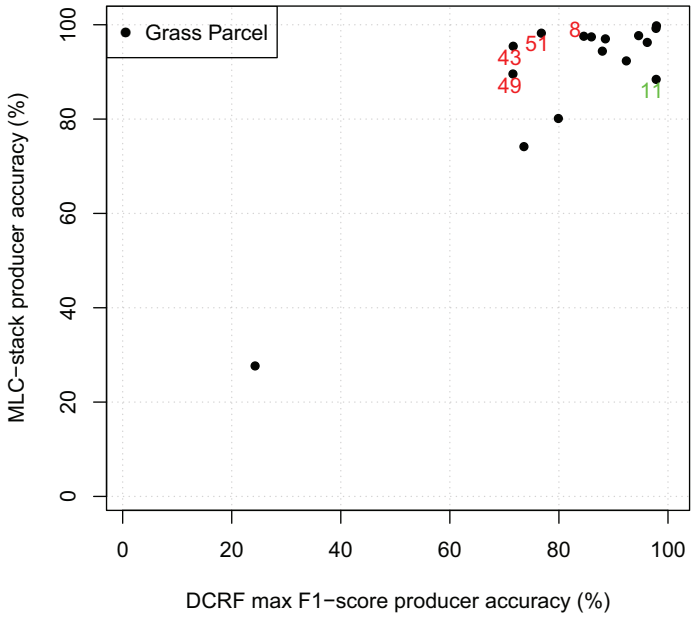


Figure 5.12.: Scatterplot of first order DCRFs max F1-score against MLC-stack producer accuracy computed from each grassland validation parcel.

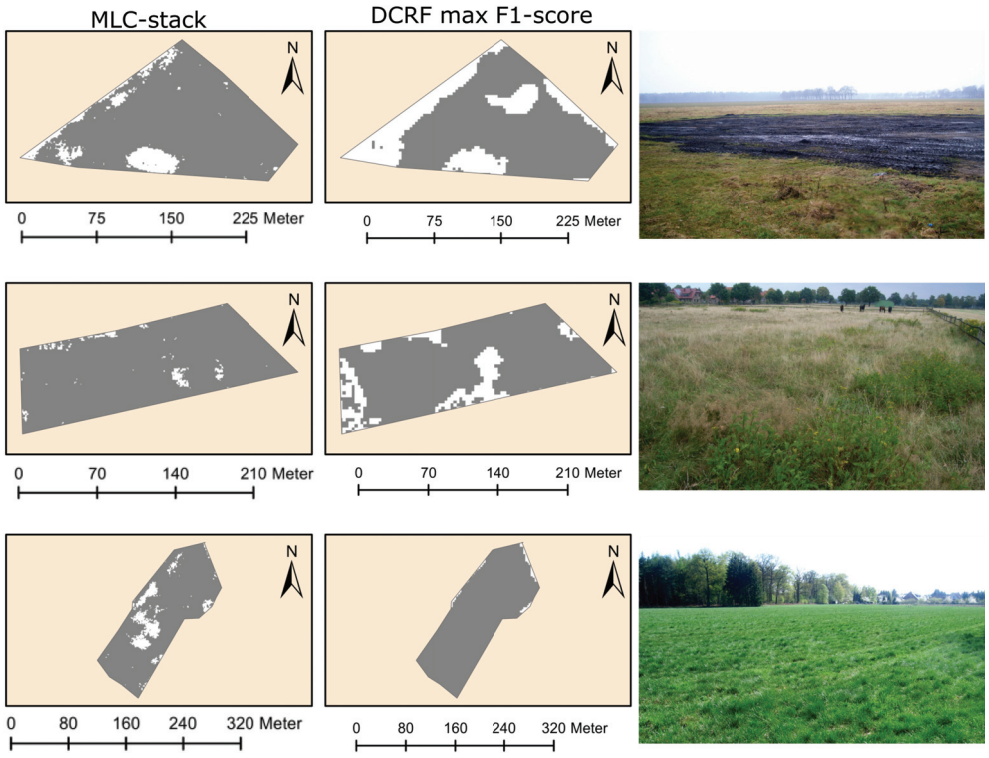


Figure 5.13.: Grassland parcels as classified by MLC-stack and first order DCRF max F1-score and corresponding ground referencing photos. Top to bottom row corresponds to parcel numbers 49, 8, and 11 respectively as shown in Figure 5.12. White areas correspond to misclassifications.

5.1.2 Kitale

Classification results

The outcome from different ensemble mapping strategies are presented in Table 5.3. Our classifier combination strategy, DCRF max F1-score, surpasses max rule, majority vote, median rule and product rule by 49.94%/0.57, 3.73%/0.06, 3.09%/0.05 and 5.08%/0.08 in OA/κ respectively. Thus, max rule has the least overall accuracy followed by majority vote, median rule and product rule respectively. It can be observed that max F1-score is robust and accurate compared to the

Table 5.3.: Accuracy of different strategies of integrating DCRFs posterior probabilities.

Method	OA	κ
Max F1-score	90.27%	0.76
Product rule	85.19%	0.68
Median rule	87.18%	0.71
Majority vote	86.54%	0.70
Max rule	40.33%	0.19

other classifier ensemble strategies. Max rule is the poorest while product rule, median rule and majority vote are competitively comparable. An examination of how accurate these methods map each crop is depicted in Figure 5.14. Max F1-score approach still has the highest mapping accuracy in all crop types except sugarcane. Median rule has the highest accuracy in mapping sugarcane followed closely by majority vote, product rule, max F1-score and max rule.

The designed DCRF max-F1 score framework is compared to RF, MLC and CRF classifiers based on stacked multitemporal images (Table 5.4). In this setup, DCRF max-F1 score outperforms RF, MLC and CRF by 13.47%/0.23, 13.03%/0.22 and 1.07%/0.01 in OA/κ respectively. These results also demonstrate the contribution of temporal information since DCRF max-F1 score performs better than CRF. In contrast, MLC surpasses RF classifier while CRF performs better than both methods. Thus, spatial interactions improve classification accuracy. An assessment of performance

Table 5.4.: Comparison of DCRF max F1-score to stacking multitemporal images for classification using RF, MLC and CRF.

Method	OA	κ
Max F1-score	90.27%	0.76
RF-stack	76.80%	0.53
MLC-stack	77.24%	0.54
CRF-stack	89.20%	0.75

of DCRF max F1-score and MLC-stack in mapping each crop type is depicted in Table 5.5. Producer and user accuracy of each crop type presents an expounded view of how each method performs as

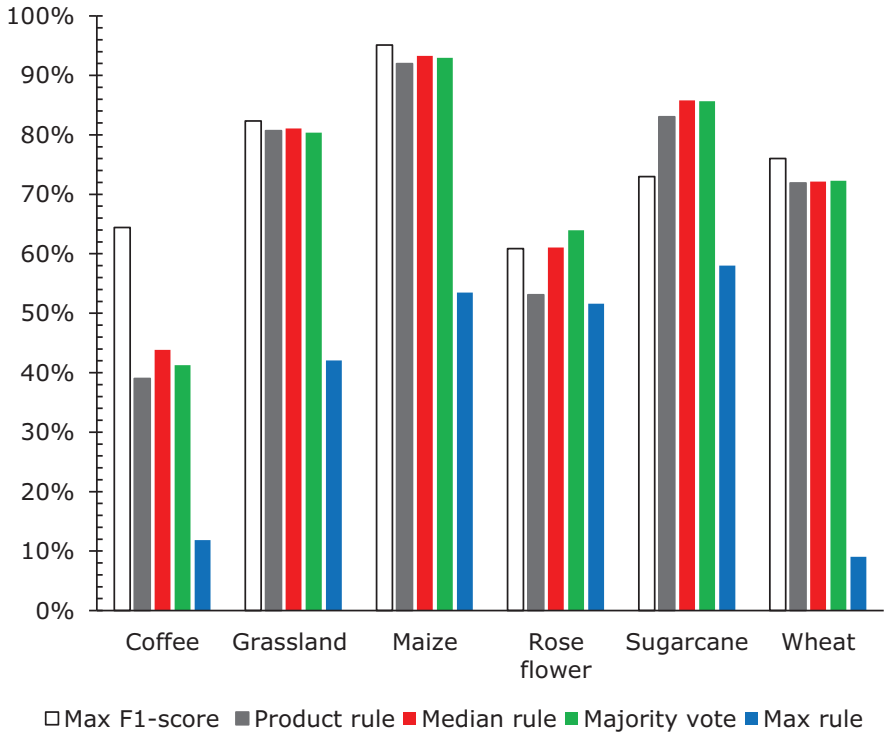


Figure 5.14.: Percentage F1-score accuracy of each crop in Kitale from different ensemble methods.

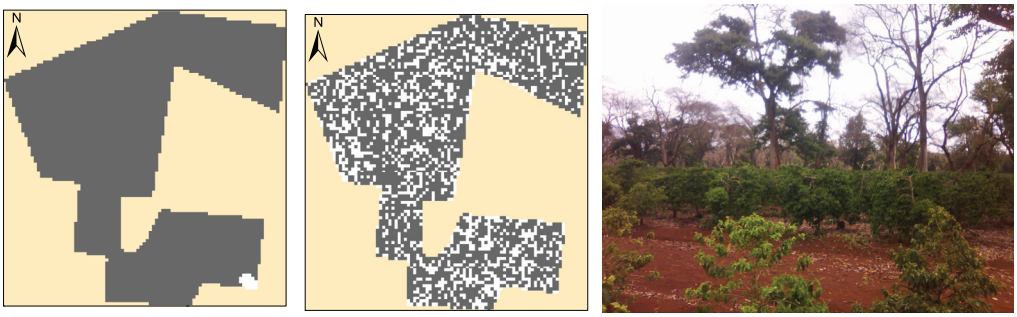
opposed to OA and κ. The producer accuracy of all crops mapped by DCRF max F1-score surpasses MLC-stack except sugarcane which dropped by -3.65%. In contrast, the user accuracy of all crops mapped by the technique outperforms MLC-stack. Complete confusion matrices for DCRF max F1-score and MLC-stack classifications are available in Tables A.5 and A.6 respectively.

Table 5.5.: Producer and user accuracy of each crop from DCRF max F1-score and MLC stack classifications.

Crop type	DCRF max F1-score		MLC-stack	
	Producer	User	Producer	User
Coffee	85.44%	51.69%	64.80%	13.63%
Grassland	93.43%	73.62%	73.89%	63.72%
Maize	92.44%	97.97%	79.17%	97.78%
Rose flower	81.88%	48.45%	69.97%	33.33%
Sugarcane	67.43%	79.54%	71.08%	64.34%
Wheat	80.37%	72.14%	67.77%	59.34%

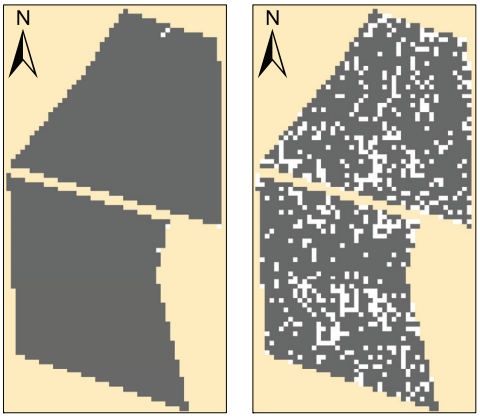
It can be noted that coffee and rose flower have low user accuracy in both methods. Basically, coffee is intercropped with trees as shown in Figure 5.15. This phenomenon challenged MLC-stack as depicted in Figure 5.15b. Coffee was mainly misclassified as maize and partly sugarcane all of which are broad leaved crops. Consequently, false negatives increased conversely decreasing

the user accuracy. Examination of confusion matrices of both techniques in Tables A.5 and A.6, indicate that most of rose flower training pixels were mapped as maize. Therefore, this indicates the difficulty of the approaches to clearly discriminate rose flower in greenhouses and maize. Some crops like sugarcane, maize and rose flower in Figure 5.17 are classified by DCRFs max F1-score better than MLC-stack. However, regions mapped as wheat by DCRFs max F1-score are false positives which caused producer accuracy of sugarcane to drop below that of MLC-stack. Some



(a) DCRF max F1-score. **(b)** MLC-stack. **(c)** Coffee intercropped with trees.

Figure 5.15.: Coffee as mapped by MLC-stack and DCRF max F1-score. The white regions indicate misclassifications mainly maize.



(a) DCRF max F1-score. **(b)** MLC-stack.

Figure 5.16.: Wheat as mapped by MLC-stack and DCRF max F1-score. The white regions are misclassifications.

farmers also intercrop maize and beans (Figure 5.18). This fact is a problem to MLC-stack as demonstrated in Figure 5.18b. On the contrary, DCRFs max F1-score manages to map the parcel better than MLC-stack (Figure 5.18a).

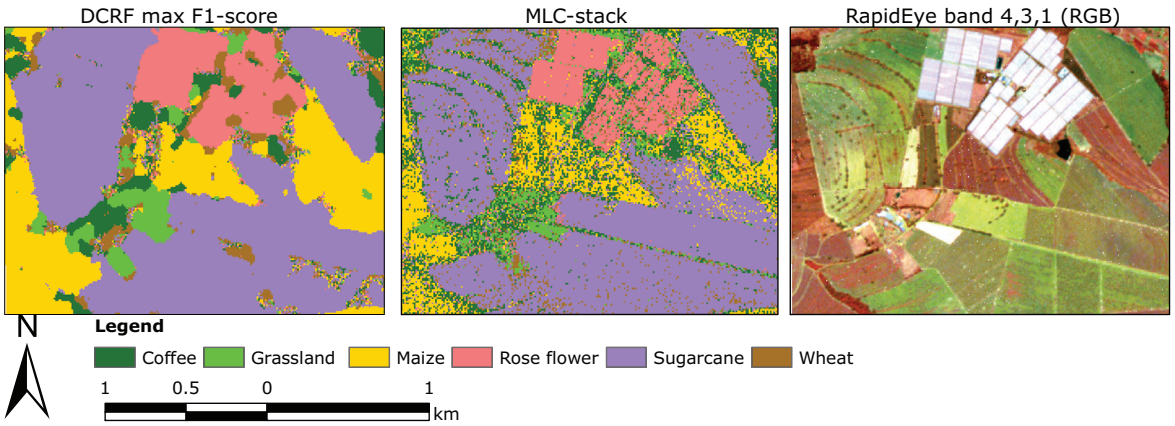
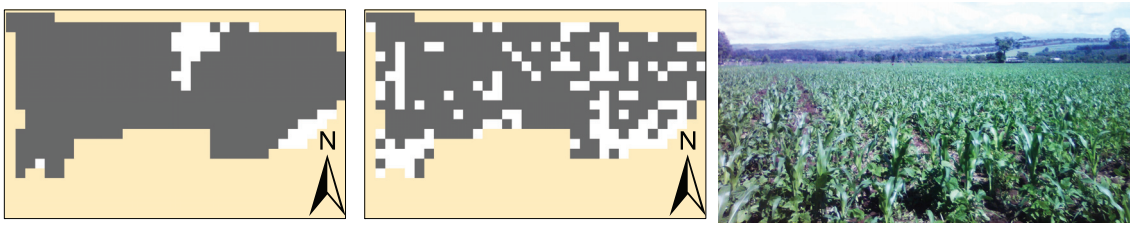


Figure 5.17.: Grassland, rose flower (in greenhouses), sugarcane and sugarcane as mapped by MLC-stack and DCRFs max F1-score compared to high resolution image ©RapidEye.



(a) DCRF max F1-score. **(b)** MLC-stack. **(c)** Maize intercropped with beans.

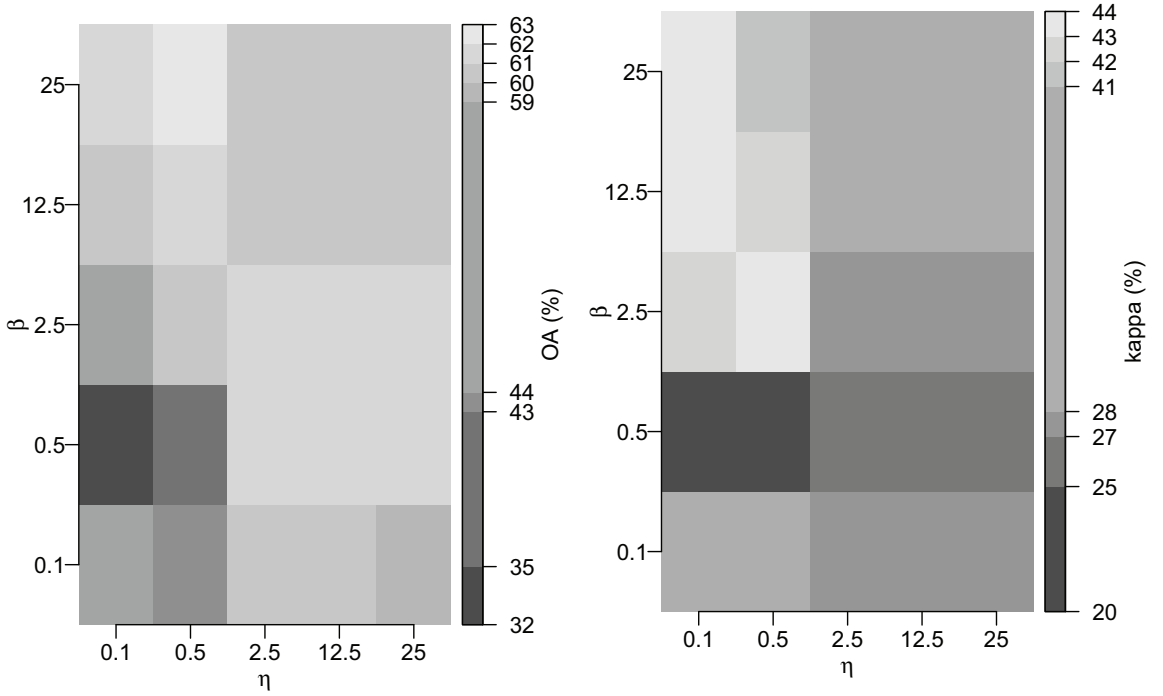
Figure 5.18.: Maize parcel mapped by MLC-stack and DCRF max F1-score. The white regions indicate misclassifications.

5.2 Results of crop type mapping from Sentinel 1 images

5.2.1 Parameter determination results

Selection of a data interaction model for I and its suitable parameters was guided by results in Figures 5.19 to 5.21. The expanded version of contrast sensitive model, Equation (3.6), surpasses standard contrast sensitive, Equation (3.4), in OA and κ over most parameters. This model gives high classification accuracy for $2.5 \leq \beta \leq 25$ and $0.5 \leq \eta \leq 12.5$. In contrast, accuracy for contrast sensitive Potts model reduces if $\beta > 0.5$ in combination with any η parameter values (Figure 5.20). Pearson correlation model in Equation (3.7) also performs better than conventional contrast sensitive model as depicted by Figure 5.21. It gives high classification accuracy if $2.5 \leq \beta \leq 25$ and $0.5 \leq \eta \leq 12.5$ like the new contrast sensitive model. These results are comparable to the new version of contrast sensitive model. However, the new version of contrast sensitive model outperforms it by a very small margin of OA.

Classification results of the three data designs (Figures 5.19 to 5.21), formed the basis for choosing a suitable interaction potential. So far, Pearson correlation and new version of contrast sensitive

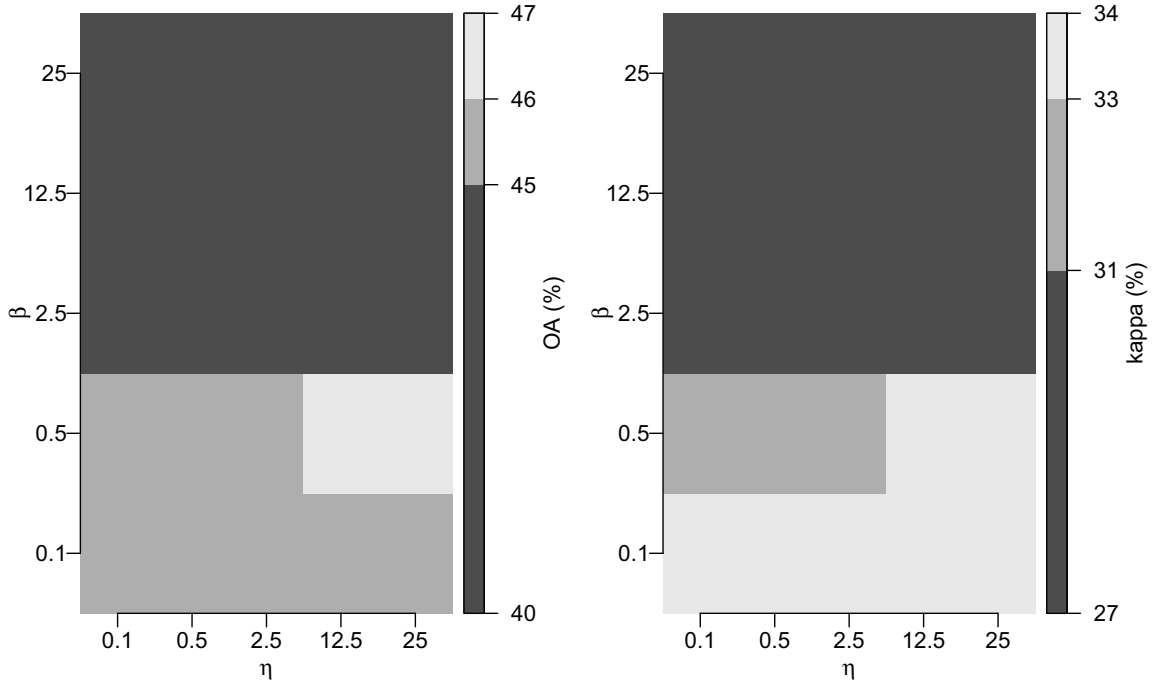


(a) Overall accuracy.

(b) Kappa coefficient in percentage.

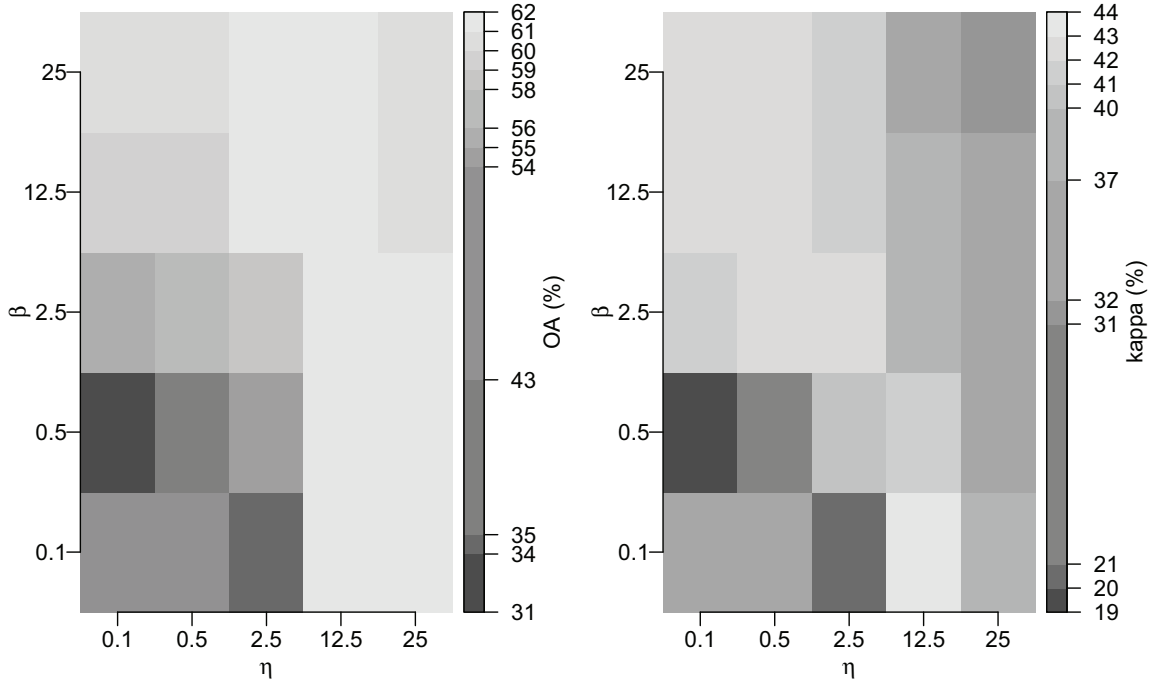
Figure 5.19.: Classification accuracy of the new version of contrast sensitive model using Sentinel 1 data.

models have similar accuracy. However, new contrast sensitive model was used for all crop classification tasks. This is because it has less computation demands than Pearson correlation. Thus, from the parameter search, we set $\beta = 5$ and $\eta = 1$ for the model. The parameters were selected as compromise between high accuracy and over-smoothing. We use these parameter values across the sequence for comparability.



(a) Overall accuracy. **(b)** Kappa coefficient in percentage.

Figure 5.20.: Standard contrast sensitive model classification accuracy using Sentinel 1 data.



(a) Overall accuracy. **(b)** Kappa coefficient in percentage.

Figure 5.21.: Pearson correlation model classification accuracy based on Sentinel 1 data.

5.2.2 Classification results

Category I classification results

This section presents results from experiments conducted based on high dimensional data in Table 4.6 and 50% training data in category I of Table 4.7. Designed DCRF epoch classification results compared to other approaches are illustrated in Figure 5.22. The results show that DCRF approach outperforms CRF, RF and MLC. In all epochs MLC has the least accuracy followed by RF and CRF respectively. The addition of temporal information also improved classification accuracy in each epoch since DCRF outperformed CRF which considers only spatial information. Spatial information also improved classification as demonstrated by CRF performance compared to RF and MLC which have lower accuracy.

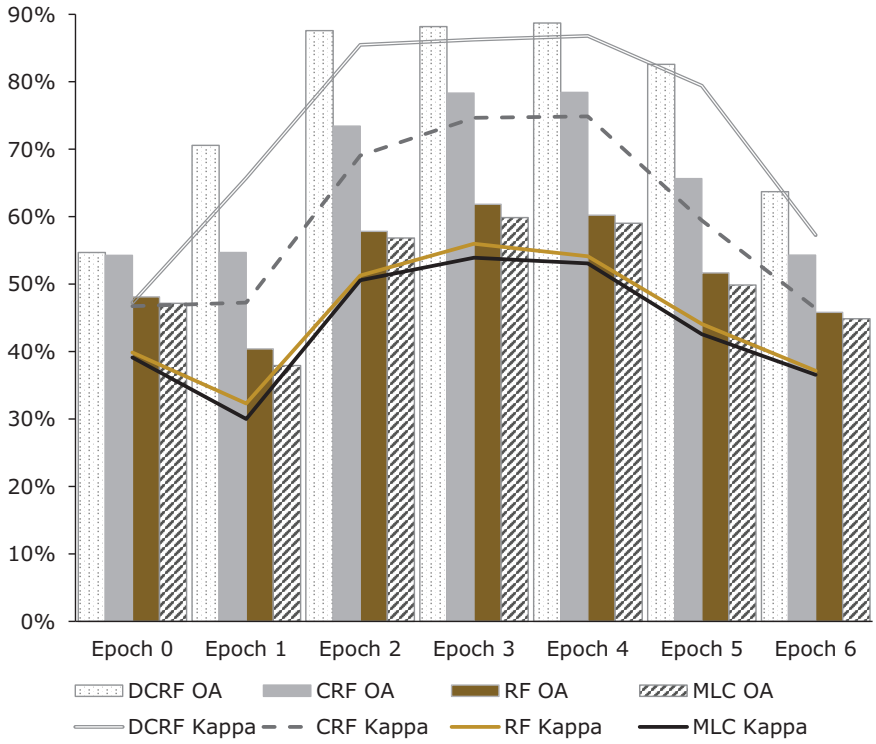


Figure 5.22.: Sentinel 1 epoch-wise classification results, overall accuracy and kappa, from different approaches.

An optimal crop map from epoch-wise DCRF posterior probabilities using different classifier combination strategies as depicted by results in Table 5.6 was generated. The technique we introduce, DCRF max F1-score, performs equal to median rule but outperforms max rule and majority vote by 27.3% and 0.98% in OA respectively. However, the product rule outperforms it by 0.21%. Nonetheless, evaluating a method based on OA and κ is limiting. Observations of each crop type F1-score accuracy measure in Figure 5.23 shows that DCRF max F1-score provides a balanced ac-

curacy in most cases. This is especially true for grassland, rye, wheat, summer and winter barley that challenged all the other techniques. Thus, the ensemble approach we introduced has stable performance in all classes.

Table 5.6.: Overall accuracies and kappa from different ensemble methods.

Method	OA	Kappa
Max F1-score	92.81%	0.92
Product rule	93.02%	0.92
Median rule	92.82%	0.92
Majority vote	91.83%	0.90
Max rule	65.51%	0.61

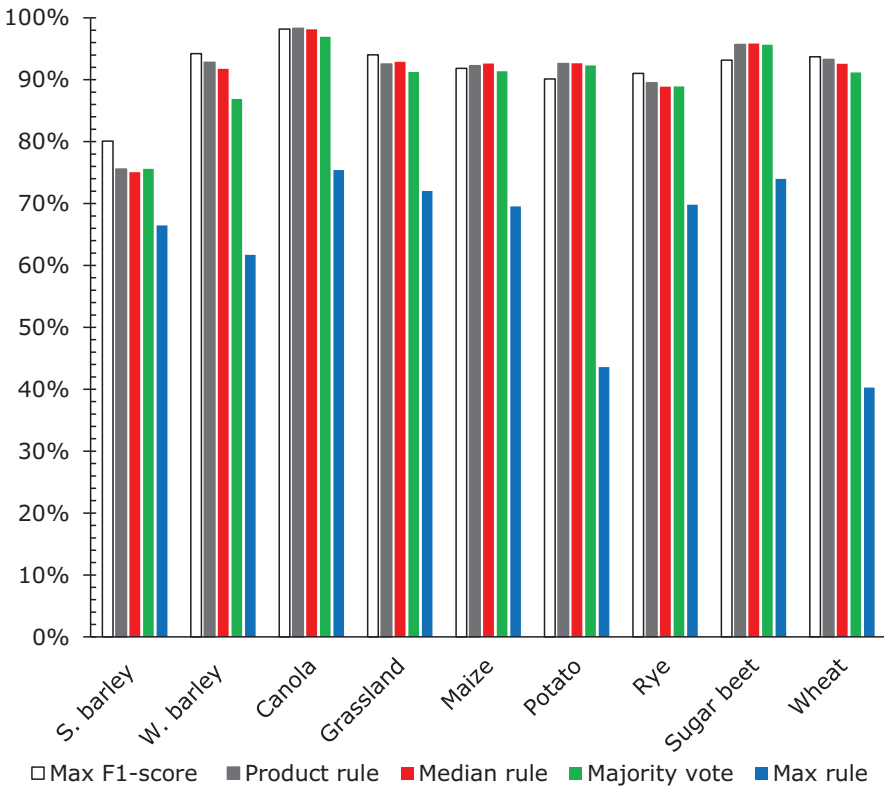


Figure 5.23.: Percentage F1-score accuracy measure of each crop from different ensemble methods. S. barley and W. barley correspond to summer and winter barley respectively.

Designed DCRF max F1-score classifier ensemble is compared to the conventional approach of stacking multitemporal images as input bands to a classifier. An additional term "stack" is added to classifiers using this approach for distinction. Table 5.7 illustrates that DCRF max F1-score outperforms MLC-stack, RF-stack and CRF-stack methods adopting stacked images by OA/κ of 4.81%/0.06, 4.43%/0.06 and 4.22%/0.05 respectively.

Table 5.7.: Comparison of DCRF max F1-score to stacking multitemporal images together as input bands for classification using MLC and CRF.

Method	OA	κ
MLC-stack	88.00%	0.86
RF-stack	88.38%	0.86
CRF-stack	88.59%	0.87
DCRF Max-F1 score	92.81%	0.92

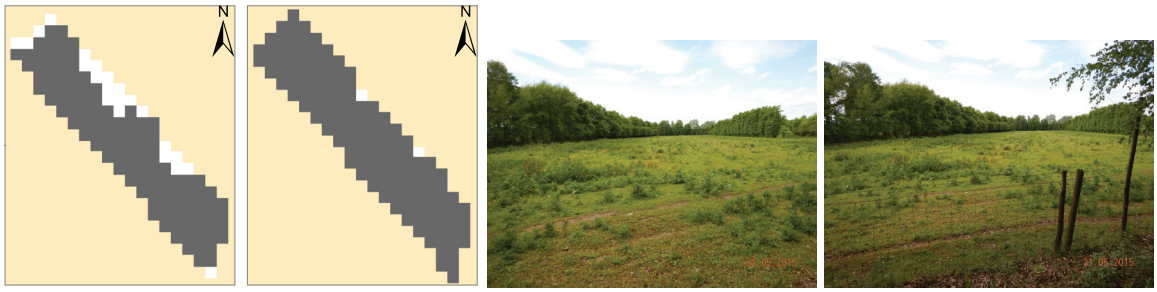
Further comparison of DCRF max F1-score and MLC-stack, which has the lowest accuracy, using producer and user accuracy of each class is made in Table 5.8. This exposes in detail how each crop is recognized by the classifiers as opposed to a limited view on overall accuracy in Table 5.7. Producer accuracy of grassland and wheat drops in our approach by -2.03% and -2.89% respectively compared to MLC-stack. In addition user accuracy of summer barley and potato drops by -16.68% and 0.83% respectively. Most of the other classes were classified better or comparable to MLC-stack approach. This is especially true for canola (producer +0.66% / user +0.38%), maize (+0.90% / +5.21%) and rye (+14.91% / +0.32%). The biggest gainers in terms of precision (producer accuracy) are potato, summer barley, winter barley and rye which increased by 15.16%, 57.3%, 21.24% and 14.91% respectively. Moreover, the capability of our approach to identify grassland and wheat based on training data improved by 15.02% and 14.01% subsequently as illustrated by user accuracy. For a complete view, confusion matrices of DCRF max-F1 score and MLC-stack are provided in Appendix A.1.3 Tables A.7 and A.8.

Table 5.8.: Producer and user accuracy of each crop based on DCRF max F1-score and MLC stack.

Crop type	DCRF max F1-score		MLC-stack	
	Producer	User	Producer	User
Summer barley	80.36%	79.78%	23.09%	96.46%
Winter barley	93.05%	95.43%	71.81%	95.80%
Canola	98.77%	97.63%	98.11%	97.25%
Grassland	93.94%	94.11%	95.97%	79.09%
Maize	93.74%	90.04%	92.84%	84.83%
Potato	91.90%	88.39%	76.74%	89.22%
Rye	96.64%	86.05%	81.73%	85.73%
Sugar beet	91.76%	94.58%	91.96%	94.73%
Wheat	89.44%	98.42%	92.33%	84.41%

An assessment of some mapped crop parcels highlights the strength of DCRF max F1-score compared to MLC-stack. The DCRFs spatial data interaction term captures ground changes better than MLC-stack. This is illustrated in Figure 5.24 where false positives in grassland parcel are due to inhomogeneity. Photos in Figures 5.24c and 5.24d shows that there are shrubs, some boundary trees and different varieties of grass which are detected in our approach as opposed to MLC-stack.

Thus, the false positives in DCRF max F1-score correspond to true ground changes. Moreover, DCRF max F1-score has a high recall rate in grassland compared to MLC-stack. This is captured in Figure 5.27 where MLC-stack (Figure 5.27b) over classified grassland compared to DCRF max F1-score (Figure 5.27a). This over classification also affected accuracy of some classes like summer barley because grass pixels dominated their parcel (see Figure 5.26b). Basically, grassland has a large variance due to inhomogeneity among most of its parcels. Consequently, it is over classified since MLC uses variance and mean during classification. Potato parcels are also mapped well com-



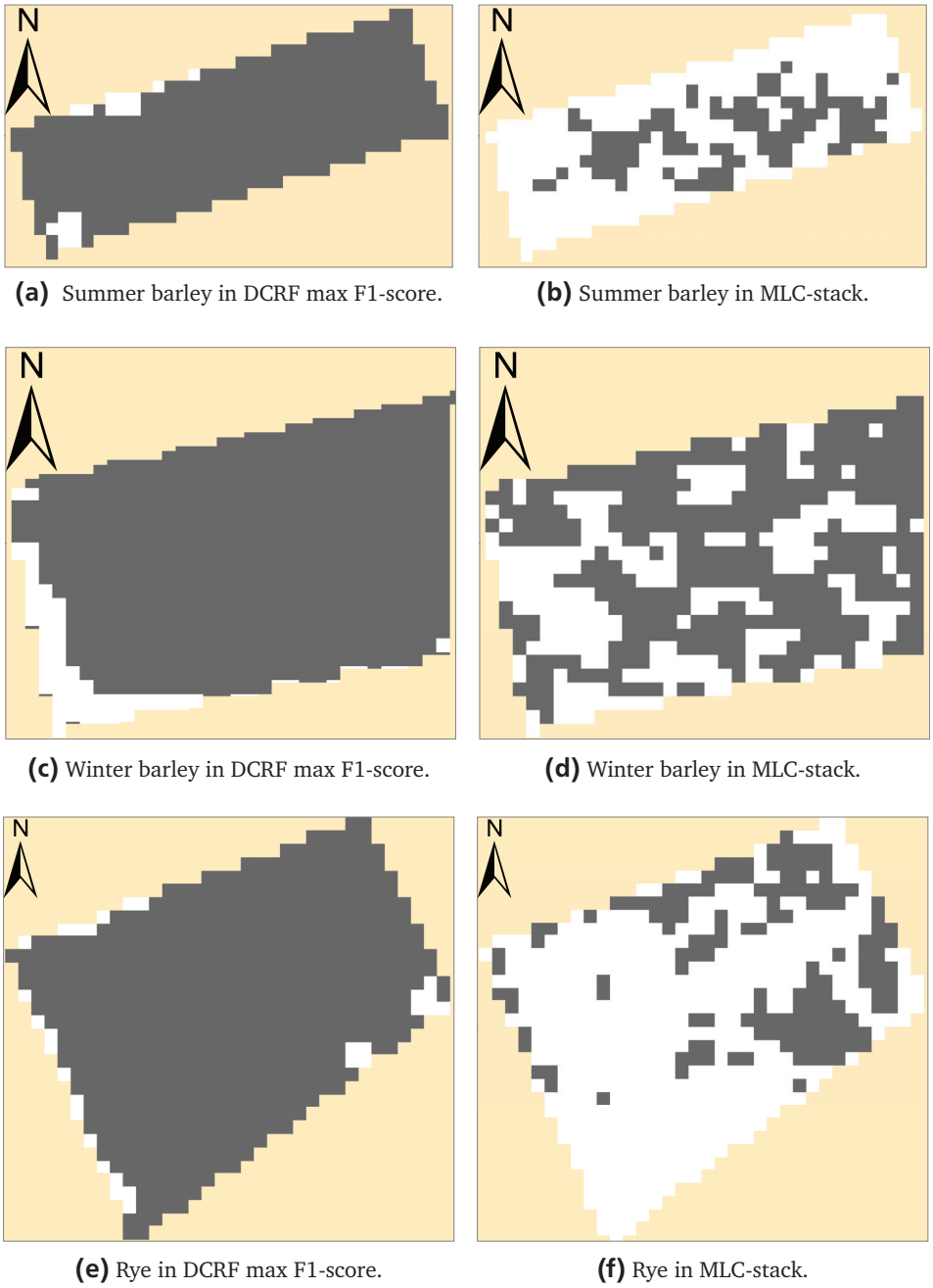
(a) DCRF max F1-score. **(b)** MLC-stack. **(c)** **(d)**

Figure 5.24.: Grassland parcel mapped by DCRF max F1-score and MLC-stack; false positives (white areas) correspond to true ground changes as shown in photos (c) and (d).

pared to MLC-stack which dominantly misclassifies them to maize and sugar beet as demonstrated in Figure 5.25. Summer barley is recognized very well in Figure 5.26a by DCRF max F1-score compared to MLC-stack approach as depicted by Figure 5.26b. MLC-stack dominantly misclassifies the parcel as wheat and grass which are similar in appearance. Equally, DCRF max F1-score correctly detected winter barley in Figure 5.26c as opposed to MLC-stack in Figure 5.26d. Here MLC-stack incorrectly classifies parts of the winter barley parcel as wheat and rye. Similarly, MLC-stack mostly misclassified rye as wheat as depicted in Figure 5.26f due to its inability to delineate the two crops. These are crops of the same family; cereals.

(a) DCRF max F1-score. **(b)** MLC-stack. **(c)** June. **(d)** July.

Figure 5.25.: A potato parcel as mapped in DCRF max F1-score and MLC-stack; false positives (white regions) consist of maize and sugar beet. Photos (c) and (d) illustrate the parcel in June and July respectively.



(a) Summer barley in DCRF max F1-score.

(b) Summer barley in MLC-stack.

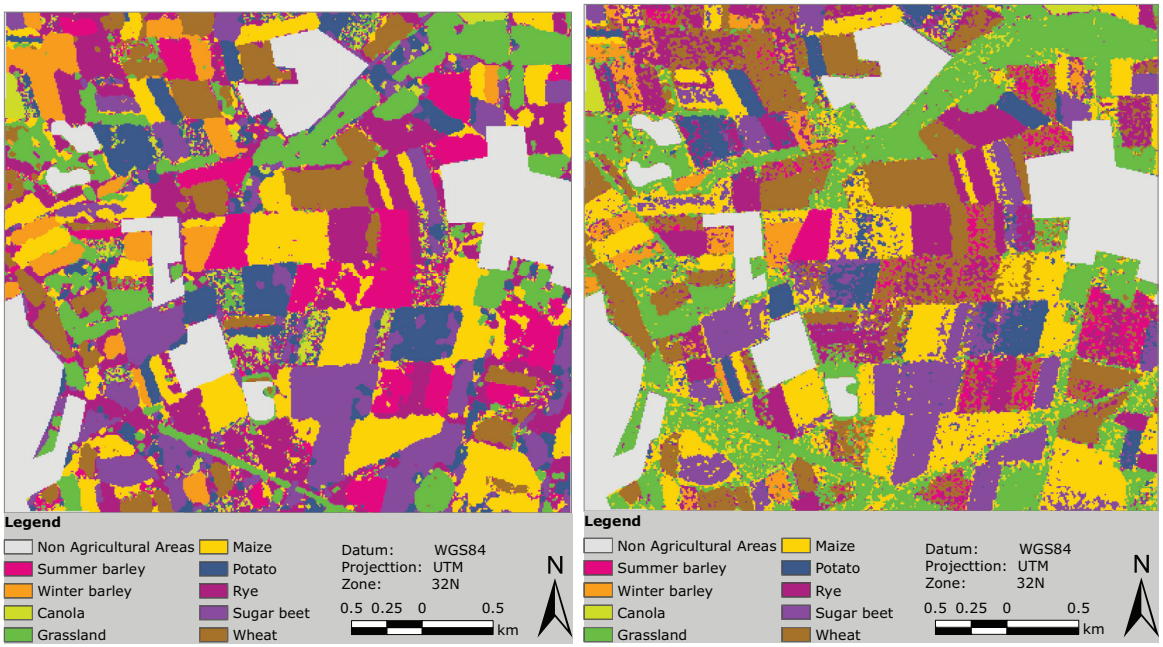
(c) Winter barley in DCRF max F1-score.

(d) Winter barley in MLC-stack.

(e) Rye in DCRF max F1-score.

(f) Rye in MLC-stack.

Figure 5.26.: Rye, summer and winter barley parcels mapped by DCRF max F1-score and MLC-stack. False positives (white areas) are dominantly grass and wheat in parcel (b), rye and wheat in parcel (d) and wheat in parcel (f).



(a) (b)

Figure 5.27.: A subset of an area covered by summer barley and other crops as mapped by DCRF max F1-score (a) and MLC-stack (b). Crop legend derived from Ebinger (2012).

Category II classification results

This section describes outcomes from experiments based on the high dimensional data in Table 4.6 and 20% training data in category II of Table 4.7. Basically, the designed DCRF Max F1-score is tested against MLC-stack and different ensemble techniques. From the experiments max F1-score ensemble approach based on DCRFs posterior probabilities maintains a high stable accuracy compared to other ensemble methods including MLC-stack as shown by Table 5.9. In this case, it surpasses product rule, median rule, majority vote, max rule and MLC-stack by 1.43%/0.02, 2.04%/0.03, 2.76%/0.04, 17.00%/0.20 and 7.73%/0.10 in OA/κ respectively. More also, a look

Table 5.9.: Overall accuracies and kappa from different ensemble techniques and MLC-stack approach based on category II data.

Method	OA	κ
Max F1-score	91.42%	0.90
Product rule	89.99%	0.88
Median rule	89.38%	0.87
Majority vote	88.66%	0.86
Max rule	74.42%	0.70
MLC-stack	83.69%	0.80

at F1-score accuracy of each class in Figure 5.28 reveals that Max F1-score supersedes the other ensemble methods in crop discrimination. For instance, it outperforms all the other approaches in all crops except for canola where its has the same F1-score with product rule and median rule. The results demonstrates the stability of max F1-score probability combination strategy.

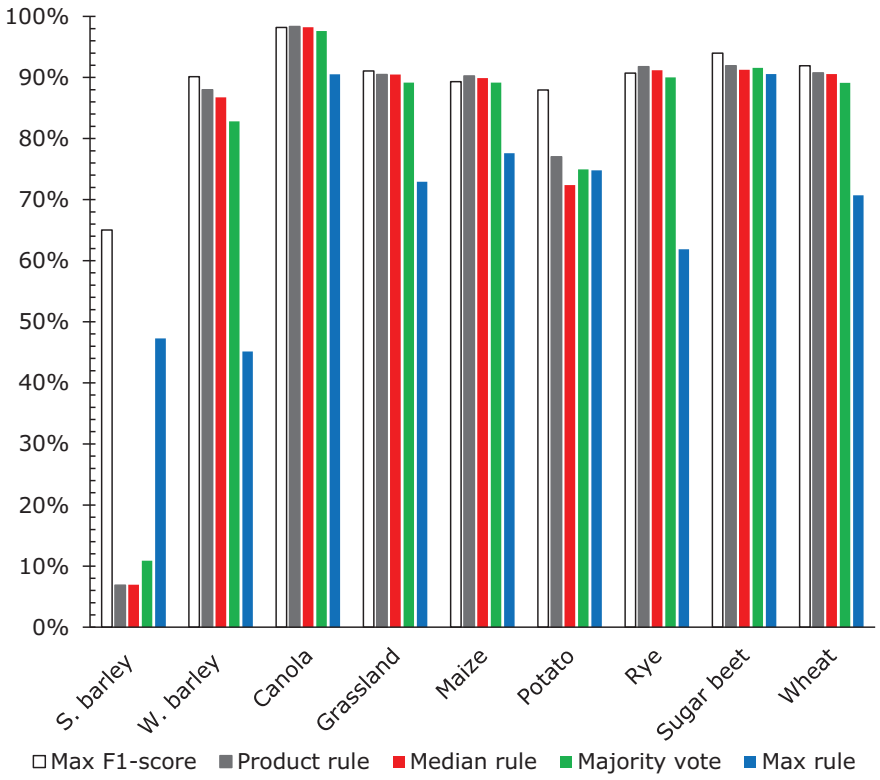


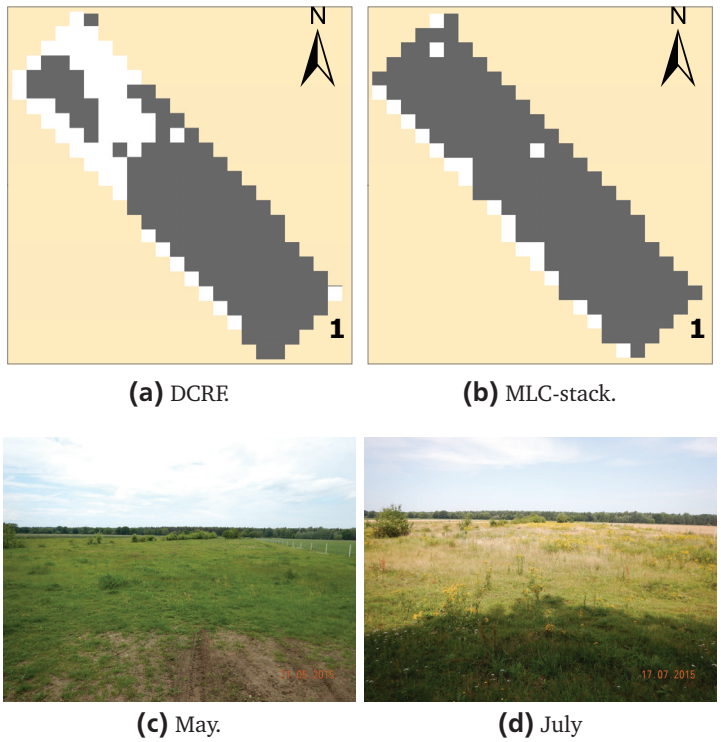
Figure 5.28.: Percentage F1-score of each crop from different ensemble methods using category II data. S. barley and W. barley correspond to summer and winter barley respectively.

An evaluation of crop mapping experiment results using DCRF max F1-score and MLC-stack is made based on producer and user accuracies and maps. Table 5.10 depicts performance of the two methods in classifying each crop when subjected to 20% training data and high dimensional data. Producer accuracy of our approach in grassland, sugar beet and wheat slightly declines by -3.14%, -2.57% and -3.21% respectively compared to MLC-stack. Grassland precision is lower than MLC-stack because of inhomogeneous parcels as highlighted by Figure 5.29. Similar observations were made using category I training data.

Most results demonstrate that DCRFs detects any inhomogeneity or abrupt changes within the crop parcels. For instance, the reason for the low producer accuracy in wheat is due to abrupt changes in some parcels (Figure 5.30). For instance, Figures 5.30a to 5.30f illustrates wheat crop felled by strong winds. This parcel is misclassified as rye because at this critical growth period, when the wheat is fallen, its head structure looks similar to rye. More also, trenches that are dug

Table 5.10.: Producer and user accuracy of each crop based on DCRF max F1-score and MLC stack using 20% training data.

Crop type	DCRF max F1-score		MLC-stack	
	Producer	User	Producer	User
Summer barley	59.77%	71.31%	1.56%	85.37%
Winter barley	86.45%	94.10%	65.13%	97.36%
Canola	97.63%	98.72%	91.90%	99.45%
Grassland	92.85%	89.32%	95.99%	74.91%
Maize	96.21%	83.32%	78.98%	87.64%
Potato	85.21%	90.86%	47.28%	88.15%
Rye	91.60%	89.86%	77.66%	84.86%
Sugar beet	93.70%	94.26%	96.27%	82.54%
Wheat	88.77%	95.30%	91.98%	79.00%



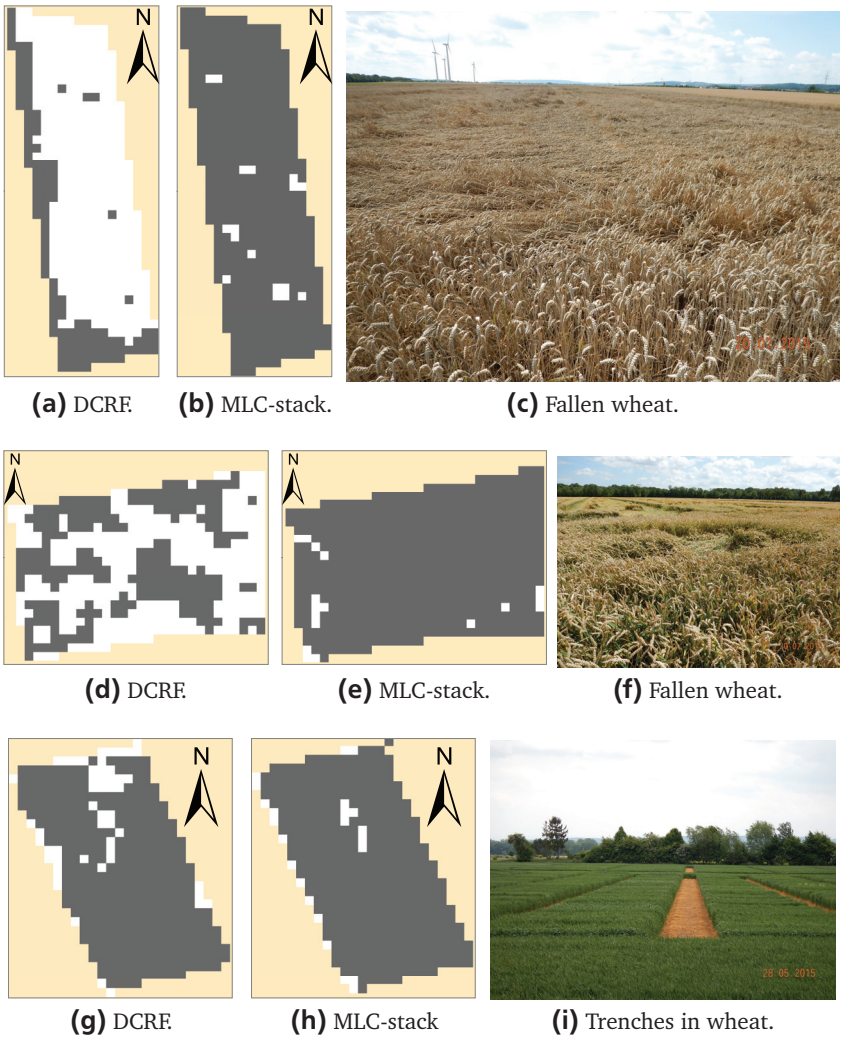
(a) DCRF. **(b)** MLC-stack.

(c) May. **(d)** July

Figure 5.29.: A grassland parcel as mapped by MLC-stack and DCRF max F1-score with corresponding ground referencing photos taken at position 1 in May and July 2015.

within wheat parcels probably due to farm management activities are detected by DCRFs as seen in Figures 5.30g to 5.30i. Hence, these are indeed true changes captured by DCRFs max F1-score as opposed to MLC-stack. Equally due to high MLC-stack false positives in summer barley, winter

barley, canola and maize by MLC-stack, its user accuracy increased by 14.06%, 3.26%, 0.73% and 4.32% respectively.



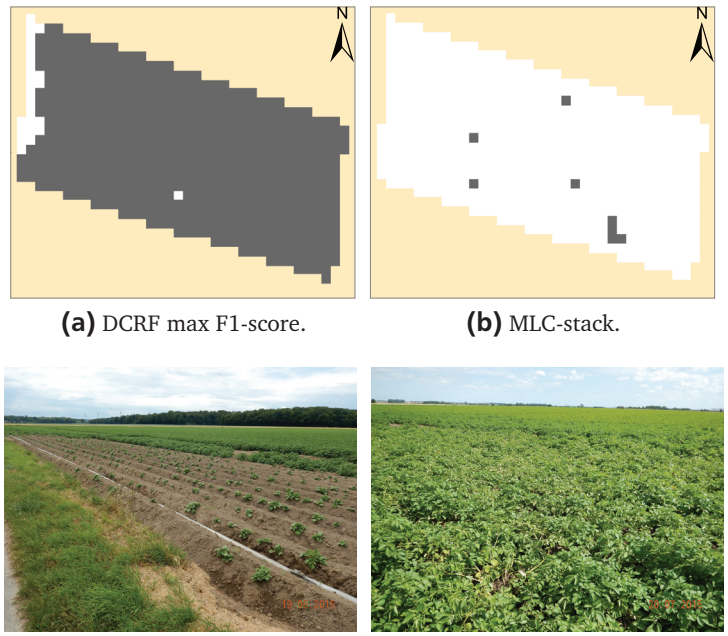
(a) DCRF. **(b)** MLC-stack. **(c)** Fallen wheat.

(d) DCRF. **(e)** MLC-stack. **(f)** Fallen wheat.

(g) DCRF. **(h)** MLC-stack **(i)** Trenches in wheat.

Figure 5.30.: Wheat as classified by DCRF max F1-score and MLC-stack. False positive pixels, white pixels, are rye mostly detected by DCRFs due to destruction by wind in (a)–(f) and dug trenches as shown by parcels (g)–(i).

Despite the aforementioned challenges, DCRF max F1-score producer accuracies in summer barley, winter barley, canola, maize, potato and rye increased by 58.21%, 21.32%, 5.73%, 17.23%, 37.93% and 13.94% respectively compared to MLC-stack. Producer accuracies of summer barley and potato are the poorest in MLC-stack. This is highlighted by Figures 5.31 and 5.32 which shows how it discriminates potato and summer barley respectively compared to our technique. In contrast, DCRF max F1-score user accuracy surpasses MLC-stack in grassland, potato, rye, sugar beet and wheat.



(a) DCRF max F1-score. **(b)** MLC-stack.

(c) June. **(d)** July.

Figure 5.31.: Potato parcel as mapped by DCRF max F1-score and MLC-stack with corresponding ground reference photos. False positives (white region) consist of maize and sugar beet in parcel (b).

(a) DCRF max F1-score. **(b)** MLC-stack. **(c)** May. **(d)** July.

Figure 5.32.: Summer barley as mapped by DCRF max F1-score and MLC-stack with corresponding ground reference photos. False positive pixels (white region) are wheat in parcel (b).

6 Discussion

This chapter gives a comprehensive analysis of experiment results and review of techniques introduced in light of performance in crop mapping. Results of experiments conducted in Fuhrberg, Kitale and Hannover region are discussed in Sections 6.1.1, 6.1.2 and 6.2 respectively. Sections 6.3 to 6.5 reviews the designed crop classification techniques highlighting their strengths and challenges.

6.1 Analysis of crop type mapping using TerraSAR-X images

6.1.1 Fuhrberg

This experiment adopted DCRFs for crop type classification using a sequence of TerraSAR-X images. In any classification, feature selection reduces computation demands. We selected four important features according to RF for crop classification. Features from VV-polarization were found important in crop discrimination compared to HH-polarization as also established in Bargiel and Herrmann (2011). We exploited their synergy for crop classification. Our DCRF framework introduces spatial and temporal interactions. To enhance better data dependent spatial interactions we designed Pearson correlation and a new version of contrast sensitive data dependent models. Experimental results established that the new version of contrast sensitive model is stable over a wide range of parameters compared to Pearson. Therefore, we chose it and set $\beta = 10$ and $\eta = 1$ based on a trade off between high classification accuracy and over-smoothing. Nonetheless, the two models gave comparable classification accuracy within selected parameters. As for temporal interactions, site-wise conditional probability matrix computed by association potential were used. In higher order DCRFs, probabilities used to generate the matrix were weighted using expert phenological knowledge.

The novel spatial-temporal interaction models were adopted for crop mapping using a sequence of TerraSAR-X images. We established that including spatial and temporal phenological information using first order DCRFs improved classification accuracy in all epochs. It was also evident that integration of expert phenological knowledge via higher order DCRFs gave the best accuracy in each epoch. Moreover, both techniques based on max F1-score ensemble outperformed MLC and CRF classification methods utilizing merged multitemporal images as bands for classification (Table 5.1). More captivating is that spatial information from stacked images in CRF classification still delivered an accuracy less than max F1-score ensemble approach. This is because like MLC, CRF association potential based its discrimination on one feature space vector which might have impacted separation of some classes. In addition, the interaction potential used a mean Euclidean

distance value computed from a vector of temporal features of adjacent nodes. This may have averaged some temporal changes in the entire sequence of a given pixel hence under-/over-classify some classes leading to lower overall accuracy. For instance, CRF over-classified barley and sugar beet resulting to lower precision than DCRFs and MLC as illustrated in Table 5.2. Hence the proposed max F1-score classifier combination scheme proved robust and accurate. Consequently, it was adopted to generate optimal crop maps using first and higher order DCRFs posterior probabilities.

Optimal maps generated from first and higher order DCRFs were compared to MLC using stacked images. An examination of user and producer accuracies exposed a detailed distribution of how each crop is recognized. The map produced by first order DCRFs has some grassland parcels poorly classified compared to MLC-stack. Inclusion of expert phenology improved precision of grassland, but is still lower than MLC-stack. However, the errors encountered by the methods correspond to true ground changes that were not detected by MLC-stack for two reasons. First, MLC-stack classification places all features in the sequence in one feature space vector from which patterns should be recognized. This may lead to a large variance that dominates small variations in a class which impact discrimination. Two, DCRF max F1-score considers data and label dependent spatial interactions. This is supported by the fact that in a homogenous parcel, e.g. parcel number 11 in Figure 5.13, DCRF max F1-score completely recognizes the parcel with higher accuracy than MLC-stack method (see Figure 5.12). In addition, artificial changes or natural changes, for example due to different variety of grassland and changes in farm management as depicted by parcel number 49 and 8 in Figure 5.13 respectively, are detected in DCRF max F1-score. In contrast, MLC-stack is a pixel based approach that ignores context hence classification results from it are accompanied by "salt and pepper" effect (Figure A.3). All other grassland parcels were classified comparably well in MLC-stack and DCRF max F1-score because they are managed in a common and unique way, and driven by economic preconditions.

Cultivation practices common to some crops can also influence discrimination ability. This was observed in two potato parcels, where row direction could have affected mapping accuracy (Figure 5.10). One of the parcels is mapped poorly by MLC-stack and first order DCRFs. Higher order expert and image phenological information in DCRFs improved recognition of the parcel. It was established that potato rows in the parcel were approximately parallel to ascending mode of TerraSAR-X path. On the contrary, the second parcel with rows perpendicular to TerraSAR-X path was detected very well by DCRFs classifiers. Ulaby and Bare (1979); Brisco et al. (1991); McNairn and Brisco (2004) highlighted that potatoes planted in ridges have significant backscatter in HH polarization especially when looking perpendicular to row direction. Generally, this phenomenon decreases from the time potato leaves cover the ridges entirely. Nonetheless, phenological information injected via higher order DCRFs minimized this assumed effect. Classification of maize also benefitted from phenological information unlike first order DCRFs and MLC-stack (Figure 5.7). Dynamic practices like early harvesting and planting of a new crop (Figure 5.11) degraded clas-

sification accuracy in all methods. Solving such challenges may require integration of advanced phenological models to crop classification tasks. In addition to cultivation practices, nature and structure of crop categories played a role in classifier performance. For instance, discrimination within broad leaved and cereals crops was a challenge. This problem affected MLC-stack technique the most compared to first order and higher order DCRFs max-F1 score ensemble.

From this experiment, it can be deduced that stacking images as bands for classification limits significant temporal phenological information inherent in images. Another disadvantage of staking images for classification is that, discrimination at each available image epoch is not perceived. In contrast, first order DCRFs incorporates temporal information at a given epoch only from the previous and the next epochs in the sequence T if $t \neq 0$ and $t \neq T$. In that case, temporal information from the entire sequence is not incorporated. Higher order DCRFs is intended for this purpose. It includes expert and image based crop phenology from the entire sequence. As already demonstrated by the results, this method has the best accuracy which is attributed to a novel, robust and accurate ensemble method; max F1-score.

6.1.2 Kitale

A total of six crops were mapped in Kitale using a sequence of TerraSAR-X strip map images. Classification tests were conducted using different designed (DCRFs max F1-score) and existing (MLC and CRF based on stacked images) techniques. The designed max F1-score technique of combining DCRFs posterior probabilities to produce an optimal map had the highest overall accuracy (Table 5.3). Product rule, median rule and majority vote have comparable accuracy while max rule was the poorest. Max F1-score classifier combination scheme still maintains high accuracy in mapping each crop (Figure 5.14). In principle, max F1-score introduced a weighting strategy to max rule. It was established that the design improved classifier combination accuracy significantly. Thus, it was adopted for optimal crop type classification based on DCRFs.

The designed DCRFs max F1-score framework was compared to RF, CRF and MLC based on stacked images. Our approach gave the highest accuracy. It was established that spatial interactions improved classification results. This is because CRF based on stacked images performed better than RF and MLC. Furthermore, MLC surpassed CRF association potential (RF) but inclusion of spatial interaction improved the accuracy beyond that of MLC. Therefore, despite using pixel blocks in MLC, spatial and temporal context are still necessary.

Analysis of MLC-stack and DCRFs max-F1 score maps highlighted challenges faced by both methods. This also exposed strengths and weaknesses of the classifiers. Some modern farming practices are a challenge to classification. For example, shade grown coffee challenged MLC-stack classification as demonstrated in Figure 5.15b. Growing coffee with trees is advocated for productive and sustainable agro-ecosystems among other benefits (Jaetzold and Schmidt, 1982). From remote sensing point of view, mixed signals are recorded from such a scene. Consequently, this impacted delineation of coffee from other crops. Despite the challenge, the designed technique

produced better maps. However, both methods had low user accuracy because maize and sugarcane were misclassified as coffee. This highlights the fact that natural structure of some crops such as broadleaves impacts classification. Maize intercropped with beans also challenged MLC-stack (Figure 5.18). Basically, farmers are encouraged to intercrop maize and legumes for soil fertility management. Despite this campaign, farmers still have their individual preferences and practices. Consequently, some farmers may intercrop legumes with maize and some prefer to plant them separately. Such sceneries definitely have mixed backscatter that challenge classification. Since our technique is based on spatial and temporal information, it minimized misclassification of crops unlike MLC-stack. However, delineation of intercropped parcels from others with individual crops was not considered. Therefore, future studies should consider classifying intercropped crops from similar crops planted individually.

6.2 Analysis of Crop type mapping from Sentinel 1 images

In this experiment, DCRF was used to map crops in Hannover region using a sequence of Sentinel 1 images. We used medium spatial resolution images compared to high resolution TerraSAR-X used in previous experiments. Difference in resolutions necessitated selection and estimation of spatial interaction model and corresponding parameters. Images with medium spatial resolution normally have small within class backscatter variations compared to ones acquired at high resolution. Therefore, high valued spatial interaction parameters tend to over-smoothen such images during classification. Consequently, our parameter search was constrained to values between 0.1 and 25 after an initial logarithmic search. Classification tests illustrated that Pearson correlation and modified version of contrast sensitive models have similar accuracy. They only differ by a small margin in overall accuracy but both performed better than contrast sensitive Potts model. However, expanded contrast sensitive model was selected for classification because it has less computation demands than Pearson correlation. We set its parameters $\beta = 5$ and $\eta = 1$ based on a trade off between high classification accuracy and over-smoothing.

The novel interaction potential model is adopted in CRF to design a DCRF sequence template classifier that considers inherent phenology in images. This framework was used in two classification test setups. The first setup was based on 50% training and 50% validation data (category I in Table 4.7). This is considered an ordinary classification with stratified random training samples generated as recommended by Stehman (2009). We established from the experiment that including spatial and temporal phenological information improved classification accuracy in all epochs as can be deduced from Figure 5.22. Max F1-score method was designed to fuse DCRFs epoch class posterior probabilities into a single optimal classification map. F1-score is used as a form of weighting in selecting optimum probabilities of each class from the sequence. This novel technique is compared to existing classifier ensemble approaches: product rule, median rule, majority vote and max rule. Thus far, it gives a stable and balanced accuracy in each crop unlike the other

ensemble approaches which fluctuate. Moreover, the method still outperformed MLC and CRF classification methods utilizing stacked multitemporal images as bands for classification (Table 5.7).

To avoid a limited judgement using overall accuracy and kappa statistic, we further compared DCRF max F1-score with MLC using stacked images. Analysis of producer and user accuracies reveals an in-depth distribution of how each crop is recognized. The designed DCRF max F1-score has slightly lower accuracy in grassland and wheat parcels compared to MLC-stack. However, the false positive errors detected by the method probably correspond to true ground changes that are not detected by MLC-stack for two reasons. First, MLC-stack classification places pixel values of all epochs in one vector which can lead to a large variance that dominates small variations in a class. Two, DCRF max F1-score considers data and label dependent spatial interactions. This is supported by the fact that DCRF max F1-score completely recognizes homogenous grass regions in Figure 5.24. For example, inhomogeneous areas of the parcels with shrubs and some boundary trees are detected by the method unlike MLC-stack which missed them (Figures 5.24c and 5.24d). Overall, DCRF max F1-score mapped most crops (see Figure 5.27a) with high classification accuracy and less "salt and pepper" common in MLC-stack (see Figure 5.27b). In particular summer barley, winter barley and rye cereal crops are discriminated with no confusion as opposed to MLC-stack as depicted in Figure 5.26. The MLC-stack could not delineate the crops. It also encounters the same challenge in broad leaved crops (Figure 5.25b).

The second experiment set up used 20% training and 80% validation data (category II in Table 4.7). This experiment aimed at testing our approach's robustness to deal with high dimensional data and less training data. It can be noted that the method still maintains high crop classification accuracy compared to other ensemble approaches and MLC-stack. More captivating, is that the product rule ensemble approach that had the highest overall accuracy in category I, dropped significantly below that of max F1-score in category II experiment. Product rule is known to be a severe rule of fusing classifier outputs because it is sufficient for a single epoch to assign a close to zero probability to a class. In such a case, the combined output will also be a value close to zero due to reduction effect of multiplication. This effectively undermines the decision of any correctly predicted class in the sequence and denies it an opportunity for its identity to be finally accepted or rejected (Kittler et al., 1998). The designed max F1-score maintains high and stable accuracy in each crop amongst other ensemble methods (Figure 5.28). Moreover, an evaluation of MLC-stack classification results still demonstrates that DCRF max-F1 score has high classification accuracy. It was observed that precision of grassland, sugar beet and wheat slightly declined compared to MLC-stack because of false positives. However, most of the false positives corresponded to true changes on the ground. Basically, DCRFs detected any inhomogeneity or abrupt changes within the crop parcels. Similar to category I, shrubs, boundary trees and different varieties in grassland were detected as demonstrated in Figure 5.29. As for wheat, abrupt changes due to destruction by strong winds were captured as demonstrated by photo evidence in Figures 5.30a to 5.30f. It is interesting that this phenomena posed a new challenge to DCRFs to de-

lineate the wheat parcel. This is because wheat head structure was altered by wind which probably changed radar backscatter. The altered head structure of wheat looks similar to rye. Therefore, their radar backscatter may subsequently be similar hence DCRFs can not delineate them. More also, changes due to, for example farm management, was detected in some wheat fields (Figures 5.30g to 5.30i). Capability of DCRFs to detect these changes is important for monitoring agricultural fields and farming practices such as early harvesting e.t.c. Such ability is a subject of future exploration. The lack of spatial-temporal phenological information inherent in images constrained MLC-stack from detecting such changes. In addition, MLC-stack could not delineate between some broad leaved crops. For instance, it misclassified almost an entire potato parcel as maize and sugar beet (Figure 5.31). It still faces the same discrimination problem with some cereal crops such as in Figure 5.32b where it classifies summer barley as wheat. These challenges were also observed using category I training data. However, DCRF max F1-score is able to delineate within broad leaved and cereal crops with high precision.

In retrospect, this experiment aimed at introducing a new method of crop classification from a sequence of Sentinel 1 radar images. Several observations can be made from the experiment. For instance, it can be deduced that stacking of multitemporal images for classification poses two major problems. First, it increases dimensionality in the feature space. This requires more training samples especially in statistical methods in order to be able to estimate parameters. Such parameters are in turn exponential with increase in data dimensionality and, for a fixed sample size, their precision reduces leading to loss of classifier efficiency (Tso and Mather, 2009, chap. 2.6). In MLC, these parameters are the mean vector and the variance-covariance matrix for each class. On the contrary, RF used in the designed technique is non-parametric because it does not involve estimation of statistical parameters. However, RF requires large enough training data to represent the characteristics of each class and should have minimal spatial autocorrelation (Millard and Richardson, 2015). The framework advocated by Stehman (2009) was used in generating training samples. Thus, requirements for both methods were adhered to. The spatial threshold of 100 m used during sampling is proposed by Van der Meer et al. (2000) to minimize spatial autocorrelated samples in vegetation related land-cover mapping. Secondly, stacking images, that bear backscattered phenological information from different periods for classification, constraints discrimination to a single feature space vector. Now, the first problem, requires more resources in terms of time and money in order to gather enough training samples such as through ground referencing campaigns. Consequently, this pushes the cost of land-cover mapping efforts using MLC higher. In contrast, the second challenge may cause overlapping of classes hence jeopardize delineation of classes from the feature space. Our approach is robust to both challenges. More also, the approach can detect any abrupt changes in crop parcels, for example due to disasters or farm management. Thus, the technique is promising for use in monitoring changes caused by disasters, early grass mowing and early crop harvesting e.t.c. in agricultural regions.

This far, the potential of Sentinel 1 radar data for crop classification by exploiting spatial-temporal information has been demonstrated. It opens up opportunities for future crop monitoring unimpeded by weather conditions. However, the high temporal resolution of Sentinel 1 introduces data dimensionality problem. Our method is robust to this problem because of class conditional independence which minimizes dimensionality while enhancing high information mining via spatial-temporal interactions. This feature also makes the framework flexible to additional information useful for classification via conditional probability matrix. In addition, our ensemble classifier can be used with any number of available images at any time of a season to get an estimate of crop coverage.

6.3 Spatial interaction potential review

One of the objectives of this research was to design a suitable spatial interaction model for DCRF crop classification. Spatial interaction potential enforces neighbourhood correlation of image sites using data and labels. This is based on the premise portrayed by intrinsic coherent and spatially connected structure of land-cover in the physical world. The same principle is well upheld by the first law of geography as defined by Tobler (1970) (see also Section 3.1.2) and spatial context (Goodchild, 1992). In CRFs, a data dependent model is commonly used to enforce data and label correlation via neighbourhood cliques. Several data dependent designs of I exist, for example concatenation of features (Zhong and Wang, 2007a,b; Li et al., 2009), absolute difference of features (He et al., 2008; Wegner et al., 2011a), contrast sensitive model (Shotton et al., 2009; Schindler, 2012) and inverse/transformed euclidean distance (Kenduiywo et al., 2014).

To enhance better spatial interactions, two data dependent models were developed and compared to commonly used contrast sensitive model (Shotton et al., 2009; Schindler, 2012). The first design is a modified version of the contrast sensitive model while the second one is based on Pearson correlation. Both models have two parameters; β and η. The role of β is to regulate smoothness while η weighs and controls inclusion ($\eta > 0$) or exclusion ($\eta = 0$) of data interactions. These two parameters were determined via empirical classification tests. The tests also served to illustrate and verify the strength of each data dependent design.

Experimental outcomes in Sections 5.1.1 and 5.2.1 demonstrate the performance of the designed models. Classification results using TerraSAR-X in Section 5.1.1 and Sentinel 1 in Section 5.2.1 shows that Pearson correlation and new version of contrast sensitive models perform comparably. However, it was established in Section 5.1.1 that Pearson correlation function is sensitive for $eta > 10^2$ parameter values. When such eta values were used, the model smoothens the entire image. Therefore, it requires a careful parameter search. For this reason, the expanded version of the contrast sensitive model was adopted for all classifications because it is robust over a range of parameters. In any case, this model outperformed Pearson correlation by a very small margin of classification accuracy. It also has less computation demands compared to Pearson model.

Both expanded version of contrast sensitive model and Pearson correlation outperforms standard contrast sensitive model. This is because, compared to contrast sensitive model, the two new designs moderated smoothing given data evidence in two scenarios. First scenario, when adjacent labels are similar and data evidence support it, smoothness weight is increased. If data evidence does not support label similarity, smoothness weight is reduced proportional to difference and dissimilarity in data respectively. The second scenario considers dissimilar adjacent labels by reducing smoothness weight if supported by data evidence. When label dissimilarity is contrary to data evidence, smoothness weight is increased inversely proportional to the magnitude of data evidence. This design realized a discontinuity adaptive model that moderates smoothing considering data evidence as suggested by Li (1995). These models are different from contrast sensitive model which only favours smoothing similar adjacent labels. Generally, the designs establishes a novel framework for including spatial context. Moreover, Pearson correlation model has been introduced for the first time in this study. It quantifies context by using correlation of adjacent image sites as opposed to Euclidean distance in contrast sensitive technique.

6.4 Temporal interaction potential and expert phenology review

This study targeted to use expert and image based crop phenology for classification. Crop phenology is a temporal phenomena. The structure and appearance of crops varies throughout its life cycle. Evolution of a crop is to a great extent unique and independent from others as its driven by natural and artificial factors. This aspect led this study to adopt conditional class independence assumption to design temporal interactions from a sequence of radar images. A temporal potential was introduced based on conditional probability matrix as a link between two temporal nodes. Bayes' conditional probability formula was applied to compute the matrix using estimated class probabilities from association potential. The use of conditional probabilities to represent temporal interactions between a pair of epochs was deemed a robust approach as opposed to feature difference function used in Hoberg and Müller (2011). This is because the epoch-wise estimated class probabilities are soft decisions that represent statistical beliefs/opinions based on prior field-work knowledge and captured images. Therefore, they do not directly suffer from prevailing image acquisition conditions like in feature difference functions which directly compares two pixels acquired in different epochs. The probabilities thus allow further higher level reasoning in order to attain acceptable class inference. Moreover, it has already been demonstrated using MLC and CRFs, that joint classification of features or images limits discrimination of classes to one feature space vector which may consequently lead to low classification accuracy. Hence the matrix of conditional probabilities directly estimated from node potentials of each epoch, was used to express pixel-wise temporal interactions instead of joint class probabilities marginalized to conditional probabilities as proposed by Liu et al. (2008). This structure was adopted to capture expert and/or image based crop phenology using two designs: first order and higher order DCRFs.

First order DCRFs only integrated image based crop phenology from a given sequence of multi-temporal images. Here, conditional probability matrix was used to capture bi-directional phenological information in images between two nodes in the sequence. For example, temporal information is added to a node at a given epoch only using nodes in the previous and the next epochs in the sequence T if $t \neq 0$ and $t \neq T$. In principle, temporal interactions are considered only from neighbouring nodes in time. This technique improved classification accuracy in all epochs compared to mono-temporal approaches as demonstrated by Figures 5.5 and 5.22. It signifies that temporal information from neighbouring nodes in time play an important role in classification. However, more temporal information could still be mined from the sequence. Basically, phenology varies throughout a crop's lifetime causing structural and canopy moisture content changes. As a result radar backscatter from crops fluctuates providing different opportune times to distinguish them (Figures 2.5 and 3.3). For that matter, temporal phenological information from the entire sequence should be incorporated. To integrate phenology information covering the lifetime of a crop we built high order DCRFs.

The developed higher order DCRFs allow n nodes (n=T) to exchange expert and image based temporal phenological information. Image based temporal interactions were still computed using conditional probability matrices like in first order DCRFs. The only difference is that probabilities used in Bayes' formula (Equation (3.10)) were weighted using expert phenology knowledge. This ensured that knowledge on crop phenology in the entire sequence is incorporated at each node. Non zero constant weights, that is 2.5, were used in epochs where the backscatter is not directly from the crops (Table 4.3). This is because epochs outside the growing phase of a crop still contain unique characteristics that can aid discrimination of a parcel. For instance, remaining stalks of maize based on different harvesting methods such as by machine or manual stooking in Figure 6.1 may return distinguishable backscatter. Moreover, certain cultivating mechanism recommended for specific crops such as ridges in potato parcels can aid classification (Figure 5.10). Such aspects were considered in designing phenology weights. The expert weights and inherent image phenology established positive contribution in higher order DCRFS compared to first order (Figure 5.5). The improved classification accuracy can be attributed to more temporal information added to each node in the sequence. To sum up, passing temporal information via conditional probability matrix produced better classification accuracy. Higher mapping accuracy was obtained especially using higher order DCRFs though computational demands increased.

6.5 Comments on optimal map generation from a sequence

Stakeholders and policy makers are interested in annual or seasonal crop statistics. The study developed a spatial-temporal sequence crop classification framework that generated maps at each epoch. This enables crop discrimination capability to be perceived at every stage. However, an optimal season crop map is more desirable to end users. Therefore, our next aim was to design an ensemble classifier. The ensemble classifier fuses crop classification information from each

Figure 6.1.: Stooking method of harvesting maize which may result in distinguishable backscatter due to clustered maize stalk. Photo credit (One Acre Fund, 2016).

epoch to produce an optimal map. Our study introduced max F1-score ensemble technique. The approach adopted F1-score weighting strategy in selecting the final probability of each crop type at a given node in the sequence. This method was investigated against four classifier combination rules namely: product rule, median rule, majority vote and max rule in (Kittler et al., 1998). Thus far, max F1-score was established as the most accurate and robust ensemble method. It is mostly followed by median rule and majority vote which have comparable accuracy while max rule is the poorest. The accuracy of product rule is in most cases below max F1-score but rapidly fluctuates in all experiments compared to the other methods.

Outcomes from two experiments in Section 5.2.2 established that product rule is a severe classifier combination strategy. In one experiment product rule ensemble approach had the highest overall accuracy which subsequently dropped significantly below that of max F1-score in a later experiment. A similar trend is observed in two experiments using standard DCRFs and higher order DCRFs in Section 5.1.1 where, the accuracy of product rule dropped in the later experiment despite increased temporal interactions. As earlier mentioned, product rule is a severe rule of fusing classifier outputs because it is sufficient for a single epoch to assign a close to zero probability to a class. In such a case, the combined output will also be a value close to zero due to reduction effect of multiplication. This effectively undermines the decision of any correctly predicted class in the sequence and denies it an opportunity for its identity to be finally accepted or rejected (Kittler et al., 1998). The other classifier ensembles also have their cons. To begin with, despite median rule being a robust estimate of mean, it can to some extent be affected by outlier probabilities. Secondly, the majority vote relies on counts from already established label decisions. This may propagate errors from initial decisions. More also, it is prone to an error if counts of two or more classes in the sequence tie. Lastly, max rule is susceptible to select a wrong probability if there

are two maximum and equal probabilities in the sequence. In addition, sometimes the highest probability in the sequence of a node is not necessarily correct. It is possible for DCRFs to make a wrong estimate for a node in the sequence which apparently ends up being the highest. The two observations motivated this study to introduce a weighting scheme in max rule. Indeed, the F1-score weighting strategy overcame these challenges because max rule accuracy improved from the lowest to the highest in all the ensemble methods. This bearing in mind the experiments were conducted in different regions with different crops hence, different challenges. It thus establishes that max F1-score technique is robust and accurate.

The designed ensemble classifier can be used with any number of available images at any time of a season to get an estimate of crop coverage. It performs better than techniques using stacked multitemporal images for classification. Moreover, compared to other classifier combination strategies, the weighting we introduced to the ensemble improved classification. Therefore, it guarantees an optimal map in terms of accuracy. This is essentially beneficial to policy makers and stakeholders in food security policy formulations and seeking alternative preventive measures to ensure balanced ecosystem services. In addition, agricultural stock market and traders can anticipate good years while insurers can accurately compute premiums and determine compensations where necessary.

6.6 Summary

In summary, this study implemented a flexible and scalable DCRFs max F1-score framework for crop mapping. The structure was designed to accommodate image tile based classification if computer memory demands are limited. Image tiles and corresponding overlaps can be adjusted by user to desired sizes. This framework mapped crops with high accuracy. Success of the design is founded on the principle that each crop has a unique period in time when their backscatter is distinguishable. For example, corn and sugar beet return high backscatter in C- or X-band during early stages of their development with little change in the rest of the growing season (Brisco and Protz, 1980; Bouman, 1988; Tso and Mather, 1999). In contrast, grains like wheat and barley have temporal backscatter variation throughout growing season (Brisco et al., 1992). This fact is not well exploited by conventional approaches that stack multitemporal images for classification. For instance, MLC-stack could not effectively discriminate broad leaved crops and cereals/grains. A study by Bargiel and Herrmann (2011) encountered the same challenge. Therefore, this study recommends integration of spatial context, expert- and image-based temporal phenological information for crop classification. The designed higher order DCRFs is well suited for this task. This is because first order DCRFs lacks expert based crop phenological information but produces results with slightly lower accuracy than higher order. Choice between the two models thus depends on availability of crop phenology knowledge.

7 Conclusion and outlook

The main goal of this study was to design a crop mapping technique that exploits spatial-temporal backscatter variation caused by plant phenology. To achieve this, a statistical framework that models spatial and temporal dynamics in a principled manner was designed. So far, such existing framework was proposed for first order DCRFs 1-D text sequence classification by Sutton et al. (2007). Therefore, our task was to extend the design for 3-D optimal crop classification from a sequence of radar images. This was accomplished through five objectives.

Our DCRFs template consist of association, spatial, and temporal potentials. The potentials can be considered as independent classifiers hence domain specific ones can be used. Therefore, our first objective was to develop a suitable spatial interaction model. Two models, namely, new version of contrast sensitive Potts model and Pearson correlation were introduced. The first design is a modification of conventional contrast sensitive Potts model to include label and data interaction both when adjacent labels are similar or dissimilar. Data interactions are incorporated via Euclidean distance function. Pearson correlation model uses the same principle except that data interactions are integrated using correlation function. We established that the two models perform comparably and better than contrast sensitive model. However, Pearson correlation model is sensitive to very high parameter values and has slightly more computation demands. Nonetheless, Pearson correlation was introduced in this study for the first time. Its performance is encouraging for future use. Furthermore, this study recommends that spatially weighted Pearson correlation coefficient from higher order neighbours of a pixel be incorporated by subsequent studies. This will enable inclusion of more spatial dependency to a given pixel as opposed to the current use of pairwise pixel dependency.

The second, third and fourth objectives saw the introduction of pixel-wise temporal interactions using DCRFs. Here, bi-directional pixel-wise temporal information was represented in matrices computed using Bayes' conditional probability formula. The approach proved to be an effective means of capturing expert and image based temporal crop phenological information. This was established using first and higher order DCRFs classification results. Only image based crop phenological information was used in first order DCRFs while higher order used both expert and image based phenological information. Basically, a weight was allocated to each crop type in each epoch based on expert phenology knowledge. These weights gave varying importance to phenological information captured by radar images. In effect, it enhanced discrimination across all nodes in the sequence as observed in classification results. However, the global weights assume similar phenology for each crop type at a given epoch. Essentially, crops in neighbouring parcels, even if similar, may at a given point in time, be in different phenology phases. This is influenced by farm

management practices, soil and weather conditions amongst other factors. Therefore, it is clear that phenology varies spatially too. Prospective studies should consider this phenomenon.

Parameter tuning and inference is a task that still requires further investigation. A weighting strategy for the three potentials (association, spatial and temporal) should be investigated. In addition, a suitable inference technique for higher order DCRFs should be adopted instead of LBP. Loopy belief propagation infers a node's probability by computing the sum product of all incoming edges (in this case spatial and temporal interactions). Due to continuous multiplication of incoming messages (probabilities) at a node during inference, reduction effect of multiplication may be encountered. Therefore, an alternative inference algorithm should be implemented in future for example tree re-weighting message passing schedule.

Finally, an optimal mapping ensemble was designed. The max F1-score performed better than other existing ensembles and also compared to stacking multitemporal images for classification. In principle, merging multitemporal images for classification increases feature space dimensionality. This reduces classifier accuracy because more training samples are required to obtain reasonable estimates from increased data dimensionality. More also, pattern recognition is constrained to an already high dimensional feature space and any errors can be costly. Some classes may overlap in a high dimensional multitemporal feature space due to large variance. This in turn affects delineation of classes. We recommend our technique which can already handle high dimensional data with less training data without a significant compromise in classification accuracy. This means that land-cover mapping applications can benefit from reduced cost of ground reference data collection and the increasing remote sensing data. In addition, the designed technique detected rapid changes in crop parcels. This presents the possibility of using it to monitor changes in agricultural areas, i.e due to farm management or natural disasters like wind destruction of crops. Another potential that should be explored in future is the delineation of intercropped parcels from parcels with similar individual crops. For instance, experiments can be conducted to discriminate maize parcels intercropped with beans from parcels with either maize or beans only.

Bibliography

Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.

Bargiel, D. and Herrmann, S. (2011). Multi-Temporal Land-Cover Classification of Agricultural Areas in Two European Regions with High Resolution Spotlight TerraSAR-X Data. *Remote Sensing*, 3(5):859–877.

Barton, D. K. and Leonov, S. A. (1998). Radar technology encyclopedia.

Besag, J. (1986). On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302.

Bishop, C. M. (2006). *Pattern recognition and machine learning*, volume 1. Springer, New York.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):2–16.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R. Q., van der Meer, F., van der Werff, H., van Coillie, F., and Tiede, D. (2014). Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:180–191.

Blaschke, T. and Strobl, J. (2001). What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GeoBIT/GIS*, 6(1):12–17.

Bouman, B. A. M. (1988). Microwave backscatter from beets, peas and potatoes throughout the growing season. In *Spectral Signatures of Objects in Remote Sensing*, volume 287, page 25.

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 1st edition.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Brisco, B., Brown, R. J., Gairns, J. G., and Snider, B. (1992). Temporal Ground-Based Scatterometer Observations of Crops in Western Canada. *Canadian Journal of Remote Sensing*, 18(1):14–21.

Brisco, B., Brown, R. J., Snider, B., Sofko, G. J., Koehler, J. A., and Wacker, A. G. (1991). Tillage effects on the radar backscattering coefficient of grain stubble fields. *International Journal of Remote Sensing*, 12(11):2283–2298.

Brisco, B. and Protz, R. (1980). Corn Field Identification Accuracy Using Airborne Radar Imagery. *Canadian Journal of Remote Sensing*, 6(1):15–25.

Brisco, B., Ulaby, F. T., and Protz, R. (1984). Improving crop classifications through attention to the timing of airborne radar acquisitions. *Photogrammetric Engineering and Remote Sensing*, 50:739–745.

Buja, K. and Menza, C. (2013). *Sampling Design Tool for ArcGIS - Instruction Manual*. NOAA, Silver Spring, MD.

Deutscher Wetterdienst (2012). "Mittelwerte der Temperatur und des Niederschlags bezogen auf den aktuellen Standort". http://www.dwd.de. (2 Oct. 2015).

Ebinger, L. (2012). "133 map categories! How the US Department of Agriculture solved a complex cartographic design problem". http://www.sco.wisc.edu/news/133-map-categories-how-the-us-department-of-agriculture-solved-a-complex-cartographic-design-problem.html. (2 Oct. 2015).

FAO (2013). Food wastage footprint-Impacts on natural resources: summary report. Technical report, Food and Agriculture Organization (FAO).

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockstrom, J., Sheehan, J., Siebert, S., Tilman, D., and Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, 478(7369):337–342.

Forkuor, G., Conrad, C., Thiel, M., Ullmann, T., and Zoungrana, E. (2014). Integration of Optical and Synthetic Aperture Radar Imagery for Improving Crop Mapping in Northwestern Benin, West Africa. *Remote Sensing*, 6(7):6472–6499.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.

Gomez-Chova, L., Fernández-Prieto, D., Calpe, J., Soria, E., Vila, J., and Camps-Valls, G. (2006). Urban monitoring using multi-temporal SAR and multi-spectral data. *Pattern Recognition Letters*, 27(4):234–243.

Goodchild, M. F. (1992). Geographical information science. *International Journal of Geographical Information Systems*, 6(1):31–45.

Gould, S., Rodgers, J., Cohen, D., Elidan, G., and Koller, D. (2008). Multi-class segmentation with relative location prior. *International Journal of Computer Vision*, 80(3):300–316.

Hammersley, J. M. and Clifford, P. (1971). Markov field on finite graphs and lattices.

Hastie, T., Tibshirani, R., and Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2nd edition.

He, W., Jäger, M., Reigber, A., and Hellwich, O. (2008). Building Extraction from Polarimetric SAR Data using Mean Shift and Conditional Random Fields. In *7th European Conference on Synthetic Aperture Radar (EUSAR)*, pages 1–4, Friedrichshafen, Germany.

He, X., Zemel, R., and Carreira-Perpiñán, M. A. (2004). Multiscale conditional random fields for image labeling. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 695–702, Washington, DC, USA.

Henderson, F. M. and Lewis, A. J., editors (1998). *Principles and Applications of Imaging Radar (Manual of remote sensing)*. Wiley, 3 edition.

Hoberg, T. and Müller, S. (2011). Multitemporal Crop Type Classification Using Conditional Random Fields and RapidEye Data. In *ISPRS Workshop*, Hannover, Germany.

Hoberg, T. and Rottensteiner, F. (2010). Classification of settlement areas in remote sensing imagery using conditional random fields. In *ISPRS TC VII Symposium-100 Years ISPRS*, Vienna, Austria.

Hoberg, T., Rottensteiner, F., Feitosa, R. Q., and Heipke, C. (2015). Conditional Random Fields for Multitemporal and Multiscale Classification of Optical Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2):659–673.

Hoberg, T., Rottensteiner, F., and Heipke, C. (2010). Classification of multitemporal remote sensing data using Conditional Random Fields. In *IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*, pages 1–4.

Hoberg, T., Rottensteiner, F., and Heipke, C. (2011). Classification of multitemporal remote sensing data of different resolution using Conditional Random Fields. In *IEEE International Conference on Computer Vision (ICCV) Workshops*, pages 235–242.

Hornegger, J., Paulus, D., and Niemann, H. (2000). Probabilistic Modeling in Computer Vision. In Jaehne, B. and Haussecker, H., editors, *Computer Vision and Applications*, pages 517–540. Academic Press, San Diego.

ITC (2010). *GI Science and Earth Observation: a process - based approach*. ITC Educational Textbook Series. University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.

Jaetzold, R. and Schmidt, H. (1982). *Farm management handbook of Kenya*, volume 2. Ministry of Agriculture Nairobi.

Jeon, B. and Landgrebe, D. A. (1992). Classification with spatio-temporal interpixel class dependency contexts. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4):663–672.

Jianya, G., Haigang, S., Guorui, M., and Qiming, Z. (2008). A review of multi-temporal remote sensing data change detection algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII(B7):757–762.

Jiao, X., Kovacs, J., Shang, J., McNairn, H., Walters, D., Ma, B., and Geng, X. (2014). Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96:38–46.

Kenduiywo, B. K., Bargiel, D., and Soergel, U. (2015). SPATIAL-TEMPORAL CONDITIONAL RANDOM FIELDS CROP CLASSIFICATION FROM TERRASAR-X IMAGES. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W4:79–86.

Kenduiywo, B. K., Bargiel, D., and Soergel, U. (2016). CROP TYPE MAPPING FROM A SEQUENCE OF TERRASAR-X IMAGES WITH DYNAMIC CONDITIONAL RANDOM FIELDS. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-7:59–66.

Kenduiywo, B. K., Tolpekin, V. A., and Stein, A. (2014). Detection of built-up area in optical and synthetic aperture radar images using conditional random fields. *Journal of Applied Remote Sensing*, 8(1):083672–1 – 083672–18.

Kenya Meteorological Department (2015). "Seasonal forecast". http://www.meteo.go.ke. (6 May 2016).

Kittler, J. and Föglein, J. (1984). Contextual classification of multispectral pixel data. *Image and Vision Computing*, 2(1):13–29.

Kittler, J., Hatef, M., Duin, R. P. W., and Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239.

Korc, F. and Förstner, W. (2008). Interpreting terrestrial images of urban scenes using discriminative random fields. In *21st Congress of the International Society for Photogrammetry and Remote Sensing*, pages 291–296, Beijing, China.

Kuenzer, C. and Knauer, K. (2013). Remote sensing of rice crop areas. *International Journal of Remote Sensing*, 34(6):2101–2139.

Kumar, S. (2005). *Models for learning spatial interactions in natural images for context-based classification*. PhD thesis, Carnegie Mellon University Institute of Robotics.

Kumar, S. (2006). Discriminative random fields. *International Journal of Computer Vision*, 68(2):179–201.

Kumar, S. and Hebert, M. (2003). Discriminative fields for modeling spatial dependencies in natural images. In *Advances in Neural Information Processing Systems*.

Lafferty, J. D., McCallum, A., and Pereira, F. (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann.

Le Toan, T., Lopes, A., and Huet, M. (1984). On the relationships between radar backscattering coefficient and vegetation canopy characteristics. In *IGARSS '84: remote sensing–from research towards operational use*, volume 1, pages 155–160, Strasbourg. ESA.

Leite, P., Feitosa, R. Q., Formaggio, A. R., da Costa, G., Pakzad, K., and Sanches, I. D. (2011). Hidden Markov Models for crop recognition in remote sensing image sequences. *Pattern Recognition Letters*, 32(1):19–26.

Li, S. (1995). On discontinuity-adaptive smoothness priors in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):576–586.

Li, Z., Ma, J., Zhang, R., and Chen, X. (2009). Urban areas detection based on conditional random field and multiscale textural features. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7498, pages 74981I–6.

Lillesand, T., Kiefer, R. W., and Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley & Sons.

Liu, D., Kelly, M., and Gong, P. (2006). A spatial-temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. *Remote Sensing of Environment*, 101(2):167–180.

Liu, D., Song, K., Townshend, J. R. G., and Gong, P. (2008). Using local transition probability models in Markov random fields for forest change detection. *Remote Sensing of Environment*, 112(5):2222 – 2231.

Long, J. A., Lawrence, R. L., Greenwood, M. C., Marshall, L., and Miller, P. R. (2013). Object-oriented crop classification using multitemporal ETM+ SLC-off imagery and random forest. *GIScience & Remote Sensing*, 50(4):418–436.

Lu, D. and Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870.

Lu, W., Murphy, K. P., Little, J. J., Sheffer, A., and Fu, H. (2009). A Hybrid Conditional Random Field for Estimating the Underlying Ground Surface From Airborne LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2913–2922.

McNairn, H. and Brisco, B. (2004). The application of C-band polarimetric SAR for agriculture: a review. *Canadian Journal of Remote Sensing*, 30(3):525–542.

McNairn, H., Kross, A., Lapen, D., Caves, R., and Shang, J. (2014). Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. *International Journal of Applied Earth Observation and Geoinformation*, 28:252 –259.

McNairn, H., Shang, J., Jiao, X., and Champagne, C. (2009). The Contribution of ALOS PALSAR Multipolarization and Polarimetric Data to Crop Classification. *IEEE Transactions on Geoscience and Remote Sensing,*, 47(12):3981–3992.

Meier, U. (2001). *Growth stages of mono-and dicotyledonous plants*. Federal Biological Research Centre for Agriculture and Forestry, Berlin and Braunschweig, 2nd edition.

Melgani, F. and Serpico, S. (2003). A Markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2478–2487.

Millard, K. and Richardson, M. (2015). On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, 7(7):8489–8515.

Moser, G. and Serpico, S. B. (2011). Multitemporal region-based classification of high-resolution images by Markov random fields and multiscale segmentation. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 102–105, Vancouver, BC, Canada.

Mosleh, M., Hassan, K., and Chowdhury, H. (2015). Application of Remote Sensors in Mapping Rice Area and Forecasting Its Production: A Review. *Sensors*, 15(1):769–791.

Murphy, K. P., Weiss, Y., and Jordan, M. I. (1999). Loopy Belief Propagation for Approximate Inference: An Empirical Study. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 467–475. Morgan Kaufmann Publishers Inc.

Niemeyer, J., Rottensteiner, F., and Soergel, U. (2013). Classification of urban LiDAR data using conditional random field and random forests. In *Joint Urban Remote Sensing Event (JURSE)*, pages 139–142.

Niemeyer, J., Wegner, J. D., Mallet, C., Rottensteiner, F., and Soergel, U. (2011). Conditional random fields for urban scene classification with full waveform LiDAR data. In *Photogrammetric Image Analysis*, pages 233–244. Springer.

One Acre Fund (2016). PHOTO ESSAY: MAIZE HARVEST AND STORAGE. https://www.oneacrefund.org/blogs/information/post/photo-essay-maize-harvest-and-storage/742. (13 May 2016).

OpenCV (2014). Random Trees. http://docs.opencv.org/modules/ml/doc/ml.html. (20 Nov. 2014).

Parikh, D. and Batra, D. (2008). CRFs for image classification. Technical report, Carnegie Mellon University.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

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*, 22(1).

Richards, J. A. (2009). *Remote Sensing with Imaging Radar*. Signals and Communication Technology. Springer-Verlag, Berlin Heidelberg.

Schindler, K. (2012). An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing,*, 50(11):4534–4545.

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*, 88(1–2):157–169.

Shotton, J., Winn, J., Rother, C., and Criminisi, A. (2009). TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context. *International Journal of Computer Vision*, 81(1):2–23.

Siachalou, S., Mallinis, G., and Tsakiri-Strati, M. (2015). A Hidden Markov Models Approach for Crop Classification: Linking Crop Phenology to Time Series of Multi-Sensor Remote Sensing Data. *Remote Sensing*, 7(4):3633–3650.

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. H., editors, *AI 2006: Advances in Artificial Intelligence*, volume 4304 of *Lecture Notes in Computer Science*, pages 1015–1021. Springer-Verlag Berlin Heidelberg.

Solberg, A. H. S., Taxt, T., and Jain, A. K. (1996). A markov random field model for classification of multisource satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 34(1):100–113.

Sonobe, R., Tani, H., Wang, X., Kobayashi, N., and Shimamura, H. (2014). Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sensing Letters*, 5(2):157–164.

Sonobe, R., Tani, H., Wang, X., Kobayashi, N., and Shimamura, H. (2015). Discrimination of crop types with TerraSAR-X-derived information. *Physics and Chemistry of the Earth, Parts A/B/C*, 83–84:2–13.

Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, 30(20):5243–5272.

Sutton, C. and McCallum, A. (2006). An introduction to conditional random fields for relational learning. In *Introduction to statistical relational learning*. The MIT Press.

Sutton, C. and McCallum, A. (2011). An introduction to conditional random fields. *Machine Learning*, 4(4):267–373.

Sutton, C., McCallum, A., and Rohanimanesh, K. (2007). Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research*, 8:693–723.

Swain, P. H. (1978). Bayesian Classification in a Time-Varying Environment. *IEEE Transactions on Systems, Man and Cybernetics*, 8(12):879–883.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234–240.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., and Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120:9–24. The Sentinel Missions - New Opportunities for Science.

Tso, B. and Mather, P. M. (1999). Crop discrimination using multi-temporal SAR imagery. *International Journal of Remote Sensing*, 20(12):2443–2460.

Tso, B. and Mather, P. M. (2009). *Classification methods for remotely sensed data*. CRC Press, Boca Raton, 2nd edition.

Tupin, F. (2010). Fusion of Optical and SAR Images. In Soergel, U., editor, *Radar Remote Sensing of Urban Areas*, volume 15 of *Remote Sensing and Digital Image Processing*, pages 133–159. Springer, Netherlands.

Ulaby, F. T. and Bare, J. E. (1979). Look direction modulation function of the radar backscattering coeffieient of agricultural fields. *Photogrammetric Engineering and Remote Sensing*, 45(11):1495–1506.

Ulaby, F. T. and Wilson, E. A. (1985). Microwave Attenuation Properties of Vegetation Canopies. *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(5):746–753.

Ulusoy, I. and Bishop, C. M. (2005). Generative versus discriminative methods for object recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 258–265.

Van der Meer, F., Bakker, W., Scholte, K., Skidmore, A., De Jong, S., Dorresteijn, M., Clevers, J., and Epema, G. (2000). Scaling to the MERIS Resolution: Mapping Accuracy and Spatial Variability. *Geocarto International*, 15(1):39–50.

Wang, L. and Yung, N. H. C. (2015). Improved hierarchical conditional random field model for object segmentation. *Machine Vision and Applications*, 26(7):1027–1043.

Wang, Y., Loe, K., and Wu, J. (2006). A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):279–289.

Webster, R. and Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. Wiley, 2 edition.

Wegner, J. D. (2011). *Detection and height estimation of buildings from SAR and optical images using conditional random fields*. PhD thesis, Leibniz Universität Hannover.

Wegner, J. D., Hänsch, R., Thiele, A., and Soergel, U. (2011a). Building Detection From One Orthophoto and High-Resolution InSAR Data Using Conditional Random Fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1):83–91.

Wegner, J. D., Soergel, U., and Rosenhahn, B. (2011b). Segment-based building detection with conditional random fields. In *Joint Urban Remote Sensing Event (JURSE)*, pages 205–208, Munich, Germany.

Woodhouse, I. H. (2006). *Introduction to microwave remote sensing*. CRC press, Boca Raton, FL.

World-Bank (2011). Food Price Watch. http://www.worldbank.org/foodcrisis/food_price_watch_report_feb2011.html. [Online accessed: 22nd September 2014].

World Food Programme (2013). Hunger Map 2013. http://cdn.wfp.org/hungermap/. (13 May 2016).

Yang, H., Tian, J., Chu, Y., Tang, Q., and Liu, J. (2008). Spatiotemporal Smooth Models for Moving Object Detection. *IEEE Signal Processing Letters*, 15:497–500.

Yin, J., Hu, D. H., and Yang, Q. (2009). Spatio-Temporal Event Detection Using Dynamic Conditional Random Fields. In *International Joint Conference on Artificial Intelligence*, volume 9, pages 1321–1327.

Zhang, L. and Qiang, J. (2008). Segmentation of Video Sequences using Spatial-temporal Conditional Random Fields. In *IEEE Workshop on Motion and video Computing*, pages 1–7.

Zhang, S., Cheng, X., Guo, H., Zhou, L., and Wu, Z. (2014). Tracking deformable parts via dynamic conditional random fields. In *IEEE International Conference on Image Processing (ICIP)*, pages 476–480.

Zhong, P. and Wang, R. (2007a). A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3978–3988.

Zhong, P. and Wang, R. (2007b). Using combination of statistical models and multilevel structural information for detecting urban areas from a single gray-level image. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1469–1482.

Index

A Appendix: Part one

A.1 Complimentary results

A.1.1 Crop type classification in Fuhrberg

Table A.1.: Higher order DCRFs max F1-score confusion matrix.

Map \ Ref.[7]	Barley	Canola	Grassland	Maize	Oat	Potato	Rye	Sugar beet	Wheat	Total	User accuracy
Barley	29521	8	418	58	22	8	10486	0	81	40602	72.71%
Canola	509	44593	860	181	455	1268	1560	214	526	50166	88.89%
Grassland	4951	133	57506	471	400	374	10290	734	924	75783	75.88%
Maize	225	171	1352	22527	175	16097	564	1338	8	42457	53.06%
Oat	184	75	39	53	16453	716	504	132	97	18253	90.14%
Potato	342	658	253	13361	899	54600	1179	9625	75	80992	67.41%
Rye	7732	91	562	86	1045	112	55960	20	8246	73854	75.77%
Sugar beet	42	24	707	291	34	1429	321	41146	6	44000	93.51%
Wheat	2931	9	334	60	35	14	8071	109	29054	40617	71.53%
Total	46437	45762	62031	37088	19518	74618	88935	53318	39017	466724	
Producer accuracy	63.57%	97.45%	92.71%	60.74%	84.30%	73.17%	62.92%	77.17%	74.46%		
OA	75.28%										
Kappa	71.73%										

Table A.2.: DCRF max F1-score confusion matrix.

Map \ Ref.	Barley	Canola	Grassland	Maize	Oat	Potato	Rye	Sugar beet	Wheat	Total	User accuracy
Barley	32350	2	687	10	4	3	10919	26	118	44119	73.32%
Canola	240	44541	536	170	781	1834	852	190	218	49362	90.23%
Grassland	2001	57	53304	6824	110	741	5350	359	209	68955	77.30%
Maize	56	294	201	18473	185	18656	132	1053	7	39057	47.30%
Oat	1283	312	3250	424	17357	1192	2958	1658	957	29391	59.06%
Potato	301	347	1518	10720	373	47154	930	2165	89	63597	74.15%
Rye	6673	159	1415	97	270	225	57259	143	10178	76419	74.93%
Sugar beet	215	45	623	307	378	4744	1005	47637	60	55014	86.59%
Wheat	3318	5	497	63	60	69	9530	87	27181	40810	66.60%
Total	46437	45762	62031	37088	19518	74618	88935	53318	39017	466724	
Producer accuracy	69.66%	97.33%	85.93%	49.81%	88.93%	63.19%	64.38%	89.35%	69.66%		
OA	73.97%										
Kappa	70.34%										

[7] Ref. is short form for ground reference data.

Table A.3.: Confusion matrix of CRFs crop classification using stacked TerraSAR-X multitemporal images.

Map \ Ref.	Barley	Canola	Grassland	Maize	Oat	Potato	Rye	Sugar beet	Wheat	Total	User accuracy
Barley	24179	0	60	0	0	0	7264	27	0	31530	76.69%
Canola	192	44834	220	19	324	119	999	136	266	47109	95.17%
Grassland	2279	454	60732	294	131	301	4409	1036	27	69663	87.18%
Maize	0	71	15	6409	65	20944	29	1587	0	29120	22.01%
Oat	0	0	0	0	18404	0	0	191	0	18595	98.97%
Potato	219	248	176	21471	279	53037	508	27852	35	103825	51.08%
Rye	16721	135	329	61	315	38	59892	78	10908	88477	67.69%
Sugar beet	51	0	0	8692	0	156	122	22410	0	31431	71.30%
Wheat	2796	20	499	142	0	23	15712	1	27781	46974	59.14%
Total	46437	45762	62031	37088	19518	74618	88935	53318	39017	466724	
Producer accuracy	52.07%	97.97%	97.91%	17.28%	9429.24%	71.08%	67.34%	42.03%	71.20%		
OA	68.07%										
Kappa	63.26%										

Table A.4.: Confusion matrix of MLC crop mapping using stacked TerraSAR-X multitemporal images.

Map \ Ref.	Barley	Canola	Grassland	Maize	Oats	Potato	Rye	Sugar beet	Wheat	Total	User accuracy
Barley	25747	1	95	3	1	2	10958	64	114	36985	69.61%
Canola	132	44455	414	70	160	235	839	233	341	46879	94.83%
Grassland	1505	398	58695	310	148	275	3928	696	66	66021	88.90%
Maize	29	70	991	9366	478	13968	90	2345	0	27337	34.26%
Oats	115	1	2	5	8031	3	387	52	854	9450	84.98%
Potato	336	668	523	19096	3707	58732	595	26813	89	110559	53.12%
Rye	15175	161	840	211	6989	161	63190	359	13458	100544	62.85%
Sugar beet	18	0	39	8014	0	1213	189	22753	0	32226	70.60%
Wheat	3380	8	432	13	4	29	8759	3	24095	36723	65.61%
Total	46437	45762	62031	37088	19518	74618	88935	53318	39017	466724	
Producer accuracy	55.45%	97.14%	94.62%	25.25%	41.15%	78.71%	71.05%	42.67%	61.76%		
OA	67.51%										
Kappa	62.42%										

Figure A.1.: Crop map of Fuhrberg, Germany, as classified by 5^{th} order DCRFs Max F1-score ensemble method. Crop legend adopted from Ebinger (2012).

Figure A.2.: Crop map of Fuhrberg, Germany, as classified by standard first order DCRFs Max F1-score ensemble method. Crop legend adopted from Ebinger (2012).

Figure A.3.: Crop map of Fuhrberg, Germany, as classified by MLC-stack method. Crop legend adopted from Ebinger (2012).

A.1.2 Crop type classification in Kitale

Table A.5.: Confusion matrix of DCRF max F1-score classification in Kitale.

Map \ Ref.	Coffee	Grassland	Maize	Rose flower	Sugarcane	Wheat	Total	User accuracy
Coffee	6903	266	5422	11	551	202	13355	51.69%
Grassland	88	19798	5133	24	103	1746	26892	73.62%
Maize	45	467	190925	0	3359	81	194877	97.97%
Rose flower	31	75	1551	1677	94	33	3461	48.45%
Sugarcane	39	73	1906	4	9737	482	12241	79.54%
Wheat	973	511	1611	332	597	10419	14443	72.14%
Total	8079	21190	206548	2048	14441	12963	265269	
Producer accuracy	85.44%	93.43%	92.44%	81.88%	67.43%	80.37%		
OA	90.27%							
Kappa	76.45%							

Table A.6.: Confusion matrix of MLC-stack classification in Kitale

Map \ Ref.	Coffee	Grassland	Maize	Rose flower	Sugarcane	Wheat	Total	User accuracy
Coffee	5235	2456	28445	215	897	1156	38404	13.63%
Grassland	554	15658	6051	141	179	1989	24572	63.72%
Maize	1069	513	163531	13	2069	41	167236	97.78%
Rose flower	149	612	1926	1433	96	83	4299	33.33%
Sugarcane	324	180	4247	29	10264	909	15953	64.34%
Wheat	748	1771	2348	217	936	8785	14805	59.34%
Total	8079	21190	206548	2048	14441	12963	265269	
Producer accuracy	64.80%	73.89%	79.17%	69.97%	71.08%	67.77%		
OA	77.24%							
Kappa	53.67%							

A.1.3 Crop type mapping using Sentinel 1 images

Table A.7.: Sentinel 1 DCRF max F1-score confusion matrix using category I training data.

Map \ Ref.	Summer barley	Winter barley	Canola	Grassland	Maize	Potato	Rye	Sugar beet	Wheat	Total	User accuracy
Summer barley	1326	1	1	14	14	1	20	4	281	1662	79.78%
Winter barley	0	2545	0	0	46	2	3	31	40	2667	95.43%
Canola	0	48	5691	2	6	0	0	81	1	5829	97.63%
Grassland	64	1	1	4824	64	2	50	9	111	5126	94.11%
Maize	9	4	12	122	8352	71	115	385	206	9276	90.04%
Potato	8	37	5	20	28	3587	59	252	62	4058	88.39%
Rye	233	87	21	109	190	61	8704	201	509	10115	86.05%
Sugar beet	10	11	15	41	132	179	37	11611	240	12276	94.58%
Wheat	0	1	16	3	78	0	19	80	12287	12484	98.42%
Total	1650	2735	5762	5135	8910	3903	9007	12654	13737	63493	
Producer accuracy	80.36%	93.05%	98.77%	93.94%	93.74%	91.90%	96.64%	91.76%	89.44%		
OA	92.81%										
Kappa	91.58%										

Table A.8.: Sentinel 1 MLC-stack confusion matrix using category I training data.

Class	Summer barley	Winter barley	Canola	Grassland	Maize	Potato	Rye	Sugar beet	Wheat	Total	User accuracy
Summer barley	381	0	1	0	0	1	6	0	6	395	96.46%
Winter barley	0	1964	8	0	12	0	1	45	20	2050	95.80%
Canola	3	42	5653	0	0	1	12	98	4	5813	97.25%
Grassland	171	61	32	4928	273	23	367	110	266	6231	79.09%
Maize	28	49	5	144	8272	460	128	487	178	9751	84.83%
Potato	6	32	2	23	76	2995	15	188	20	3357	89.22%
Rye	137	428	41	32	94	36	7361	42	415	8586	85.73%
Sugar beet	2	20	1	0	96	383	1	11636	144	12283	94.73%
Wheat	922	139	19	8	87	4	1116	48	12684	15027	84.41%
Total	1650	2735	5762	5135	8910	3903	9007	12654	13737	63493	
Producer accuracy	23.09%	71.81%	98.11%	95.97%	92.84%	76.74%	81.73%	91.96%	92.33%		
OA	88.00%										
Kappa	85.86%										