Mario Gilcher

vom Fachbereich VI
(Geographie/Geowissenschaften)
der Universität Trier
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

Remote Sensing Based Crop Classification of Maize

Improving Model Robustness in State-of-the-Art Machine Learning Models

Betreuender:

Univ.-Prof. Dr. Thomas Udelhoven

Berichterstattende:

Univ.-Prof. Dr. Thomas Udelhoven

Apl.-Prof. Dr. Christoph Emmerling

Dr. Gilles Rock

Datum der wissenschaftlichen Aussprache:

12.05.2022

Trier, 2022

Acknowledgements

Writing these final words of my thesis, I would like to dedicate my thanks first and foremost to my daughter Annouk. When you were born almost 18 years ago, I was not on a path to a career in geosciences. I could not write a single line of python code. I had no idea who Humboldt was, Mercator or Tobler. I had neither Bachelor's nor Master's degree; in fact, I was not even qualified to attend University. While I was holding you on my belly, I made the decision to get my Abitur, and while we went to the playground together every day, I was preparing myself for the exam. In my Bachelor and Master years, you filled my heart with joy every time I came home from the University and I could hear you getting excited from the first floor. You accompanied me almost every single day on my way to work at the University until the pandemic started. You were my first and cutest student, and you are my most honest teacher. Without you, not a single word in this thesis would have been written, so I thank you now and forever. You and your brother are everything to me.

I would like to thank my supervisor, Prof. Dr. Thomas Udelhoven, for giving me this unique opportunity. You believed in me relentlessly while I needed, probably more so than other PhD students, time to contemplate and get things on paper. At the same time, you were always there with technical and methodical advice when necessary. Furthermore, I would like to thank Dr. Thorsten Ruf and Prof. Dr. Christoph Emmerling for contributing to the core ideas of this thesis in a major way, and also for being great coauthors.

Furthermore, I would like to thank my always supportive colleagues for being genuinely amazing people. In no particular order:

- Dr. Henning Buddenbaum for proofreading and being the kind of neighbor everybody wishes for. No matter the problem you have, Henning has the right tool in his shed or a smart suggestion in his head.
- Dr. Achim Röder and Dr. Johannes Stoffels for being fantastic teachers while I was still a student, and great colleagues after I joined the department.
- Nicole Gellner-Aschhoff for being much more than a good secretary.
- JProf. Dr. David Frantz and Dr. Sebastian Mader for answering countless technical questions about statistics and atmospheric correction.
- Dr. Gilles Rock and Dr. Rebecca Retzlaff for joining me on the field, during the early days of my PhD.
- Sascha Nink, Melanie Brauchler, Dr. Max Gerhards, Asmaa Abdelbaki and many other colleagues for joining me on daily trips to the Mensa and listening to my constant complaints about the vegetarian options.

This has not been the quickest of PhDs, therefore I also want to thank all the colleagues I have forgotten to mention here. It was a great time, and I have not met a single colleague who was less than extraordinarily supportive. Thusly, I am beyond thankful for the time I have been sharing with you in this department.

Summary

Agricultural monitoring is necessary. Since the beginning of the Holocene, human agricultural practices have been shaping the face of the earth, and today around one third of the ice-free land mass consists of cropland and pastures. While agriculture is necessary for our survival, the intensity has caused many negative externalities, such as enormous freshwater consumption, the loss of forests and biodiversity, greenhouse gas emissions as well as soil erosion and degradation.

Some of these externalities can potentially be ameliorated by careful allocation of crops and cropping practices, while at the same time the state of these crops has to be monitored in order to assess food security. Modern day satellite-based earth observation can be an adequate tool to quantify abundance of crop types, i.e., produce spatially explicit crop type maps. The resources to do so, in terms of input data, reference data and classification algorithms have been constantly improving over the past 60 years, and we live now in a time where fully operational satellites produce freely available imagery with often less than monthly revisit times at high spatial resolution. At the same time, classification models have been constantly evolving from distribution based statistical algorithms, over machine learning to the now ubiquitous deep learning.

In this environment, we used an explorative approach to advance the state of the art of crop classification. We conducted regional case studies, focused on the study region of the Eifelkreis Bitburg-Prüm, aiming to develop validated crop classification toolchains. Because of their unique role in the regional agricultural system and because of their specific phenologic characteristics we focused solely on maize fields.

In the first case study, we generated reference data for the years 2009 and 2016 in the study region by drawing polygons based on high resolution aerial imagery, and used these in conjunction with RapidEye imagery to produce high resolution maize maps with a random forest classifier and a gaussian blur filter. We were able to highlight the importance of careful residual analysis, especially in terms of autocorrelation. As an end result, we were able to prove that, in spite of the severe limitations introduced by the restricted acquisition windows due to cloud coverage, high quality maps could be produced for two years, and the regional development of maize cultivation could be quantified.

In the second case study, we used these spatially explicit datasets to link the expansion of biogas producing units with the extended maize cultivation in the area. In a next step, we overlayed the maize maps with soil and slope rasters in order to assess spatially explicit risks of soil compaction and erosion. Thus, we were able to highlight the potential role of remote sensing-based crop type classification in environmental protection, by producing maps of potential soil hazards, which can be used by local stakeholders to reallocate certain crop types to locations with less associated risk.

In our third case study, we used Sentinel-1 data as input imagery, and official statistical records as maize reference data, and were able to produce consistent modeling input data for four consecutive years. Using these datasets, we could train and validate different models in spatially

and temporally independent random subsets, with the goal of assessing model transferability. We were able to show that state-of-the-art deep learning models such as UNET performed significantly superior to conventional models like random forests, if the model was validated in a different year or a different regional subset. We highlighted and discussed the implications on modeling robustness, and the potential usefulness of deep learning models in building fully operational global crop classification models.

We were able to conclude that the first major barrier for global classification models is the reference data. Since most research in this area is still conducted with local field surveys, and only few countries have access to official agricultural records, more global cooperation is necessary to build harmonized and regionally stratified datasets. The second major barrier is the classification algorithm. While a lot of progress has been made in this area, the current trend of many appearing new types of deep learning models shows great promise, but has not yet consolidated. There is still a lot of research necessary, to determine which models perform the best and most robust, and are at the same time transparent and usable by non-experts such that they can be applied and used effortlessly by local and global stakeholders.

Table of Contents

List of Abbreviations

AFOLU	Agriculture, Forestry and Other Land Use
API	Application Programming Interface
BPU	Biogas Producing Units
CDL	Crop Data Layer
CLC	Corine Land Cover
CM	Confusion Matrix
CNN	Convolutional Neural Networks
DEM	Digital Elevation Model
ECHO	Extraction and Classification of Homogenous Objects
EO	Earth Observation
ESA	European Space Agency
FCNN	Fully Convolutional Neural Network
GDP	Gross Domestic Product
GEE	Google Earth Engine
GEOBIA	Geographic Object-Based Image Analysis
GLM	Generalized Linear Model
GRD	Ground Range Detected
GSD	Ground Sampling Distance
JECAM	Joint Experiment for Crop Assessment and Monitoring
LPIS	Land Parcel Information System
LSTM	Long Short-Term Memory
LULCC	Land Use/Land Cover Change
MAP	Mean Annual Precipitation
MAT	Mean Annual Temperature
ML	Maximum Likelihood
MLC	Maximum Likelihood Classifier
MODIS	Moderate Resolution Imaging Spectroradiometer
MSS	Multispectral Scanner
NDVI	Normalized Difference Vegetation Index
NN	Neural Network
OA	Overall Accuracy
OLI	Operational Land Imager
OOB	Out of Bag
PR	Precision
RC	Recall
RESA	RapidEye Science Archive
RF	Random Forests
RK	Regression Kriging
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SD	Standard Deviation
SVM	Support Vector Machines
USLE	Universal Soil Loss Equation
VNIR	Visible and Near Infrared

1 Introduction

1.1 The Agricultural Sector

1.1.1 Global Dynamics of the Agricultural Sector

Human societies began to transform the landscapes they inhabited millennia before the advent of concrete, cars and combine harvesters. There is archaeological and palaeoecological evidence of forest cleanings, megafaunal extinctions and human induced soil erosion at least since the late Pleistocene (Ellis et al. 2013). By the beginning of the Holocene, 12,000 years ago, transformative land use practices like cultivation and domestication were adopted by all human societies (Ellis et al. 2021). Later on, in more advanced civilizations such as the Roman cultures, the idea of the noble human transforming wild and unorganized natural landscapes into cleanly structured lands of plenty evolved. It established itself deeply within the European ideology and was consequently spread around the globe through the paths of imperialism (Kirch 2005).

The transformation of landscapes has continued ever since, albeit with greatly varying speed in different areas of the world. In the last 3 centuries, wildlands with little or no human influence receded from around 50% of the ice-free global landmass to around 25%. As a result, the majority of earth's biosphere is now covered by croplands, rangelands and settlements (Ellis, Beusen, and Goldewijk 2020). In the year 2000, the global amount of cropland area is estimated to be around 15 million km^2, which accounts for 12% of the ice-free land area. Pastures and rangelands are even more land-intensive at 28 million km^2 and about 22% of global land mass (Ramankutty et al. 2008). Between 1985 and 2005 alone, the global agricultural area increased by 1.53 million km^2 (Foley et al. 2011).

1.1.2 Negative Externalities of the Agricultural Sector

Landscape transformations on this scale do not occur without a cost. Croplands, pastures and settlements are not simply added to our ecosystem, they are replacing preceding land cover types. This is usually accompanied by a trade-off of ecosystem services and natural resources, where forest production and habitat biodiversity is sacrificed for the sake of yield maximization (Foley et al. 2005). The issues that arise on a global scale are manifold. Some non-exhaustive examples are:

Freshwater resources are consumed and transformed irreversibly. Agriculture withdraws around 3,800 km^3 of freshwater per year, which is 76% of the total human consumption. Yet, the global water cycle is not only altered by directly removing resources from the system. Rivers and lakes are also changed by chemical inputs, temperature and riverbed alterations as well as harvesting and agriculture (Carpenter, Stanley, and Vander Zanden 2011)

Forests are especially impacted for various reasons. Firstly, they are, as a natural land cover type, one of the major targets of land cover transitions and often replaced by other cover types such as agriculture. Secondly, other cover types depend on forest resources such as timber. As a result, agricultural activities and timber production have caused an estimated loss of 7 to 11 million km^2 of forests in the past 300 years (Foley et al. 2005). From 2000 to 2012 alone, 2.3 million km^2 of

forests were cleared, while only 0.8 million km² were gained. This is especially problematic in the tropics, where the forest loss is still increasing by around 2000 km² per year (Hansen et al. 2013).

Global rapid loss of biodiversity has been widely reported, with species extinction rates at least 10 times higher than the average during the past 10 million years (IPBES 2019). Though this is driven by various other human factors like climate change, pollution and direct exploitation, Land cover changes are a major factor as well. One example are oil palm plantations, which are considered one of the largest contributors to global deforestation. The land cover transformation from rain forest to oil palm plantation reduces the animal density and diversity by roughly 50% (Barnes et al. 2014).

Changes in land use also have a large impact on **climate change**. Between 1750 and 2011, changes in land use released between 100 and 240 Petagrams of carbon into the atmosphere, mainly due to deforestation. It is therefore the second largest contributor to the global Greenhouse Gas concentration behind fossil fuel combustion (Ciais et al. 2013). In 2010, Agriculture, Forestry and Other Land Use (AFOLU) accounted for 24% of all global greenhouse gas emissions, roughly 11.76 billion tons of CO_2 equivalent (Parry 2015).

Soil erosion and degradation as a direct consequence of agricultural land use has been causing the decline of entire civilizations since at least 4000 years, when salt and silt accumulated as a side effect of large scale irrigation systems in what is now the Iraq (Thorkild 1958). Contemporarily, many different physical, chemical and organic threats to soil ecosystem functions are described and analyzed in global frameworks (FAO & ITPS 2015). In 2012 alone, an estimated 35.9 billion tons of soil were lost to erosion globally (Borrelli et al. 2017). Though soil erosion rates are generally very high on croplands compared to other land cover types, it can be influenced significantly by crop management and environmental protection policies (Wuepper, Borrelli, and Finger 2020)

1.1.3   Feeding the World

The previously described externalities cannot be completely avoided. We need food to survive. Of the roughly 2900 calories the average human consumes each day in 2017, only about 1.2% are fish and seafood. The remaining 98.8% are produced by agricultural systems, with cereals alone accounting for 45%. Agriculture, while economically rather insignificant at only 4% of the global GDP, employs 27% of the global workforce, and feeds every single human on earth (FAO 2020). Though the monetary value of the sector is relatively low, while the percentage of people employed by it is shrinking, human society cannot exist without planetary scale agriculture and gradual changes in food availability can cause largescale political instability (Soffiantini 2020).

The precise extent to which our agricultural systems are able to support us, has been subject of heated academic discussions for a very long time. The World population conference in Geneva in 1927 (List et al. 1995) started to coin the term overpopulation which was later often instrumentalized. At the time, world population was about 2 billion, long before the works of Malthus and the Club of Rome. In 2020, two or three years from when the world population will hit 8 billion (United Nations, Department of Economic and Social Affairs 2019), 768 millions of

people are undernourished. While the absolute and relative number of undernourished people have been in a steady decline since 2005, both numbers started to plateau after 2010 around 600 million or 8 % of the global population. After a slow increase in the past decade, 2020 saw a very sharp increase of this number of more than 100 million undernourished people (Food and Agriculture Organization of the United Nations 2021). While COVID-19 played a huge role by causing market disruptions and economic decline in households (Laborde et al. 2020), the key drivers behind the increase of undernourished people are climate extremes, conflicts and economic downturns which often occur in combination. Especially the increasing amounts of severe droughts in many countries in Africa have caused changes in trade flows from export to import, which triggered economic decline in already low-income countries, which in turn results in a decline in food security.

1.1.4   Agricultural Monitoring and Earth Observation

The previous paragraphs can be summarized as one big positive and one big negative aspect of planetary scale agriculture: Feeding the world and at the same time degrading its ecosystems. Striking a sustainable balance between these two does not come naturally. Options that benefit the community as a whole might at the same time not be the easiest and most profitable for the individual farmer. This phenomenon is called "Tragedy of the Commons" (Hardin 1968), a special case of the Prisoners Dilemma where informed individuals could theoretically produce the optimal outcome if they cooperated. Practically, since there is no communication, individuals don't act altruistically, and the consequence is the worst possible outcome for everybody: reduction and eventual depletion of natural habitats as direct consequence of overuse and exploitation.

In the following manuscript, we will argue that earth observation (EO) through remote sensing (RS) in general, and land use/ land cover change (LULCC) classification in particular, can be an essential part of the solution to this dilemma. The aforementioned externalities of LULCC can be quantified, and risks can be allocated in spatial context through careful, consequent and global monitoring (Estoque 2020). At the same time we can observe the productivity of our agricultural systems to adjust to the consequences of climate change, most notably the increase of severe drought events.

This is only possible because of a constant increase in quality of RS products over the last decades, as a consequence of better algorithms, better sensors, and better data availability. The improvements that enabled LULCC monitoring on the level which we are used today will be described in the following section.

1.2   Advancements in Remote Sensing based Agricultural Land Use Monitoring

1.2.1   Advancements in Satellite Remote Sensing Hardware

1.2.1.1   Advancements in passive Satellite Remote Sensing

Considering Satellite Imagery, we can simplify sensor characteristics in trade-offs between sensor resolution (often described as spatial, spectral and temporal) and data availability. High spatial resolution was technically possible since the early days of the space age, since analogue film was

used to take the images, which were then physically dropped and recovered. The CORONA project, conducted by the US under the Keyhole Program between 1960 and 1972 is one of the first imaging satellite mission (Philip et al. 2002). Spatial resolutions were as high as 2m, which is relatively high even by modern day standards, but availability was very limited in many ways. Since these were military missions, mostly images from certain areas in Russia, China and the Middle East were acquired. Additionally, the images were only declassified in the nineties.

The next step into satellite-based earth observation was the Landsat program. In its now almost 50-year history, it perfectly represents the advancements achieved in Satellite Remote Sensing in a nutshell:

- **Spatial resolution** has increased quickly from about 80m on the Multispectral Scanner (MSS) to 30m on the Thematic Mapper (TM) (Wulder et al. 2019). From there it has plateaued, and it is still the resolution of most bands in the Operational Land Imager (OLI) of Landsat 8 and Landsat 9 (NASA 2019).
- **Spectral resolution** has increased as more bands have been added and the configuration has changed since the original MSS.
- **Temporal Resolution** has started out with an 18-day revisit cycle, which changed to 16 days on Landsat 4. Since the orbit of Landsat 9 is similar to Landsat 8, but with an 8-day offset, the revisit cycle will be effectively cut in half, since the sensors are very similar.
- **Data availability** massively increased. After a severe price reduction from around 4000 USD in the 80s to about 600 USD per scene in the 90s, a new distribution policy made the entire Landsat archive as well as all newly acquired images available for free starting in January 2008 (Woodcock et al. 2008).

In contrast to wide swath operational satellite missions that acquire images continuously, there are commercial satellites and satellite constellations that acquire images on a much narrower swath, but with a much higher spatial resolution. One of the first examples for commercial remote sensing satellites is the US based IKONOS, which launched in 1999 and had a multispectral sensor with a 4m resolution (1m panchromatic) (Ünsalan and Boyer 2011). In the last 20 years many similar satellites were launched by different countries, with multispectral resolution in the range of single digit meters, and panchromatic resolutions often around or below one meter, with WorldView-3 even as low as 0.3 m (Toth and Jóźków 2016). Theoretical revisit times are usually very low, and some satellites can acquire multiple images per day from a single site. Since swath widths are so narrow, sometimes as low as 2 km (SkySat), specific orders have to be placed in advance to make use of these high revisit times. While these orders usually cost money, using already acquired and archived images from these commercial programs is sometimes free for scientific use (Borg 2014). As a consequence, the data availability for these kinds of satellites is slightly misleading. While revisit times are theoretically very high, they usually cannot be used for intensive time series analysis, unless the potentially very costly data acquisition is planned many years in advance.

On the opposite end of the swath width spectrum are sensors like the Moderate Resolution Imaging Spectroradiometer (MODIS) which is on board of the Terra (1999) and Aqua (2002) Satellites (Wenxue, Jianwen, and Pei 2020). While the spatial resolution is very low (1 km in most bands), it provides 36 spectral bands ranging from 459 up to 14385 nm. But its most prominent feature is a uniquely high data availability, as a result of the very high swath width of 2300 km, and a revisit time of two observations per day. As a consequence, it provides a freely available very dense time series of highly refined earth observation products like surface temperature and vegetation indices. There are other sensors with a similar focus on high swath and low spatial resolution, like AVHRR with a 1km pixel size and 3000 km swath (cf. (Loveland et al. 2000)). They often have a focus on meteorology, where low revisit times and the coverage of large areas are more important than high spatial detail. Nonetheless, these sensors can still be used for large scale vegetation monitoring, especially where long term time series are needed.

Apart from MODIS, there have been other platforms exploring the use of an increased number of bands with a decreased average bandwidth, starting with Hyperion, launched in 2000 (Vangi et al. 2021). The sensor had 196 bands with a 10 nm bandwidth and was operating until 2017. For more than 10 years, since its launch it had been the only operating spaceborne hyperspectral imager covering the spectral range from 400 to 2500 nm. Even today there are only very few hyperspectral sensors operating (e.g., the Italian PRISMA (Labate et al. 2009), the Japanese HISUI (Matsunaga et al. 2017)), and the characteristics in terms of spatial resolution (around 30 m), swath width (around 30 km) are comparable. In comparison to multispectral satellite remote sensing, the evolution of hyperspectral sensors has been rather slow, mainly due to technical challenges and low signal to noise ratios (Transon et al. 2018).

In 2015 and 2017, Sentinel 2A and B were launched by ESA, marking the beginning of a new era in global land surface monitoring. Their 13 bands, with resolutions ranging from 10 m to 60 m, were specifically designed to guarantee continuity with Landsat imagery, while their 290 km swath and twin constellation enables a revisit time of 5 days for most areas around the globe (Drusch et al. 2012). The sensor is specifically designed with LULCC classification in mind, and many studies with this topic were published since its launch (Phiri et al. 2019). The low revisit time specifically enables the use of dense vegetation index time series, which give a detailed perspective on the phenology of the vegetation cover, which can then be used to use state of the art classification algorithms (Campos-Taberner et al. 2020). The sensor is the current state of the art for production of large-scale land cover mapping, as a direct consequence of its high-quality sensor characteristics, low revisit time and free data access under the Copernicus program (ESA 2014).

1.2.1.2 *Advancements in active Satellite Remote Sensing*

Active sensors in the microwave ranges above 1mm wavelength use synthetic apertures (SAR) in order to detect signals that were sent by the instrument and reflected back by ground objects (Toth and Jóźków 2016). While surface reflections in the visible and near infrared (VNIR) domains are determined mostly by the biochemical composition of objects (e.g., chlorophyll), the long waves in the microwave domain interact with objects in complex ways that depend heavily

on the relative length of the wave and the object geometry. As a consequence, SAR sensors are classified in subdomains called bands, with very different properties concerning the objects they interact with and are penetrating and reflecting from, and thusly very different application possibilities. They have letter designations, and range from wavelengths of 1cm up to 1m. In general the bands are following the pattern of short wave, low penetration, high spatial resolution on one end, to long wave, high penetration and low spatial resolution on the other (NASA 2020). The most popular bands in spaceborne SAR are the L band (15 – 30 cm), the C band (3.8 – 7.5 cm) and the X band (2.4 – 3.8 cm). The L band was historically used in the first spaceborne SAR sensors, starting with SEASAT (USA) in 1978 (Gade 2015). It can be used for large scale geophysical mapping, and mostly penetrates vegetation cover. As a consequence it is especially suitable to observe through the canopy, for example, mapping floods covered by standing vegetation (Tsyganskaya et al. 2018). C and X band sensors are the most common SAR types, with the European Space Agencies (ESA) European Remote Sensing Satellite (ERS) starting in 1992, Canada's Radarsat in 1995, Germany's TanDEM-X in 2010 and again ESA's Sentinel-1 (Tsai et al. 2019) in 2014. The penetration capabilities within these bands are moderate, and the applications in vegetation monitoring largely depend on the vegetation density (Tsyganskaya et al. 2018).

With optical remote sensing, only one very well understood parameter is derived for a given wavelength, the reflectance which is ratio of incoming versus reflected energy. In SAR remote sensing, there are many more factors to consider. Within a given wavelength there are multiple polarization modes, which describe the horizontal or vertical polarization of the incoming and receiving signals. The results are unitless backscattering coefficients with different polarization, the amplitudes of the signals, influenced by sensor characteristics like wavelength, polarization and incidence angle, as well as surface characteristics like roughness and dielectric properties.

Advanced processing techniques like interferometry and polarimetry are able process the backscattering amplitude and phase into more useful information. Interferometry uses the phase shift in radar signals to produce very high resolution surface models, and can be used to detect even small changes of ground height after earthquakes and landslides (Xu, Xu, and Wen 2018). Polarimetric SAR uses different decomposition techniques to analyze the scattering matrix of SAR images, and derives parameters such as the entropy, which is a measure that can be directly interpreted as the randomness of the scattering (Moreira et al. 2013). These parameters are much more interpretable, and correlate well with other crop phenology parameters (Prudente et al. 2019). They can be used in further classification and modeling steps, as well as inversion of crop growth models.

1.2.2 Advancements in Satellite Remote Sensing Software

Progress in the availability of satellite remote sensing imagery has not only happened on the hardware side. Since the beginning of the 21st century, services like GloVis (Houska and Johnson, A.P. 2012) and the Earth Explorer (USGS 2013) by the US Geological Survey (USGS) provide access to satellite imagery. They feature visualizations of imagery and footprints as well as search queries based on location, sensor and time frame. Based on these search queries, access URLs can

be produced, and the low-level satellite data products can then be downloaded and further processed locally.

Today, the functionality of online platforms has completely shifted from purely browsing datasets, to infrastructure offering tools for all parts of the remote sensing toolchain, ranging from dataset selection, data preprocessing and processing, algorithm application, analysis, and finally visualization, presentation and dissemination. These are private and proprietary platforms, that offer free as well as paid services. The Sentinel Hub for example, allows free browsing and download of Sentinel, Landsat and MODIS datasets. In addition, paid access enables RESTful API and more advanced data processing (Gomes, Queiroz, and Ferreira 2020). The Google Earth Engine (GEE) offers its complete functionality for free, though commercial licenses exist to increase processing capacities. It offers a complete cloud computing package on top of a massive data catalogue, all of which is freely accessible through a JavaScript code editor (Gorelick et al. 2017). Besides these spatial service platforms that provide solutions for the complete remote sensing toolchain, it is also now much easier to integrate data access in existing toolchains through APIs, packages and frameworks. Through open source packages available in the most popular data science programming languages R and python, data from freely available sensors like MODIS, Landsat and Sentinel-1/-2 can be queried and downloaded directly from scripts, to integrate seamlessly into any existing program (Murray et al. 2010; Ranghetti et al. 2020). As a consequence, data access to the most popular modern sensors is now easier than ever, with many different options integrated into accessible online platforms as well as powerful programming language APIs.

Besides the literal access of data, data preprocessing has historically been a major barrier in remote sensing. It kept non-remote sensing scientists from using remote sensing data, since it is not always clear which preprocessing level is adequate. It also usually involves many different steps accounting for sensor, atmospheric and topographic effects, and thusly introduces lots of potential for errors even before any analysis algorithm is applied (Young et al. 2017). Here the user now has many different options. Firstly, a lot of platforms like the GEE offer the datasets in different processing levels, ranging from raw level 1 data, to atmospherically corrected surface reflectances with cloud masks. Secondly, high quality preprocessing frameworks like FORCE (Frantz 2019) or SEN2COR (Louis et al. 2016) are freely available and can be included seamlessly in any kind of satellite data analysis.

1.2.3   Advancements in Reference Data

High quality and high sample size reference data is crucial for every supervised modeling approach in general. Remote sensing classification models in particular need georeferenced in situ data that contains either point or polygon geometry as well as a class label for a given time stamp. Historically a lot of this data was generated with field level surveys, meaning that the researchers had to get out by car, go to the sites and record geometry as well as labels themselves. This approach is very expensive, and also means that the reference data is clustered in one area defined by the driving range. Another downside is that researchers can only get reference data starting from the year the research starts, since they cannot retroactively see what was on a given

field a year ago. A cheaper alternative that also deals with this problem is to directly approach farmers and ask permission to use their personal records, since they often have to keep track of their parcels and respective crops. This bears the risk that reference data is highly spatially clustered or correlated in other ways, since farmers often use similar crop types or are equal in other farming practices.

While large scale, balanced reference datasets remain a major challenge in machine learning, the advent of big data brought a lot of advancements in this area. In the last decades, many different remote sensing-based global land cover products have been developed, starting with the GLC2000 product, a global 1km by 1km map based on the Spot Vegetation sensor (Bartholomé and Belward 2005). More recent, products include ESAs Copernicus Land Cover Layers, with a much better spatial resolution at 100 m, and yearly availability (Buchhorn et al. 2020). Most of these global products use the FAOs classification scheme with 22 land cover classes (Pérez-Hoyos et al. 2017). For the purpose of agricultural monitoring, these classes are too broad, and much more specific thematic detail is necessary, especially the crop type (Tsendbazar, de Bruin, and Herold 2015). With the massive spread of machine learning the need for large scale training and benchmarking data intensified, and global labeled datasets are now available. BigEarthNet (cf. (Sumbul et al. 2019)) for example offers 590,326 Sentinel-2 sample patches, with CORINE Land Cover (CLC) based labels. These classes offer some information about the crop, and distinguish certain very unique crop types such as olives and rice, but other crops are just summarized under non-irrigated and irrigated arable land (Copernicus Land Monitoring Service 2010). CropHarvest (see (Tseng et al. 2021)) is another set of labeled Sentinel-1 and -2 products. Here the focus is solely agriculture, and 348 granular labels from around the globe were used to specify the type of crop or pasture in a detailed way. The reference data was collected from regional and global datasets with public licenses. The labels could not be perfectly harmonized between the regions, and some regions are not covered by the highly granular classes, e.g., Australia, East Asia, or only sparsely covered like Europe and South America. (Laso Bayas et al. 2017) used a crowd-based approach to generate 36000 labeled sample geometries. Though most of the 80 voluntary participants were from North America, Europe and India, they were able to produce well distributed samples from all around the globe.

The vast majority of referenced data for detailed agricultural monitoring exists on national or regional level. In the European Union, accurate field level records have to exist as part of the Common Agricultural Policy. Farmers have to report to the Land Parcel Identification System (LPIS), which offers a spatially and temporally dense dataset of crop types (Bailly et al. 2018; Giordano et al. 2018). While LPIS data is not generally available to the public, the National Agricultural Statistics Service in the United States offers crop type data records ranging back to 1997 (NASS 2009) which can be accessed online and used as reference for crop modeling (Csillik et al. 2019). An alternative to statistical records is to use data that is independent from the remote sensing imagery, but still gives information about the ground cover and crop type. One example is the evaluation of street level imagery. Platforms like Mapillary (d'Andrimont et al. 2018) provide a dense database with geotagged and timestamped ground level imagery, which can then

be used to identify land use or even crop types by means of crowd funding, image based machine learning or manual image classification.

To summarize, remote sensing researchers have many different ways to access field data, compared to the situ surveys that have been done historically. Especially on regional levels in North America and Europe, geometrically and thematically dense training sets are available. Globally, the agricultural structure is very heterogenous, in terms of what is grown and how it is recorded. As such, the available reference data is still sparse, and denser recordings of in situ data will be required in the future. Nonetheless, researchers have various different options to access reference data to train the models. In the next section we will outline how these models have evolved from pixel-based statistical classifiers to deep learning-based object detection.

1.2.4 Advancements in Supervised Classification Algorithms

Historically, Maximum Likelihood Classification (MLC) has been the most popular classification algorithm in Remote Sensing. It is a parametric statistical algorithm, which means it is based on the assumption of a particular distribution for the values of each band and a given class, which is usually the gaussian distribution (Sisodia, Tiwari, and Kumar 2014). Based on this multivariate distribution, a given pixel takes the most likely class. As one of the first classification methods, this algorithm has been successfully employed for the classification of crops based on satellite imagery since the 1980s (see also (Dawbin and Evans 1988; Ulaby, Li, and Shanmugan 1982; Zenzo et al. 1987)), and still sees frequent use today. In comparison to more modern approaches, it brings a lot of advantages. It is very transparent and the decision-making process can be comprehended with very minimal prior knowledge. The training process is also very simple, and consists solely of the estimation of the distribution parameters, which is just the variable means and the variable covariance matrix. The training of the model also requires no hyperparameters, and thusly requires no expert knowledge by the researcher. The only major drawback of the approach is the strict requirement of a particular probability distribution.

Though the groundwork for machine learning was laid much earlier (Rosenblatt 1958), two major milestone papers were published in 1995. Cortes and Vapnik (Cortes and Vapnik 1995) published their works which led to the Support Vector Machines (SVM) we know today, at the same time as Ho (Ho 1995), which led to the Random Forest (RF) classifier (Breiman 2001). A decade later, these machine learning techniques were beginning to be widely implemented in remote sensing classification research, while they both had their peak between 2013 and 2018 (Sheykhmousa et al. 2020). One reason for this popularity was that the increase in classification performance compared to statistical classifiers. Many studies specifically compared machine learning algorithms like RF and SVM with classical MLC (Mondal et al. 2012; Nitze, Schulthess, and Asche 2012), and in further review studies (Khatami, Mountrakis, and Stehman 2016; Yu et al. 2014) it was proven that this modern approach showed significant improvements compared to Maximum Likelihood. Another advantage of these models is that they do not rely on specific distributions of the predictors, which makes them much more flexible. However, they do require varying degrees of expert knowledge, mainly due to the introduction of hyperparameters. These hyperparameters can be considered as second order model parameters. The first order model

parameters are optimized in model training e.g., the distribution parameters in MLC or the coefficients in a linear model. Second order model parameters have to be selected by the expert building the model. Their influence on model performance, and therefore the effort that needs to be put into optimization varies greatly from model to model. This tunability is generally considered to be lowest in RF models, and highest in SVM models where the kernel function and coefficients have a large influence on model performance (Probst, Boulesteix, and Bischl 2019). Another factor is interpretability of modeling results, which means the degree of understanding a human can have about why a model made a decision in favor of a specific class. In simple machine learning models, like a decision tree, this can be very transparent as the boundaries for a given variable can be accessed directly. In more complex models, such as SVMs, this is not as straightforward, which is why they are also called "black box models". The development of machine learning algorithms in the past has been strongly pointing towards high performance, low interpretability models, which require an increasing amount of expert knowledge to build and train (Sheykhmousa et al. 2020).

Assuming that each pixel is independent from all other pixels is a direct violation of the first law of geography (Tobler 1970). Two major approaches have been pursued to deal with this problem and acknowledge the spatiality of remote sensing raster data. Firstly, we can use a pixel-based classifier to produce an a priori predicted class, and then update the classification based on the immediate context. An early example of this are probability relaxation concepts (cf. (Zenzo et al. 1987)), where class probabilities were updated iteratively based on class compatibilities of the neighborhood. The second major approach is to split the classification into two parts. An image segmentation, and an object classifier. Early implementations of this are for example the Extraction and Classification of Homogenous Objects (ECHO) algorithm, which used an iterative cell merging approach to segment the image, and a maximum likelihood classifier to classify the resulting object. This approach eventually culminated in the Geographic Object-Based Image Analysis (GEOBIA) which rose in popularity around and after 2010 (Hossain and Chen 2019). While many approaches for the image segmentation part exist, the commercial solution eCognition is the most popular, and used in about half of all GEOBIA related publications (see (Blaschke 2010)). After the segmentation, features are extracted and non-spatial classification models are applied, like decision trees (Li et al. 2015) or SVMs (Peña et al. 2014).

Convolutional Neural Networks (CNN) solve the problem of spatiality in a very direct way, by incorporating neighborhood information with kernel convolutions on nested spatial levels (Chen et al. 2018). They were originally developed by Yann Lecun in the Nineties (cf. (Lecun and Bengio 1995), but it took over a decade until the computational capabilities were enough to develop the complex models with millions of parameters (e.g. (Krizhevsky, Sutskever, and Hinton 2012)) that are everywhere today (Ma et al. 2019). CNNs are only one of the many factors that enabled the sudden rise of popularity of neural networks at the end of the last decade. Techniques like perceptrons (Rosenblatt 1958) and backpropagation (Rumelhart, Hinton, and Williams 1986) were around for a long time, but it took until the availability of large scale labeled training datasets like ImageNet (Deng, J. and Dong, W. and Socher, R. and Li, L.-J. and Li, K. and Fei-Fei

2009) coincided with the advancement of modern Graphics Processing Units (GPUs) in personal computers, for deep learning to get widely accepted as the toolset of the future. Today, many well established deep learning-based approaches exist, specifically in the case of agricultural monitoring and land use mapping. The most suitable deep learning approach for land use mapping is semantic segmentation, where each pixel of a given input image or extracted patch-like array is assigned a class or label, which in this case could be a crop or land use type. Image segmentation is the most common approach in vegetation monitoring related remote sensing literature. The most popular algorithms have an encoding part of the model, which takes the input data and gradually transforms it with convolutional layers, and a decoding part, which takes the previous steps in the other direction and fuses them together to a fully connected output layer with the desired class probabilities for each pixel (Zhang, Tang, and Zhao 2021). Examples include SegNet (Badrinarayanan, Kendall, and Cipolla 2017) and U-Net (Ronneberger, Fischer, and Brox 2015) which is widely used in the remote sensing community (see (Garg et al. 2019; Khan, Fraz, and Shahzad 2021)). The advantage here is that spatiality is considered in a very flexible way, due to the cascading nature of the encoders. Neighborhood is not only defined by a 3 by 3 kernel, but by an efficient learner who is looking for connections on many levels.

A simple statistical model has usually 2 parameters per class, which are directly estimated on the empirical distributions. If there is a lot of overlap between the value distributions of two classes, e.g., pastures and wheat fields are both relatively green at the same time, there is not much that can be done about it, from the model building experts' perspective. In machine learning the expert can chose different kernel functions, adjust hyperparameters to increase or decrease overfitting. In deep learning, an arbitrary amount of model architectures exists, because layers can be connected and nested freely. Additionally, each individual layer comes with hyperparameters like input shape, activation function and number of output kernel filters, on top of arbitrarily complex data augmentation techniques, that can extend a limited amount of training data. As a consequence, right now, the performance of deep learning models is mostly limited by the model building experts' knowledge, rather than the capabilities of the models themselves. In theory, modelling agricultural growth and crop type is well within the capabilities of modern classification algorithms, and near-real-time satellite imagery is available globally for free.

1.3 Current Trends in Remote Sensing Based Crop Classification

While the previous section described the advancements of satellite based agricultural modeling in a broader historical sense, the following section aims to describe much more recent trends in detail. With this goal in mind, a large sample of recent case studies has been reviewed. For each year, starting from 2015, the first 20 google scholar results with the search query "crop type classification remote sensing" have been screened. From these 140 search results, 41 have been excluded because of various reasons. Most of these were not focused on satellite based remote sensing, or were review papers. The 99 articles that were left were systematically analyzed, with a focus on study area, input data and the classification algorithm.

Most articles covered by this review focus on a single study area. China (20 articles), USA (15 articles), Germany (10 articles) and Spain (8 articles) are the most popular study areas, while

with very few exceptions smaller subsets of these countries were analyzed. One such exception is a recent study by (Blickensdörfer et al. 2021), where crops in the entire country of Germany were classified in 2017, 2018 and 2019. While the exact footprints of the studies were not available in most cases, there still seemed to be a clear trend towards bigger areas in more recent years which is very plausible considering the increasing availability of datasets. Within this sampled subset of the literature body of the past 7 years, only 4 articles covered study areas including more than one country. Two connected studies (cf. (J. Inglada et al. 2015; Jordi Inglada et al. 2015)) have been conducted in 2015 and used simulated Sentinel-2 datasets to classify crops from 12 different test sites which were part of the Joint Experiment for Crop Assessment and Monitoring (JECAM) network (cf. (JECAM 2015)). Another study from 2019 (cf. (Defourny et al. 2019)) applied ESAs Sen2-Agri product, which aims to classify crops based on Sentinel-2 operationally for entire countries, to Ukraine, Mali and South Africa. (Rustowicz et al. 2019) specifically focused on two countries in Sub-Saharan Africa with its agricultural structure focusing on smallholder farms. While the overall sizes of the study areas seemed to be increasing, there was no clear trend towards the development of regional transferability of models.
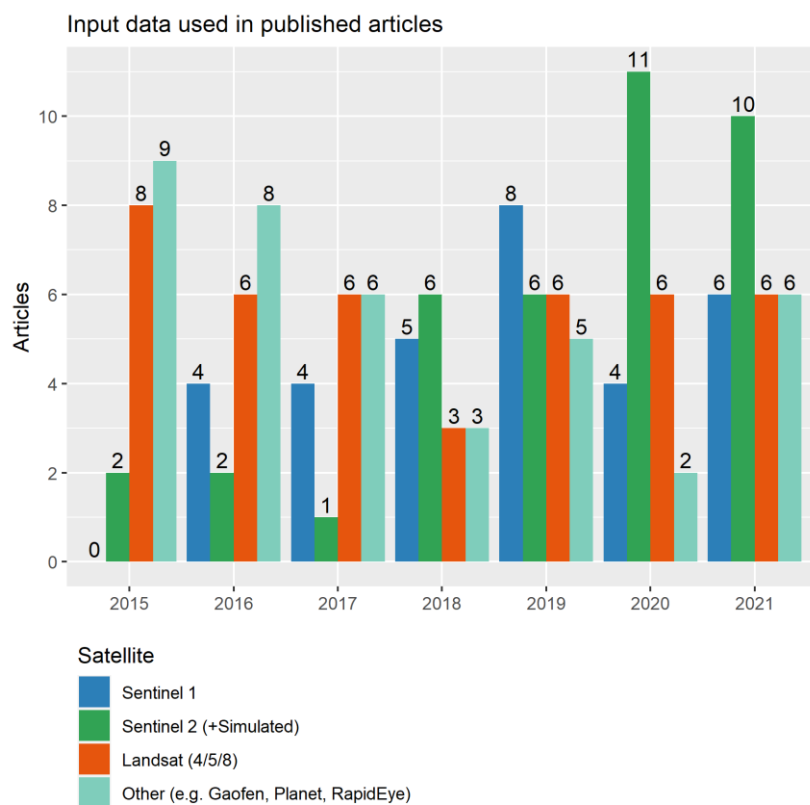

Figure 1: Input data used in crop classification

The remote sensing input data in the analyzed years is illustrative for the immense success of the Copernicus program (see Figure 1). Starting from 2018, the majority of crop classification studies use either optical Sentinel-2 or SAR Sentinel-1 imagery, often even both, and they almost completely replaced other sensors. This is largely due to the fact that data from these sensors is of high spectral, temporal and spatial resolution, as well as available around the globe for free

(see section 1.2.1.). The temporal aspect is especially important, since most of the studies in the sample use time series datasets, as a result of the dynamic nature of agricultural vegetation growth. This trend, which was already apparent in 2015 increased further in the last seven years, and in 2020 there was only a single article which did not use time series data.

Considering the reference data, there is a clear regional gap. Studies conducted in China exclusively rely on field surveys to determine the standing crops. Studies conducted in Europe also do these kinds of surveys, but additionally use cadaster data in the form of LPIS (see 1.2.3). Studies conducted in the USA exclusively use CDL data, which has proven to be a very rich source in agricultural monitoring. It is particularly beneficial here, since it enables a long term temporal scope which would never be possible with manual ground truth acquisitions. As a consequence, these are the only long term studies in the sample, covering over a decade of agricultural development based on MODIS (cf. (Massey et al. 2017)) or Landsat data (cf. (Cai et al. 2018)). This shows the enormous potential of operationally available satellite imagery in conjunction with long term georeferenced labels, while at the same time highlighting the main limitation of studies where no institutional long-term records are public, which is the case in most countries.

Another consequence of this is that the transferability of models is rarely analyzed. Since regional field surveys are expensive, and traveling would make them even more so, assessing the generalization requires usually a collaborative effort and is very rare, as seen in the JECAM studies. Generalization across multiple years is even more difficult, because it would require long term studies. Therefore, this is also rare, and the only case studies in the sample which test performance between years are studies where long term official records like CDL and LPIS exist (cf. (Johnson and Mueller 2021)).
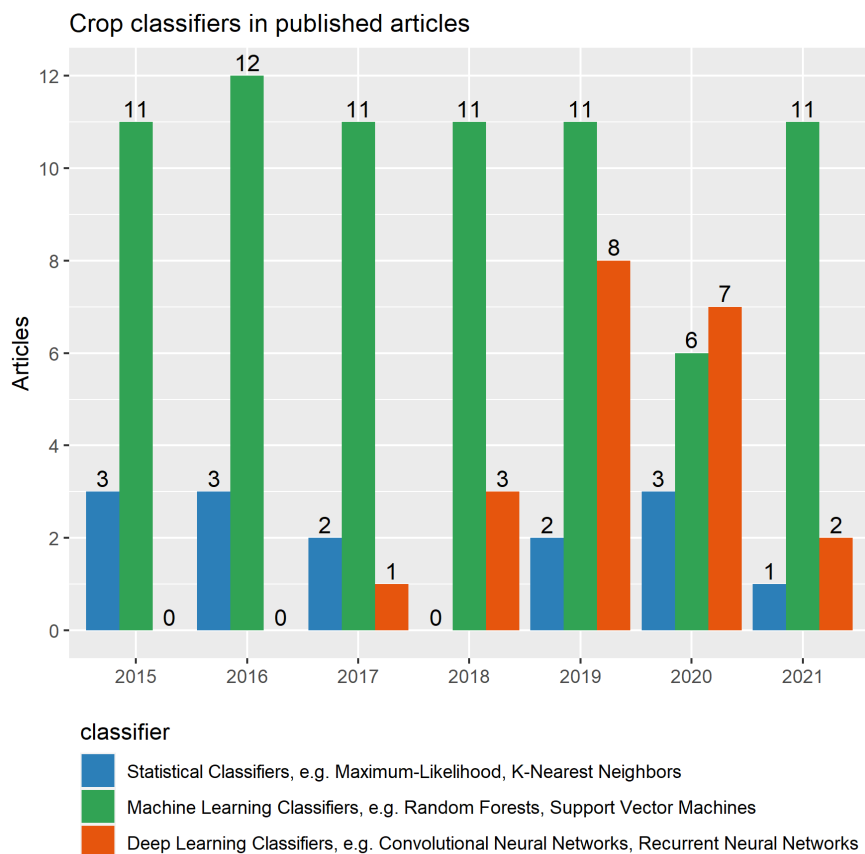
Crop classifiers in published articles

Figure 2: Classifiers used in published articles

The development of classification algorithms in the analyzed sample of published articles illustrates the unbroken popularity of machine learning models in the current remote sensing community. Particularly Random Forests and Support Vector Machines are used in the majority of all publications, and very often even in this exact combination. While the role of statistical models, mostly consisting of the Maximum Likelihood Classifier, is certainly minor, they have not completely vanished and show a constant but low usage through the past 7 years. Deep learning models show a very sharp increase in usage in the past 5 years in particular, though there is a notable low number in the year 2021. Of these deep learning models, many variations and architectures are used, most notably CNNs and Recurrent Neural Networks, as well as combinations and variations of both. In comparison, the machine learning models consist mostly of SVMs and RFs.

In addition to the classification algorithm, we checked if the models had any spatial component (see also section 1.2.4). From the 99 studies in our sample, 70 did not have a spatial component, and treated each pixel as an independent vector, mostly to feed into a machine learning classifier. Some studies enhanced the raw pixel values with parcel based extracted texture features (e.g. (Inglada et al. 2016)) or included neighborhood information with Conditional Random Fields (e.g. (Kenduiywo, Bargiel, and Soergel 2018)). The most popular approach to include spatial context was the CNN, which was used in a total of 15 studies, mainly in 2019 and 2020.

In summary, deep learning, while very popular in the past 5 years, has not yet taken hold in the remote sensing community, in the same way as machine learning models have. This is mostly evident since no particular architecture is proven to be superior, and there is still a lot of room to assess robustness and performance in comparison to more established algorithms.

1.4 Thesis Objectives and Structure

1.4.1 Thesis Objectives

The premises of this thesis are the following three conclusions of the preceding sections:

- Yearly monitoring of crop type with high spatial resolution is a crucial factor in global environmental protection as well as food security.
- Yearly monitoring of crop type with high spatial resolution is reasonably achievable with currently freely available satellite imagery. There is a clear consensus that Sentinel-1 and -2 are capable of delivering adequate input data to classify crop types.
- The algorithms which could achieve this are very capable but not mature enough. There is high research demand to explore details in:
    - The composition of reference data
    - The preprocessing of input data
    - The model type and architecture

The following thesis describes an explorative approach to advance the state of the art of crop classification. Regional case studies are conducted, aiming to develop validated crop classification toolchains. Because of its unique role in the regional agricultural system and because of its specific phenologic characteristics we will focus solely on maize fields. The specific objectives and research questions are the following:

1. Explore capabilities and practical limitations of optical remote sensing-based crop classification

    1.1. Can optical remote sensing data be used for operational crop classification? Though optical remote sensing data has been extensively used to classify land cover, it has some severe limitations in terms of temporal resolution. This could prove to be a crucial issue, since individual crops follow a specific phenological cycle which might be missed if the window for image acquisition is not met. This objective aims to verify this limitation by applying the classification to as many years as possible, and check if it is still possible to quantify crop extent.

    1.2. **Can high resolution imagery be used to generate reference data?** While high resolution imagery is often widely available, the feasibility of using it to replace in situ field surveys depends on other factors as well. The crop needs to be well distinguishable from other crops, and the imagery also needs to meet a consistent acquisition window. This objective aims to assess these practicalities in a specific regional setting.

    1.3. **How can the problem of residual autocorrelation be addressed?** While established machine learning classifiers are able to distinguish crop types based on pixel values, they are usually treating each observation as independent. Since pixels are spatially

dependent, a solution must be found to deal with the spatial dependency and autocorrelation.

2. <u>Explore regional applications of crop classification</u>

   2.1. **Can binary crop classification be used to link the expansion of biogas producing units to extended maize cultivation?** In a setting where no official records are available to the researcher, remote sensing-based crop classification can be used if spatially explicit vegetation cover is needed. Keeping in mind the limitations in terms of availability, we explore the question whether biogas plants are a major driver behind the extension of maize cultivation in the region.

   2.2. **Can binary crop classification in combination with existing soil data be used to identify spatially explicit soil hazard potential?** While the first objective focusses on the usefulness of remote sensing-based crop classification data considering the limitations in terms of availability over a longer period of time, this objective highlights the usefulness with respect to the high spatial resolution. We establish an approach to estimate soil hazard potential in the form of erosion, which inherently requires high spatial resolution of the vegetation cover product as well as the 3D surface model.

3. <u>Explore spatial and temporal transferability of upcoming state of the art models</u>

   3.1. **Can operational SAR time series be used to classify maize?** Optical remote sensing imagery will always be limited by acquisition gaps. If regional and temporal transferability of crop classification models is the ambition, SAR time series are one major way to get datasets in a consistent way across multiple years. Though there is some information loss in contrast to optical sensors, because SAR imagery has no information about canopy pigments, we analyse the hypothesis that the development of the SAR backscatter coefficient as a function of phenology is sufficient to distinguish maize pixels.

   3.2. **Are state of the art convolution based neural networks superior to established pixel-based machine learning classifiers in terms of spatial and temporal generalization?** The main objective is to assess the spatial and temporal generalization of classification algorithms. As such, a randomized framework is necessary that trains many models and measures performance in different regions and years. The resulting gap in classification performance will be used to assess and compare model transferability for a subset of machine learning and deep learning classifiers.

1.4.2 Thesis Structure

The presented objectives will be worked on in three different studies. The first two studies are connected by the same crop classification dataset, while the third study is about the exact same study region, but analyzes a different dataset.

Chapter 2, *Remote Sensing Based Binary Classification of Maize. Dealing with Residual Autocorrelation in Sparse Sample Situations* deals with objective one. An optical remote sensing pipeline is established, and reference data to classify maize in the study region of Eifelkreis Bitburg-Prüm is generated.

Chapter 3, *Implications of Bioenergy Cropping for Soil: Remote Sensing Identification of Silage Maize Cultivation and Risk Assessment Concerning Soil Erosion and Compaction* deals with objective two. The datasets from the first study are used in a case study about the expansion of bioenergy in the study region, and the associated risks in the form of soil erosion and compaction.

Chapter 4, *Field Geometry and the Spatial and Temporal Generalization of Crop Classification Algorithms—A Randomized Approach to Compare Pixel Based and Convolution Based Methods* deals with objective three. With better and more dense input imagery, crop labels based on statistical records, as well as new optimized image segmentation algorithms based on deep learning, the situation in the same study area is reassessed with a more ambitious modeling approach. The introduction of modeling scopes serves as a toolset to focus on model transferability, as a major milestone towards an operational and global crop classification framework.

Chapter 5 will summarize the main findings of each of the three studies, as well as describe an outlook into potential future research scenarios.

2 Remote Sensing Based Binary Classification of Maize. Dealing with Residual Autocorrelation in Sparse Sample Situations

The following manuscript is the final version of the accepted article. The paper has been through peer review. The contribution of Mario Gilcher as the main author is the development of the experimental design, data access and screening as well as the algorithm implementation and analysis. The article can be accessed online: https://doi.org/10.3390/rs11182172

MDPI

Article

Remote Sensing Based Binary Classification of Maize. Dealing with Residual Autocorrelation in Sparse Sample Situations

Mario Gilcher [1,*]**, Thorsten Ruf** [2]**, Christoph Emmerling** [2] **and Thomas Udelhoven** [1]

[1] Department of Remote Sensing and Geoinformatics, Faculty of Regional and Environmental Sciences, University of Trier, Campus II, D-54286 Trier, Germany; udelhoven@uni-trier.de
[2] Department of Soil Science, Faculty of Regional and Environmental Sciences, University of Trier, Campus II, D-54286 Trier, Germany; ruf@uni-trier.de (T.R.); emmerling@uni-trier.de (C.E.)
[*] Correspondence: gilcher@uni-trier.de; Tel.: +49-(0)-651-201-4607

check for
updates

Abstract: In order to discuss potential sustainability issues of expanding silage maize cultivation in Rhineland-Palatinate, spatially explicit monitoring is necessary. Publicly available statistical records are often not a sufficient basis for extensive research, especially on soil health, where risk factors like erosion and compaction depend on variables that are specific to every site, and hard to generalize for larger administrative aggregates. The focus of this study is to apply established classification algorithms to estimate maize abundance for each independent pixel, while at the same time accounting for their spatial relationship. Therefore, two ways to incorporate spatial autocorrelation of neighboring pixels are combined with three different classification models. The performance of each of these modeling approaches is analyzed and discussed. Finally, one prediction approach is applied to the imagery, and the overall predicted acreage is compared to publicly available data. We were able to show that Support Vector Machine (SVM) classification and Random Forests (RF) were able to distinguish maize pixels reliably, with kappa values well above 0.9 in most cases. The Generalized Linear Model (GLM) performed substantially worse. Furthermore, Regression Kriging (RK) as an approach to integrate spatial autocorrelation into the prediction model is not suitable in use cases with millions of sparsely clustered training pixels. Gaussian Blur is able to improve predictions slightly in these cases, but it is possible that this is only because it smoothes out impurities of the reference data. The overall prediction with RF classification combined with Gaussian Blur performed well, with out of bag error rates of 0.5% in 2009 and 1.3% in 2016. Despite the low error rates, there is a discrepancy between the predicted acreage and the official records, which is 20% in 2009 and 27% in 2016.

Keywords: crop classification; spatial autocorrelation; Regression Kriging

1. Introduction

1.1. Motivation

Expanded silage maize cultivation, as a direct consequence of Germany's energy turnaround and more precisely the governmentally guaranteed feed-in-tariffs [1,2], comes with several potentially negative externalities. The high management intensity of silage maize and the elevated threat of soils towards erosion, observed losses of agrobiodiversity as well as the trend towards monotonous landscape structures are calling the sustainability of such cultivation systems into question [3]. Consequently, spatially explicit monitoring is necessary, which may facilitate a risk assessment towards

erosion rates. Defining critical limits for soil erosion may help to make silage maize cultivation subject to conditions of erosion protection methods to distinctly reduce the risk of soil loss or exclusion of highly vulnerable sites. Moreover, such analysis may be used to support political decisions. Although these kinds of datasets are tracked by administrative bodies on the state level, they are rarely available to researchers. Thus, land-use classification approaches using optical remote sensing imagery could allow for a more independent and conclusive research. While the paper by Ruf [4] focuses more on the application and the implications of such an analysis, this paper addresses the major methodological obstacles from this study, which are very specific for spatially explicit pixel based classification of satellite imagery.

1.2. Monitoring Land Use Dynamics with Pixel Based Classification of Optical Remote Sensing Imagery

Land cover classification is one of the most popular applications for remote sensing scientists. Yu et al. [5] analyzed the popularity of land cover classification algorithms, and assessed that the majority of publications used the Maximum Likelihood (ML) classification, which is comparatively simple and widely implemented in remote sensing software packages. At the same time, more advanced classification algorithms like Support Vector Machines (SVM) and ensemble based methods perform land cover classification considerably better. A meta study analyzing the performance of supervised pixel-based classification algorithms specifically focusing on studies where performances of algorithms were compared to compute a quantitative synthesis of 266 articles [6]. It specifically identified SVMs, Neural Networks (NN) and Random Forest (RF) classifiers as superior to statistical tools like ML classification. However, the median differences of Overall Accuracy (OA) between SVM, NN and RF were all less than 2% or equal, which means that these rather modern machine-learning algorithms performed comparably in most cases.

In contrast to a complete landcover classification of a certain region, in this study, only the detection of maize fields was of importance. This means that only one class exists on a conceptual level, and therefore only one type of training data. One approach to deal with this problem is to adjust the classification algorithm. For instance, several adjusted versions exist of the SVM algorithm, which was initially conceived as a binary classifier [7]. Another approach is to sample from the unlabeled observations, to get two sets of training data, which are either positive or unlabeled, which is why this approach is called PU learning [8]. The downside here is that the unlabeled observations can contain a varying amount of positive pixels. Since the reference data are generated based on orthophotos (see Section 2.3), the distinction of maize and other pixels can be made quite clearly; therefore, two classes can be produced. One of these classes is very homogeneous (maize), the other is very heterogeneous and can contain pixels like pastures which are very close to maize, and bare soil which is very different from maize.

The intended pixel based classification of the multispectral imagery misspecifies the problem insofar, as it treats each pixel independently. Every observation is a numeric vector, consisting of a number of reflectance values from the satellite imagery, and a categorical dependent variable that indicates the presence of maize at that particular location. The location itself, along with the implications of proximity and neighborhood to other pixels, is not included. Since maize fields are on average between 0.5 and 2 hectares, which is much larger than a single pixel, the presence of maize in one pixel increases the probability of maize for all pixels in close proximity. This phenomenon is called spatial autocorrelation [9].

There are many different angles from which this issue can be approached [10], beginning with the sample selection [11] and sample refinement [12], via the classification algorithm itself [13] and ending with adjustments made post-classification [14]. In spite of that, spatial relationships are still rarely analyzed in remote sensing classification research. The aforementioned review papers on classification of remote sensing imagery ([5,6]) do not cover the topic at all. Another recent review paper [15] specifically on RF based classifiers only mentions the issue of spatially autocorrelated training data.

Therefore, while many different ways exist to incorporate spatial relationships of pixels, their usage is not widespread and most applications are still aspatial.

1.3. Objectives

This study aimed to establish a validated toolchain to prove the suitability of optical remote sensing as a primary source to monitor the prevalence of maize in a given region, which has potential sustainability implications for medium scale regions. For this purpose, well understood and performant classification algorithms were employed in conjunction with two different methods to incorporate pixel relationships with their immediate neighborhood. This includes the following objectives:

1. **Systematic assembly of reference data for the study area**:
   To produce uniformly distributed reference data, aerial images from 2009 and 2016 were used to digitize maize and non-maize polygons for the entire study region. For both years, two kinds of reference datasets have been produced. A sparse dataset, with a minimum of six polygons per 25 km^2, covers the entire study area. Additionally, five dense datasets for 1.5 by 1.5 km subsets were produced, where all fields were digitized manually.
2. **The preprocessing of remote sensing imagery**:
   In the given years, two RapidEye scenes from late August were available. The tiles were preprocessed, in order to get one multispectral observation for each pixel of the study area.
3. **Small scale model evaluation**:
   The dense reference dataset for the five subsets was then used to analyze and optimize model behavior on a small scale. Three classification techniques (Generalized Linear Model, Random Forests, Support Vector Machines) were combined with two methods to account for spatial autocorrelation (Simple Kriging, Gaussian Blur).
4. **Medium scale model application for the entire dataset**:
   The optimized modeling approach was then trained with the sparse medium scale reference dataset, and applied to the entire region. Performance was analyzed and results were aggregated and visualized.

2. Materials and Methods

2.1. Study Area

The study was conducted in the "Eifelkreis Bitburg-Prüm" (in the following referred to as 'study area'), which is an administrative district in the southwest of Germany, bordering Luxembourg and Belgium. It has a size of 1627 km^2 and mostly consists of low mountain ranges, with a slow gradient in elevation from roughly 300 masl. in the south up to around 700 masl. in the northern part of the study area. Its population is sparsely distributed, and the rural landscape mostly consists of a considerable share of forests, while 53.2% of the total area is used by agriculture (Statistisches Landesamt Rheinland-Pfalz [16]).The agriculture of the region is mostly pastureland, with some cereals, and a substantial increase in silage maize cultivation during the past 20 years. This increase coincides with the progressive installation of biogas plants, starting with a first plant built in 2000, to 57 plants with a total installed capacity of 19,152 kW in the year 2016 [17]. As a result, the study area now shows the highest capacity of agricultural biogas producing units in Rhineland-Palatinate. The comparatively homogeneous rural structure in conjunction with the clear regional shift in cultivation strategies as a direct consequence of policymaking offers a unique opportunity to illustrate the capabilities of optical remote sensing in detecting land cover changes.

2.2. Image Data

The selection of image datasets was constrained by requirements of resolution, availability, phenology window, cloud coverage and time frame, since reference data was only available in 2009

and 2016 (see Section 2.3). The spatial resolution has to be adequate, with many fields smaller than a few hectares. The RapidEye Science Archive [18] offers a unique opportunity to researchers by providing multispectral imagery for research projects. RapidEye is a constellation of five satellites in a 630 km sun-synchronous orbit, with a nadir Ground Sampling Distance of 6.5 m, which is orthorectified to a pixel size of 5 m. The sensor has three bands in the visible range, one red edge band, as well as one band in the near infrared range [19]. The revisit time of approximately six days increases the chances to get imagery on cloud free days, while the general timing window of the phenologic cycle of maize is met. Of all the crops used in the study area, maize reaches the peak vegetation cover the latest, around the end of August. At that time, rape and cereals have typically been already harvested, and maize is largely the only agricultural crop still standing [20]. This leaves a timing window of about four weeks between mid-August until mid-September.

Within these constraints, two scenes in late August could be identified, in the years 2009 and 2016. They were entirely cloud free, with only negligible amounts of missing pixels in the far northwest in 2009. The imagery was provided as 25 by 25 km tiled Level 3A Geotiffs, which means they were orthorectified and radiometrically corrected. These tiles were then atmospherically corrected based on the 5S algorithm [21] with the software AtCPro (Version 6.0, Department of Remote Sensing and Geoinformatics, Trier, Germany) [22,23], mosaicked and cropped. In the atmospheric correction process, a digital elevation model (DEM) [24] was used to compensate for topographic effects. It was also used in the prediction models, since elevation has a significant influence on local climate and therefore phenology. After the atmospheric correction, the mosaic was masked in two steps. Firstly, this was to only include pixels within the borders of the study area (Nomenclature of Territorial Units for Statistics (NUTS)-3 Region Eifelkreis Bitburg Prüm). Secondly, this was to exclude roads, settlements and forests, i.e., only contain pixels in agricultural use (agricultural areas acc. to the German land appraisal, provided by the LGB-RLP).

2.3. Reference Data

Since the classification of around 87,000 hectares of agricultural land requires a large amount of evenly distributed training data, a visual survey approach was chosen, based on Google Earth aerial imagery (see Figure 1). The high resolution aerial images were used to identify maize fields, which was feasible because of the unique texture of grown maize fields. Their shape was then roughly digitized on top of the RapidEye images. At the same time, polygons were digitized with no maize fields inside. To guarantee an even distribution of training fields across the study site, it was split in subdivisions using a 9 by 12 grid, with each cell spanning 5 by 5 km. The main goal was to select a small set of no less than three polygons for both maize and non-maize land cover types separately. This was not always possible, especially since some cells around the fringes of the study area were mostly covered by forests. In addition, some cells in the east and west of the scene in 2016 were left empty because there was no high-resolution Google Earth imagery available for that date. Overall, around 2500 polygons were digitized for both years combined, amounting to over 1000 hectares of maize fields per year. This is a substantial percentage of the overall acreage, and should guarantee that the selected pixels were representative for the entire study area. In the following sections, this dataset is referred to as sparse medium scale reference.

To analyze spatial autocorrelation and use it to the classifiers' advantage, smaller subsets have been digitized independently from the grid based sparse reference dataset, but still based on the high-resolution imagery. For each of both years, five 1.5 km by 1.5 km squares were subsetted. They specifically included mostly pixels without settlements and very little roads, so that most pixels were covered either by maize, bare soil or vegetation different from maize. These ten square subsets were then digitized completely, aside from transitional pixels on the edges of fields and dirt roads. This lead to dense datasets with a moderate amount of overall observations, which were used to monitor spatial correlation effects especially for the model probability residuals. In the following sections, this dataset is referred to as dense small scale reference.
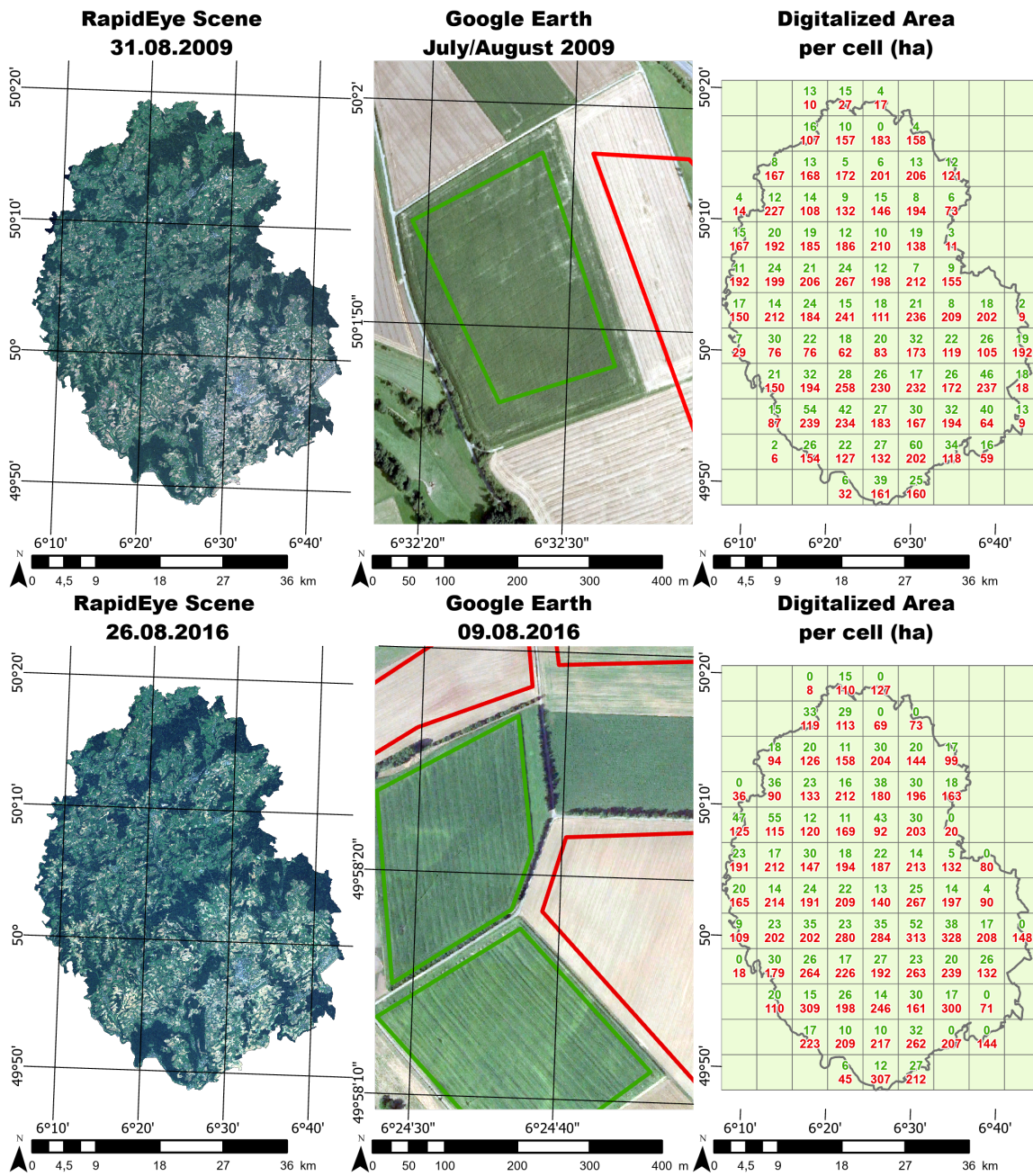
Figure 1. RapidEye coverage of the study area and manual digitization of reference data. The left column shows the two RapidEye scenes. The middle column shows the high resolution aerial imagery, and some digitized polygons. The green polygons are maize fields and the red polygons are a mixture of mostly bare soil and green lands. The right column shows the distribution of digitized area for each cell.

2.4. Modeling

2.4.1. Modeling Overview

The main goal of this study was to find ways to deal with spatial autocorrelation of residuals in the context of independent pixel based predictions. To achieve this, the modeling process was split up in two parts. The first part was the independent pixel based model, where a probability for a given pixel to be maize was estimated based on the covariates for this particular pixel, disregarding all

neighboring pixels. The second part took the spatial autocorrelation between neighboring pixels into account, and adjusted the probabilities for a given pixel accordingly.

2.4.2. Spatial Autocorrelation

Every observation is a numeric vector, consisting of five reflectance values from the satellite imagery, the altitude of the DEM, and a categorical dependent variable, which indicates the presence of maize at that particular location. The location itself, along with the implications of proximity and neighborhood to other pixels were not included. The Regression Kriging approach [25] allows for using spatial autocorrelation to the advantage of the classification process. By splitting each modelling problem into an independent pixel based modeling part (deterministic) and a spatial part (stochastic), many different pixel based modeling approaches can be combined with the Simple Kriging (SK) of their residuals, and therefore incorporate inherent spatial effects in an otherwise spatially independent modelling approach. The most common way to describe the relationship of probabilities being increasingly similar with proximity is the variogram. The empirical variogram describes the variance between pairs of observations with a fixed distance [26]. Figure 2 shows five different variograms from the small-scale digitized sites in 2016 (see Section 2.3). On all five sites, the variance increases steadily for the first few hundred meters, and plateaus at a distance between 300 and 500 m.
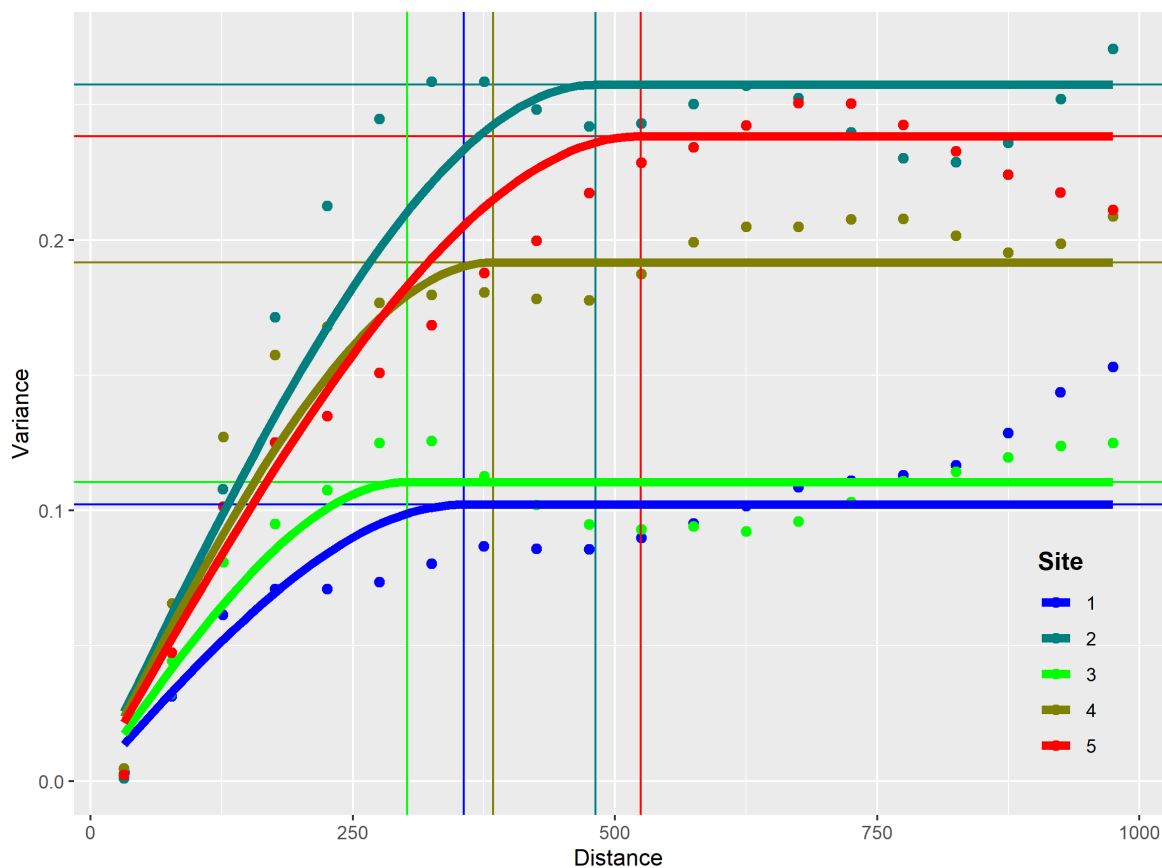

Figure 2. Variograms of five different sites in 2016. The points show the empirical variance of data points for the given distances. The bold lines show the fitted spherical models for each of the sites. The horizontal lines indicate the sill of the model, while the vertical lines indicate the range.

2.5. Research Design

The research design (see Figure 3) consisted mainly of three parts. The first part included all the preprocessing steps that were necessary to establish the datasets, to train, validate and apply the model. The results were the raster files with the predictors, a small scale reference dataset with dense

polygons for the five subsets (see Section 2.3), and a medium scale dataset with sparse polygons for the entire study area, each of which were available for both years.
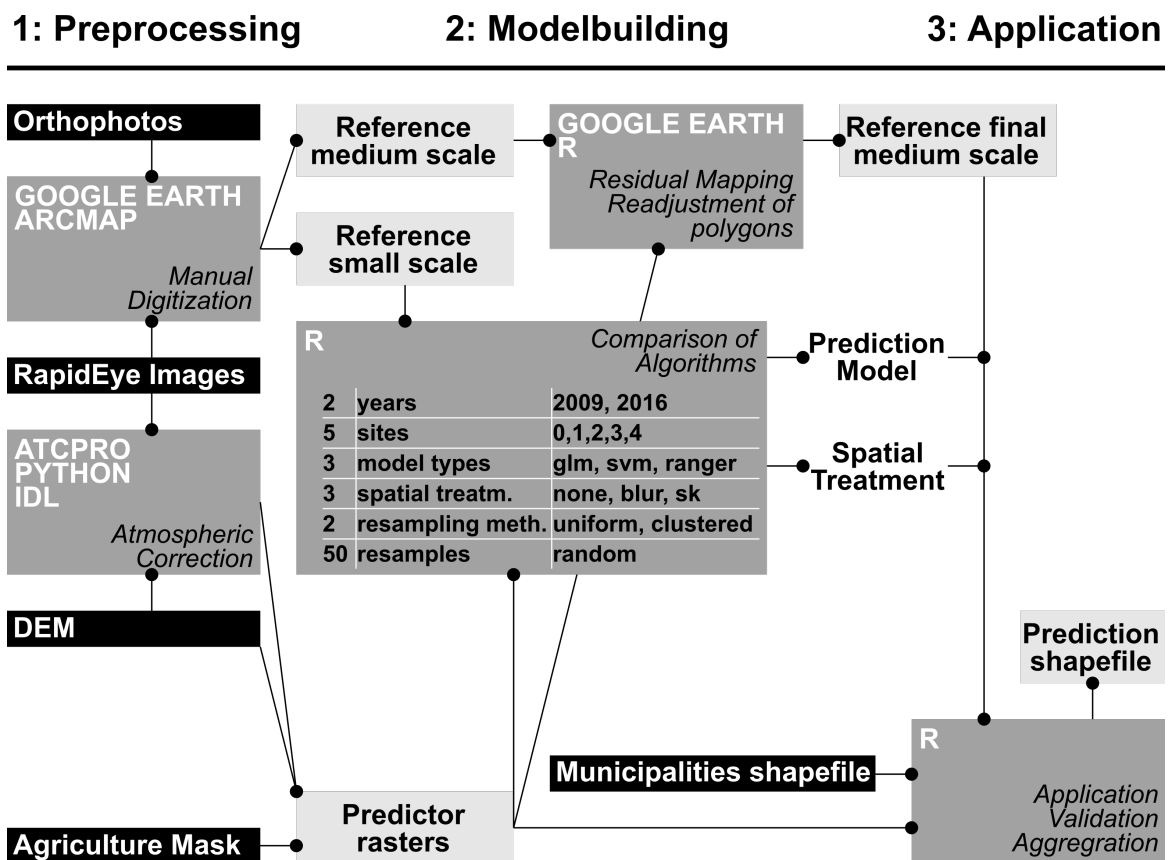
1: Preprocessing   2: Modelbuilding   3: Application

Figure 3. Research design. **1** : Preprocessing. The left side of the diagram shows the two major preprocessing steps of atmospheric correction and manual digitization, based on the given datasets shown in the black rectangles. **2**: Modelbuilding. The prediction model and spatial treatment were chosen with an extensive algorithm comparison based on the small scale reference data. The medium scale dataset was used to evaluate itself by mapping prediction residuals and adjusting critical areas. **3**: Application. The prediction model and the chosen spatial treatment was then applied to the final medium scale dataset, and the results were aggregated based on municipalities.

The second part of the research design, was the actual modeling. It was split into two distinct parts:

1. Small scale model evaluation
2. Visual evaluation and optimization of the medium-scale grid based sparse digitization of maize and non-maize polygons

In the first step, three different model types were compared, in combination with three (two + one control classification with independent pixels) different approaches to deal with spatial autocorrelation. The pixel based model types used were firstly the Generalized Linear Model (GLM) with binomial error distribution, also called Logistic Regression. Since the monitoring of maize occurrence is a binary classification problem, the GLM is a transparent and robust statistical approach that can then be compared to the two state-of-the-art machine-learning techniques of SVM and RF (also sometimes referred to as ranger, after the implementation algorithm in R). All three algorithms have the advantage that they are able to output probabilities, which is important to apply the techniques to reduce residual autocorrelation. Hyperparameters were not tuned at this stage because gridsearches would greatly inflate the runtime. Therefore, standard hyperparameters were used, with a radial SVM kernel. The classification probabilities predicted by these algorithms were then combined with three approaches of

dealing with residual autocorrelation. The first approach was a control treatment, focusing on the extent of the residual autocorrelation with just independent pixel based predictions. The second approach was the Gaussian blurring, where the probability of each pixel is a weighted sum of surrounding pixel based on a Gaussian kernel function. The third approach was Regression Kriging where the residuals of initially predicted probabilities are interpolated with Simple Kriging. The three classification models were chosen because they are popular, well understood and generally very performant, both in runtime and in accuracy. The RK approach was chosen to test one demanding geostatistical model (RK) which solves the problem of residual autocorrelation in an explicit way. It was compared to the very simple and runtime-economical Gaussian blurring.

Since clustering of training samples is a problem in Regression Kriging [25], another layer was added to the research design, as illustrated in Figure 4. The computation of these models, even in the context of these smaller subsamples is very time-consuming, which means that training samples have to be drawn randomly. For each individual run, 1000 sample pixels were drawn, 500 being maize, and 500 being not maize. The rest of the, on average, roughly 55,000, reference pixels were used to validate that particular model run. In order to truly test the applicability of the Regression Kriging approach, this clustering was simulated by randomly picking five polygons with and five polygons without maize to draw the random sample of 1000 training pixels from. This resampling was done 50 times for each combination of factors, which led to a total of 9000 model runs. For each run, performance indicators were computed to assess model performance. Additionally, the variogram objects were stored in order to visualize the spatial autocorrelation of residuals for each of the influencing factors.

Figure 4. Uniform versus clustered sampling.

The second step of the model building process was the visual evaluation and optimization of the Medium-scale grid based sparse digitization of maize and non-maize polygons. Once the suitability of pixel based models was established in step 1, these were used to map uncertainty by visualizing probabilities that deviate substantially from 0 or 1. In cases where many of these pixels were clustered together, mistakes in digitizing the polygons were likely, and each of these cases was investigated closely.

Once a cleaned up reference dataset was finalized, a suitable model was applied to the predictor raster datasets. Since the predicted pixels are not useful enough on their own, statistics were aggregated on two levels. One single total area was computed for each year, in order to compare these predicted values to official data [27]. This is another important step to validate the prediction beyond the performances measured with the training data. Aside from performance measures purely based on the training data like kappa, error rates, sensitivity and specificity, independent official records are used to test model plausibility. In a last step, the maize area was aggregated spatially based on municipality

polygons [28], and relative values were calculated by dividing the maize area by the total agricultural area of each individual administrative unit.

3. Results

3.1. Small Scale Model Evaluation

Of the theoretical 9000 model computations, 8747 had meaningful outputs. This difference is a result of the clustered resampling because, in some resampling cases, the variogram could not be fit successfully, thus the Residual Kriging could not be performed. Figure 5 illustrates one random example of these runs. On this particular transect, two problematic areas with high residuals were identified. A closer look showed that these were small patches with less vegetation cover in the big field in the middle, which showed spectral properties closer to soil or grass.
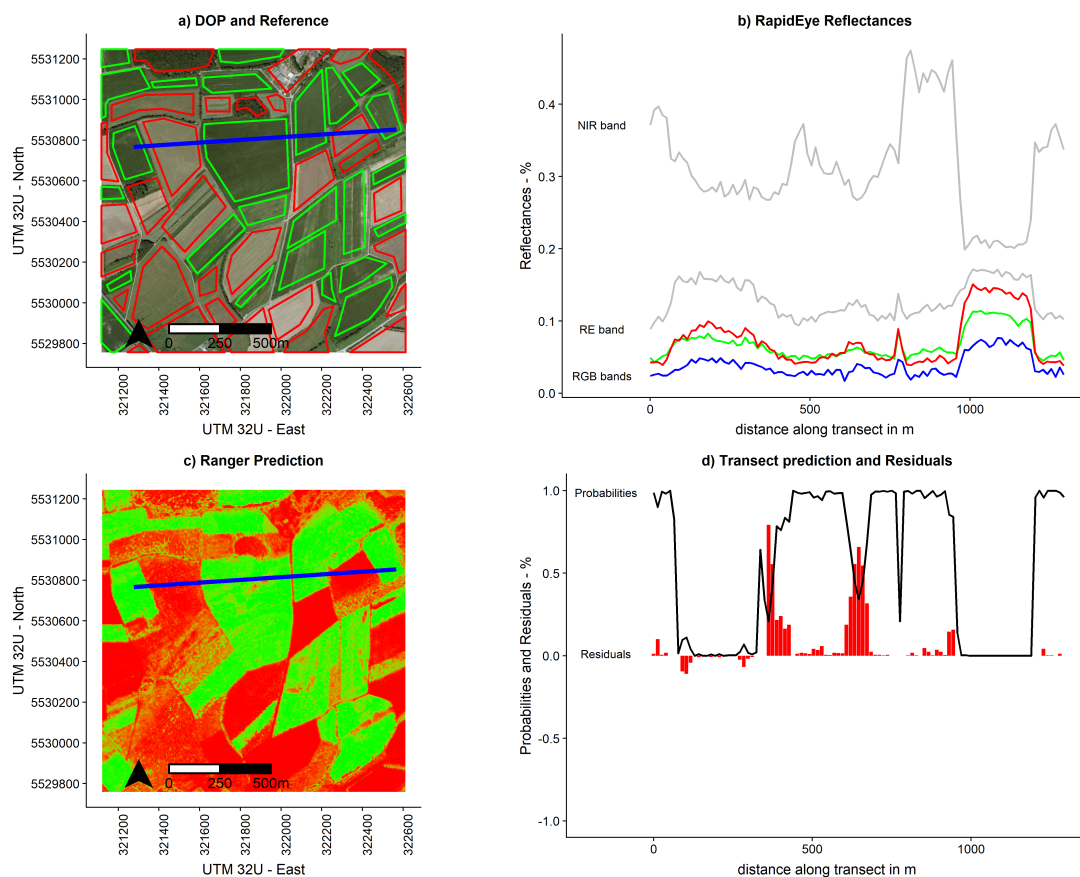

Figure 5. Example result of subset 2–2009–Random Forest (**a**) The high-resolution aerial image can be seen in the background, with the reference polygons for maize (green) and non-maize (red) on top. The blue line indicates the location of a transect, which is used to plot values in a 2D plot in (**b**). (**b**) The transect plot shows reflectance values for all five bands. In this plot, maize can be easily identified by looking at the peaks in the NIR band. (**c**) Probabilities of the random forest prediction are shown, with the green values being close to 1, and the red values being close to zero. (**d**) This 2D transect plot shows the probabilities and residuals along the transect.

All complete cases are summarized in Table 1, where the average kappa values for both years and sampling approaches were aggregated by the model. The kappa value is used to assess classification performance purely based on its popularity. Overall Accuracy was computed as well, with very similar results. To analyze the algorithms further, especially their impact on residual autocorrelation, empirical residual variograms were computed for all model residuals. A spherical model was then fitted to all empirical variograms, and the parameters were aggregated based on model type and sampling

approach for both years. While most of the fitted variograms showed no substantial nugget effect, the partial sill can be used to describe the impact of proximity on residual probabilities (see Table 2).

Table 1. Average kappa of small scale model runs.

Approach	Uniform—2009	Clustered—2009	Uniform—2016	Clustered—2016
glm	0.865	0.760	0.781	0.631
glm + blur	0.906	0.791	0.846	0.668
glm + sk	0.941	0.765	0.918	0.637
ranger	0.930	0.766	0.897	0.65
ranger + blur	0.958	0.784	0.937	0.668
ranger + sk	0.947	0.769	0.923	0.645
svm	0.931	0.769	0.893	0.678
svm + blur	0.949	0.784	0.925	0.694
svm + sk	0.964	0.779	0.949	0.666
average pure	0.909	0.765	0.857	0.653
average blur	0.938	0.786	0.902	0.677
average sk	0.951	0.771	0.930	0.650
average glm	0.904	0.772	0.848	0.645
average ranger	0.945	0.773	0.919	0.655
average svm	0.948	0.777	0.922	0.679
average total	0.932	0.774	0.896	0.660

Table 2. Averaged partial sills—fitted spherical variogram.

Approach	Uniform—2009	Clustered—2009	Uniform—2016	Clustered—2016
glm	0.0290	0.0642	0.0433	0.0865
glm + blur	0.0276	0.0617	0.0387	0.0821
glm + sk	0.0088	0.0639	0.0124	0.0872
ranger	0.0152	0.0613	0.0168	0.0753
ranger + blur	0.0148	0.0599	0.0165	0.0724
ranger + sk	0.0092	0.0605	0.0100	0.0763
svm	0.0168	0.0673	0.0186	0.0787
svm + blur	0.0163	0.0657	0.0172	0.0754
svm + sk	0.0052	0.0660	0.0071	0.0824

The kappa value is ranging from −1 to 1, and is intended to describe the performance of the prediction, with higher values indicating better performance. The partial sill, defined as the difference between the minimum and maximum variance of the fitted semivariogram, is intended to measure the residual autocorrelation of the prediction. It describes the difference in variability between two points that are close, and two points that are far away. Since we aimed to minimize spatial autocorrelation, lower values were preferred. As a first observation, all kappa values were reasonably high, ranging from 0.631 to 0.964. Furthermore, all partial sills were reasonably low, ranging from 0.0052 to 0.0872, which means that the difference in variance between points that are nearby and points that are far away is never above 0.1. There was also a significant difference between the two years, with the classification of the 2016 dataset having consistently lower kappa values and higher partial sills. RF and SVM classification performed substantially better than the GLM.

3.2. Visual Evaluation and Optimization of the Medium-Scale Grid Based Sparse Digitization of Maize and Non-Maize Polygons

The small scale model evaluation has shown that both SVM and RF classification were suitable to perform the necessary binary classification of maize and non-maize pixels. Since the RF approach does not rely on extensive hyperparameter tuning as much as the SVM, it is used in the further process.

In the next step, classification was performed without the Gaussian Blur on the dataset, with all bands and elevation as predictors of maize. Elevation was used as a proxy of the climatic conditions, which change slightly from north to south (see Section 2.1). Afterwards, errors in the training polygons were visualized as red pixels in a raster format, and as a circle in the center of each polygon with the size indicating the number of error pixels in each polygon. The size of the circles was used to get a good indicator of areas with critical error levels from the large scale view of the entire map, while the red pixels can then be used to get a detailed view of the spatial distribution in each digitized polygon. Once a problematic polygon is identified, the field ID can be used to track the polygon in the reference datasets and make adjustments if necessary.

In the example of Figure 6, fields with the number 39, 565 and 36 appeared to be problematic. The misclassifications in 39 and 36, maize pixels that are classified as non-maize, were vague in shape, and upon closer look were just maize pixels which have some bare soil showing because of worse growth conditions. These pixels were fine to be digitized that way because they were part of the inherent heterogeneity of the crop. The errors in field 565, however, non-maize pixels that were classified as maize, had a clear edge, which seemed to be part of a field to the north of the polygon. Upon closer look, it was clear that these pixels were actually maize pixels, and that polygon 565 was too large. With over 2500 digitized polygons, mistakes like this were quite common, but they could be easily identified and corrected with this approach.
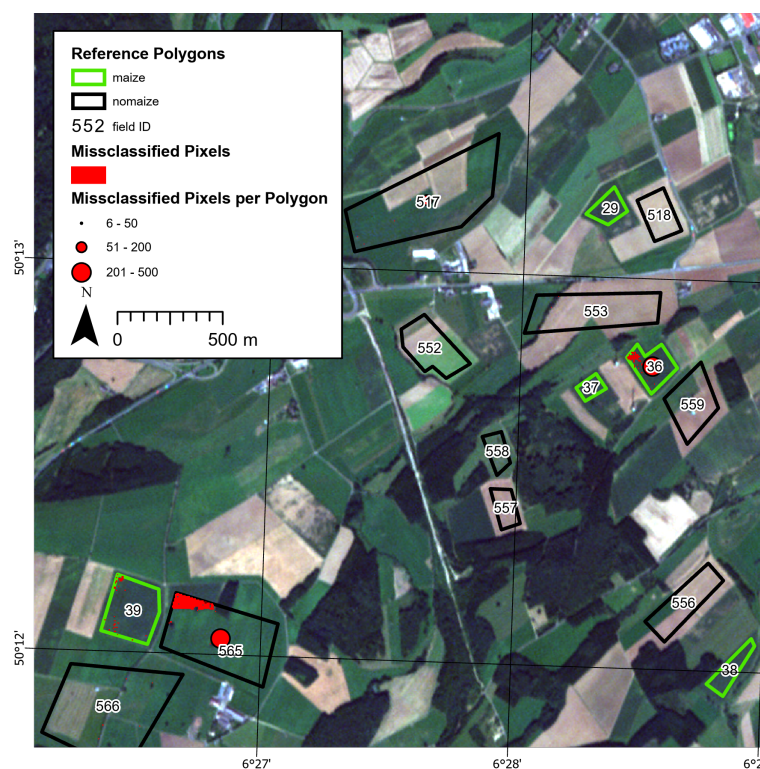

Figure 6. Digitization errors. The green and black polygons represent the digitized reference areas. The red circles indicate the number of misclassifications for any given polygon, while red pixels illustrate the location and spatial pattern of these misclassifications.

3.3. Pixel Based Mapping of Maize for the Entire Study Area

After the modeling approach was established, and the reference dataset was reevaluated, the RF classification and Gaussian Blur was applied to the entire region. The RF hyperparameters of mtry (number of variables used in a tree) and ntree (number of trees) were tested but had very little influence on model performance. Roughly 35 million pixels were classified, amounting to a total of 87,406 hectares of agricultural area. The performances are reported below in Table 3. The error rates

are Out-Of-Bag (OOB) meaning that these are the fractions of trees that were not trained with a given pixel, and predicted it incorrectly. Error rates and Kappa values were consistently good, while the classification in 2009 performed better than 2016, which was already a pattern in Section 3.1. Specificity was approximately 1, which means that there are very little false positives. Sensitivity, on the other hand, was quite a bit lower, which means that a lot of maize pixels went undetected. The total area of all maize pixels was then compared to official records [27], which showed an underestimation of 20% in 2009 and 27% in 2016. This discrepancy was substantially larger than the validation of the reference data would suggest.

Table 3. Overall classification results.

Year	Error (OOB)	Kappa	Spec.	Sens.	Area (ha) Estimated	Area (ha) Official	Rel. Diff. (%)	Abs. Diff. (ha)
2009	0.005	0.97	1	0.97	7305	9147	20	1842
2016	0.013	0.93	1	0.91	8447	11496	27	3049

In a final step, the prediction rasters were aggregated based on municipality polygons, and relative percentages were calculated based on the ratio of maize area and agricultural area for any given municipality. The final result (see Figure 7) was then mapped as the changes in relative percentages of maize area for a given municipality. This way, specific key regions could be separated from most areas with very little relative changes. Furthermore, a closer look at the prediction raster gave a detailed picture of the development in certain regions—for instance, the red area south of Prüm, where relative changes were larger than 20%, and many larger maize fields defined the landscape in 2016.
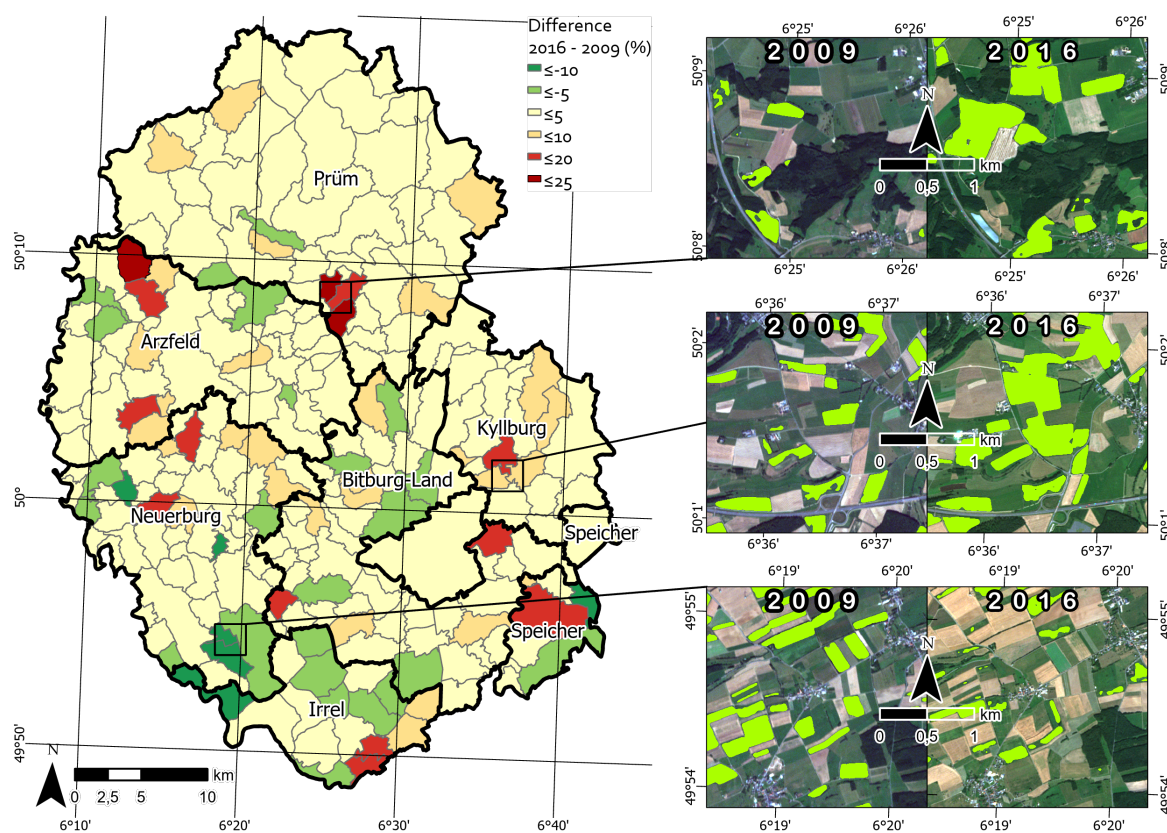

Figure 7. Changes in maize abundance. The six smaller images are true color RapidEye composites, with predicted maize fields in green. The map of the study area shows differences in the percentage of maize fields of the total agricultural area.

4. Discussion

4.1. Small Scale Model Evaluation

Key finding 1: *RF and SVM perform similarly well, and comparatively better than GLM.*

Both SVM and RF classification performed substantially better than the GLM, with Kappa values usually well above 0.9 in the case of uniform sampling distribution. This aligned well with the findings of methodical classification reviews, as described in Section 1.2, which mostly pointed out the strength of machine learning classifiers in comparison to statistical tools, which usually rely on assumptions about the distributions of the covariates. RF and SVM do not rely on these kind of assumptions, and are able to optimize nonlinear separation problems. While the maize class was mostly homogeneous (see also Section 2.3), the non-maize class consisted of pixels as diverse as bare soil, which is very dark in the infrared, and vegetation pixels like pastureland, which is quite bright in the infrared. This leads to possible bimodal distributions for some of the covariates, which is not a problem for machine learning algorithms.

The choice made towards RF classification done later in the study was mostly practical because of the more complex hyperparameter tuning SVMs require. While RFs also have hyperparameters, they are not as critical, optimizing them is rather a trade-off between performance and runtime, as opposed to over and underfitting the training data. This is usually the case when optimizing the cost parameter in Support Vector Classification. It is quite possible that a finely tuned SVM classification might have led to better results, as would have other potentially more intricate machine learning algorithms. The conclusion is not that RF classification is the best possible approach, but rather that it is highly adequate—in particular, since it is questionable if a perfect classification is possible or even desirable due to natural noise in the training data (see also Section 2.3).

Key finding 2: *Regression Kriging increases performance and decreases spatial autocorrelation when samples are uniformly distributed, but not when they are clustered.*

The RK approach worked as intended when samples were uniformly distributed, and was able to produce the highest kappa values, as well as very flat variogram models with the lowest overall sills. When clustered sampling was simulated, Kappa values did not improve as much and were even worse in comparison to the pure pixel based classifier. The same can be said for the estimated sills of the variogram models, where little to no improvement could be measured. That observation in itself was no surprise, since clustering of samples has been indicated as a problem by [14] as well.

Figure 8 shows why this is the case. Simple Kriging heavily alters the results in the direction of high residuals that are nearby, and does almost nothing for uncertain pixels if there are no high residuals around.

The results also showed that the classification performance was generally a lot worse in the clustered case, meaning that residuals are much higher. This is because five polygons were chosen in each run and each class, some of which were potentially very small or not representative for other reasons. Therefore, while in the clustered and uniform case, the same number of training pixels was drawn each run; the overall variability, especially of the maize training data, was potentially distinctly lower.
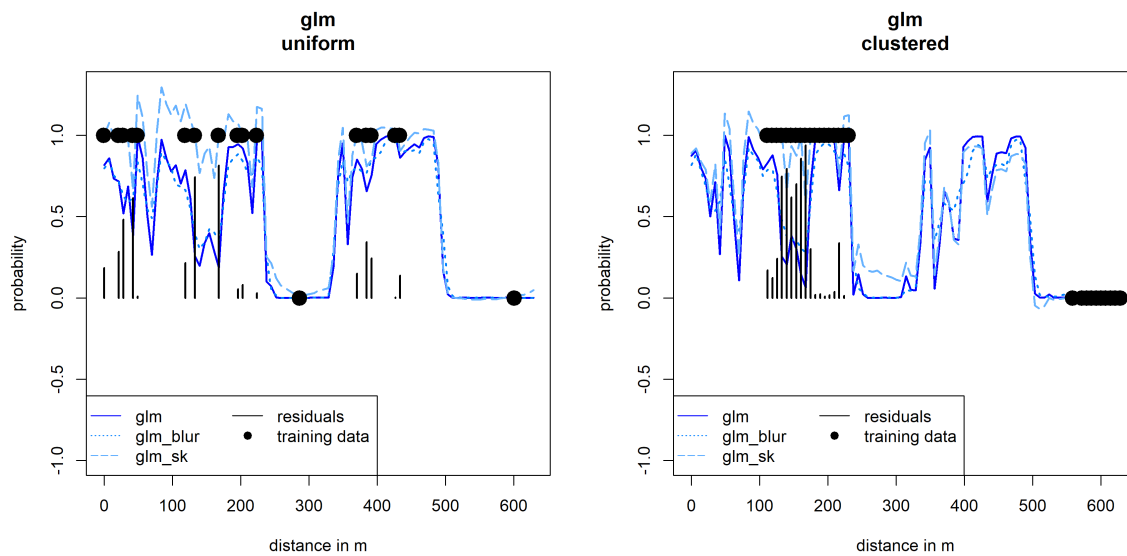
Figure 8. Transect of maize probabilities. In both plots, the solid blue lines indicate the predicted probabilities along the same transect. Training pixels are visualized with the black dots, uniform on the left, clustered on the right. The light blue line with the longer dashes shows the probabilities adjusted with the Simple Kriging. The difference is very large near large residuals. When no residuals are present, or when residuals are close to zero, Regression Kriging and glm probabilities are very similar. Gaussian blur probabilities, indicated by the lines with shorter dashes, work independently.

Key finding 3: *Gaussian Blur increases performance and decreases spatial autocorrelation when samples are uniformly distributed and when they are clustered.*

Clustered samples are not as problematic for the Gaussian Blur approach, since all neighboring pixels influence the probability equally depending on their distance, and not on whether they contain high residuals, or residuals at all. Only slight adjustments were made, as is shown in Figure 8. This mostly improved the classification result by removing noise in the training data, and only slightly decreased sills but for both clustered and uniform samples. It is likely that the decreased sills were a result of overall lower residuals, and that the Gaussian Blur does not lower residual autocorrelation at all in itself. The two issues of model performance and residual autocorrelation were hard to isolate, since it was not possible to change one without the other.

4.2. Pixel Based Mapping of Maize for the Entire Study Area

Key finding 1: *Classification of both years performs quite well, but 2009 is better than 2016.*

Maize is generally easy to classify, since most agricultural areas in the region were either bare soil, harvested cereal fields and pastures or still growing maize in late august. Difficulties arose when drought stress severely influenced pigments and vegetation cover in maize. Analysis of red and near infrared reflectances in both years, as shown in Figure 9, suggested that the training pixels of maize in 2016 contained substantially more soil and less chlorophyll content, since it is spectrally more heterogeneous. The figure also illustrates the distribution of false negatives and false positives of the out of bag predictions. It shows that, in both years, false negatives and false positives look surprisingly similar, with no real indication as to why model performance was worse in 2016.
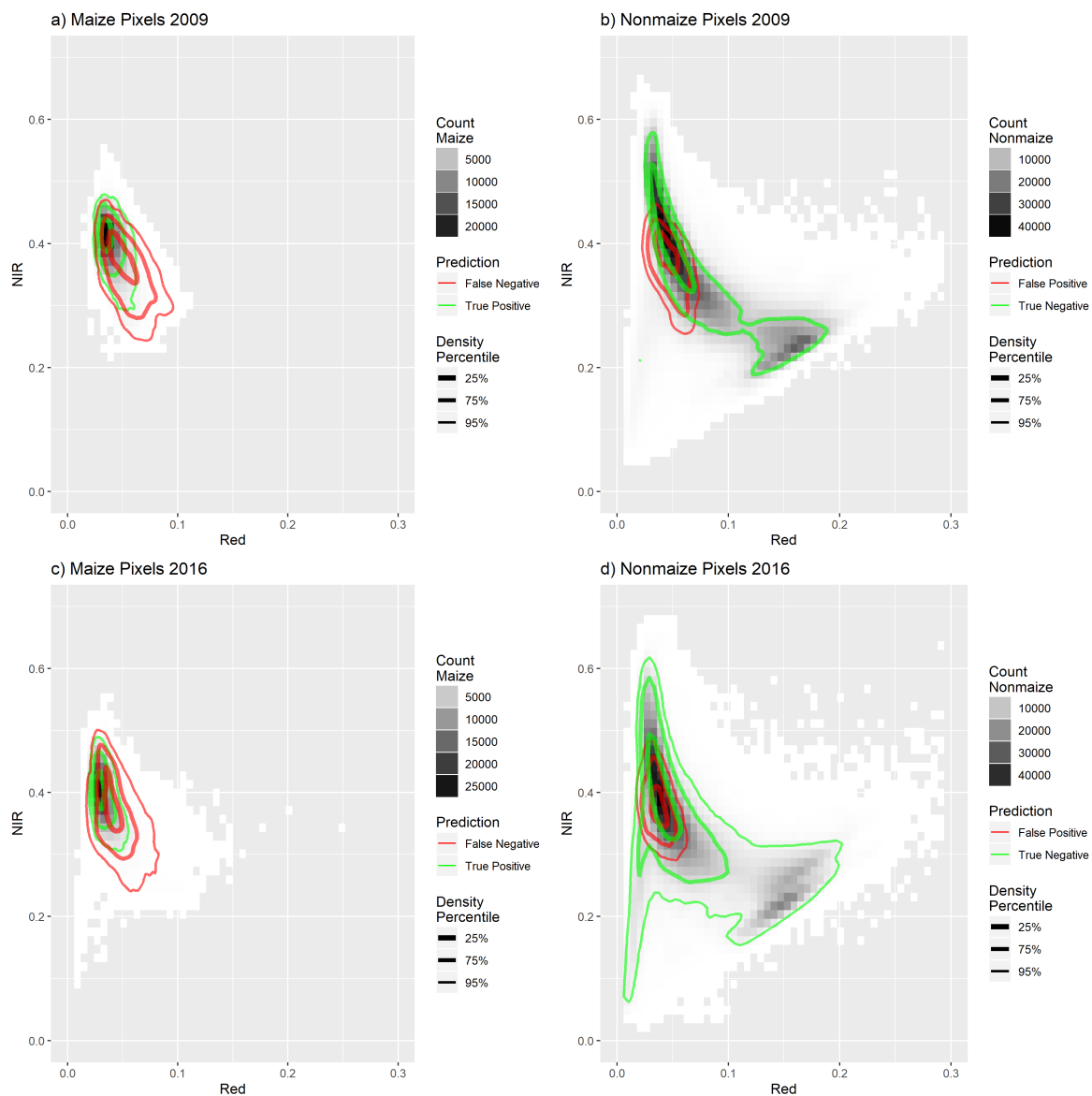
Figure 9. Spectral comparison of maize pixels 2009–2016.

This issue is ultimately impossible to deal with while using optical remote sensing alone, since it relies on the pigments of vegetation cover. Even if algorithms were applied to the satellite imagery to detect inhomogeneities in the spectra, excluding them would exclude a potentially large amount of total maize area in dry areas.

Key finding 2: *Disparity in estimated area and statistical records is substantially larger than validated modeled performances suggest.*

In Section 3.3, the estimated maize abundances were validated in two ways by comparing them to the reference, and by comparing them to the statistical record. Since the difference between estimation and reference was by far smaller than the difference between the estimation and the statistical record, there must be additional inconsistencies to consider. The disparity between the estimation necessarily consisted of two parts, error of the classification and error of the statistical data. The potential sources for errors in the classification were discussed above. The exact methodology of statistical records is unknown, but relies on farmers self-reporting their cultivation types based on the polygons of their fields. However, as mentioned in Section 2.3, some fields were missing a substantial amount of vegetation cover due to drought and other crop failures, so there naturally was a difference between

area with active maize vegetation cover versus area declared as maize by the farmer. This difference is very hard to determine, since the statistical record is not transparent and the data that it's based on is unavailable to researchers and the public. However, based on what is known about the quality of our estimation, it is likely that that the classification was closer to the truth than the statistical record.

5. Conclusions and Outlook

This study has shown that optical remote sensing is an adequate tool to monitor maize abundance in a spatially explicit way. Reference data could be generated reliably and efficiently by digitizing digital orthophotos. Optical remote sensing imagery could then be used to predict maize fields for any given pixel. To take the spatial relationships of all these pixels into account, we used two different approaches to incorporate spatial autocorrelation into our model. Regression Kriging, while very capable in theory, is not as beneficial when reference data are highly clustered. It is also technically not feasible when the sample count is very high, since it scales cubically. Gaussian Blur can help with residual noise, and scales only linearly, but is ultimately no systematic solution to the problem of residual autocorrelation. While we were not able to present a perfect way to deal with residual autocorrelation, we still underlined the importance of thorough residual analysis, since spatial classification has reliability indicators beyond overall accuracy measures or kappa values. The performance of machine learning classifiers on independent observations is well understood, but how to incorporate spatial relationships is still up for discussion. Combining pixels to objects and treating them as discrete entities might be a direct solution to the problem, known as Object Based Classification. It relies on segmentation techniques and high spatial resolution since homogeneous pixels are grouped together and treated as objects, while their aggregated attributes are then used with traditional classifiers like SVMs and Neural Networks [29]. Alternatively, computer vision approaches like Convolutional Neural Networks can be used to detect objects in remote sensing imagery, by computing additional kernel based feature layers [30].

The main technical limitation of this approach was the reliance on the reflectance of sunlight on leaf pigments. Firstly, this means there is simply no data when clouds cover the area of interests, which severely limits the availability of imagery, and makes it impossible to produce a time series for larger areas. Secondly, this makes it hard to include mostly brown or yellow canopies that are under severe drought stress. One way to deal with this is to use active microwave remote sensing with the combination of synthetic aperture radar (SAR) with optical remote sensing techniques. Due to the wavelength ranges of SAR sensors, they are mostly unaffected by clouds and other atmospheric attributes that are problematic for optical remote sensing. This enables the composition of gapless time series, not only from year to year, but with many observation points within one year [31].

Author Contributions: Conceptualization, M.G. and T.R.; methodology, M.G.; software, M.G.; validation, M.G.; formal analysis, M.G.; investigation,M.G.; resources, M.G. and T.R.; data curation, M.G.; writing—original draft preparation, M.G.; writing—review and editing, M.G.; visualization, M.G.; supervision, T.U. and C.E.; project administration, T.U. and C.E.

References

1. Lupp, G.; Albrecht, J.; Darbi, M.; Bastian, O. Ecosystem Services in Energy Crop Production—A Concept for Regulatory Measures in Spatial Planning? *J. Landsc. Ecol.* **2012**, *4*, 49–66, doi:10.2478/v10285-012-0043-y. [CrossRef]

2. Erneuerbare-Energien-Gesetz (EEG). Gesetz für den Ausbau erneuerbarer Energien vom 21. Juli 2014 (BGBl. I S. 1066), das durch Artikel 2 des Gesetzes vom 22. Dezember 2016 (BGBl. I S. 3106), 2016. Available online: www.gesetze-im-internet.de/eeg_2014/BJNR106610014.html (accessed on 16 September 2019).

3. Blanco-Canqui, H. Energy crops and their implications on soil and environment. *Agron. J.* **2010**, *102*, 403–419, doi:10.2134/agronj2009.0333. [CrossRef]

4. Ruf, T.; Gilcher, M.; Udelhoven, T.; Emmerling, C. Extended silage maize cultivation for bioenergy production—A case study in a low mountain range and its implications for soil health. *Agric. Syst.* **2019**, Submitted.

5. Yu, L.; Liang, L.; Wang, J.; Zhao, Y.; Cheng, Q.; Hu, L.; Yu, L.; Wang, X.; Zhu, P.; Li, X.; et al. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *Int. J. Remote Sens.* **2014**, *1161*, doi:10.1080/01431161.2014.930206. [CrossRef]

6. Khatami, R.; Mountrakis, G.; Stehman, S.V. Remote Sensing of Environment A meta-analysis of remote sensing research on supervised pixel-based land-cover image classi fi cation processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100, doi:10.1016/j.rse.2016.02.028. [CrossRef]

7. Muñoz-marí, J.; Bovolo, F.; Gómez-chova, L.; Bruzzone, L.; Camps-valls, G. Semisupervised One-Class Support VectorMachines. *IEEE Trans. Geosc. Remote Sens.* **2010**, *48*, 3188–3197. [CrossRef]

8. Chen, X.; Yin, D.; Chen, J.; Cao, X. Effect of training strategy for positive and unlabelled learning classification: Test on Landsat imagery. *Remote Sens. Lett.* **2016**, *7*, 1063–1072, doi:10.1080/2150704X.2016.1217437. [CrossRef]

9. Getis, A. Reflections on spatial autocorrelation. *Reg. Sci. Urban Econ.* **2007**, *37*, 491–496, doi:10.1016/j.regsciurbeco.2007.04.005. [CrossRef]

10. Wang, L.; Shi, C.; Diao, C.; Ji, W.; Yin, D. A survey of methods incorporating spatial information in image classification and spectral unmixing. *Int. J. Remote Sens.* **2016**, *37*, 3870–3910, doi:10.1080/01431161.2016.1204032. [CrossRef]

11. Millard, K.; Richardson, M. On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. *Remote Sens.* **2015**, *7*, 8489–8515. doi:10.3390/rs70708489. [CrossRef]

12. Lv, Z.; Li, G.; Benediktsson, J.A.; Zhang, Z.; Yan, J. Training sample refining method using an adaptive neighbor to improve the classification performance of very high-spatial resolution remote sensing images. *J. Appl. Remote Sens.* **2019**, *13*, 1, doi:10.1117/1.jrs.13.036501. [CrossRef]

13. Georganos, S.; Grippa, T.; Niang Gadiaga, A.; Linard, C.; Lennert, M.; Vanhuysse, S.; Mboga, N.; Wolff, E.; Kalogirou, S. Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* **2019**, 1–16, doi:10.1080/10106049.2019.1595177. [CrossRef]

14. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.M.; Gräler, B. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **2018**, *6*, e26693v3, doi:10.7287/peerj.preprints.26693v3. [CrossRef]

15. Belgiu, M.; Dragu, L. ISPRS Journal of Photogrammetry and Remote Sensing Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–26. [CrossRef]

16. Statistisches Landesamt Rheinland-Pfalz (Statistical Office Rhineland-Palatinate). Statistisches Jahrbuch Rheinland-Pfalz 2017, 2017. Available online: www.statistik.rlp.de/fileadmin/dokumente/jahrbuch/Jahrbuch2017.pdf (accessed on 16 September 2019).

17. DLR Mosel. Biogasanlagen in Rheinland-Pfalz 2017, 2018. Available online: api.edoweb-rlp.de/resource/edoweb:7019416 (accessed on 16 September 2019).

18. Borg, E. *Geodaten auf Bestellung—Das RapidEye Science Archive RESA*; German Aerospace Center: Neustrelitz, Germany, 2014.

19. Planet. Rapideye™ Imagery Product Specifications; 2016; p. 50. Available online: www.planet.com/products/satellite-imagery/files/160625-RapidEye%20Image-Product-Specifications.pdf (accessed on 16 September 2019).

20. Löhnertz, M.; Schlerf, M.; Seeling, S. Description of Vegetation Cover during the Growth Period and Crop Classification With Multitemporal High Resolution Spot Images. In *Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover 80 Land Cover and Land Use*; Technical Report; Center for Remote Sensing of Land Surfaces: Bonn, Germany, 2006.

21. Tanré, D.; Deroo, C.; Duhaut, P.; Herman, M.; Morcrette, J.J.; Perbos, J.; Deschamps, P.Y. Description of a computer code to simulate the satellite signal in the solar spectrum: The 5S code. *Int. J. Remote Sens.* **1990**, *11*, 659–668, doi:10.1080/01431169008955048. [CrossRef]

22. Hill, J.; Sturm, B. Radiometric correction of multitemporal thematic mapper data for use in agricultural land-cover classification and vegetation monitoring. *Int. J. Remote Sens.* **1991**, *12*, 1471–1491, doi:10.1080/01431169108955184. [CrossRef]

23. Hill, J.; Mehl, W.; Radeloff, V. Improved forest mapping by combining corrections of atmospheric and topographic effects in Landsat TM imagery. In *Sensors and Environmental Applications of Remote Sensing*; Balkema: Rotterdam, The Netherlands, 1995; pp. 143–151.

24. Landesamt für Vermessung und Geobasisinformation Rheinland-Pfalz (LVermGeo). *Digital Elevation Model with a Grid Width of 5 m (DEM5)*; State Office for Measurement and Geo Base Information Rhineland Palatinate: Koblenz, Germany, 2018.

25. Hengl, T.; Heuvelink, G.B.; Rossiter, D.G. About regression-kriging: From equations to case studies. *Comput. Geosci.* **2007**, *33*, 1301–1315, doi:10.1016/j.cageo.2007.05.001. [CrossRef]

26. Atkinson, P.M.; Lewis, P. Geostatistical classification for remote sensing: An introduction. *Comput. Geosci.* **2000**, *26*, 361–371. [CrossRef]

27. DLR Mosel. Daten zum Silomaisanbau in der Region Trier und in Rheinland-Pfalz. Available online: https://www.dlr-mosel.rlp.de/Internet/global/themen.nsf/d0e5087e9e1e8b79c1257abf0060c5df/b7c65bbd218ffd94c1257c1c004c2c99?OpenDocument (accessed on 16 September 2019).

28. *Administrative Geometries NUTS and LAU;* State Office for Measurement and Geo Base Information Rhineland Palatinate: Koblenz, Germany, 2011.

29. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. ISPRS Journal of Photogrammetry and Remote Sensing A review of supervised object-based land-cover image classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293, doi:10.1016/j.isprsjprs.2017.06.001. [CrossRef]

30. Cheng, G.; Yang, C.; Yao, X.; Guo, L.; Han, J. When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2811–2821, doi:10.1109/TGRS.2017.2783902. [CrossRef]

31. Joshi, N.; Reiche, J.; Kuemmerle, T.; Jepsen, M.; Mitchard, E.; Grogan, K.; Meyfroidt, P.; Hostert, P.; Ehammer, A.; Ryan, C.; et al. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sens.* **2016**, *8*, 70, doi:10.3390/rs8010070. [CrossRef]

3 Implications of Bioenergy Cropping for Soil: Remote Sensing Identification of Silage Maize Cultivation and Risk Assessment Concerning Soil Erosion and Compaction

Ruf, T., Gilcher, M., Emmerling, C., & Udelhoven, T. (2021). Implications of Bioenergy Cropping for Soil: Remote Sensing Identification of Silage Maize Cultivation and Risk Assessment Concerning Soil Erosion and Compaction. Land, 10(2073-445X).

land

Article

Implications of Bioenergy Cropping for Soil: Remote Sensing Identification of Silage Maize Cultivation and Risk Assessment Concerning Soil Erosion and Compaction

Thorsten Ruf [1,*] **, Mario Gilcher** [2]**, Thomas Udelhoven** [2] **and Christoph Emmerling** [1]

[1] Department of Soil Science, Faculty of Regional and Environmental Sciences, University of Trier, 54296 Trier, Germany; emmerling@uni-trier.de

[2] Department of Environmental Remote Sensing and Geoinformatics, Faculty of Regional and Environmental Sciences, University of Trier, 54296 Trier, Germany; gilcher@uni-trier.de (M.G.); udelhoven@uni-trier.de (T.U.)

[*] Correspondence: ruf@uni-trier.de; Tel.: +49-651-201-3581

Abstract: Energy transition strategies in Germany have led to an expansion of energy crop cultivation in landscape, with silage maize as most valuable feedstock. The changes in the traditional cropping systems, with increasing shares of maize, raised concerns about the sustainability of agricultural feedstock production regarding threats to soil health. However, spatially explicit data about silage maize cultivation are missing; thus, implications for soil cannot be estimated in a precise way. With this study, we firstly aimed to track the fields cultivated with maize based on remote sensing data. Secondly, available soil data were target-specifically processed to determine the site-specific vulnerability of the soils for erosion and compaction. The generated, spatially-explicit data served as basis for a differentiated analysis of the development of the agricultural biogas sector, associated maize cultivation and its implications for soil health. In the study area, located in a low mountain range region in Western Germany, the number and capacity of biogas producing units increased by 25 installations and 10,163 kW from 2009 to 2016. The remote sensing-based classification approach showed that the maize cultivation area was expanded by 16% from 7305 to 8447 hectares. Thus, maize cultivation accounted for about 20% of the arable land use; however, with distinct local differences. Significant shares of about 30% of the maize cultivation was done on fields that show at least high potentials for soil erosion exceeding 25 t soil ha^{-1} a^{-1}. Furthermore, about 10% of the maize cultivation was done on fields that pedogenetically show an elevated risk for soil compaction. In order to reach more sustainable cultivation systems of feedstock for anaerobic digestion, changes in cultivated crops and management strategies are urgently required, particularly against first signs of climate change. The presented approach can regionally be modified in order to develop site-adapted, sustainable bioenergy cropping systems.

Keywords: biomethantion; biogas; environmental impact; land use; risk assessment; soil management strategy

1. Introduction

The cultivation of energy crops on arable land aiming to substitute fossil fuels has significantly increased during the past decade [1]. In the course of Germany's energy transition strategy, numerous farmers have taken the opportunity to integrate agricultural energy production as a second mainstay with long-term governmentally guaranteed feed-in-tariffs [2,3].

Since silage maize was identified to be the most economically important crop for anaerobic digestion under temperate conditions, the cultivation area for energy crops has been significantly expanded; in Germany for example, from 200,000 to more than 900,000 hectares (ha) in the last decade [4–8]. However, introducing energy crops as a substrate for anaerobic digestion has implied changes in the traditional, cereal, and

grassland dominated agricultural system. As a consequence, the share of silage maize in crop rotations has increased and it is being cultivated on arable land, which has formerly never been used for the growing of row crops.

In the course of rising awareness about negative externalities associated with energy crop cultivation, the sustainability of silage maize cultivation has become a lively debate. Aspects, such as depletion of soil organic matter, which would certainly counteract the carbon dioxide saving targets, loss of agrobiodiversity, and the increased vulnerability of maize stands to soil erosion and compaction, are currently discussed, mainly based on statistical data about expanded silage maize cultivation [9–12]. However, using statistical data referring to large-scale areas like national territories or federal states to describe changes in the agricultural system and to evaluate consequences for the environment is hardly helpful and largely ignores contributing factors governed by site-specific conditions. For example, the erosion potential depends on the cultivated crop and specific management operations like tillage intensity but also on the terrain inclination, texture-governed soil erodibility, slope length, and erosivity of rainfall. Thus, a fact-based debate about hazards and implications of silage maize cultivation necessarily requires spatially-explicit data of the cultivated areas and additional information, including soil and climatic data, as well as structural parameters of the landscape and the agroecosystem. Simultaneously, such an approach requires to be conducted on a scale that allows matching spatially heterogeneity of environmental factors.

The present study aimed to reach conclusions on the implications of expanded silage maize cultivation for soil threats, particularly erosion and compaction in a low mountainous agricultural region. However, spatially-explicit census data were not available. Thus, firstly, a model application for detecting maize cultivation using RapidEye satellite images (Planet Labs Inc., San Francisco, CA, USA) was established for the years 2009 and 2016. We hypothesized that the established remote sensing-based classification approach allows tracking changes in the maize cultivation area with progressing expansion of agricultural biogas producing units (BPU). Therefore, additional data sources like statistical data and registers about BPUs were systematically screened. Secondly, a target-specific analysis of existing soil data was conducted to (i) identify pedotopes that are susceptible for soil compaction during harvest of silage maize in late autumn and (ii) to model the susceptibility of the soils for erosion by water (hereinafter referred to as 'erosion'). In order to achieve the overarching goal of this study, the results of the maize classification and the evaluation of the hazard potential caused by maize cultivation on the soil resource were combined. We hypothesized that the increased demand of feedstocks for anaerobic digestion led to pressure on soil resources and an expansion of the maize cultivation on sites that appear not suitable in a sustainable manner.

2. Materials and Methods

2.1. Characterization of the Study Area

The study was conducted in the administrative district 'Eifelkreis Bitburg-Prüm' (in the following referred to as 'study area'), which is located in the Federal State of Rhineland-Palatinate in the southwestern part of Germany. The rural nature of the study area is underlined by a population density of 59.8 inhabitants km^{-2} and the dominance of agricultural land use, which accounts for 53.2% of the total area [13]. The study area was selected because statistical data indicate an intense development of the biogas sector from two BPUs in the year 2000 to 62 BPUs in the year 2016; associated with a distinct increase in silage maize cultivation (Figure 1) [14–16].
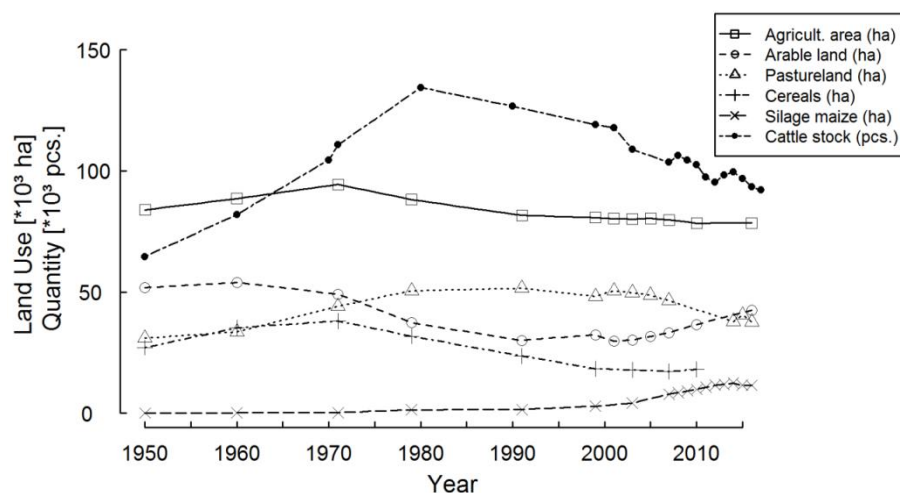
Figure 1. Development of the land-use (cereals excluding silage maize) and of the agricultural structure in the study region during the last decades based on census data [14,15].

The study area has a size of 1626 km^2 and spans an inclined plateau; the elevation climbs from 300 to 550 m a.s.l. in a south–north direction and peaks in the northern parts with hills of nearly 700 m a.s.l. The plateau is pervaded by numerous deep valley incisions with the main flow direction to the south. This pattern leads to a wavy, tightly structured landscape [17,18].

As a result of the elevation profile, the climatic conditions in the study area show a large latitudinal gradient (see Supplementary Material Table S1). Its southernmost parts are influenced by the viticultural climate of the Moselle valley with mild winters (mean annual temperature (MAT): 9.0 °C; mean annual precipitation (MAP): 861 mm a^{-1}), whereas the northern part shows a harsh low mountain range climate with distinctly lower MAT, as well as cold and long-lasting, precipitation-rich winters (MAT: 7.0 °C; MAP: 1250 mm a^{-1}) [19]. This general pattern is disrupted by the deep valleys which are characterized by considerably milder winters and warmer summers.

In the northern part of the study area, Devonian rocks are dominating whereas in the southern part, a small-scale co-existence of Triassic and Jurassic rocks can be found which are moreover partially overlaid by varying amounts of Quaternary sediments (loess) [20]. From these parent materials, in the northern part of the study area, silty to clayey Regosols and Cambisols developed. Soils that developed from Lower Triassic sediment rocks in the central and eastern part of the study area can be classified as sandy and clayey Regosols and Cambisols with alternating levels of loess. The small scale variability of parent materials in the southern part results in a mosaic of soils; the textures are mainly silt-loam, occasionally sandy and clayey. A wide range of soil types are present: Regosols, Cambisols, Stagnosols, Luvisols, and Vertisols whereby the both last mentioned soil types frequently tend to show signs of stagnant water [20–22].

Due to the gradient in climatic conditions and soils, the agricultural land in the northern part of the study area was traditionally dominated by grassland use whereas the more southern parts were characterized by cereal-based crop cultivation. Grain maize is not cultivated in the study area.

2.2. Remote Sensing Based Classification of Maize Fields

Farmers themselves, as well as governmental bodies, generally track crop types for cultivated areas very thoroughly. This kind of data is, unfortunately, rarely available to the public, nor to researchers, mostly because of privacy issues. Surveying the crop types manually by going out to the field or by asking farmers to access their records leads to a sparse or heavily spatially biased set of reference data. In order to have spatial explicit data about silage maize cultivation at hand, a remote sensing-based classification was conducted.

2.2.1. Selection and Pre-Processing of Remote Sensing Data

The classification approach was designed for a binary classification of maize and non-maize cultivation. The availability of suitable, gap-free coverages of the entire study area was of vital importance. The acquisition window was limited by the phenologic cycle of maize, which reaches the peak vegetation cover around the end of August. Using earlier imagery would result in emerging maize fields that show large soil fractions, while after August, there is the potential of early harvests. The third major constraint was the availability of high-resolution aerial photos on GoogleEarth, since these were used to produce reference data. After reviewing all available products, two RapidEye scenes in 2009 and 2016 met all three requirements. These two scenes were then atmospherically corrected based on the 5S algorithm [23] using the software AtCPro [24,25] and mosaicked together to one single image per year. A flowchart showing the steps of data processing and modeling is presented as Supplementary Material Figure S1.

2.2.2. Preparation of Reference Data

Since the classification of around 87,000 hectares of agricultural land requires a large amount of uniformly distributed training data, a visual survey approach was pursued. Based on GoogleEarth aerial imagery, it was trivial to distinguish maize fields from other agricultural crops, mostly cereals, and grassland. This is because maize is a high-growing row crop and has thus a very apparent texture which facilitates visual object recognition.

To guarantee an even distribution of training fields across the study area, it was split in subdivisions with 9 by 12 grids, with each cell spanning 5 by 5 km (Figure 2). The main goal was to select a representative set of training fields with at least three polygons, each with and without maize in every grid. Overall 2566 polygons were digitized for both years combined, amounting to almost 2000 hectares of maize fields per year. This represents about 20% of the overall acreage; the grid-cell approach in selecting training field should guarantee that the selected pixels are representative for the entire study area.

2.2.3. Pixel-Based Classification of Imagery

The classifications of all, roughly, 35,000,000 RapidEye pixels was done individually for both years using the random forest implementation of the ranger package in R programming language version 3.3.2 [26,27]. The random forest approach is a classifier that is based on the votes of randomly generated decision trees [28]. This particular approach was chosen because it scales well with a high number of features, as well as observations [27]. Furthermore, it is widely used for pixel-based land cover classification, because of its high accuracy and insensitivity to overfitting [29]. The classifications of both years trained with relatively high out-of-bag Cohens kappa values (index derived from expected agreement by chance, and observed agreement of classification with ground truth), but the classifier showed a slightly weaker performance in 2016 (Table 1). The high sensitivity values (true positive rate: quotient of pixels correctly classified as maize and pixels that are truly maize) mean that essentially all digitized maize fields where identified. The comparatively low specificity (true negative rate: quotient of pixels correctly classified as nonmaize and pixels that are truly nonmaize) indicates that there were a lot of false positives, meaning pixels that were classified as maize, but digitized as non-maize. After the classification with ranger, the probabilities where smoothed with a Gaussian kernel (see also Supplementary Material Figure S1). In a nutshell, the Gaussian kernel is a probability function. In the process of classifying a certain pixel, it takes into account the classification of the surrounding pixels in order to reduces the amount of noise in the prediction.
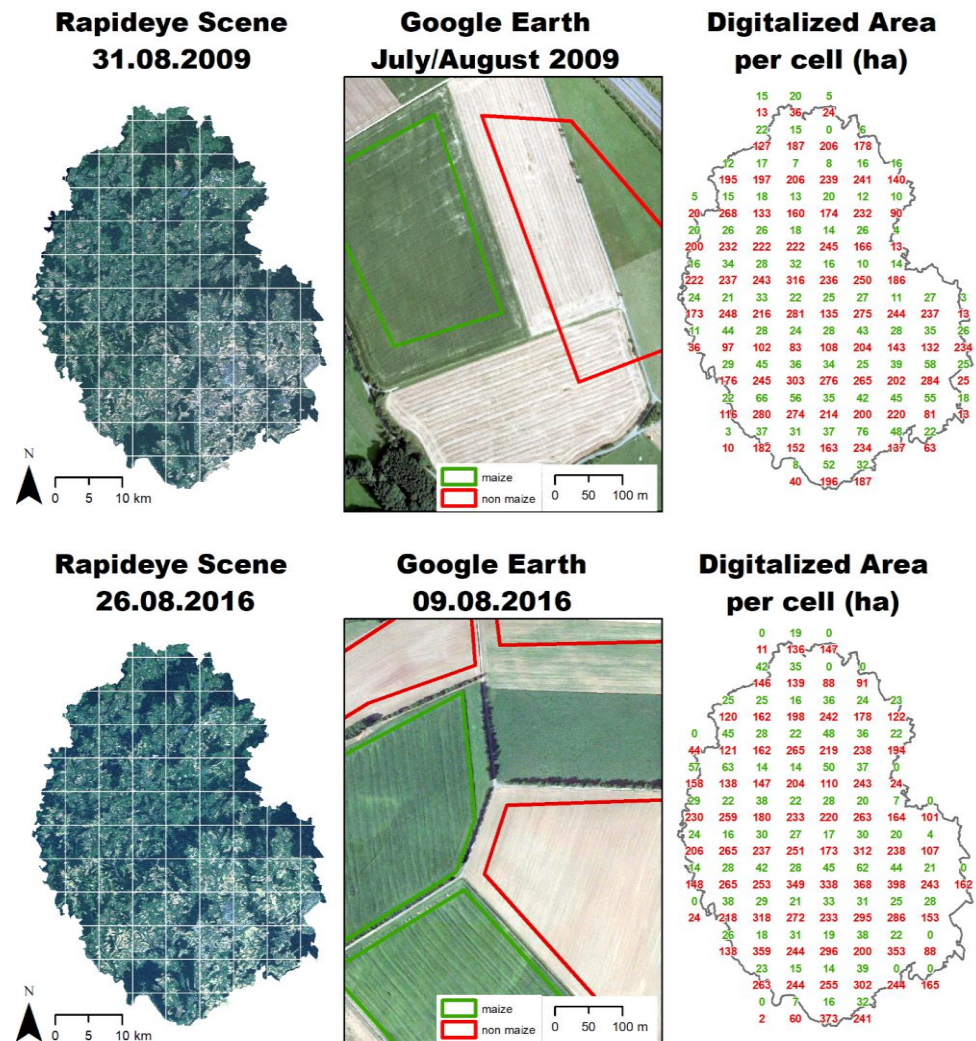
Figure 2. RapidEye coverage of the study area and manual digitization of reference data for the years 2009 (**top**) and 2016 (**bottom**).

Table 1. Summary of parameters specifying the efficiency of the selected approach for classification of maize fields.

Year	Sensitivity	Specificity	Kappa
2009	>0.99	~0.97	~0.97
2016	>0.99	~0.91	~0.93

2.3. Agricultural Biogas Producing Units

As defined in our hypotheses, we assumed that agricultural biogas plants are the main drivers for expanded silage maize cultivation. To be able to relate silage maize cultivation and feedstock demands, the locations of all agricultural BPUs, as well as their installed capacities in the course of time, were compiled based on (i) administrative data, (ii) 'grey' directories, and (iii) surveys of plant operators from third parties for internal use (references cannot be stated due to confidentiality agreement) due to the fact that no dataset was complete and up-to date, and contradictory data existed. The installed capacities of combined heat and power units located at remote places to supply local demands were ascribed to the digestion units since the demand for substrates is given there.

2.4. Determining of Pedotopes Susceptible for Compaction during Maize Harvesting

Soils that are particularly susceptible for soil compaction, with respect to maize harvesting operations in late autumn typically belong to the classes of Stagnosols and Planosols. These soils show horizons with very low permeability for water in the subsoil ('Sd' or 'Bg' horizon, acc. to [30] and [31], respectively) and tend to show waterlogging resulting from abundant autumn precipitation and low evaporation rates. Thus, in the period of silage maize harvesting, i.e., in late autumn, Stagnosols and Planosols frequently show high soil water contents resulting in low mechanical carrying capacities. These circumstances commonly result in serious soil compaction and structural damages caused by the heavy equipment used for silage maize harvesting.

For the purpose of this study, we performed a risk assessment for the soils concerning their vulnerability for soil compaction. Classification of soil types was based on the site specific sequence of soil horizons recorded in the framework of German land appraisal, which still represents the data basis for soils with the highest resolution and coverage (national law on farm land appraisal: [32]). The size of soil units in this framework is about one hectare. The sequence of the soil horizons was compared to a self-developed, predefined classification key using RStudio programming language version 3.3.2 [26]. In case that the sequence of soil horizons was in compliance with the key, a soil type was classified. Sequences of soil horizons that were not listed in the key were, as far as they were interpretable and meaningful, manually added to the key in iterative steps (flowchart of the classification presented as Supplementary Material Figure S2). However, data were only available for 55.3% of the agricultural area and for 36.3% the data basis was sufficient for soil type classification. For further analysis, soil types were grouped according to their specific vulnerability for soil compaction during harvesting operations of silage maize in late autumn as shown in Table 2.

Table 2. Classification of soil types regarding their susceptibility for soil compaction during harvesting operations of silage maize.

Susceptibility for Soil Compaction	Soil Types
High	Stagnosols, Planosols
Medium	• Transition soil types that show stagnic or gleyic properties but are not assigned to the class of Stagnosols or Planosols such as e.g., stagnic Cambisols, stagnic Fluvisol or stagnic Anthrosols. • Soils that are influenced by groundwater like Gleysols
Low	All other soil types

2.5. Determining of Site-Specific Susceptibility for Soil Erosion by Water

Natural proneness of the agricultural sites for soil erosion (E_{nat}) was modeled on pixel basis (grid resolution: 5 m × 5 m) according to [33], using a simplified version of the universal soil loss equation (USLE) [34]. Modeling was done using ESRIs ArcMap 10.5. Input variables were:

- Slope factor (S): Slopes were calculated based on a digital elevation model with a resolution of 5 m (DEM5) provided by the [17]. The slope allowed calculating the slope factor using equation 7 given in [33].
- Soil erodibility factor (K): Derived from the soil texture of the topsoil determined in the framework of the German land appraisal. Textures were then assigned to K factors and corrected by soil skeleton and organic carbon contents using Equations (3)–(6), given in [33]. Values of the K factor for the study area with a raster resolution of 5 m by 5 m were provided by the State Office for Geology and Mining Rhineland Palatinate.
- Rainfall and runoff factor (R): R factors were calculated using the following regression equation: R = 0.0788 * mean annual precipitation (in mm) * −2.82 which was valid for the Federal Republic of Germany. Values of the R factor for the study area with a

raster resolution of 1 km by 1 km were provided by the State Office for Geology and Mining Rhineland Palatinate.

S, K, and R factors were multiplied for each raster cell. E_{nat} values were classified using the scheme presented in Table 3. Compared to the [33], the classes of erosion potential were considerably enlarged in order to take into account the fact that the study region is located in a low mountain region (consider also Figure 7).

Table 3. Assignment of calculated soil losses (E_{nat}) to classes of erosion potential.

Soil Erosion Potential	Soil Loss acc. to E_{nat} $t\ ha^{-1}\ a^{-1}$
Very low	<5
Low	5–10
Medium	10–25
High	25–50
Very high	>50

2.6. Data Analysis and Mapping

The analysis of data was done using RStudio programming language version 3.3.2 [26] and ArcMap 10.5 (ESRI Inc.). Maps were created using ArcMap 10.2.

3. Results and Discussions

3.1. Maize Classification

The classification approach to detect silage maize cropping sites has estimated a cultivation area of 7305 ha in 2009 and 8447 ha in 2016. In contrast to that, the official statistical records state values of 9147 ha and 11,496 ha in 2009 and 2016, respectively [15]. This means that the classifier underestimated the total maize cultivation area by roughly 2000 (2009) and 3000 ha (2016) compared to the official statistics (Table 4).

Table 4. Results of the silage maize classification approach using remote-sensing data compared to statistical data.

Year	Remote-Sensing Based Approach	Statistical Data	Difference between stat. Data and Remote-Sensing Based Approach	
	ha	ha	ha	%
2009	7305	9147	−1842	−20.1
2016	8447	11496	−3049	−26.5

The substantial 'underestimation' of the classifier compared to the statistical data can be traced back to different methodological approaches. Whereas statistical data are field based, the remote-sensing classification approach provides pixel-based results. Thus, areas in the maize fields exhibiting spots of bare soil due to late sowing dates, erosion processes, damages by wild boars, and weak field emergence were (correctly) not classified as maize but were included in the official statistics as maize (Figure 3).

Moreover, due to the south to north gradient in vegetative development (see Supplementary Material Table S1), there is a quite short time slot suitable for classification based on optical satellite images. This weakness of the chosen approach could be overcome by using radar data, which makes image recording independent from cloud coverage. Moreover, a complete time series of maize cultivation sites would also allow answering the questions of mono-cultural growth of silage maize. However, for the years of this investigation, no suitable data, such as nowadays, e.g., provided by the Sentinel 2 platforms, were available.
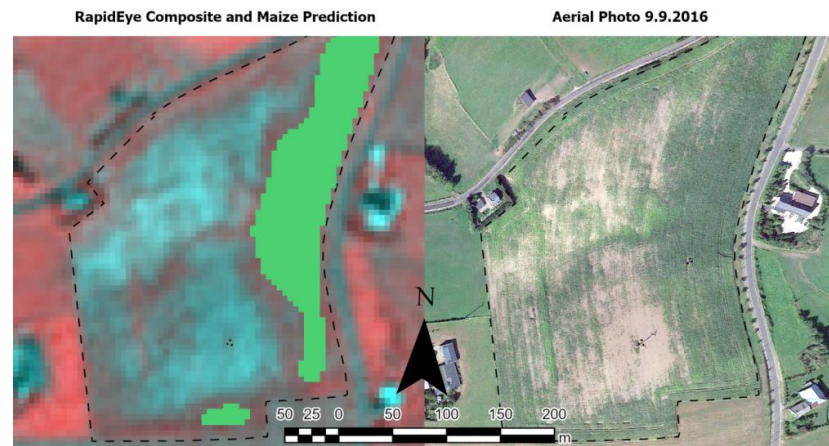
Figure 3. Visual clarification of different approaches of maize classification. Data included in the official statistics include the whole area of the maize field as shown by the dashed lines including large spots of bare soil as visible in the aerial photo. In contrast, the remote sensing-based classification approach only classifies pixel as 'maize' that are in fact maize-covered as shown by the prediction in green colour in the RapidEye composite.

3.2. Extent and Changes in Silage Maize Cultivation in the Study Area

In 2009 and 2016, 8.4 and 9.7% of the agricultural area were covered with maize, respectively. By the fact that more than 50% of the agricultural area was grassland and the size of arable land was increased from 36,610 ha to 42,439 ha in the same period (Figure 1), the share of silage maize on arable land amounted to about 20% in both years. However, large regional differences were observed (Figure 4). In 14 (2009) and 15 (2016) municipalities, silage maize was not cultivated. In the municipality with the highest share of silage maize, the cultivation area accounted for 29.9 (2009) and 34.2% (2016) of the agricultural area. Unfortunately, the share of arable land on the agricultural area is currently not available at municipal level, resulting in an inaccuracy in describing the relative proportion of silage maize cultivation in landscape.
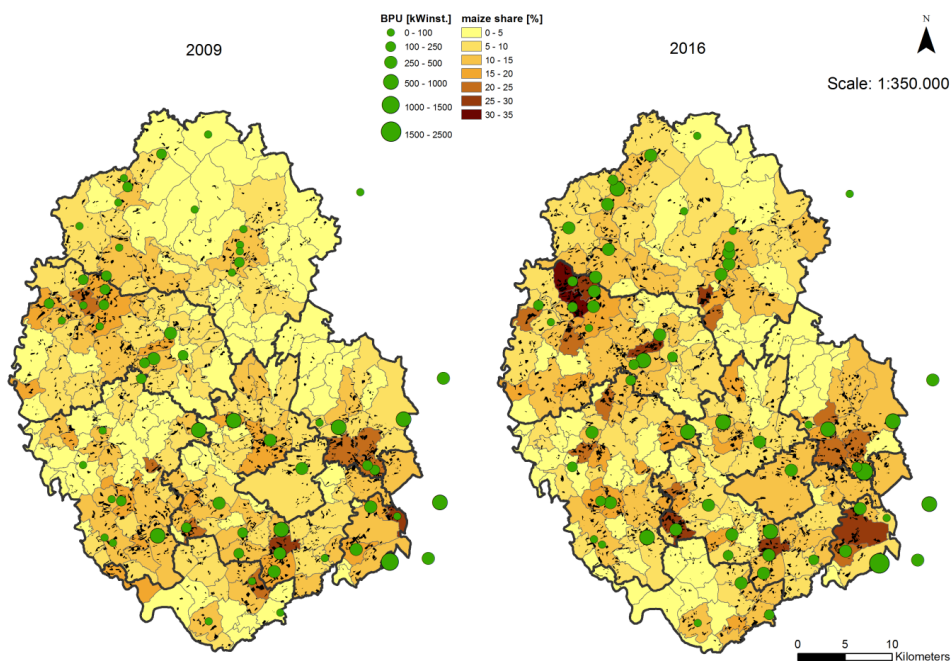

Figure 4. Representation of silage maize cultivation (black spots) and of the share of silage maize cultivation on agricultural land use based on the municipalities of the study area. Moreover, the location and size of the biogas producing units (BPU) is presented.

The changes in silage maize cultivation at municipal level apparently correlate with newly constructed or increased installed capacity of existing BPUs. Although, silage maize cultivated as feedstock for BPUs cannot be gathered separately from silage maize cultivated as fodder for ruminant production, there is evidence that the demands of BPUs are the drivers for expanded cultivation. The cattle stock in the study area steadily declines for around three decades (Figure 1). Moreover, regardless the final uses of the silage maize, the effects on soils are similar as there is no difference in the cultivation system. Figure 5 also implies that the feedstock sources for a certain BPU overlapped the borders of municipalities. Hot spots of increased silage maize cultivation and BPU capacity from 2009 to 2016 were detected in municipalities in the north-western part of the study region. These municipalities also showed the highest shares of silage maize on agricultural soil use in 2016 (Figure 4).
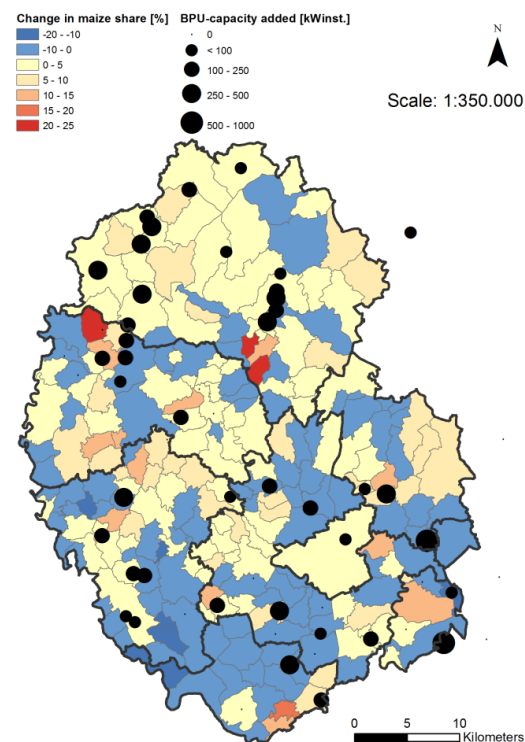

Figure 5. Representation of the changes in silage maize cultivation and the installed capacity of biogas producing units (BPU) from 2009 to 2016.

3.3. Maize Cultivation on Pedotopes Susceptible for Soil Compaction

Overall, 36.3% or 31,706 ha of the total agricultural area could be classified according to their susceptibility for soil compaction (Figure 6). Under the assumption that available data are representative for the whole study area, this would mean that 90.7, 7.0, and 2.3% show a low, medium, and high vulnerability for soil compaction, respectively (Table 5). Although the approach for classification of soils that are particularly vulnerable for compaction was suitable for an unbiased derivation of soil types from sequences of soil horizons given in the data of Germanys' land appraisal, the final results were not satisfying at all. The dataset was incomplete, since numerous soil profile descriptions were not available in a digital form or do not exist (M. Goldschmitt, Landesamt für Geologie und Bergbau Rheinland-Pfalz, personal communication, 02.07.2015). Moreover, lots of site descriptions showed apparent mistakes; thus, a classification was also not possible for these sites.
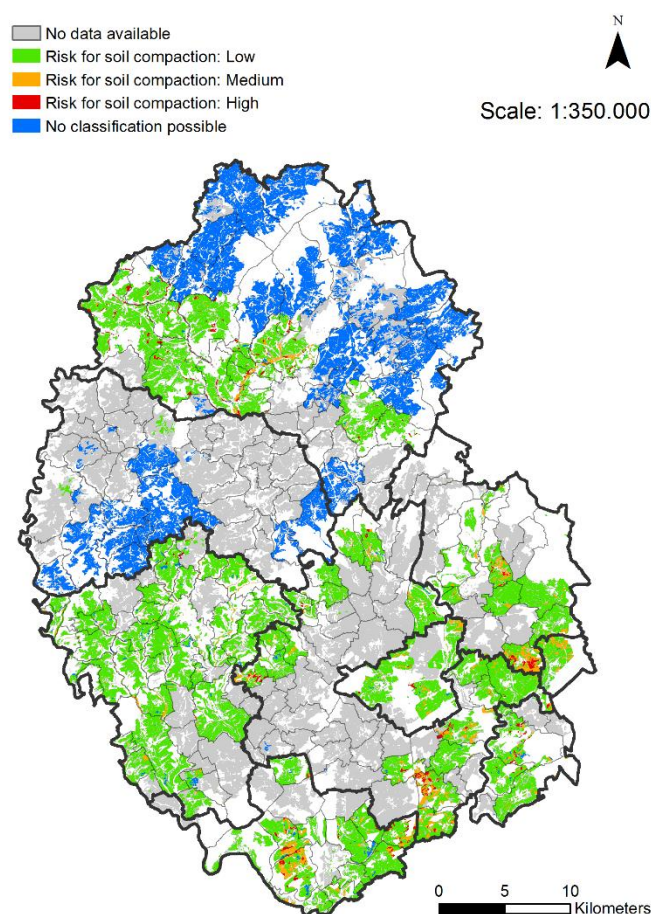
Figure 6. Representation of the susceptibility of the agricultural sites for soil compaction during harvesting action of silage maize. 'Classification not possible' refers to sites for which the soil profile descriptions showed apparent mistakes. When soil profile descriptions were not existing or not available in digital form, they were categorized as 'No data available'. Areas shown in white colour are non-agricultural sites.

Table 5. Parameters specifying the quality of the data basis for the soil type classification approach. Assignment of the agricultural area as well as maize cultivation in 2009 and 2016 to the classified susceptibility for soil compaction.

		Agricultural Area		Maize 2009		Maize 2016	
		ha	%	ha	%	ha	%
	Total	87,422	100	7305	100	8447	100
	Classifiable	31,706	36.3	2867	39.2	3249	38.5
Vulnerability for soil compaction	**Low**	28,758	90.7	2567	89.5	2997	92.2
	Medium	2226	7.0	252	8.8	195	6.0
	High	722	2.3	48	1.7	57	1.8

Based on the authors' knowledge about soil type distribution in the study area, it appears that the soil profile descriptions from German land appraisal poorly represent the spatial heterogeneity of soil types. This becomes evident from large areas characterized as a single soil type and may result from the fact that the soil units in the framework of the German land appraisal should have a size of about one hectare (M. Beck, valuator of land appraisal in the Federal District of St. Wendel, personal communication, 18.10.2018). Likely, this particularly leads to an underestimation of Stagnosols, which show a patchy, small-scale distribution in low-mountain regions, particularly in dips, on planes, and

at foots of slopes [35,36], which may not be well represented in the soil valuation units. Thirdly, the borders of the local authorities (thick lines in Figure 6) also mark breaking lines. It may be hypothesized that this indicates different pedological knowledge of the land valuators. Generally, valuators of land appraisal in Germany are no experts in soil science and may thus overlook redoximorphous features indicating stagnant water in soils [32].

In contrast to other approaches that coupled soil physical and hydrological models to derive information about the trafficability of soils in the course of time [37], the approach used in this study used a single information base aiming to conclude for the potential susceptibility for compaction based on specific soil types associated with harvesting operations of silage maize. Thus, the amount of input data necessary to conclude for soil types is much lower compared for example to the approach of [38], who used pedotransfer functions and data about bulk density and the pre-compression stress in order to estimate the soil compaction hazard. Furthermore, sophisticated modeling approaches of the soils' carrying capacity at specific points of time were not meaningful in this case because of (i) the insufficient data basis, (ii) the large size of study area, and (ii) the impossibility to validate the modelled data. Moreover, such approaches would add a temporal dimension to the data analysis as soil water contents show a high temporal dynamics.

However, similar to the concept perused in this study, also the more sophisticated approach of [38] does not allow deriving time-resolved maps of the risk for soil compaction.

As a result of the fragmentary data basis (Section 3.3, Figure 6), less than 40 %, or 2867 (2009) and 3249 ha (2016), of silage maize cultivation could be overlapped with the map of soil compaction risk (Table 5). Thereof, 8.8 (2009) and 6.0% (2016) were cultivated on soils with a medium risk for compaction and slightly less than 2.0 % in both years on soils showing a high risk for compaction. Extrapolated to the whole study area, this would mean that approximately 500 to 600 ha of silage maize cultivation was done on soils that show a medium risk for soil compaction during silage maize harvesting. Another 150 to 160 ha of silage maize was grown on sites that are highly susceptible for compaction; both assuming that the available data are representative for the whole study area.

Due to the above mentioned deficiencies in raw data, these values should be regarded as a rough estimate. Furthermore, by extrapolating to the whole study area, the spatial explicit character of the study approach gets lost. Nonetheless, the presented example impressively shows the difficulties in conducting rather simple risk evaluations due to limitations in suitable and available input data. Although comprehensive soil maps exist, they were also finally based on the same basis of data and regionalized by geostatistical methods. Thus, subject to these conditions, more complex and virtually more precise methods, such as pedotransfer functions, are therefore currently not applicable in a landscape context.

To reduce the risk for soil compaction, a substitution from silage maize to perennial energy crops (PECs) can be suggested. In this aspect the benefits of PECs result from (i) earlier harvesting under dryer soil conditions and (ii) the permanent, intense routing system both enhancing the carrying capacity of the soil [39,40].

3.4. Natural Susceptibility of the Soils for Erosion and Threats to Soil Due to Maize Cultivation

The modeling of the natural susceptibility of the soils for erosion processes revealed a small-scale pattern typical for the conditions in a low mountain landscape with a mean E_{nat} value of 31.1 t ha^{-1} a^{-1} (Figure 7). A total of 20% of the agricultural sites were classified to have a low or very low susceptibility for soil erosion with natural erosion rates <10 t ha^{-1} a^{-1} (Table 6). At the same time, nearly 50% of the agricultural area were classified to have a high to very high susceptibility for soil erosion (E_{nat} > 25 t ha^{-1} a^{-1}).
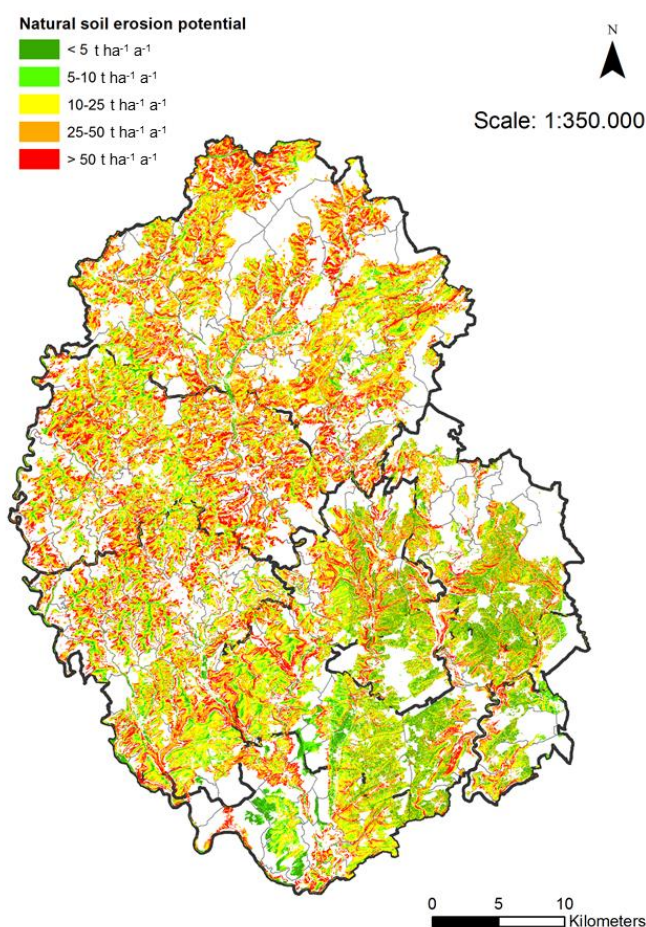
Figure 7. Map showing the calculated natural potential for soil erosion (E_{nat}) in the study area. Areas shown in white colour are non-agricultural sites and were thus excluded from erosion modeling.

Table 6. Assignment of the total agricultural area as well as maize and non-maize (all agricultural soil uses except maize) cultivation in 2009 and 2016 to the classified natural soil erosion potential (E_{nat}).

Erosion Potential	Agricultural Area		2009				2016			
			Maize		Non-Maize		Maize		Non-Maize	
	ha	%	ha	%	ha	%	ha	%	ha	%
Very low	8335	9.5	724	10.4	7611	9.5	791	9.7	7545	9.5
Low	9268	10.6	1034	14.8	8233	10.2	1147	14.1	8121	10.2
Medium	29,414	33.7	3045	43.6	26,369	32.8	3508	43.2	25,906	32.7
High	25,126	28.8	1846	26.5	23,280	29.0	2249	27.7	22,877	28.9
Very high	15,256	17.5	330	4.7	14,926	18.6	426	5.3	14830	18.7

The rainfall ('R') factor showed, similar to the topography ('S') factor, a distinct south to north gradient finally leading to a high natural soil erosion potential throughout the northern part of the study region. In contrast to that, the erodibility of the soils (K factor) in the southern part of the study region is considerably elevated due to higher silt contents of the soils. Nonetheless, in the southern part, the susceptibility for soil erosion is generally lower than in the northern part resulting from the more gentle slopes. For the arable land in Rhineland Palatinate, [41] calculated mean actual erosion rates (E_{nat} values multiplied by the cropping management factor C) of 11.2 ± 22.9 t ha^{-1} a^{-1}. Thus, the values calculated for the study region appear valid considering the low mountain character of the study region and the large standard deviation stated by [41].

In 2009 and 2016, more than 30% of the total silage maize cultivation was done on sites showing at least a high potential for soil erosion, whereby the area of silage maize cultivated on these sites increased from 2176 to 2675 ha. The mean E_{nat} value of fields cultivated with silage maize increased from 21.1 to 21.9 t ha^{-1} a^{-1}. Thus, it can be concluded that the silage maize cultivation was distinctly expanded on sites that show high or very high erosion potential—sites that are subject to the cross compliance regulations of the European Union (CC$_{water}$1: E_{nat} 15–27.5 t ha^{-1} a^{-1}; CC$_{water}$2: E_{nat} > 27.5 t ha^{-1} a^{-1}) and national laws in order to 'prevent soil erosion' [42–44].

Finally, this indicates that the cultivation of silage maize is done largely ignoring the site conditions. On the one hand, this confirms the statement of [45] that "farmers clearly did not consider erosion in their management decisions like field size or selection of crops". On the other hand, it expresses the scarcity of suitable agricultural land. Political incentives for feedstock cultivation have made a crucial contribution to expand silage maize cultivation to inadequate sites thus raising external costs of bioenergy production. Although no data about the specific management practice are available, strategies to prevent soil erosion are to our experience no common practice in the study area. For example, undersown crops, minimum tillage methods, or changes in cultivated crops may potentially reduce soil erosion rates by factors between two and ten [33,46,47].

Even worse, against the background of progressing climate change, with increased frequency and intensity of heavy precipitation events, particularly in spring [48], when maize fields show mostly bare soil, the calculated values for soil erosion rates are likely no more than a conservative estimate. In particular, the erosivity of the rainfall seems to be a weak point in soil erosion modeling likely leading to a significant underestimation of the currently prevailing erosion rates. Recent research has shown that presently used rain erosivity factors significantly underestimate the conditions by (i) methodological reasons and (ii) resulting from climate changes. Firstly, state of the art methods (rain radar), have revealed that R factors are about 60% higher than those presently used [49], which were based on the analysis of ombrometers with a limited spatial data density [50]. Secondly, in the course of climate change, erosivity of the rainfall has significantly increased compared to data collection periods of [49] or [51] as a result of (i) higher change of thunderstorms, (ii) larger size of the drops, (iii) reduced wind speed, and, thus, (iv) larger amounts of precipitation per rain event [52]. Accordingly, [52] state that rain erosivity increases by a rate of 10% every 6 years.

Summarizing, cultivation of silage maize at sites that show a high susceptibility for soil erosion, as presented in Table 6, was underpinned, on one hand, by observing a huge number of maize fields that visibly showed signs of erosion in aerial photos (Figure 3) and on the other hand by the discrepancy of the maize cultivation area compared to statistical data by remote-sensing based classification approach (Table 5). Areas of bare soil, like areas of erosion, transport, and accumulation, were classified as non-maize pixels due to the soil signal (Figure 3).

4. Conclusions

This spatial study depicts an illustrative example in which way political measures, such as subsidies, may influence the agricultural land-use implying negative consequences for soil health. The additional demand for feedstocks in the study area has led to land-use pressure and significant problems in sustaining soil health in the study region. The combination of remote sensing-based maize classification and target-specific analysis of the soils' vulnerability allowed for a spatially explicit estimation of the site-specific hazard potential. Unfortunately, the data quality for estimating the risk for soil compaction was insufficient to a large extent. Nonetheless, soil erosion seems to represent the more important threat to soil health. However, by the fact that eroded patches and areas with weak field emergence were not detected as maize by the classification, the results certainly present conservative estimates, particularly when considering the predicted impacts of climate change.

The chosen approach to identify the site-specific vulnerability for soil compaction and erosion presents a valuable tool for decision support, aiming for a strategically planned and thus more sustainable energy crop cultivation. It can be used to initiate countermeasures in order to meet the demands of both valuable feedstock production and soil protection, as well as to sensitize farmers for the long-term consequences of soil degradation. A further development of the presented concept may support farmers in their crop selection and management strategy, e.g., in the form of a smartphone app. However, the availability of comprehensive and small-scale valid soil data presents the bottleneck of the approach.

In the long term, sustaining or improving soil health is inevitable in order to maintain the production function of agricultural soils.

References

1. Braun, R.; Weiland, P.; Wellinger, A. Biogas from Energy Crop Digestion. IEA Bioenergy Task 37: Energy from Biogas and Landfill Gas. 2010. Available online: http://task37.ieabioenergy.com/files/daten-redaktion/download/energycrop_def_Low_Res.pdf (accessed on 12 September 2018).
2. Lupp, G.; Albrecht, J.; Darbi, M.; Bastian, O. Ecosystem services in energy crop production: A concept for regulatory measures in spatial planning? *J. Landsc. Ecol.* **2011**, *4*, 49–66. [CrossRef]
3. Erneuerbare-Energien-Gesetz (EEG), Gesetz für den Ausbau Erneuerbarer Energien vom 21. Juli 2014 (BGBl. I S. 1066, Berlin), das Durch Artikel 2 des Gesetzes vom 22. Dezember 2016 (BGBl. I S. 3106, Berlin) Geändert Worden Ist. Available online: https://www.gesetze-im-internet.de/eeg_2014/BJNR106610014.html (accessed on 14 September 2020).
4. Deutsches Biomasseforschungszentrum (DBFZ). *Anlagenbestand Biogas und Biomethan—Biogaserzeugung und–Nutzung in Deutschland*; Report Nr. 30; DBFZ: Leipzig, Germany, 2017.
5. Fachagentur Nachwachsende Rohstoffe (FNR). *Basisdaten Bioenergie Deutschland 2014. Festbrennstoffe, Biokraftstoffe, Biogas*; Bestell-Nr. 469; FNR: Gülzow-Prüzen, Germany, 2014.
6. Fachagentur Nachwachsende Rohstoffe (FNR). *Basisdaten Bioenergie Deutschland 2018. Festbrennstoffe, Biokraftstoffe, Biogas*; FNR: Gülzow-Prüzen, Germany, 2018. Available online: http://www.fnr.de/fileadmin/allgemein/pdf/broschueren/Basisdaten_Bioenergie_2018.pdf (accessed on 11 September 2018).
7. KTBL (Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V.). *Energiepflanzen. Daten für die Planung des Energiepflanzenanbaus*, 2nd ed.; KTBL: Darmstadt, Germany, 2012.
8. Weiland, P. Biogas production: Current state and perspectives. *Appl. Microbiol. Biotechnol.* **2010**, *85*, 849–860. [CrossRef] [PubMed]
9. Adler, P.R.; Del Grosso, S.J.; Parton, W.J. Life-cycle assessment of the net greenhouse-gas flux for bioenergy cropping systems. *Ecol. Appl.* **2007**, *17*, 675–691. [CrossRef] [PubMed]
10. Blanco-Canqui, H. Energy crops and their implications on soil and environment. *Agron. J.* **2010**, *102*, 403–419. [CrossRef]
11. Herrmann, A. Biogas production from maize: Current state, challenges and prospects. 2. Agronomic and environmental aspects. *BioEnergy Resear.* **2013**, *6*, 372–387.

12. Immerzeel, D.J.; Verweij, P.A.; van der Hilst, F.; Faaij, A.P.C. Biodiversity impacts of bioenergy crop production: A state-of-the-art review. *Glob. Chang. Biol. Bioenergy* **2014**, *6*, 183–209. [CrossRef]

13. Statistisches Landesamt Rheinland-Pfalz (Statistical Office Rhineland-Palatinate). *Statistisches Jahrbuch 2017, Rheinland-Pfalz*; Statistisches Landesamt Rheinland-Pfalz: Sinzig, Germany, 2017.

14. DLR Mosel, Daten zum Silomaisanbau in der Region Trier und in Rheinland-Pfalz. 2018. Available online: http://www.dlr-mosel.rlp.de/Internet/global/themen.nsf/d0e5087e9e1e8b79c1257abf0060c5df/b7c65bbd218ffd94c1257c1c004c2c99?OpenDocument (accessed on 12 September 2018).

15. Statistisches Landesamt Rheinland-Pfalz (Statistical Office Rhineland-Palatinate). *Statistische Bände: Die Landwirtschaft mit Vergleichszahlen seit 1949*; Statistisches Landesamt Rheinland-Pfalz: Bad Ems, Germany, 1949; multiple years.

16. Von Francken-Welz, H.; Wenghoefer, V.; Hamilton, V. Biogasanlagen in Rheinland-Pfalz 2017. 5. Betriebserhebung Biogas. Dienstleistungszentrum Ländlicher Raum Eifel (publ.), Germany. 2017. Available online: http://www.dlreifel.rlp.de/Internet/global/themen.nsf/3377acc1e3e31fb8c12579f000322e61/c661a290b8aef10ec12581f300343635/$FILE/Biogaserhebung2017_final.pdf (accessed on 18 July 2018).

17. Landesamt für Vermessung und Geobasisinformation Rheinland-Pfalz (LVermGeo). *Digital Elevation Model with a Grid Width of 5 m (DEM5)*; Landesamt für Vermessung und Geobasisinformation Rheinland-Pfalz (LVermGeo): Koblenz, Germany, 2018.

18. Steingötter, K. *Geologie von Rheinland-Pfalz. Landesamt für Geologie und Bergbau Rheinland-Pfalz (Hrsg.)*; E. Schweitzerbart'sche Verlagsbuchhandlung (Nägele u. Obermiller): Stuttgart, Germany, 2005.

19. German Meteorological Service, Climate Data Center. Available online: ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/multi_annual/mean_81-10/ (accessed on 20 July 2018).

20. Landesamt für Geologie und Bergbau Rheinland-Pfalz (LGB-RLP) (State Office for Geology and Mining Rhineland-Palatinate): Geologische Übersichtskarte von Rheinland-Pfalz: Online Karte GÜK 300. 2018. Available online: http://www.lgbrlp.de/karten-und-produkte/online-karten/online-karte-guek-300.html (accessed on 18 July 2018).

21. Landesamt für Geologie und Bergbau Rheinland-Pfalz (LGB-RLP) (State Office for Geology and Mining Rhineland-Palatinate) Bodenflächendaten 1:200,000. 2018. Available online: http://www.lgb-rlp.de/karten-und-produkte/wms-dienste.html (accessed on 18 July 2018).

22. Wagner, W.; Negendank, J.F.W.; Fuchs, G.; Mittmeyer, H.G. *Geologische Übersichtskarte: Rheinisches Schiefergebirge SW-Teil 1:100,000*; Geologisches Landesamt Rheinland-Pfalz, Ed.; Geologisches Landesamt Rheinland-Pfalz: Mainz, Germany, 1983.

23. Tanré, D.; Deroo, C.; Duhaut, P.; Herman, M.; Morcrette, J.J.; Perbos, J.; Deschamps, P.Y. Technical note. Description of a computer code to simulate the satellite signal in the solar spectrum: The 5S code. *Int. J. Remote Sens.* **1990**, *11*, 659–668. [CrossRef]

24. Hill, J.; Sturm, B. Radiometric correction of multitemporal Thematic Mapper data for use in agricultural land-cover classification and vegetation monitoring. *Int. J. Remote Sens.* **1991**, *12*, 1471–1491. [CrossRef]

25. Hill, J.; Mehl, W.; Radeloff, V. Improved forest mapping by combining corrections of atmospheric and topographic effects in Landsat TM imagery. In *Sensors and Environmental Applications of Remote Sensing, Proceedings of the 14th EARSel Symposium, Goteborg, Sweden, 6–8 June 1994*; Askne, J., Ed.; Chalmers University of Technology: Göteburg, Sweden, 1995; pp. 143–151.

26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016. Available online: https://www.R-project.org/ (accessed on 19 January 2021).

27. Wright, M.N.; Ziegler, A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [CrossRef]

28. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

29. Belgiu, M.; Draăgut, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

30. Ad-hoc-AG Boden. *Bodenkundliche Kartieranleitung*, 5th ed.; E. Schweizerbart'sche Verlagsbuchhandlung: Stuttgart, Germany, 2005.

31. Food and Agriculture Organization of the United Nations (FAO). *Guidelines for Soil Description*, 4th ed.; FAO: Rome, Italy, 2006.

32. Bodenschätzungsgesetz (BodSchätzG), Gesetz zur Schätzung des landwirtschaftlichen Kulturbodens vom 20. Dezember 2007 (Bundesgesetzblatt I S. 3150, 3176, Berlin), das durch Artikel 232 der Verordnung vom 31. August 2015 (Bundesgesetzblatt I S. 1474, Berlin) Geändert Worden ist. Available online: http://www.gesetze-im-internet.de/bodsch_tzg_2008/BodSchätzG.pdf (accessed on 14 September 2020).

33. Deutsches Institut für Normung (DIN). *19708:2017-08: Soil Quality: Predicting Soil Erosion by Water by Means of the USLE (Bodenbeschaffenheit: Ermittlung der Erosionsgefährdung von Böden Durch Wasser Mithilfe der ABAG)*; Beuth-Verlag: Berlin, Germany, 2017.

34. Wischmeier, W.; Smith, D. *Predicting Rainfall Erosion Losses: A Guide to Conservation Planning*; Agriculture Handbook No. 537; U.S. Department of Agriculture: Hyattsville, MD, USA, 1978.

35. Blume, H.-P.; Brümmer, G.W.; Fleige, H.; Horn, R.; Kandeler, E.; Kögel-Knabner, I.; Kretzschmar, R.; Stahr, K.; Wilke, B.-M. *Scheffer/Schachtschabel. Soil Science*; Springer: Berlin/Heidelberg, Germany, 2016.

36. Zech, W.; Schad, P.; Hintermaier-Erhard, G. *Böden der Welt. Ein Bildatlas*, 2nd ed.; Springer Spektrum: Berlin/Heidelberg, Germany, 2014.

37. Kuhwald, M.; Dörnhöfer, K.; Oppelt, N.; Duttmann, R. Spatially explicit soil compaction risk assessment of arable soils at regional scale: The SaSCiA-Model. *Sustainability* **2018**, *10*, 1618. [CrossRef]

38. Lebert, M. Entwicklung Eines Prüfkonzeptes zur Erfassung der tatsächlichen Verdichtungsgefährdung Landwirtschaftlich Genutzter Böden. UBA-Texte 51/2010, Förderkennzeichen: 370771202. Available online: https://www.umweltbundesamt.de/sites/default/files/medien/461/publikationen/4027.pdf (accessed on 1 April 2019).
39. Ruf, T.; Audu, V.; Holzhauser, K.; Emmerling, C. Bioenergy from periodically waterlogged cropland in Europe: A first assessment of the potential of five perennial energy crops to provide biomass and their interactions with soil. *Agronomy* **2019**, *9*, 374. [CrossRef]
40. Schoo, B.; Schroetter, S.; Kage, H.; Schittenhelm, S. Root traits of cup plant, maize and lucerne grass grown under different soil and soil moisture conditions. *J. Agron. Crop Sci.* **2017**, *203*, 345–359.
41. Saggau, P.; Bug, J.; Gocht, A.; Kruse, K. Aktuelle Bodenerosionsgefährdung durch Wind und Wasser in Deutschland. *Bodenschutz* **2017**, *22*, 120–125.
42. European Union. *EU Regulation No 1306/2013, Regulation of the European Parliament and of the Council of 17 December 2013 on the Financing, Management and Monitoring of the Common Agricultural Policy and Repealing Council Regulations (EEC) No 352/78, (EC) No 165/94, (EC) No 2799/98, (EC) No 814/2000, (EC) No 1290/2005 and (EC) No 485/2008*; EU: Brussels, Belgium, 2013.
43. AgrarZahlVerpflG Gesetz zur Regelung der Einhaltung von Anforderungen und Standards im Rahmen Unionsrechtlicher Vorschriften über Agrarzahlungen. Agrarzahlungen-Verpflichtungengesetz vom 2. Dezember 2014 (Bundesgesetzblatt I S. 1928, Berlin). Available online: https://www.gesetze-im-internet.de/agrarzahlverpflg/AgrarZahlVerpflG.pdf (accessed on 14 September 2020).
44. AgrarZahlVerpflV Verordnung über die Einhaltung von Grundanforderungen und Standards im Rahmen Unionsrechtlicher Vorschriften über Agrarzahlungen. Agrarzahlungen-Verpflichtungenverordnung vom 17. Dezember 2014 (BAnz AT 23.12.2014 V1, Berlin), die Zuletzt Durch Artikel 2 der Verordnung vom 27. September 2018 (BAnz AT 28.09.2018 V1, Berlin) Geändert Worden ist. Available online: https://www.gesetze-im-internet.de/agrarzahlverpflv/AgrarZahlVerpflV.pdf (accessed on 14 September 2020).
45. Auerswald, K.; Fischer, F.K.; Kistler, M.; Treisch, M.; Maier, H.; Brandhuber, R. Behavior of farmers in regard to erosion by water as reflected by their farming practices. *Sci. Total Environ.* **2018**, *613–614*, 1–9. [CrossRef]
46. Panagos, P.; Borrelli, P.; Meusburger, K.; Alewell, C.; Lugato, E.; Montanarella, L. Estimating the soil erosion cover-management factor at the European scale. *Land Use Policy* **2015**, *48*, 38–50. [CrossRef]
47. Schwertmann, U.; Vogl, W.; Kainz, M. *Bodenerosion Durch Wasser. Vorhersage des Abtrags und Bewertung von Gegenmaßnahmen*; Verlag Eugen Ulmer: Stuttgart, Germany, 1987.
48. European Environmental Agency (EEA). *Climate Change, Impacts and Vulnerability in Europe 2016: An Indicator-Based Report (EEA Report No 1/2017)*; European Environment Agency: Copenhagen, Denmark, 2017.
49. Sauerborn, P. Die Erosivität der Niederschläge in Deutschland—Ein Beitrag zur quantitativen Prognose der Bodenerosion durch Wasser in Mitteleuropa. In *Bonner Bodenkundliche Abhandlungen*; Band 13; Universität Bonn: Bonn, Germany, 1994.
50. Auerswald, K.; Fischer, F.K.; Winterrath, T.; Brandhuber, R. Rain erosivity map for Germany derived from contiguous radar rain data. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 1819–1832. [CrossRef]
51. Rogler, H.; Schwertmann, U. Erosivität der Niederschläge und Isoerodentkarte Bayerns. *Z. Kult. Flurberein.* **1981**, *22*, 99–112.
52. Auerswald, K.; Fischer, F.; Winterrath, T.; Elhaus, D.; Maier, H.; Brandhuber, R. Klimabedingte Veränderung der Regenerosivität seit 1960 und Konsequenzen für Bodenabtragsschätzungen. In *Bodenschutz, Ergänzbares Handbuch der Maßnahmen und Empfehlungen für Schutz, Pflege und Sanierung von Böden, Landschaft und Grundwasser (Loseblattsammlung)*; Bachmann, G., König, W., Utermann, J., Eds.; Erich Schmidt Verlag: Berlin, Germany, 2019.
53. Gilcher, M.; Ruf, T.; Emmerling, C.; Udelhoven, T. Remote sensing based binary classification of maize. Dealing with residual autocorrelation in sparse sample situations. *Remote Sens.* **2019**, *11*, 2172. [CrossRef]

4 Field Geometry and the Spatial and Temporal Generalization of Crop Classification Algorithms—A Randomized Approach to Compare Pixel Based and Convolution Based Methods

Gilcher, M., & Udelhoven, T. (2021). Field geometry and the spatial and temporal generalization of crop classification algorithms—a randomized approach to compare pixel based and convolution based methods. Remote Sensing, 13(4), 1–20.

Article

Field Geometry and the Spatial and Temporal Generalization of Crop Classification Algorithms—A Randomized Approach to Compare Pixel Based and Convolution Based Methods

Mario Gilcher * and Thomas Udelhoven

Department of Remote Sensing and Geoinformatics, Faculty of Regional and Environmental Sciences, University of Trier, Campus II, D-54286 Trier, Germany; udelhoven@uni-trier.de
* Correspondence: gilcher@uni-trier.de; Tel.: +49-(0)-651-201-4607

Abstract: With the ongoing trend towards deep learning in the remote sensing community, classical pixel based algorithms are often outperformed by convolution based image segmentation algorithms. This performance was mostly validated spatially, by splitting training and validation pixels for a given year. Though generalizing models temporally is potentially more difficult, it has been a recent trend to transfer models from one year to another, and therefore to validate temporally. The study argues that it is always important to check both, in order to generate models that are useful beyond the scope of the training data. It shows that convolutional neural networks have potential to generalize better than pixel based models, since they do not rely on phenological development alone, but can also consider object geometry and texture. The UNET classifier was able to achieve the highest F1 scores, averaging 0.61 in temporal validation samples, and 0.77 in spatial validation samples. The theoretical potential for overfitting geometry and just memorizing the shape of fields that are maize has been shown to be insignificant in practical applications. In conclusion, kernel based convolutions can offer a large contribution in making agricultural classification models more transferable, both to other regions and to other years.

Keywords: deep learning; sentinel 1; image segmentation

1. Introduction

1.1. State of the Art

The remote sensing community in general, and the land use/land cover (LULC) classification community in particular, is currently in a stage where potential applications seem endless [1], high resolution satellite imagery at national levels is available for free [2], and sophisticated classification algorithms perform better with each new generation [3]. This is especially true for the field of agricultural crop classification, where many studies were published in the past 5 years, that implemented one of many different iterations of deep learnig algorithms. In these studies, the model input has been rather stable, mostly consisting of multispectral global sensors like Landsat [4], Synthetic Aparture Radar (SAR) satellites like Sentinel-1 [5] and very high resolution optical sensors like Gaofen [6] as well as combinations of optical and SAR sensors [7]. The SAR domain has seen some advances in data processing techniques, where the phase information of polarimetric radar imagery can be used to derive parameters of the scattering. Parameters like anisotropy/entropy [8] and double bounce/surface/volume scattering [9] can be used to classify LULC and even radar based vegetation indices can be derived [10]. While there were some advances in the processing of the input datasets, model algorithms have almost completely switched from traditional machine learning and classical neural network algorithms, over Convolutional Neural Networks (CNN) and autoencoders [11] towards advanced derivates of the CNN approach like UNET [12] or the Recurrent Neural Network (RNN) approach like Long Short Term Memory (LSTM) models [13].

Another visible trend in the past 5 years, is that an increasing number of papers start to assess temporal model generalization. In the beginning, for example in case studies in India [14], the Ukraine [15] and Brazil [16], validation data was split spatially. This means that pixels or fields from one single scene, or one single mosaick were taken out of the training dataset and split into two or three sets of training, validation and testing samples. Cai et al. [17] were successfully able to test the temporal transferability of a Neural Network based corn and soybean classifier. They used long term Landsat derived vegetation index time series, and were able to predict years that were not part of the training datasets with high accuracy. Momm, ElKadiri and Porter [18] also used Landsat NDVI time series to classifiy crop types in the United States, and systematically tested models based on 2005 data with data from the year 2000 and vice versa. Zhang et al. [6] went a different approach and did not rely on index based phenlogy curves. They used high resolution satellite imagery to delineate cropland and non-cropland in several chinese regions, while testing both spatial and temporal transferability of their CNN models. The temporal transferability was tested by applying models from 2016 to 2017, and the spatial transferability was tested on different regions. Ajadi et al. [19] classified soybean and corn crops in brasil based on a composite of optical and SAR imagery, and also tested spatial and temporal transferability of their models by testing a model trained in summer 2017–2018 with a dataset from summer 2018–2019 and from a different state.

Though there is a clear trend in these case studies, to value transferability considerations in model assessment, recent review papers focused on the topics of deep learning in remote sensing and crop classification [20–22] do not cover transferability considerations at all. Olofsson et al. [23] mention the importance of stratified randomization in the sampling process, but do not specificy the nature of these strata any further. Though generalization has been an important concept in remote sensing based classifaction for more than a decade [24], and strategies exist to asses it on a technical level (sampel splitting, crossvalidation), there is no clear definition or discussion about the way the validation and testing datasets need to be different from the training datasets, in order to say that a model generalizes well.

1.2. Objectives

The goal of this study is to add to this discussion by introducing a novel way to frame the model validation process. It will be described by a number of different scopes on a similarity spectrum where distance is implied, but not quantifiable. Its premise will be, that a high performance classification in the closest scope (within-scene) is achievable by many different simple and complex algorithms, but is not enough to evaluate a models usefulness. Good performance in scopes that are very far from the training scopes on the other hand are increasingly impossible, as summarized in the No Free Lunch Theorem [25,26]. The concept of temporal and spatial validation scopes will be used to compare the transferability and generalization of classical pixel based algorithms to contemporary convolution based approaches in the context of remote sensing analysis, using the example of binary maize classification. The quantification and localization of silage maize expansion has been shown to be a very important part of the environmental monitoring process in southwest Germany [27]. While pixel based statistical and machine learning algorithms are well suited to classify agricultural use based on optical remote sensing imagery [28], a reliable modeling approach that is able to classify across the entire region and for any given year is still not yet identified. In this study, we try to take advantage of the aforementioned progress in the field of deep learning and CNNs, while at the same time using multitemporal SAR composites as a way to consistently acquire data for the entire study region for 4 years from 2016 to 2019. While many studies were able to produce deep learning models with high performance and highlight their superiority over classifcal machine learning, a fully randomized design is going to improve the understanding of how well this translates to improved performance in out of training samples. The goals of this study in more detail:

1. Input Data:

   In order to validate temporal transferability, there is a requirement for consistency between years. For this reason, Sentinel-1 scenes will be processed in a way to generate monthly average based timeseries to reflect crop phenology.

2. Validation Scopes:

   Scopes will be defined as subsets of the entire dataset. Aside from the training scope, there will be scopes that test spatial and temporal transferability, as well as both combined. To check for potential overfitting of field shapes and memory effects, special within geometry scopes will be defined and tested for predictive inconsistencies.

3. Model Training and Evaluation:

   Based on the defined scopes, 500 sets of samples are randomly drawn. Two representative pixel based and two representative convolution based classifiers are trained, and the model performances are measured for each scope independently. Finally, the sampled model performances are evaluated with parameter free U-tests, in order to infer which algorithms perform better in any given scope.

2. Materials and Methods

2.1. Study Area

The administrative district "Eifelkreis Bitburg-Prüm" (in the following referred to as 'study area', see Figure 1) has been shown to be a perfectly suitable region for crop classification studies before [28]. It is located in the center of Europe, but still shows very little segmentation by settlements, and offers a comparatively homogenous landscape dominated by agriculture (about 53% of the total area [29]) and forests. While there is an altitude gradient from south to north (from 300 masl. to 700 masl.) and some larger population hotspots, agricultural fields in general and maize in particular are still distributed almost uniformly across the entire study area. Therefore, the area is ideal for a study like this, where lots of subsets are randomized that should be comparable in spatial structure and land use, but not so similar, that model transferability would be trivial.
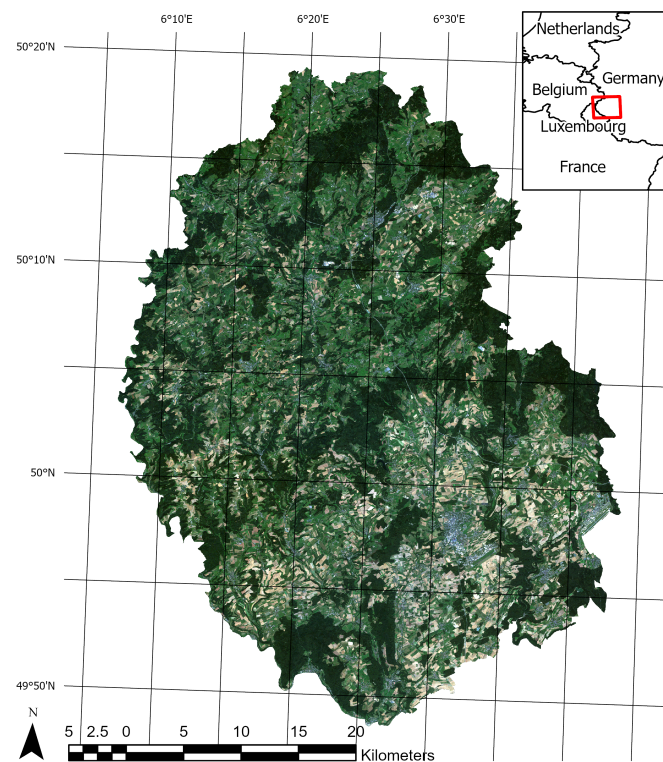

Figure 1. The figure shows an overview of the study area. The background shows a RapidEye true color composite from 2016.

2.2. Classification Algorithms

　　The selection of algorithms to apply and validate in the presented study, is characterized by theoretical and practical considerations. The goal is to have a small set of both pixel based and convolution based classifiers, that give a realistic representation of classical as well as state of the art implementations in the field. The focus of this study is to provide conceptual considerations about model transferability, rather than developing and optimizing novel implementations. Thusly, for each modeling approach, one basic implementation, and one advanced iteration was selected (see also Table 1).

Table 1. Overview of modeling algorithms.

Approach	Pixel Based	Convolution Based
basic	ranger	FCNN
advanced	xgboost	UNET

　　Machine learning classifiers like Support Vector Machines and Random Forests have long been proven to be superior to statistical techniques like Maximum Likelihood (cf. [1]), and are well suited for binary land use classification tasks. Each single observation is a multivariate vector of independent variables, while the dependent variable is a binary outcome. Random Forest classifiers (cf. [30]) are particularly useful in the context of this study, because there are very few and mostly inconsequential hyperparameters required. It is a well established algorithm, and many different implementations exist. Within the context of this study, ranger was used, a well optimized and popular implementation of the random forest algorithm (cf. [31]). In addition to ranger, the Extreme Gradient Boosting (XGBoost, cf. [32]) algorithm was selected, to represent a more state of the art ensemble based machine learning method. It is very prominent in contemporary remote sensing studies (see [33,34]), well optimized for large datasets and implementations exist for R (cf. [35]). Both algorithms are representative for the two major resampling strategies in ensemble based machine learning, boosting and bagging [36,37]. Both aim to reduce variance in weak learners (i.e., single trees) by training many learners on resampled datasets, but boosting focusses more on reducing bias by iteratively adjusting new models to the residuals of the previous models [38].

　　In contrast to the aforementioned pixel based machine learning classifiers, Convolutional Neural Networks (CNN, cf. [39]) are able to derive arbitrary spatial features with kernel convolutions. With the rising popularity of deep learning algorithms since roughly 2016, they quickly saw widespread use in remote sensing applications, and are the most frequently applied deep learning algorithm by a big margin (cf. [21]). One simple way to implement a CNN is to stack many convoluted layers on top of each other, and merge them with a sigmoid activated fully connected output layer. This simple fully connected convolutional network (FCNN) is used as a baseline neural network in this study. It is implemented in keras (see [40]). A vast amount of neural network architectures have been developed and applied in the field of image analysis. In this study, the U-Net architecture (cf. [41]) has been chosen, because it is able to process geometric relationships on multiple levels. It has its origin in medical image segmentation, and uses a nested succession of downsampling with maxpooling layers, as well transposed convolutions to upsample to the original resolution again. Due to this varying and flexible implementation of scale, it is especially suited to make use of field geometries in large scale datasets.

　　Optimizing models in terms of tuning hyperparameters requires advanced strategies like grid search, which in turn rely on measuring validation performance. Since the main focus of this study is to illustrate that measuring validation performance is not trivial and involves a clear definition of scope, the same reasoning has to be applied when optimizing parameters for model tuning. The focus is not to optimize the performance of a model in a given scope, the focus is to compare performances of models when changing scope. Therefore, default hyperparameters were kept whenever possible. If not, models were

optimized manually, by trying out different combinations on a larger subset in order to find parameters that give good training results. The ranger model performed well out of the box, and the default parameters (number of trees and number of variables in each split) were kept. The xgboost model comes with additional hyperparameters specific to the chosen booster. We were also able to keep the default parameters of the tree booster here, since they performed well on the training data.

Hyperparameter selection for deep learning algorithms is much more complicated, since they can have arbitrary complexity in the network architecture, while each part of the network can have parameters and functions to adjust. The shape of the input tensor is the first part of the architecture that needs to be defined, and is the same for both fcnn and unet. It is defined by the dimensions of the input dataset, and the edgesize of one sample cell. We decided to set the edgesize rather small, to keep the number of layers low. The edgesize has to have 2 as multiple factors to allow for downsampling in the unet algorithm, so 32 was chosen in both convolution based algorithms. The rest of the fcnn model is rather simple. In total, 20 convolution layers with rectified linear unit activation functions were found to be enough to segment the input images. These layers were then combined in a final sigmoid activated output layer. The unet model was initialized with 10 convolution layers before the first downsampling step, and four downsampling steps with 50% dropout layers total. In addition to the model specific hyperparameters, the training of a deep learning algorithm itself is defined by an additional set of parameters as well as the optimizer and the loss function. We again tested several approaches with a training subset of the study area and found that the very common RMSprop optimizer in combination with a binary crossentropy loss function worked well. Both models were then trained with a batch size of 10, for 150 epochs.

2.3. Data Sources

2.3.1. Reference Data

As a direct result of the common agricultural policy of the European Union, all administrative units are keeping track of field geometries and crops in their municipalities. The integrated administration and control system (InVeKos [42]) is used to register all official records, in exhaustive geodatasets. In this study, we use shapefiles from 2016 till 2019 each containing roughly 50,000 discrete field geometries. The field geometries have distinct registered crop types, of which around 100 different types exist in the area. Of these crop types, the ones that cover the different type of maize were selected, converted to a binary variable, and rasterized in a grid aligned to the image data (see Figure 2).

2.3.2. Image Data

The focus of this study is the transferability of image based models on a regional scale, both temporally and spatially. Consequently, SAR imagery is well suited as a prediction input. While it is much harder to interpret visually and the signal of the phenology is not as clear as in the reflectances of a vegetation spectrum, it is largely unaffected by atmospheric conditions. While it is affected by topography and soil water content, monthly timeseries can be derived consistently, which show distinct temporal patterns for certain crop types [43]. Along with the benefit of temporal consistency, SAR imagery comes with other downsides, most importantly the complete lack of information about leaf pigments [22] and generally lower ground sampling distance (GSD). The sensor used in this study is the C-band SAR mounted on the Sentinel-1 satellites A and B. We used the ground range detected (GRD) product with the instrument mode interferometric wide swath. GRD imagery provides the advantage of speckle and noise reduction through multi-looking (5 looks in this case), at the cost of losing the phase information of the signal, as well as an increased GSD. We used the level 1 GRD high resolution product with vertical transmit/horizontal receive polarization, which has a GSD of 10 m [44]. VH crosspolarized SAR data benefits from the increasing interaction of transmitted and received waves, as the canopy height develops [45]. The data was processed in and downloaded from the google

earth engine (GEE, cf. [46]). GEE Sentinel-1 datasets are pre-processed by the Sentinel-1 Toolbox, which means that thermal noise removal, radiometric calibration and terrain correction algorithms were applied. Since the Sentinel-1 constellation has a revisit time of roughly 6 days, and 12 days in 2016 because there was only Sentinel-1 A, monthly means could be computed for every single pixel. This lead to a consistent 12 band image with 10 m GSD for all 4 years. In a last step, each image was scaled. Figure 3 shows the distributions of both classes in the images. Though SAR is invariant to pigment features of the canopy, even on this aggregated level the growth curve of maize, starting around June and ending in September/October, is visible. One example year is visualized in Figure 2, where the maize fields have a distinct green/turquoise hue, which indicates a peak in Augst/September, as opposed to the red crops which are also visible.
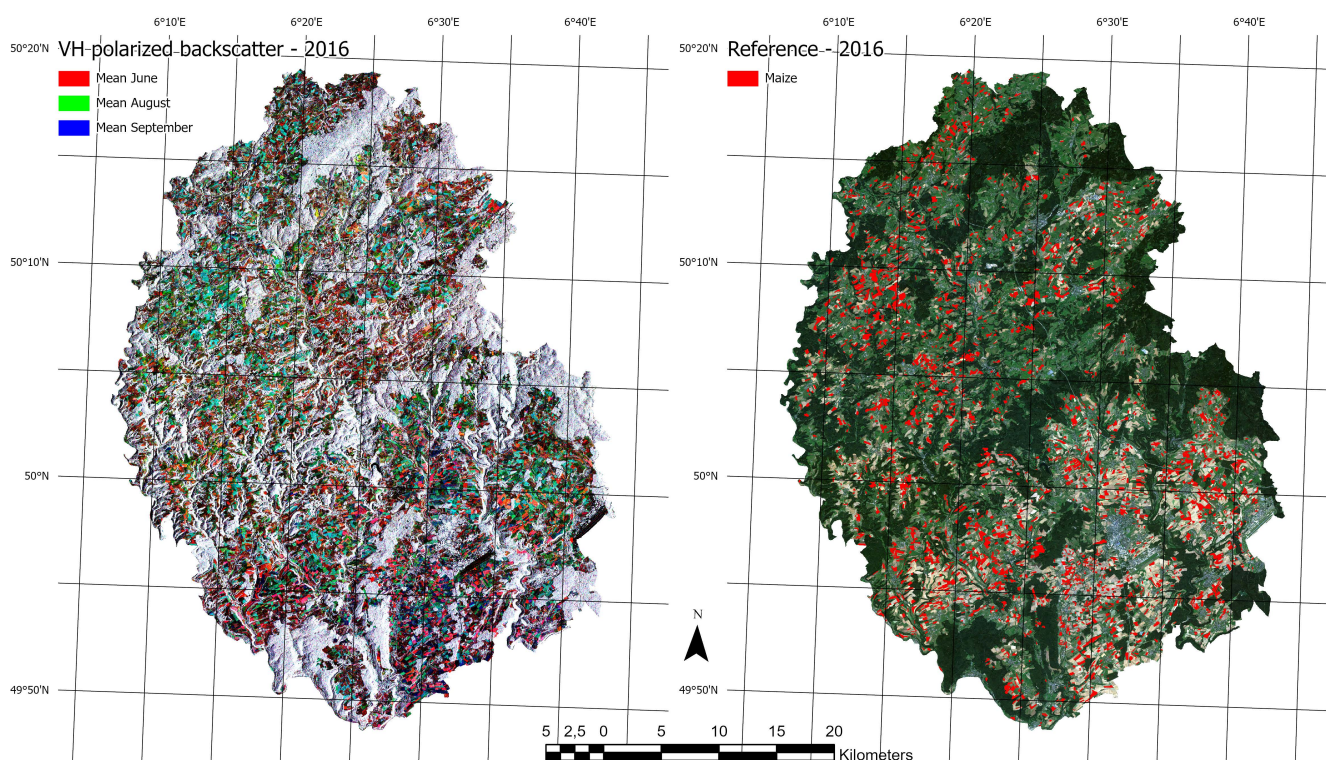

Figure 2. The figure shows an overview of input data. The left side shows an RGB composite of three different monthly means from the Sentinel-1 input data. The right side shows a rasterized version of all maize fields from the the InVeKos dataset with a RapidEye RGB composite in the background.
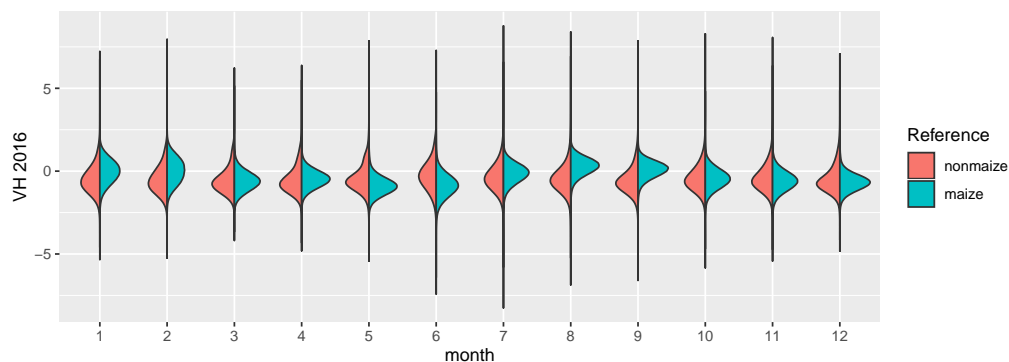
Figure 3. *Cont*.
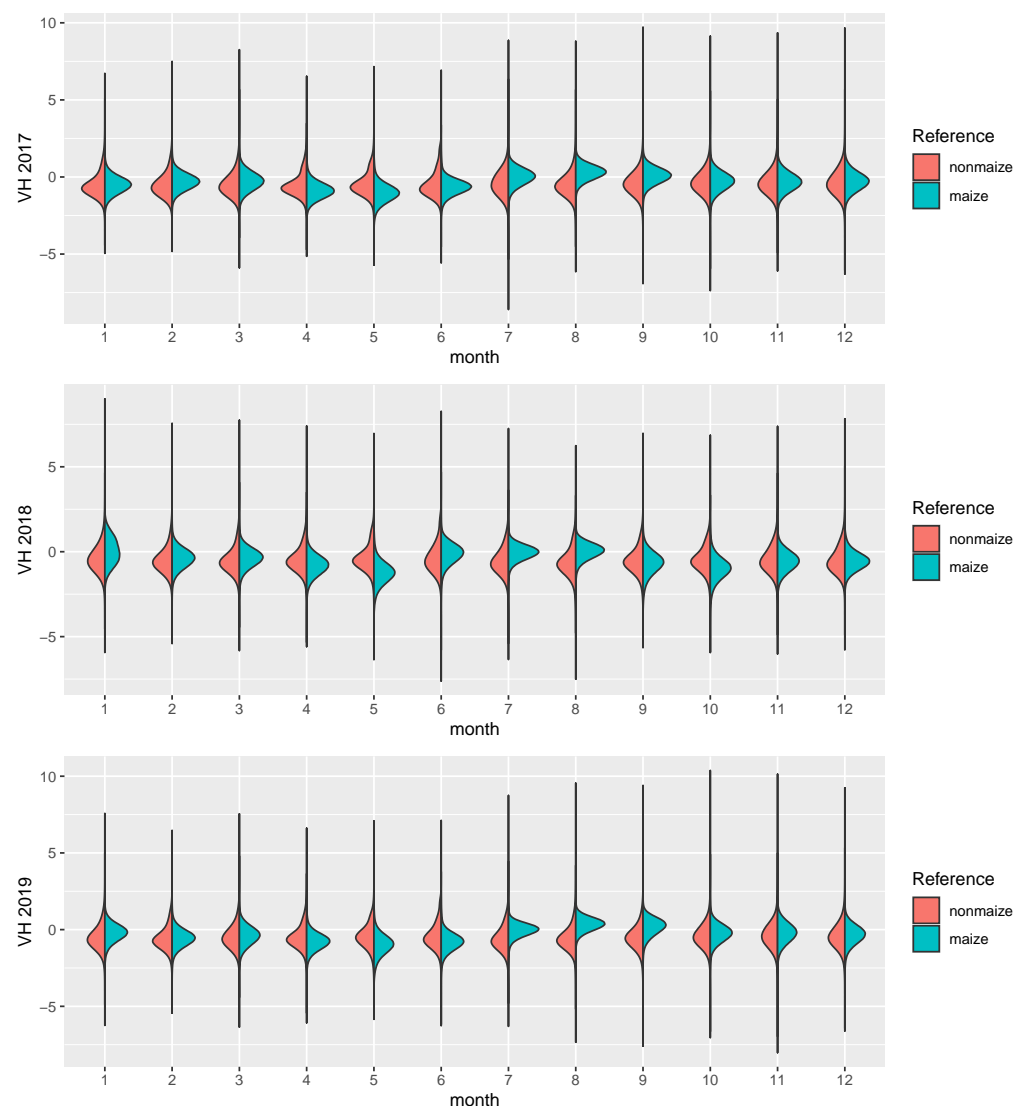
Figure 3. The figure shows distribution density estimates of the 12 monthly means for 4 years, split by maize (turquoise) and nonmaize (red) pixels.

2.4. Research Design

2.4.1. Spatiotemporal Generalization

The basis of this study is a repeated resampling of 5 by 5 km tiles (500 by 500 pixels) by changing location within the Eifelkreis Bitburg-Prüm, and the year ranging from 2016–2019. One sample contains a total of 4 tiles, from two different locations (hereafter called m and k), and two different years (hereafter called year 1 and year 2). For each run, all models are trained based on tile m_1 (see Table 2). From this tile, 500 sample cells with 32 by 32 pixels are drawn, while overlap is explicitly allowed. The result is a $500 \times 32 \times 32 \times 12$ tensor that is used as an input for the deep learning models. Since the cell shape is irrelevant for the pixel based algorithms, the input is then reshaped into observation rows with 12 variables. As a consequence of the input cell overlap, a lot of the observations are identical, so the amount unique observations is much lower than $500 \times 32 \times 32$. Based on this sample, the four models are trained and all four tiles are then predicted.

For the pixel based models, this prediction is simple. Since each pixel is treated independently, it is just a prediction with 250,000 observations and 12 variables. The convolution based predictions are a bit more complex, since inputs are not pixels, but cells. To predict a complete tile, each pixel is treated as a corner pixel of a cell, which creates overlap for pixels that are not at the borders of a tile. The predictions are then averaged for

each pixel based on the values of all cells including that given pixel. As a consequence, the input datasets for pixel based and convolution based methods are not directly comparable since the raw amount of input data is, by virtue of nature, much higher for the convolution based algorithms.

Each tile then represents one of four basic generalization scopes. Tile m_1 represents the *training scope*. It still has pixels not sampled in the actual training dataset, but they are in close proximity and from the same year. Tile k_1 represents *spatial* generalization. Its dataset is from the same year, but the field geometry is relatively independent. Tile m_2 is at the exact same location as m_1 but from a different year, and the validation scope is therefore called *temporal*. Field geometries are not always identical, but changes are very minor. Tile k_2 represents *spatiotemporal validation*, because it is different in both location and year.

2.4.2. Geometrical Generalization

Kernel convolutions, which are the basis of CNNs, compute artificial features based on a given neighborhood. This means, that features like angled edges and texture can be included in the semantic segmentation optimized by the perceptrons. These features can be summarized as "geometry" of the field, and add an entirely new set of information to an observation, that pixel based algorithms cannot include. This layer of information can be useful in the classification of agricultural fields, because they are often defined by long straight edges and comparatively similar in size. It can also be harmful in temporal generalization, because in a field with a very specific shape, crop types can change from year to year, while the shapes stay mostly the same.

This problem is visualized in Figure 4, where training cells were artificially generated with a very specific star shape. Pixels that are known to be maize are randomly drawn from the image data and used to fill the star. Pixels that are known to be not maize are used to fill the rest of the cell, as well as an additional artificial control cell. The star shaped cells are then used to train a UNET CNN as well as a ranger classification model. The figure shows one prediction of a training cell, and a control cell for each classifier. The training prediction works well for both models, and the star shape is clearly visible. The control cell however, contains no pixels from maize fields, and the ranger classifier shows only a small amount of noise and misclassified pixels. The UNET classifier however, shows almost the exact same picture, which means that it was overfitting the geometry of the field, and seemingly did not include the numeric information of the pixel features at all.
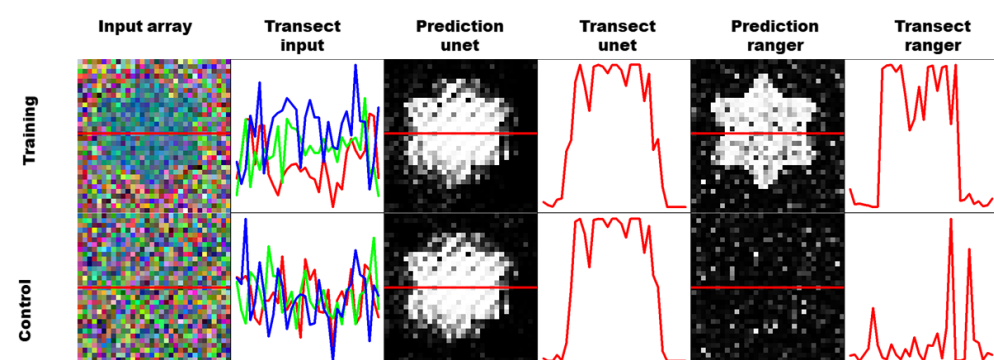
Figure 4. The figure is visualizing geometric overfitting by training based on an artificial sample with a star shaped maize field (**top row**) and validating on an artificial sample without maize (**bottom row**).

This artificial example is very far from field conditions, and overfitting of this kind would be easily detectable in simple crossvalidation or split-data validation samplings. It is still possible that the perceptrons could show a similar memory effect when trained on maize fields with unique geometric features. If the model is validated with spatial validation, just with image data from another location, this overfitting will not be apparent, because the classifier will never see the same shape again. If validated with temporal

validation datasets, which contain the very same field from another year, the geometry will be the same, but the pixel features and potentially the crop type are different. A convolution based classifier could potentially be more likely to classify this as a maize field, because it has been trained to classify this geometry, albeit with different pixel features, as maize before. A pixel based classifier is invariant to the geometry, and therefore the classification result should not be biased.

Resampling from a large dataset spanning an entire region for multiple years, provides the unique opportunity to specifically assess this kind of overfitting problems. Thusly, we introduce four additional scopes, that are a subset of the temporal scope defined in Table 2 since they only apply to tile m_2, which has the same geometry but data from a different year. One caveat of this approach is that the geometry between two years is not exactly the same. There are some differences between polygon geometries from all four years, but they are usually very minor, and this simplification should have little impact on the overall conclusion of the study. The scopes themselves are described in Figure 5. The scope *bothmaize* includes fields that have been maize in the training year, and still are maize in the validation year, while the scope *frommaize* includes fields that were maize, but are not anymore.
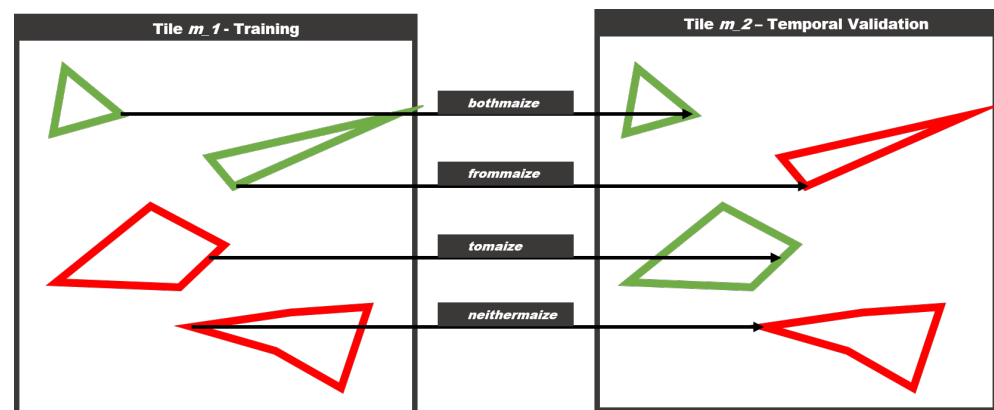
Figure 5. The figure describes the geometric scopes for the validation approach. Assuming the same geometry in a different year, there are four possible scenarios.

The scope *tomaize* includes fields that have not been maize in the training year, and changed to maize in the validation year, while the scope *neithermaize* includes fields that are not maize in both the training and validation year. Conceptually the performances in the *bothmaize* and *tomaize* scopes should be similar, as well as the performances of *frommaize* and *neithermaize*.

2.4.3. Performance Metrics

Binary classification is a concept known in many different domains, and as such, a large amount of classification performance metrics exist. All of them have in common, that they are based on the confusion matrix (CM) (cf. [47]). Overall accuracy is a metric directly derived from the CM, but has the problem that it is misleading if the no information rate is high, i.e., if there is a large imbalance between the two classes. Other base metrics, like sensitivity/specificity and precision/recall give a better idea of the overall performance because they are not focused on the diagonal of the CM. Additionally, there are higher level metrics like Cohens kappa and the F1 score. Cohens kappa has been very popular in remote sensing publications in the past, but has been heavily criticized (cf. [23,48]). It has also the problem that it is a very domain specific metric, and implementations in contemporary deep learning frameworks is not common. In this study, F1 (see Tables 2 and 3) will be used as a comprehensive and widespread higher level metric, in addition to the base metrics precision and recall. In some geometric scopes there are no true positives or negatives, therefore sensitivity and specificity will be reported if applicable.

Table 2. Overview of Validation Scopes.

Tiles	Handle	Metrics
m_1	training	F1/Precision/Recall
m_2	temporal	F1/Precision/Recall
k_1	spatial	F1/Precision/Recall
k_2	spatiotemporal	F1/Precision/Recall
m_2	bothmaize	Sensitivity
m_2	tomaize	Sensitivity
m_2	frommaize	Specificity
m_2	neithermaize	Specificity

Table 3. Overview of Comparison Scopes.

Comparison	Handle	Metrics
training-temporal	temporal	F1/Precision/Recall
training-spatial	spatial	F1/Precision/Recall
training-spatiotemporal	spatiotemporal	F1/Precision/Recall
bothmaize-tomaize	tomaize	Sensitivity
neithermaize-frommaize	frommaize	Specificity

In a final step, Mann–Whitney U-tests will be used to compare the performances of each model against each other. Since the metrics used range from 0 to 1, they are not normally distributed, therefore a nonparametric two sample test will be used to test the alternative, that a given test yields better results on average than the test it is compared against. Between the geometric scopes, two sided tests will be used in oder to test the null hypothesis that bothmaize and tomaize yield on average similar performances. In other words, that the sensitivity of maize fields is indifferent to the status of that particular field in training.

3. Results

3.1. Illustrative Sample Run

Figures 6 and 7 show the input data of one sample run. Tile m shows a heterogenous landscape, structured by riverbeds and disperse forest areas. It has a moderate amount of maize fields compared to tile k, which shows a very high density of maize fields divided into small sections in the north, and a dense forest in the south. The fields in tile k are also very regularly shaped compared to the fields in tile m.
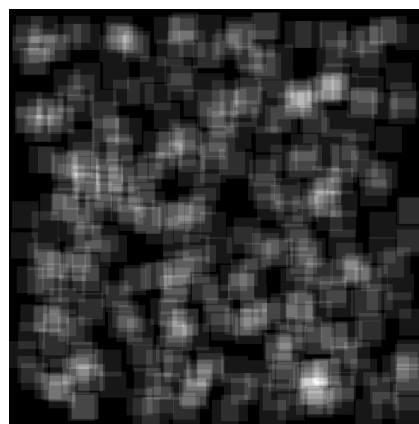

Figure 6. The figure shows the spatial distribution of training cells in a tile. A substantial amount of overlap is visible and necessary, since cells that contain the same field but slightly translated still provide new information to the classifiers. For the pixel based algorithms only unique pixels were counted, in this 207,595 observations, roughly 84% of the entire tile.
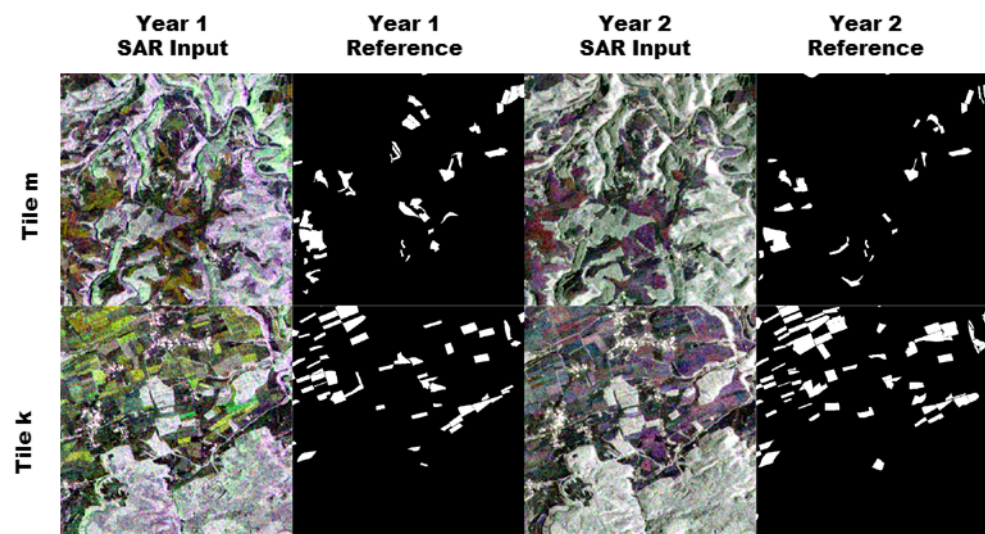
Figure 7. The figure shows a sample for one complete run. The two rows show tile m and k. Columns 1 and 3 show false color composites of the 12 band monthly mean Sentinel-1 images. Columns 2 and 4 show the position of the maize pixels in white.

All four algorithms predicted the training tile m_1 rather well (see Figure 8), and did not miss any of the fields. The pixel based classifiers produce spatially rather evenly distributed residuals. The residuals produced by the convolution based algorithms are more clustered and are often missing entire fields or sections of fields. Sometimes the residuals of both FCNN and UNET form clearly structured spatial patterns around fields, indicating that the geometry of the reference data is slightly bigger than the prediction. Overall none of the four classifiers are able to adequately perform the binary segmentation in the second tile k, especially in the second year. However, there are some minor differences in how the classifiers predict, and it is not obvious which one performs better. To asses this, this random sample was repeated 500 times, and the results were summarized in the upcoming section.

3.2. Spatiotemporal Generalization

Table 4 and Figure 9 show the model summaries of all 500 resampling runs, grouped by algorithm and validation scope. They show a very clear pattern of differences for the four different scopes. In the training scope, all four classifiers are able to achieve relatively high F1 scores (mostly above 0.75), with relatively low standard deviations (below 0.05). In all the other scopes, performances are both much lower on average, and much more varied. The spatial scope has still relatively high scores, ranging from 0.68 (ranger) to 0.77 (UNET), and also much higher standard deviations. The temporal validation scopes show even lower performances, while spatiotemporal scopes show the lowest overall performances, and the highest variability. In general, the true positive rate, or recall, seems to be a big problem in most models. While still above 0.9 for the pixel based models, and above 0.8 for the convolution based models, the detection rates drop mostly below 0.5 in the temporal and spatiotemporal validation scopes (see Table 5). The algorithm ranger detects on average just about one third (36.6%) of maize pixels in the spatiotemporal validation scope. Precision rates (see Table 6) in comparison are much higher, and also much less variable within models, between models, and between scopes. This means that false positives are rather uncommon with all four classifiers (see also Figure 8). XGBoost seems to have the highest false positive rate, and performs considerably worse than the other classifiers, and ranger in particular.
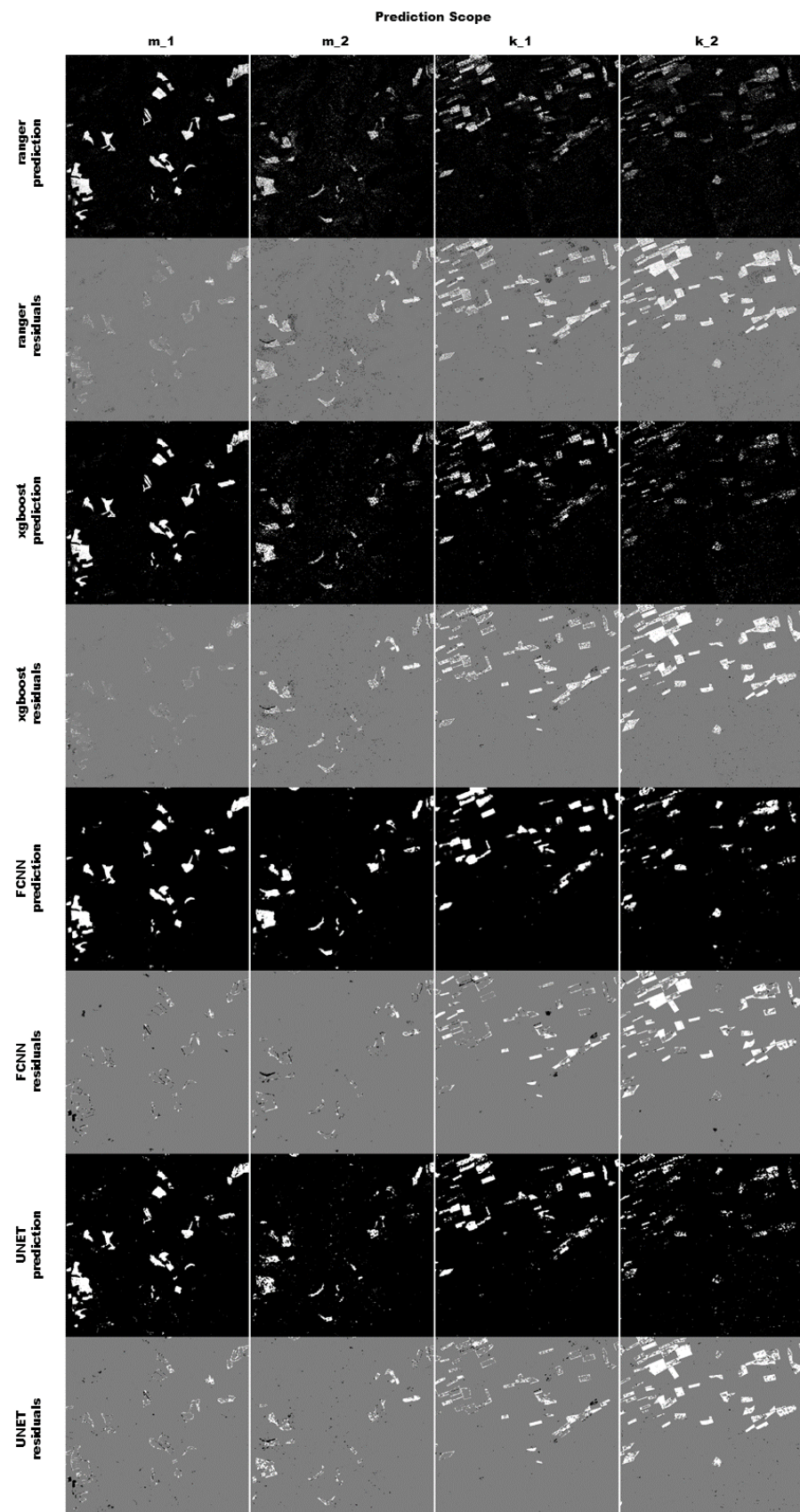
Figure 8. The figure shows a result for a complete run. The four different scopes are shown in the columns, while the model predictions and residuals are shown in the rows. The prediction pixels are colored in white (predicted to be maize) and black (predicted to be not maize). The residuals are colored in white (false negative), black (false positive) and grey (true positive or true negative).

The same pattern, spatial generalization being better than temporal, and temporal generalization being better than spatiotemporal, can be observed when just looking at the differences in scores (see Figure 9). The validation score differences show, compared to the absolute performance scores, relatively low standard deviations. Both pixel based methods and both convolution based methods show similar value ranges in the spatial comparison scope, with convolution based methods losing much less performance. The performance difference in the temporal and spatiotemporal validation is much higher, and the difference between all models gets more pronounced.
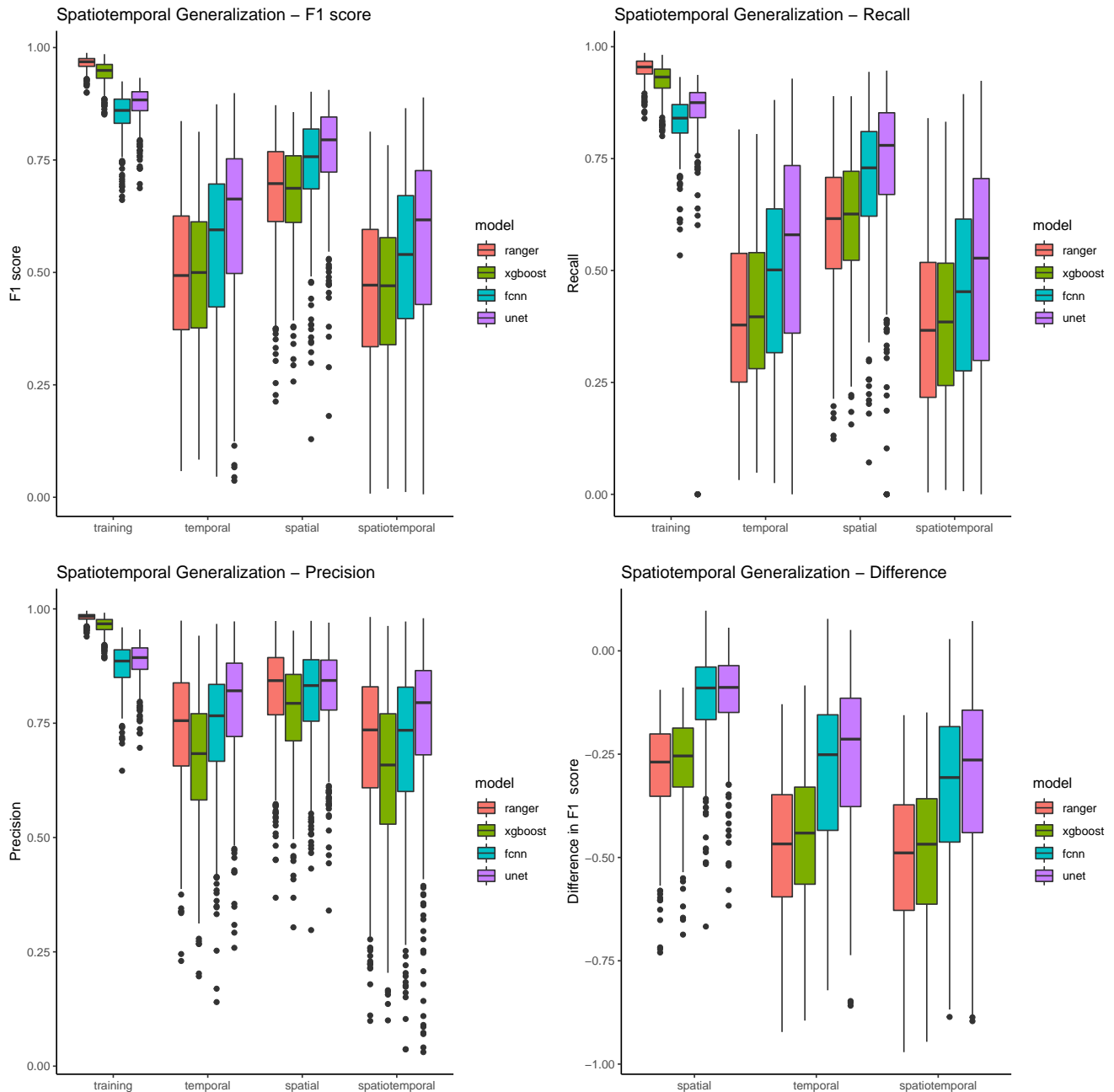
Figure 9. The figure shows boxplots for all three performance metrics plus the performance difference between scopes. Boxplots are grouped by scope and colored by model.

Table 4. Summary Spatiotemporal Generalization—F1.

	Ranger (SD)	Xgboost (SD)	FCNN (SD)	UNET (SD)
training (F1)	0.965 (0.014)	0.944 (0.025)	0.853 (0.045)	0.875 (0.038)
temporal (F1)	0.489 (0.171)	0.488 (0.156)	0.553 (0.183)	0.612 (0.189)
spatial (F1)	0.679 (0.115)	0.674 (0.106)	0.738 (0.109)	0.771 (0.102)
spatiotemporal (F1)	0.45 (0.191)	0.449 (0.172)	0.508 (0.204)	0.561 (0.216)

Table 5. Summary Spatiotemporal Generalization—Recall.

	Ranger (SD)	Xgboost (SD)	FCNN (SD)	UNET (SD)
training (RC)	0.95 (0.024)	0.925 (0.035)	0.832 (0.054)	0.854 (0.105)
temporal (RC)	0.393 (0.182)	0.409 (0.173)	0.479 (0.209)	0.539 (0.231)
spatial (RC)	0.6 (0.145)	0.615 (0.136)	0.701 (0.146)	0.738 (0.163)
spatiotemporal (RC)	0.366 (0.196)	0.38 (0.184)	0.443 (0.224)	0.495 (0.251)

Table 6. Summary Spatiotemporal Generalization—Precision.

|	Ranger (SD)	Xgboost (SD)	FCNN (SD)	UNET (SD)
training (PR)	0.981 (0.008)	0.963 (0.018)	0.877 (0.046)	0.887 (0.039)
temporal (PR)	0.739 (0.132)	0.669 (0.137)	0.735 (0.139)	0.789 (0.124)
spatial (PR)	0.819 (0.101)	0.775 (0.109)	0.809 (0.105)	0.823 (0.096)
spatiotemporal (PR)	0.701 (0.177)	0.636 (0.174)	0.699 (0.177)	0.752 (0.171)

The statistical analysis in Table 7 shows that ranger is the superior classifier in the training scope, and both pixel based classifiers are significantly better than the convlution based models. In all other scopes, the convolution based classifiers perform significantly better, with UNET being the most accurate of them. The difference between ranger and xgboost is much less significant, than the difference between UNET and xgboost.

Table 7. Summary of Mann–Whitney U tests.

H1	Training	Temporal	Spatial	Spatiotemporal
H1: ranger > xgboost	1.07×10^{-56}	3.95×10^{-1}	1.07×10^{-1}	3.21×10^{-1}
H1: ranger > FCNN	7.60×10^{-165}	1.00×10^{0}	1.00×10^{0}	1.00×10^{0}
H1: ranger > UNET	5.44×10^{-163}	1.00×10^{0}	1.00×10^{0}	1.00×10^{0}
H1: xgboost > ranger	1.00×10^{0}	6.05×10^{-1}	8.93×10^{-1}	6.79×10^{-1}
H1: xgboost > FCNN	2.86×10^{-149}	1.00×10^{0}	1.00×10^{0}	1.00×10^{0}
H1: xgboost > UNET	5.85×10^{-133}	1.00×10^{0}	1.00×10^{0}	1.00×10^{0}
H1: FCNN > ranger	1.00×10^{0}	4.22×10^{-10}	3.21×10^{-19}	1.39×10^{-7}
H1: FCNN > xgboost	1.00×10^{0}	1.55×10^{-11}	1.20×10^{-24}	5.51×10^{-9}
H1: FCNN > UNET	1.00×10^{0}	1.00×10^{0}	1.00×10^{0}	1.00×10^{0}
H1: UNET > ranger	1.00×10^{0}	3.31×10^{-29}	3.94×10^{-42}	1.89×10^{-21}
H1: UNET > xgboost	1.00×10^{0}	5.87×10^{-32}	2.33×10^{-49}	2.94×10^{-24}
H1: UNET > FCNN	3.88×10^{-19}	1.09×10^{-8}	1.43×10^{-8}	8.63×10^{-7}

3.3. Geometric Generalization

Figure 10 shows that the patterns established in the spatiotemporal generalization also appear in sensitivity performances in maize fields, with the pixel based classifiers being relatively similar, FCNN being much better, and UNET outperforming all of them. At the same time, the value ranges and medians of the performances are very similar for the scopes *bothmaize* and *tomaize*. The value ranges of the scopes *nonmaize* and *frommaize* are relatively low, but *nonmaize* shows a much smaller variability, and a much more consistent specificity (see also Table 8). From the geometric validation scopes, *nonmaize* is special in a way that it is the only scope that contains pixels that are not agricultural. The fraction of these pixels is

rather high (around 87% in comparison to 3/5/5% for *bothmaize*/*tomaize*/*frommaize*), and they are at the same time much easier to discriminate. Looking at differences between the algorithms, the specificity values show very little difference, mostly because most values are consistently high. The right side of Figure 10 summarizes the performance metrics by calculating the differences in performance for maize pixels (to maize) and nonmaize pixels (from maize). This difference is around 0 in all cases, with again higher variance in the sensitivities compared to the specificities.
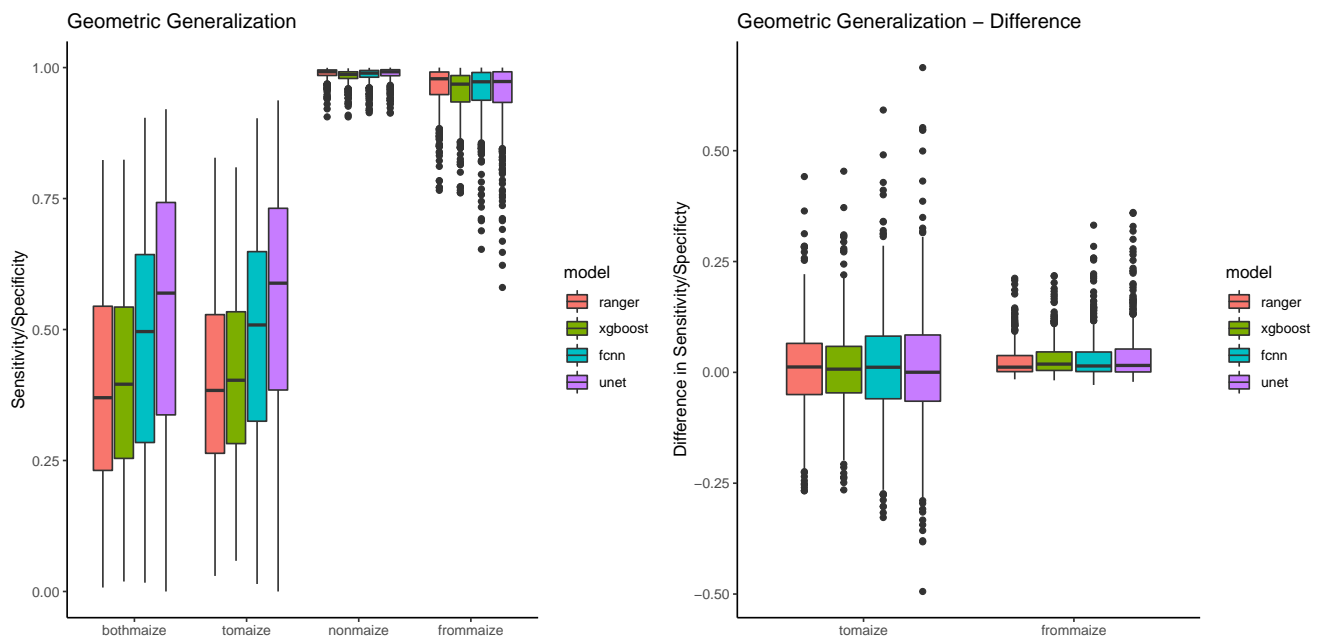

Figure 10. The figure shows a boxplot for the performance difference between scopes colored by model.

Table 8. Overview of Spatiotemporal Generalization Scopes.

	Ranger (SD)	Xgboost (SD)	FCNN (SD)	UNET (SD)
bothmaize (Sens)	0.386 (0.194)	0.402 (0.184)	0.471 (0.221)	0.532 (0.249)
tomaize (Sens)	0.396 (0.183)	0.412 (0.174)	0.485 (0.212)	0.543 (0.233)
nonmaize (Spec)	0.988 (0.011)	0.984 (0.013)	0.986 (0.013)	0.988 (0.013)
frommaize (Spec)	0.963 (0.042)	0.953 (0.045)	0.954 (0.053)	0.951 (0.063)

The statistical analysis (see Table 9) verifies that there is no significant difference between the *bothmaize* and *tomaize* scope. This means, that none of the four classifiers of maize fields is biased by the status of the field in the training year. If a field is maize in the training year, its not more or less likely to be detected in the validation year than a field that is not maize in the training year. On the other hand, the analysis shows significant differences between the *nonmaize* and *frommaize* scope. This is again due to the fact that *nonmaize* is special because of a much larger variety of pixels. The scope signifiance is however relatively consistent for all four models.

Table 9. Summary of Mann–Whitney U tests—geoscopes.

H1	Ranger	Xgboost	FCNN	UNET
H1: bothmaize != tomaize	3.35×10^{-1}	3.48×10^{-1}	4.05×10^{-1}	6.95×10^{-1}
H1: nonmaize != frommaize	2.93×10^{-32}	1.21×10^{-41}	2.37×10^{-29}	3.01×10^{-32}

4. Discussion

4.1. Input Data

The focus of this study was to assess advantages of convolution based models in model generalization. To find and optimize a novel combination of deep learning model and multitemporal image data was outside of the scope of this study. Thusly, some simplifications have been made which provide ample opportunity to improve the absolute performances of the models, while at the same time preserving the relative differences between the validation scopes. The Sentinel-1 imagery was processed out the box and monthly averages were calculated. Improvements to data quality can be largely categorized in three groups, noise reduction, adding information and reducing redundancy. Thermal noise removal is already implemented in the Sentinel-1 Toolbox, but bias introduced by different incidence angles could be corrected with empirical coefficients. Information could simply be added, either from the same sensor but with a different cross-polarization mode, or from other sensors with high revisit times like Sentinel-2. Furthermore and within the SAR domain, data preprocessing algorithms like polarimetric decomposition could be used to gain more information about the exact nature of the scattering. Furthermore, finally, redundancy could be reduced with data transformation algorithms like the principal component analysis or even implemented in the deep learning framework itself with autoencoders.

One major caveat is that the drop in sensitivity/recall rates in all the validation scopes was very high for most models. However, two counterpoints suggest that this is a modeling problem and not an input data problem. First, the variability with a given model is high, which indicates a strong random effect of tile selection, for both validation and training tiles. Furthermore, high variability means that there is still a lot of potential to increase model sensitivities by increasing the size of training datasets, which is easily possible by several orders of magnitude. Secondly, the variability between models is very high, which in a similar way suggest that there is potential to delineate maize fields, which could be developed with further algorithm Improvements. As a consequence it can be concluded that SAR imagery provides a good basis for temporally and spatially transferable crop classification models.

4.2. Validation Scopes

Validating within the training scope is the most common step of each modeling process, and for deep learning algorithms even required in each iteration as part of the loss function. High performance in the training scope is necessary for each model, but not a sufficient criterion for a useful model in most research contexts. All models in this study were able to achieve reasonably high performances in the training scope, but the spatiotemporal scopes were able to show a much more complete picture. Here, the convolution based algorithms performed better by a very significant margin.

Changing the location or year of a sample is a simplification for changing the many factors that result in a different phenology with the same crop type. These factors can be anything, from different altitudes and microclimate in neighboring regions, different precipitation and drought patterns in two years, to different management practices of two farmers in fields right next to each other. Ideally, we would have complete spatial information about all of them, and we would do a stratified sampling based on altitude, water availability and crop management. While altitude and precipitation strata could be implemented, crop management variables are nearly impossible to get for larger datasets and multiple years. Furthermore, there are many more variables that needed to be accounted for in a stratified sampling. Some of these factors are more likely to be different in different years, while some of them are more likely to be different in different locations. Consequently, changing both allows for a more complete assessment of model performance, without the requirement of other spatial datasets necessary for a stratified sampling, a concept that can be easily applied to other crop classification studies.

The geometric scopes however, are most likely not very useful in other contexts. They were only included to analyse a very specific aspect of convolution based classifiers, and were not able to lead to any further conclusions. Potential overfitting and memory problems could be also analysed with spatial validation scopes, but geometric scopes were more precise in that regard.

4.3. Model Training and Evaluation

Since optimizing the absolute performance of the models was not the focus of the study, the pixel based models in this study were trained with default hyperparameters wherever possible, and it is expected that the impact on their performance is very low. The same cannot be said about the convolution based models, where model performance can be optimized by many different parameters, activation functions, choosing the right optimizer and adjusting the network architecture. Wherever possible default parameters were used, but there is without a doubt a lot of potential to find better models than the ones used in this study. This does not affect the findings, but could in the future be used to build on the validation framework established in this study.

Despite this limitation, we were able to show, that models are much more performant if applied to different regions than compared to different years. In SAR remote sensing, there are less reasons to assume this would be the case, compared to optical remote sensing where the image acquisition is very time critical and sporadic because of atmospheric constraints that apply much less to radar data. Instead we have consistent monthly means that cover the entire year and are largely invariant to atmospheric conditions. However, the phenology development curve is still potentially very different in different years because of precipitation and sunlight. Additionally, there might be SAR specific factors that cause short term fluctuations of the backscatter coefficients. Surface wetness, for example, can cause short term spikes in VH backscatter of vegetation coverage [49]. Because of these uncertainties, it is unclear if this particular finding can be transferred to other crop classification applications. It is however, evidence that spatial validation (i.e., randomized samples from one year) gives an incomplete picture of model transferability.

In addition we provide evidence that including field geometry (i.e., with convolutional neural networks) is beneficial to model generalization. Pixel based models, while being consistently more performant in the training subset, showed clearly less sensitivity to maize pixels than the convolution based models, while at the same time being unable to offer more precision as a tradeoff. Geometric overfitting has been shown to not be a problem in any of the models. Though it cannot be ruled out that geometry based memory effects could happen in certain models, there is still enough evidence that convolution based models generalize much better, and it is hard to imagine future crop classification research without it.

5. Conclusions

- SAR imagery provides a good basis for temporally and spatially transferable crop classification models

Despite the simplifications made in the preprocessing of the datasets, and despite the fact that it is also affected by topography and soil wetness, the information represented by the SAR imagery was shown to be sufficient to detect between 95% (ranger) and 85% (UNET) of maize fields in the training scope. The main benefits is independence from atmospheric conditions and consistently available image products as a result. The main drawback is insensitivity to leaf pigments, and therefore major limitiations in other aspects of vegetation monitoring.

- Within a region, temporal generalization is harder than spatial generalization.

Most studies in the past have focused on spatial validation by splitting their training datasets. With freely available datasets and a growing body of highly capable classifiers, temporal generalization is a designated goal for a growing body of LULC classification

research. We were able to highlight the importance of temporal validation, by showing that there was a significant drop in model quality if a model was validated with data from a different year, compared to data from a different region. F1 scores dropped to values between 0.49 (xgboost) and 0.61 (UNET) in the temporal validation scope, compared to values between 0.67 (xgboost) an 0.77 (UNET).

- Validation scopes can be helpful in assessing model quality

Analyzing these models showed that a comparatively high modeling performance was achieved within the training scope, even though the amount of training samples was relatively low. Switching scopes gave a much more complete picture. By analyzing the drop in performance observed when validating with samples from these different scopes, some algorithms were shown to be significantly more robust, even though their performance in the training scope lower. Therefore, the concept of validation scopes has been proven to be a helpful tool in assessing the quality of a model.

- Including field geometry is helpful and geometric overfitting is not a problem

The main goal of this study was to compare classical machine learning classifiers to state of the art deep learning algorithms. By using a robust harmonized dataset for these years, as well as a randomized, scope based sampling approach, we were able to show that convolution based algorithms are clearly superior to pixel based algorithms. Albeit none of the tested algorithms in this study was able to produce a well generalizable classifier, it was shown that kernel based convlutions showed much more promise, and have potential to be a stable entity in all future crop classification research.

Author Contributions: Conceptualization, M.G.; methodology, M.G.; software, M.G.; validation, M.G.; writing—original draft preparation, M.G.; writing—review and editing, M.G. and T.U.; visualization, M.G.; supervision, T.U.; All authors have read and agreed to the published version of the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
RF	Random Forest
masl	meters above sea level
FCNN	Fully Convolutional Neural Network
SAR	Synthetic Aperture Radar
GSD	Ground Sampling Distance
GRD	Ground Range Detected
CM	Confusion Matrix

References

1. Khatami, R.; Mountrakis, G.; Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classi fi cation processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [CrossRef]
2. Song, X.-P.; Huang, W.; Hansen, M.C. An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. *Sci. Remote Sens.* **2021**, 102560. [CrossRef]
3. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Exp. Syst. Appl.* **2021**, *169*, 114417. [CrossRef]
4. Wang, S.; Chen, W.; Xie, S.M.; Azzari, G.; Lobell, D.B. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sens.* **2020**, *12*, 207. [CrossRef]

5. Lavreniuk, M.; Kussul, N.; Novikov, A. Deep learning crop classification approach based on sparse coding of time series of satellite data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 4812–4815. [CrossRef]

6. Zhang, D.; Pan, Y.; Zhang, J.; Hu, T.; Zhao, J.; Li, N.; Chen, Q. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sens. Environ.* **2020**, *247*, 111912. [CrossRef]

7. Rustowicz, R.; Cheong, R.; Wang, L.; Ermon, S.; Burke, M.; Lobell, D. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods. In Proceedings of the CVPR Workshops, Long Beach, CA, USA, 16–20 June 2019; Volume 1, pp. 75–82.

8. Harfenmeister, K.; Itzerott, S.; Weltzien, C.; Spengler, D. Agricultural Monitoring Using Polarimetric Decomposition Parameters of Sentinel-1 Data. *Remote Sens.* **2021**, *13*, 575. [CrossRef]

9. Parida, B.R.; Mandal, S.P. Polarimetric decomposition methods for LULC mapping using ALOS L-band PolSAR data in Western parts of Mizoram, Northeast India. *SN Appl. Sci.* **2020**, *2*. [CrossRef]

10. Sonobe, R. Parcel-based crop classification using multi-temporal TerraSAR-X dual polarimetric data. *Remote Sens.* **2019**, *11*, 1148. [CrossRef]

11. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]

12. Wei, S.; Zhang, H.; Wang, C.; Xu, L.; Wu, F.; Zhang, B. Large-scale rice mapping of Thailand using sentinel-1 multi-temporal SAR data. In Proceedings of the 2019 SAR in Big Data Era, BIGSARDATA 2019, Beijing, China, 5–6 August 2019. [CrossRef]

13. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [CrossRef]

14. Kumar, P.; Gupta, D.K.; Mishra, V.N.; Prasad, R. Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data. *Int. J. Remote Sens.* **2015**, *36*, 1604–1617. [CrossRef]

15. Skakun, S.; Kussul, N.; Shelestov, A.Y.; Lavreniuk, M.; Kussul, O. Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3712–3719. [CrossRef]

16. Castro, J.D.B.; Feitoza, R.Q.; Rosa, L.C.L.; Diaz, P.M.A.; Sanches, I.D.A. A Comparative Analysis of Deep Learning Techniques for Sub-Tropical Crop Types Recognition from Multitemporal Optical/SAR Image Sequences. In Proceedings of the 30th Conference on Graphics, Patterns and Images, SIBGRAPI 2017, Niterói, Brazil, 17–20 October 2017; pp. 382–389. [CrossRef]

17. Cai, Y.; Guan, K.; Peng, J.; Wang, S.; Seifert, C.; Wardlow, B.; Li, Z. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* **2018**, *210*, 35–47. [CrossRef]

18. Momm, H.G.; ElKadiri, R.; Porter, W. Crop-type classification for long-term modeling: An integrated remote sensing and machine learning approach. *Remote Sens.* **2020**, *12*, 449. [CrossRef]

19. Ajadi, O.A.; Barr, J.; Liang, S.z.; Ferreira, R.; Kumpatla, S.P. Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *97*, 102294. [CrossRef]

20. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**. [CrossRef]

21. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]

22. Orynbaikyzy, A.; Gessner, U.; Conrad, C. Crop type classification using a combination of optical and radar remote sensing data: A review. *Int. J. Remote Sens.* **2019**, *40*, 6553–6595. [CrossRef]

23. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]

24. Pax-Lenney, M.; Woodcock, C.E.; Macomber, S.A.; Gopal, S.; Song, C. Forest mapping with a generalized classifier and Landsat TM data. *Remote Sens. Environ.* **2001**, *77*, 241–250. [CrossRef]

25. Wolpert, D.H.; Macready, W.G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]

26. Su, T.; Zhang, S. Local and global evaluation for remote sensing image segmentation. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 256–276. [CrossRef]

27. Ruf, T.; Gilcher, M.; Emmerling, C.; Udelhoven, T. Implications of Bioenergy Cropping for Soil: Remote Sensing Identification of Silage Maize Cultivation and Risk Assessment Concerning Soil Erosion and Compaction. *Land* **2021**, *10*, 128. [CrossRef]

28. Gilcher, M.; Ruf, T.; Emmerling, C.; Udelhoven, T. Remote sensing based binary classification of maize. Dealing with residual autocorrelation in sparse sample situations. *Remote Sens.* **2019**, *11*, 2172. [CrossRef]

29. Statistical Office Rhineland-Palatinate. Statistisches Jahrbuch Rheinland-Pfalz 2017. 2017. Available online: www.statistik.rlp.de/fileadmin/dokumente/jahrbuch/Jahrbuch2017.pdf (accessed on 19 February 2021).

30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

31. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [CrossRef]

32. Chen, X.; Yin, D.; Chen, J.; Cao, X. Effect of training strategy for positive and unlabelled learning classification: Test on Landsat imagery. *Remote Sens. Lett.* **2016**, *7*, 1063–1072. [CrossRef]
33. Saini, R.; Ghosh, S.K. Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery. *Geocarto Int.* **2019**, 1–19. [CrossRef]
34. Memon, N.; Patel, S.B.; Patel, D.P. *Comparative Analysis of Artificial Neural Network and XGBoost Algorithm for PolSAR Image Classification*; Springer: Cham, Switzerland, 2019; Volume 1941, pp. 452–460.
35. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco, CA, USA, 13–17 August 2016.
36. Saini, R.; Ghosh, S.K. Ensemble classifiers in remote sensing: A review. In Proceedings of the IEEE International Conference on Computing, Communication and Automation (ICCCA 2017), Greater, Noida, 5–6 May 2017; pp. 1148–1152. [CrossRef]
37. Briem, G.J.; Benediktsson, J.A.; Sveinsson, J.R. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2291–2299. [CrossRef]
38. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl. Soft Comput. J.* **2020**, *86*, 105837. [CrossRef]
39. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
40. Allaire, J.J.; Chollet, F. Keras: R Interface to 'Keras'. 2020. Available online: https://cran.r-project.org/web/packages/keras/index.html (accessed on 19 February 2021).
41. Falbel, D.; Zak, K. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2020. Available online: https://github.com/r-tensorflow/unet (accessed on 19 February 2021).
42. Federal Ministry of Justice and Consumer Protection. InVeKoSV. Available online: https://www.gesetze-im-internet.de/invekosv_2015/index.html (accessed on 19 February 2021).
43. Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.F.; Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* **2017**, *199*, 415–426. [CrossRef]
44. Aulard-Macler, M. Sentinel-1 Product Definition. 2012. Available online: https://sentinels.copernicus.eu/documents/247904/1877131/Sentinel-1-Product-Definition.pdf/6049ee42-6dc7-4e76-9886-f7a72f5631f3?t=1461673251000 (accessed on 19 February 2021).
45. Prudente, V.H.R.; Oldoni, L.V.; Vieira, D.C.; Cattani, C.E.V.; Del'Arco Sanches, I. Relationship between SAR/Sentinel-1 polarimetric and interferometric data with biophysical parameters of agricultural crops. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Arch.* **2019**, *42*, 599–607. [CrossRef]
46. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**. [CrossRef]
47. Canbek, G.; Temizel, T.T.; Sagiroglu, S.; Baykal, N. Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. In Proceedings of the 2nd International Conference on Computer Science and Engineering (UBMK 2017), Antalya, Turkey, 5–8 October 2017; pp. 821–826. [CrossRef]
48. Pontius, R.G.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]
49. Molijn, R.A.; Iannini, L.; Dekker, P.L.; Magalhães, P.S.G.; Hanssen, R.F. Vegetation Characterization through the Use of Precipitation-Affected SAR Signals. *Remote Sens.* **2018**, *77*, 1–17. [CrossRef]

5 Synthesis

5.1 Main Findings

In the thesis at hand, we were able to progress the scientific field of remote sensing-based crop classification by:

- adding onto the already existing body of evidence that operational crop mapping is practically achievable
- elaborating on some very important technical details in terms of:
    - input data
    - post-processing of predictions
    - real world use of prediction maps
    - robustness of novel classification algorithms

Objective 1: Explore capabilities and practical limitations of optical remote sensing-based crop classification

In chapter 2 "Remote Sensing Based Binary Classification of Maize. Dealing with Residual Autocorrelation in Sparse Sample Situations", we generated two sets of reference data for the years 2009 and 2016 in the study region Eifelkreis Bitburg-Prüm by drawing polygons based on high resolution aerial imagery. A sparse set covered the entire region with sample polygons and large unlabeled gaps in-between. Additionally, a second set of dense polygons was drawn for small square shaped subsets. The dense set was necessary to analyze residual autocorrelation, and evaluate the influence of Regression Kriging and Gaussian kernel filtering on model performance, while the best approach was used to train and apply random forest classification models on the sparse medium scale dataset, to get model predictions for the entire study area.

We showed that in spite of clouds making it impossible to acquire images covering the entire area at a consistent phenological state for each individual year, we were able to produce two high quality dense prediction maps for about 87,000 hectares of agricultural land based on RapidEye imagery. Data from the years in-between were also available, but not cloud free. Another problem of optical remote sensing data is the reliance on consistent leaf pigmentation. Though we managed to get two scenes from a near perfect acquisition window, some fields with drought stress where not recognized because of the loss of chlorophyll.

Additionally, the approach of using high resolution aerial imagery to label reference data showed great promise, but the success was highly dependent on data availability. It is a basic requirement that both the acquisition of the aerial, and of the satellite imagery have to be within the preferred window of phenological development. Though both data products rely on similar circumstances, mostly a cloud free late August or early September, this further narrows down the years in which this approach could feasibly be pursued.

In addition to the input data, the chosen modeling approach has proven to be successful. Random forest classification has been able to distinguish maize with high accuracies. While Kriging was

able to deal with residual autocorrelation, Gaussian blur has proven to be much more practical. Firstly, it does not depend on a set of dense samples in the same way as regression kriging does to build up and predict errors based on variograms. Secondly, it is much more runtime efficient, since it is just a kernel function.

Objective 2: Explore regional applications of crop classification

In chapter 3 "Implications of Bioenergy Cropping for Soil: Remote Sensing Identification of Silage Maize Cultivation and Risk Assessment Concerning Soil Erosion and Compaction", we used the spatially explicit maize predictions to assess potential negative externalities of political measures.

We compared the location of newly constructed biogas producing units (BPU) with the ongoing expansion of maize fields in the study area of Eifelkreis Bitburg-Prüm. We were able to show that the increase in agricultural area covered by maize fields from 2009 to 2016 correlates with an increase in BPUs over that timeframe. Consequently, we were able to show that governmental subsidies for biogas production can introduce land use pressure and therefore act as a driver for major land use changes.

In a next step we analyzed the implications of these artificially driven land use changes on soil health, specifically focused on soil compaction and soil erosion. The analysis in terms of compaction was severely limited by data availability, since plausible assessment of pedotope susceptibility to compaction was only possible for about one third of the total agricultural area. In spite of this limitation, about 195 hectares of maize fields with particularly high risk of soil compaction could be identified in 2016. The analysis of soil erosion potential was much more complete, since it was not as much limited by data availability, and only required a complete DEM to compute runoff potential. It was shown that, in 2016, 2677 hectares of maize were placed on soil with high or very high erosion potential, which could result in a net loss of potentially over 25 tons per hectare per year.

Objective 3: Explore spatial and regional transferability of upcoming state of the art models

In chapter 4 "*Field Geometry and the Spatial and Temporal Generalization of Crop Classification Algorithms—A Randomized Approach to Compare Pixel Based and Convolution Based Methods*" the focus went back to more technical aspects of crop classification, based on newly available input data, and upcoming state of the art models. We built a consistent dataset based on monthly Sentinel-1 backscatter averages. As a result, we obtained 12-band images for four years between 2016 and 2019 for the entire study area of Eifelkreis Bitburg-Prüm. In our sampling approach, we trained 500 models based on a random year and random 5 km by 5 km subsets of the study region, with different machine learning and deep learning classifiers. We then computed validation metrics for these different scopes, ranging from the same year and same location as the training data, to different year and different locations.

We were able show that our monthly SAR composite provided a sufficient basis for crop classification, since the training performance was consistently high, averaging a recall of 0.95 with the random forest classification. Additionally, we outlined several approaches to improve

the input data, such as incidence angle correction or the inclusion of optical imagery such as Sentinel-2. The main result of this objective is that convolution-based deep learning models showed a significant improvement of model generalization, both over time and space. Temporal generalization was more difficult in our study region, as performance decreased much more if a model was validated in a different year. While the average F1 score of random forest classification fell from 0.97 to 0.49 when generalized temporally, the scores of the UNET classifier only decreased from 0.88 to 0.61. Model performances from all scope/model combinations were compared with U-tests, which showed a significant superiority of all deep learning models over the machine learning models. The hypothesis that convolution-based models could overfit on field geometry was tested and could not be verified.

5.2   Conclusion and Outlook

In this study we showed substantial evidence that the past 60 years of advancements in environmental remote sensing paved the way for a time when operational global crop monitoring is technically possible, and this time is now. We outlined how advances in sensor hardware and data processing can give us crucial information about the canopy of standing vegetation with very good spatial and temporal resolution. In a medium-scale case study we showed how spatially explicit crop abundance data can be generated and validated based on freely available remote sensing satellite and aerial imagery. We gave an example of how this data can be used to assess regionally specific environmental risks in the form of soil compaction and erosion. We also introduced a novel way to structure validation data in distinct scopes, and used this framework to assess the generalization capabilities of upcoming state of the art classification algorithms.

The two major bottlenecks towards an operational global crop model are reference data and model architecture. In situ field surveys are useful and sufficient for regional studies, but they cannot be used to produce models that are able to work well for different years, and even less so in different regions of the globe. Statistical records existent in European countries and the USA are great tools to assess temporal generalization, but are still regionally limited and might produce models that do not work at all in Asia or Africa, where the agricultural structure is quite different in terms of field geometries and phenology cycles. Generating agricultural labels based on high resolution imagery is feasible for certain crops that are highly distinguishable by texture, like maize, but many crops might not be detected by human vision as easily. While many tools exist, it is clear that establishing a globally stratified and harmonized dataset of crop type reference data can only be accomplished through international cooperation. First attempts at concerted efforts exist, and have been outlined in section 1.2.3, and it is clear that more steps are necessary in this direction, to counteract the heavy regional bias in crop classification research towards European countries, China and the USA.

The improvement of modeling architecture in deep learning is the focus of research in many scientific fields, taking many different directions. Due to the fact that these models can get arbitrarily complex in terms of fitted parameters, and still show scaling in terms of performance, the consolidation of these architectures could potentially take very long. Specifically in the field of remote sensing, it is much too early to make predictions about the direction this development

will go, or how long it might take. A conceptually very simple modeling approach like random forests took decades to find its now very prominent place in the remote sensing community. Deep learning models in their current iteration have been around for not even one decade, and are more complex than any machine learning model, by several orders of magnitude. Given their ongoing popularity in many fields of science, engineering and general technology, it is important to not get overly optimistic about their role in this specific subset of remote sensing applications. Though, a steadily increasing body of research, this thesis included, implies that these algorithms are here to stay, and that they are a huge opportunity to improve satellite-based crop monitoring.

Sources

Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. 2017. "Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(12):2481–95.

Bailly, S., S. Giordano, L. Landrieu, and N. Chehata. 2018. "Crop-Rotation Structured Classification Using Multi-Source Sentinel Images and LPIS for Crop Type Mapping." *International Geoscience and Remote Sensing Symposium (IGARSS)* 2018-July:1950–53. doi: 10.1109/IGARSS.2018.8518427.

Barnes, Andrew D., Malte Jochum, Steffen Mumme, Noor Farikhah Haneda, Achmad Farajallah, Tri Heru Widarto, and Ulrich Brose. 2014. "Consequences of Tropical Land Use for Multitrophic Biodiversity and Ecosystem Functioning." *Nature Communications* 5:1–7. doi: 10.1038/ncomms6351.

Bartholomé, E., and a. S. Belward. 2005. "GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data." *International Journal of Remote Sensing* 26(February 2015):1959–77. doi: 10.1080/01431160412331291297.

Blaschke, T. 2010. "Object Based Image Analysis for Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65(1):2–16. doi: 10.1016/j.isprsjprs.2009.06.004.

Blickensdörfer, Lukas, Marcel Schwieder, Dirk Pflugmacher, Claas Nendel, Stefan Erasmi, and Patrick Hostert. 2021. "Mapping of Crop Types and Crop Sequences with Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data for Germany." *Remote Sensing of Environment* 269(April 2021):112831. doi: 10.1016/j.rse.2021.112831.

Borg, Erik. 2014. "Geodaten Auf Bestellung - Das RapidEye Science Archive RESA." (March 2010).

Borrelli, Pasquale, David A. Robinson, Larissa R. Fleischer, Emanuele Lugato, Cristiano Ballabio, Christine Alewell, Katrin Meusburger, Sirio Modugno, Brigitta Schütt, Vito Ferro, Vincenzo Bagarello, Kristof Van Oost, Luca Montanarella, and Panos Panagos. 2017. "An Assessment of the Global Impact of 21st Century Land Use Change on Soil Erosion." *Nature Communications* 8(1). doi: 10.1038/s41467-017-02142-7.

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

Buchhorn, Marcel, Bruno Smets, Luc Bertels, Myroslava Lesiv, Nandin-Erdene Tsendbazar, D. Masiliunas, L. Linlin, Martin Herold, and S. Fritz. 2020. "Copernicus Global Land Service: Land Cover 100m: Collection 3: Epoch 2015: Globe (Version V3.0.1)." *Zenodo* 1–14.

Cai, Yaping, Kaiyu Guan, Jian Peng, Shaowen Wang, Christopher Seifert, Brian Wardlow, and Zhan Li. 2018. "A High-Performance and in-Season Classification System of Field-Level Crop Types Using Time-Series Landsat Data and a Machine Learning Approach." *Remote Sensing of Environment* 210(January):35–47. doi: 10.1016/j.rse.2018.02.045.

Campos-Taberner, Manuel, Francisco Javier García-Haro, Beatriz Martínez, Emma Izquierdo-Verdiguier, Clement Atzberger, Gustau Camps-Valls, and María Amparo Gilabert. 2020. "Understanding Deep Learning in Land Use Classification Based on Sentinel-2 Time Series." *Scientific Reports* 10(1):1–12. doi: 10.1038/s41598-020-74215-5.

Carpenter, Stephen R., Emily H. Stanley, and M. Jake Vander Zanden. 2011. "State of the World's Freshwater Ecosystems: Physical, Chemical, and Biological Changes." *Annual Review of Environment and Resources* 36(1):75–99. doi: 10.1146/annurev-environ-021810-094524.

Chen, Gang, Qihao Weng, Geoffrey J. Hay, and Yinan He. 2018. "Geographic Object-Based Image Analysis (GEOBIA): Emerging Trends and Future Opportunities." *GIScience and Remote Sensing* 55(2):159–82. doi: 10.1080/15481603.2018.1426092.

Sources

Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C. Jones, C. Le Quéré, R. B. Myneni, S. Piao, P. Thornton, Philippe Ciais France, Jan Willem, Pierre Friedlingstein, and Guy Munhoven. 2013. *2013: Carbon and Other Biogeochemical Cycles*.

Copernicus Land Monitoring Service. 2010. "CORINE Land Cover Nomenclature Conversion to Land Cover Classification System." *Report* (Clc):6.

Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks." *Machine Learning* (20):273–97. doi: 10.1109/64.163674.

Csillik, Ovidiu, Mariana Belgiu, Gregory P. Asner, and Maggi Kelly. 2019. "Object-Based Time-Constrained Dynamic Time Warping Classification of Crops Using Sentinel-2." *Remote Sensing* 11(10). doi: 10.3390/rs11101257.

d'Andrimont, Raphaël, Momchil Yordanov, Guido Lemoine, Janine Yoong, Kamil Nikel, and Marijn van der Velde. 2018. "Crowdsourced Street-Level Imagery as a Potential Source of in-Situ Data for Crop Monitoring." *Land* 7(4):1–26. doi: 10.3390/land7040127.

Dawbin, Ken W., and John C. Evans. 1988. "Large Area Crop Classification in New South Wales, Australia, Using Landsat Data." *International Journal of Remote Sensing* 9(2):295–301. doi: 10.1080/01431168808954853.

Defourny, Pierre, Sophie Bontemps, Nicolas Bellemans, Cosmin Cara, Gérard Dedieu, Eric Guzzonato, Olivier Hagolle, Jordi Inglada, Laurentiu Nicola, Thierry Rabaute, Mickael Savinaud, Cosmin Udroiu, Silvia Valero, Agnès Bégué, Jean François Dejoux, Abderrazak El Harti, Jamal Ezzahar, Nataliia Kussul, Kamal Labbassi, Valentine Lebourgeois, Zhang Miao, Terrence Newby, Adolph Nyamugama, Norakhan Salh, Andrii Shelestov, Vincent Simonneaux, Pierre Sibiry Traore, Souleymane S. Traore, and Benjamin Koetz. 2019. "Near Real-Time Agriculture Monitoring at National Scale at Parcel Resolution: Performance Assessment of the Sen2-Agri Automated System in Various Cropping Systems around the World." *Remote Sensing of Environment* 221(October 2018):551–68. doi: 10.1016/j.rse.2018.11.007.

Deng, J. and Dong, W. and Socher, R. and Li, L.-J. and Li, K. and Fei-Fei, L. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." *IEEE Computer Vision and Pattern Recognition* 20(11):1221–27.

Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini. 2012. "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services." *Remote Sensing of Environment* 120:25–36. doi: 10.1016/j.rse.2011.11.026.

Ellis, Erle C., Arthur H. W. Beusen, and Klein Kees Goldewijk. 2020. "Anthropogenic Biomes: 10,000 BCE to 2015 CE." *Land* 9(5):8–10. doi: 10.3390/LAND9050129.

Ellis, Erle C., Nicolas Gauthier, Kees Klein Goldewijk, Rebecca Bliege Bird, Nicole Boivin, Sandra Díaz, Dorian Q. Fuller, Jacquelyn L. Gill, Jed O. Kaplan, Naomi Kingston, Harvey Locke, Crystal N. H. McMichael, Darren Ranco, Torben C. Rick, M. Rebecca Shaw, Lucas Stephens, Jens Christian Svenning, and James E. M. Watson. 2021. "People Have Shaped Most of Terrestrial Nature for at Least 12,000 Years." *Proceedings of the National Academy of Sciences of the United States of America* 118(17):1–8. doi: 10.1073/pnas.2023483118.

Ellis, Erle C., Jed O. Kaplan, Dorian Q. Fuller, Steve Vavrus, Kees Klein Goldewijk, and Peter H. Verburg. 2013. "Used Planet: A Global History." *Proceedings of the National Academy of Sciences of the United States of America* 110(20):7978–85. doi: 10.1073/pnas.1217241110.

ESA. 2014. "Copernicus Open Access Hub." Retrieved (https://scihub.copernicus.eu/).

Sources

Estoque, Ronald C. 2020. "A Review of the Sustainability Concept and the State of SDG Monitoring Using Remote Sensing [Remote Sens., 12 (2020) (1770)] DOI:10.3390/Rs12111770." *Remote Sensing* 12(16). doi: 10.3390/RS12162512.

FAO. 2020. *World Food and Agriculture - Statistical Yearbook 2020*.

FAO & ITPS. 2015. *Intergovernmental Technical Panel on Soils. Status of the World's Soil Resources.*

Foley, Jonathan A., Navin Ramankutty, Kate A. Brauman, Emily S. Cassidy, James S. Gerber, Matt Johnston, Nathaniel D. Mueller, Christine O'Connell, Deepak K. Ray, Paul C. West, Christian Balzer, Elena M. Bennett, Stephen R. Carpenter, Jason Hill, Chad Monfreda, Stephen Polasky, Johan Rockström, John Sheehan, Stefan Siebert, David Tilman, and David P. M. Zaks. 2011. "Solutions for a Cultivated Planet." *Nature* 478(7369):337–42. doi: 10.1038/nature10452.

Foley, Jonathan a, Ruth Defries, Gregory P. Asner, Carol Barford, Gordon Bonan, Stephen R. Carpenter, F. Stuart Chapin, Michael T. Coe, Gretchen C. Daily, Holly K. Gibbs, Joseph H. Helkowski, Tracey Holloway, Erica a Howard, Christopher J. Kucharik, Chad Monfreda, Jonathan a Patz, I. Colin Prentice, Navin Ramankutty, and Peter K. Snyder. 2005. "Global Consequences of Land Use." *Science (New York, N.Y.)* 309(5734):570–74. doi: 10.1126/science.1111772.

Food and Agriculture Organization of the United Nations. 2021. *Food Security and Nutrition in the World Security , Improved Nutrition and Affordable Healthy Diets for All*.

Frantz, David. 2019. "FORCE-Landsat + Sentinel-2 Analysis Ready Data and Beyond." *Remote Sensing* 11(9). doi: 10.3390/rs11091124.

Gade, Martin. 2015. "Synthetic Aperture Radar Applications in Coastal Waters." *12th International Conference on the Mediterranean Coastal Environment, MEDCOAST 2015* 2(October):965–76.

Garg, Lakshya, Parul Shukla, Sandeep Kumar Singh, Vaishangi Bajpai, and Utkarsh Yadav. 2019. "Land Use Land Cover Classification from Satellite Imagery Using MUnet: A Modified UNET Architecture." *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* 4(Visigrapp):359–65. doi: 10.5220/0007370603590365.

Giordano, Sébastien, Simon Bailly, Loic Landrieu, Nesrine Chehata, Sébastien Giordano, Simon Bailly, Loic Landrieu, Nesrine Chehata, and Temporal Structured. 2018. "Temporal Structured Classification of Sentinel 1 and 2 Time Series for Crop Type Mapping."

Gomes, Vitor C. F., Gilberto R. Queiroz, and Karine R. Ferreira. 2020. "An Overview of Platforms for Big Earth Observation Data Management and Analysis." *Remote Sensing* 12(8):1–25. doi: 10.3390/RS12081253.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." *Remote Sensing of Environment*. doi: 10.1016/j.rse.2017.06.031.

Hansen, M. C., P. V Potapov, R. Moore, M. Hancher, S. a Turubanova, and A. Tyukavina. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342(November):850–53. doi: 10.1126/science.1244693.

Hardin, Garret. 1968. "The Tragedy of the Commons." *Science* 162(3859):1243–48. doi: 10.1201/9781351071765.

Ho, Tin Kam. 1995. "Random Decision Forests." *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1:278–82.

Hossain, Mohammad D., and Dongmei Chen. 2019. "Segmentation for Object-Based Image

Analysis (OBIA): A Review of Algorithms and Challenges from Remote Sensing Perspective." *ISPRS Journal of Photogrammetry and Remote Sensing* 150(February):115–34. doi: 10.1016/j.isprsjprs.2019.02.009.

Houska, T. R., and 2012 Johnson, A.P. 2012. "GloVis: U.S. Geological Survey General Information Product." 6(January):2.

Inglada, J., M. Arias, B. Tardy, D. Morin, S. Valero, O. Hagolle, G. Dedieu, G. Sepulcre, S. Bontemps, and P. Defourny. 2015. "Benchmarking of Algorithms for Crop Type Land-Cover Maps Using Sentinel-2 Image Time Series." *International Geoscience and Remote Sensing Symposium (IGARSS)* 2015-Novem:3993–96. doi: 10.1109/IGARSS.2015.7326700.

Inglada, Jordi, Marcela Arias, Benjamin Tardy, Olivier Hagolle, Silvia Valero, David Morin, Gèrard Dedieu, Guadalupe Sepulcre, Sophie Bontemps, Pierre Defourny, and Benjamin Koetz. 2015. "Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery." *Remote Sensing* 7(9):12356–79. doi: 10.3390/rs70912356.

Inglada, Jordi, Arthur Vincent, Marcela Arias, and Claire Marais-Sicre. 2016. "Improved Early Crop Type Identification by Joint Use of High Temporal Resolution Sar and Optical Image Time Series." *Remote Sensing* 8(5). doi: 10.3390/rs8050362.

IPBES. 2019. *Global Assessment Report on Biodiversity and Ecosystem Services*.

JECAM. 2015. "Joint Experiment for Crop Assessment and Monitoring." Retrieved (http://jecam.org/).

Johnson, David M., and Richard Mueller. 2021. "Pre- and within-Season Crop Type Classification Trained with Archival Land Cover Information." *Remote Sensing of Environment* 264(February):112576. doi: 10.1016/j.rse.2021.112576.

Kenduiywo, Benson Kipkemboi, Damian Bargiel, and Uwe Soergel. 2018. "Crop-Type Mapping from a Sequence of Sentinel 1 Images." *International Journal of Remote Sensing* 39(19):6383–6404. doi: 10.1080/01431161.2018.1460503.

Khan, Asim Hameed, Muhammad Moazam Fraz, and Muhammad Shahzad. 2021. "Deep Learning Based Land Cover and Crop Type Classification: A Comparative Study." *2021 International Conference on Digital Futures and Transformative Technologies, ICoDT2 2021* 3–8. doi: 10.1109/ICoDT252288.2021.9441483.

Khatami, Reza, Giorgos Mountrakis, and Stephen V Stehman. 2016. "Remote Sensing of Environment A Meta-Analysis of Remote Sensing Research on Supervised Pixel-Based Land-Cover Image Classi Fi Cation Processes : General Guidelines for Practitioners and Future Research." *Remote Sensing of Environment* 177:89–100. doi: 10.1016/j.rse.2016.02.028.

Kirch, Patrick V. 2005. "ARCHAEOLOGY AND GLOBAL CHANGE: The Holocene Record." *Annual Review of Environment and Resources* 30(1):409–40. doi: 10.1146/annurev.energy.29.102403.140700.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems* 25:1097–1105. doi: 10.1201/9781420010749.

Labate, Demetrio, Massimo Ceccherini, Andrea Cisbani, Vittorio De Cosmo, Claudio Galeazzi, Lorenzo Giunti, Mauro Melozzi, Stefano Pieraccini, and Moreno Stagi. 2009. "The PRISMA Payload Optomechanical Design, a High Performance Instrument for a New Hyperspectral Mission." *Acta Astronautica* 65(9–10):1429–36. doi: 10.1016/j.actaastro.2009.03.077.

Laborde, David, Will Martin, Johan Swinnen, and Rob Vos. 2020. "COVID-19 Risks to Global Food

Security." *Science* 369(6503):500–502. doi: 10.1126/science.abc4765.

Laso Bayas, Juan Carlos, Myroslava Lesiv, François Waldner, Anne Schucknecht, Martina Duerauer, Linda See, Steffen Fritz, Dilek Fraisl, Inian Moorthy, Ian McCallum, Christoph Perger, Olha Danylo, Pierre Defourny, Javier Gallego, Sven Gilliams, Ibrar Ul Hassan Akhtar, Swarup Jyoti Baishya, Mrinal Baruah, Khangsembou Bungnamei, Alfredo Campos, Trishna Changkakati, Anna Cipriani, Krishna Das, Keemee Das, Inamani Das, Kyle Frankel Davis, Purabi Hazarika, Brian Alan Johnson, Ziga Malek, Monia Elisa Molinari, Kripal Panging, Chandra Kant Pawe, Ana Pérez-Hoyos, Parag Kumar Sahariah, Dhrubajyoti Sahariah, Anup Saikia, Meghna Saikia, Peter Schlesinger, Elena Seidacaru, Kuleswar Singha, and John W. Wilson. 2017. "A Global Reference Database of Crowdsourced Cropland Data Collected Using the Geo-Wiki Platform." *Scientific Data* 4:1–10. doi: 10.1038/sdata.2017.136.

Lecun, Yann, and Yoshua Bengio. 1995. "Convolutional Networks for Images, Speech and Time-Series." in *The handbook of brain theory and neural networks*. Vol. 1.

Li, Qingting, Cuizhen Wang, Bing Zhang, and Linlin Lu. 2015. "Object-Based Crop Classification with Landsat-MODIS Enhanced Time-Series Data." *Remote Sensing* 7(12):16091–107. doi: 10.3390/rs71215820.

List, Martin, Maria Behrens, Wolfgang Reichardt, and Georg Simonis. 1995. *Internationale Politik. Probleme Und Grundbegriffe*.

Louis, Jérôme, Vincent Debaecker, Bringfried Pflug, Magdalena Main-Knorn, Jakub Bieniarz, Uwe Mueller-Wilm, Enrico Cadau, and Ferran Gascon. 2016. "Sentinel-2 SEN2COR: L2A Processor for Users." *European Space Agency, (Special Publication) ESA SP* SP-740(May):9–13.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant. 2000. "Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1 Km AVHRR Data." *International Journal of Remote Sensing* 21(6):1303–30. doi: 10.1080/014311600210191.

Ma, Lei, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. 2019. "Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 152(April):166–77. doi: 10.1016/j.isprsjprs.2019.04.015.

Massey, Richard, Temuulen T. Sankey, Russell G. Congalton, Kamini Yadav, Prasad S. Thenkabail, Mutlu Ozdogan, and Andrew J. Sánchez Meador. 2017. "MODIS Phenology-Derived, Multi-Year Distribution of Conterminous U.S. Crop Types." *Remote Sensing of Environment* 198:490–503. doi: 10.1016/j.rse.2017.06.033.

Matsunaga, Tsuneo, Akira Iwasaki, Satoshi Tsuchida, Jun Tanii, Osamu Kashimura, Hirokazu Yamamoto, and Shuichi Rokugawa. 2017. "CURRENT STATUS OF HYPERSPECTRAL IMAGER SUITE (HISUI) National Institute for Environmental Studies (NIES), Japan The University of Tokyo , Japan National Institute of Advanced Industrial Science and Technology (AIST), Japan Japan Resources Observatio." 3:443–46.

Mondal, Arun, Sananda Kundu, Surendra Kumar Chandniha, Rituraj Shukla, and P. K. Mishra. 2012. "Comparison of Support Vector Machine and Maximum Likelihood Classification Technique Using Satellite Imagery." *International Journal of Remote Sensing and GIS* 1(2):116–23.

Moreira, Alberto, Pau Prats-iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou. 2013. "SAR-Tutorial-March-2013." *IEEE Geoscience and Remote Sensing Magazine* 1(1):6–43.

Murray, Nicholas, North Korea, When Murray, and Yellow Sea. 2010. "ECOLOGY ' S REMOTE-

SENSING REVOLUTION." *Nature.*

NASA. 2019. "Landsat 9." *Fact Sheet* 2. Retrieved (http://pubs.er.usgs.gov/publication/fs20193008).

NASA. 2020. "What Is SAR?" Retrieved (https://earthdata.nasa.gov/learn/backgrounders/what-is-sar).

NASS. 2009. "CropScape - Cropland Data Layer." Retrieved (https://nassgeodata.gmu.edu/CropScape/).

Nitze, I., U. Schulthess, and H. Asche. 2012. "Comparison of Machine Learning Algorithms Random Forest, Artificial Neuronal Network and Support Vector Machine to Maximum Likelihood for Supervised Crop Type Classification." *Proceedings of the 4th Conference on GEographic Object-Based Image Analysis – GEOBIA 2012* (April 2015):35–40.

Parry, Ian. 2015. "Summary for Policymakers." *Implementing a US Carbon Tax: Challenges and Debates* xxiii–xxxiii. doi: 10.4324/9781315071961-11.

Peña, José M., Pedro A. Gutiérrez, César Hervás-Martínez, Johan Six, Richard E. Plant, and Francisca López-Granados. 2014. "Object-Based Image Classification of Summer Crops with Machine Learning Methods." *Remote Sensing* 6(6):5019–41. doi: 10.3390/rs6065019.

Pérez-Hoyos, Ana, Felix Rembold, Hervé Kerdiles, and Javier Gallego. 2017. "Comparison of Global Land Cover Datasets for Cropland Monitoring." *Remote Sensing* 9(11). doi: 10.3390/rs9111118.

Philip, G., D. Donoghue, A. Beck, and N. Galiatsatos. 2002. "CORONA Satellite Photography: An Archaeological Application from the Middle East." *Antiquity* 76(291):109–18. doi: 10.1017/S0003598X00089869.

Phiri, Darius, Matamayo Simwanda, Serajis Salekin, Vincent R. Ryirenda, Yuji Murayama, Manjula Ranagalage, Nadya Oktaviani, Hollanda A. Kusuma, Tianxiang Zhang, Jinya Su, Cunjia Liu, Wen Hua Chen, Hui Liu, Guohai Liu, M. Cavur, H. S. Duzgun, S. Kemec, D. C. Demirkan, Radhia Chairet, Yassine Ben Salem, Mohamed Aoun, Zolo Kiala, Onisimo Mutanga, John Odindi, and Kabir Peerbhay. 2019. "Remote Sensing Sentinel-2 Data for Land Cover / Use Mapping : A Review." *Remote Sensing* 42(3):14.

Probst, Philipp, Anne Laure Boulesteix, and Bernd Bischl. 2019. "Tunability: Importance of Hyperparameters of Machine Learning Algorithms." *Journal of Machine Learning Research* 20:1–22.

Prudente, Victor Hugo Rohden, Lucas Volochen Oldoni, Denis Corte Vieira, Carlos Eduardo Vizzotto Cattani, and Ieda Del'Arco Sanches. 2019. "Relationship between SAR/Sentinel-1 Polarimetric and Interferometric Data with Biophysical Parameters of Agricultural Crops." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42(3/W6):599–607. doi: 10.5194/isprs-archives-XLII-3-W6-599-2019.

Ramankutty, Navin, Amato T. Evan, Chad Monfreda, and Jonathan a. Foley. 2008. "Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000." *Global Biogeochemical Cycles* 22(1):n/a-n/a. doi: 10.1029/2007GB002952.

Ranghetti, Luigi, Mirco Boschetti, Francesco Nutini, and Lorenzo Busetto. 2020. "'Sen2r': An R Toolbox for Automatically Downloading and Preprocessing Sentinel-2 Satellite Data." *Computers and Geosciences* 139(March):104473. doi: 10.1016/j.cageo.2020.104473.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *Lecture Notes in Computer Science (Including Subseries*

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28.

Rosenblatt, F. 1958. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review* 65(6):386–408. doi: 10.1037/h0042519.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* 323(6088):533–36. doi: 10.1038/323533a0.

Rustowicz, Rose, Robin Cheong, Lijing Wang, Stefano Ermon, Marshall Burke, and David Lobell. 2019. "Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods." *CVPR Workshops* 1:75–82.

Sheykhmousa, Mohammadreza, Masoud Mahdianpari, Hamid Ghanbari, Fariba Mohammadimanesh, Pedram Ghamisi, and Saeid Homayouni. 2020. "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:6308–25. doi: 10.1109/JSTARS.2020.3026724.

Sisodia, Pushpendra Singh, Vivekanand Tiwari, and Anil Kumar. 2014. "Analysis of Supervised Maximum Likelihood Classification for Remote Sensing Image." *International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2014* 9–12. doi: 10.1109/ICRAIE.2014.6909319.

Soffiantini, Giulia. 2020. "Food Insecurity and Political Instability during the Arab Spring." *Global Food Security* 26(July):100400. doi: 10.1016/j.gfs.2020.100400.

Sumbul, Gencer, Marcela Charfuelan, Begum Demir, and Volker Markl. 2019. "Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding." 5901–4. doi: 10.1109/igarss.2019.8900532.

Thorkild, J. 1958. "Salt and Silt in Mesopotamia.Pdf." *Science*.

Tobler, Waldo Rudolph. 1970. "A COMPUTER MOVIE SIMULATING URBAN GROWTH IN THE DETROIT REGION." *Economic Geography* 46:234–40. doi: 10.1126/science.ns-13.332.462.

Toth, Charles, and Grzegorz Jóźków. 2016. "Remote Sensing Platforms and Sensors: A Survey." *ISPRS Journal of Photogrammetry and Remote Sensing* 115:22–36. doi: 10.1016/j.isprsjprs.2015.10.004.

Transon, Julie, Raphaël d'Andrimont, Alexandre Maugnard, and Pierre Defourny. 2018. "Survey of Hyperspectral Earth Observation Applications from Space in the Sentinel-2 Context." *Remote Sensing* 10(2):1–32. doi: 10.3390/rs10020157.

Tsai, Ya Lun S., Andreas Dietz, Natascha Oppelt, and Claudia Kuenzer. 2019. "Remote Sensing of Snow Cover Using Spaceborne SAR: A Review." *Remote Sensing* 11(12). doi: 10.3390/rs11121456.

Tsendbazar, N. E., S. de Bruin, and M. Herold. 2015. "Assessing Global Land Cover Reference Datasets for Different User Communities." *ISPRS Journal of Photogrammetry and Remote Sensing* 103:93–114. doi: 10.1016/j.isprsjprs.2014.02.008.

Tseng, Gabriel, Catherine Nakalembe, Ivan Zvonkov, and Hannah Kerner. 2021. "CropHarvest : A Global Satellite Dataset for Crop Type Classification." (NeurIPS).

Tsyganskaya, Viktoriya, Sandro Martinis, Philip Marzahn, and Ralf Ludwig. 2018. "SAR-Based Detection of Flooded Vegetation–a Review of Characteristics and Approaches." *International Journal of Remote Sensing* 39(8):2255–93. doi: 10.1080/01431161.2017.1420938.

Ulaby, Fawwaz T., Robert Y. Li, and K. S. Shanmugan. 1982. "Crop Classification Using Airborne

Radar and Landsat Data." *IEEE Transactions on Geoscience and Remote Sensing* 20(1):42–51. doi: 10.1109/TGRS.1982.4307519.

United Nations, Department of Economic and Social Affairs, Population Division. 2019. *World Population Prospects 2019, Online Edition. Rev. 1.*

Ünsalan, Cem, and Kim L. Boyer. 2011. "Multispectral Satellite Image Understanding." *Multispectral Satellite Image Understanding*. doi: 10.1007/978-0-85729-667-2.

USGS. 2013. "Earth Explorer." Retrieved (https://earthexplorer.usgs.gov/).

Vangi, Elia, Giovanni D'amico, Saverio Francini, Francesca Giannetti, Bruno Lasserre, Marco Marchetti, and Gherardo Chirici. 2021. "The New Hyperspectral Satellite Prisma: Imagery for Forest Types Discrimination." *Sensors (Switzerland)* 21(4):1–19. doi: 10.3390/s21041182.

Wenxue, Fu, Ma Jianwen, and Chen and Fang Chen Pei. 2020. "Remote Sensing Satellites for Digital Earth." Pp. 55–123 in *Manual of Digital Earth*.

Woodcock, CURTIS E., RICHARD Allen, MARTHA Anderson, ALAN Belward, ROBERT Bindschadler, WARREN Cohen, FENG Gao, SAMUEL N. Goward, DENNIS Helder, EILEEN HELMER, RAMA NEMANI, LAZAROS OREOPOULOS, JOHN SCHOTT, PRASAD S. THENKABAIL, ERIC F. VERMOTE, JAMES VOGELMANN, MICHAEL A. WULDER, and RANDOLPH WYNNE. 2008. "Free Access to Landsat Imagery." *Science* 320(May):1011–12.

Wuepper, David, Pasquale Borrelli, and Robert Finger. 2020. "Countries and the Global Rate of Soil Erosion." *Nature Sustainability* 3(1):51–55. doi: 10.1038/s41893-019-0438-4.

Wulder, Michael A., Thomas R. Loveland, David P. Roy, Christopher J. Crawford, Jeffrey G. Masek, Curtis E. Woodcock, Richard G. Allen, Martha C. Anderson, Alan S. Belward, Warren B. Cohen, John Dwyer, Angela Erb, Feng Gao, Patrick Griffiths, Dennis Helder, Txomin Hermosilla, James D. Hipple, Patrick Hostert, M. Joseph Hughes, Justin Huntington, David M. Johnson, Robert Kennedy, Ayse Kilic, Zhan Li, Leo Lymburner, Joel McCorkel, Nima Pahlevan, Theodore A. Scambos, Crystal Schaaf, John R. Schott, Yongwei Sheng, James Storey, Eric Vermote, James Vogelmann, Joanne C. White, Randolph H. Wynne, and Zhe Zhu. 2019. "Current Status of Landsat Program, Science, and Applications." *Remote Sensing of Environment* 225(November 2018):127–47. doi: 10.1016/j.rse.2019.02.015.

Xu, Guangyu, Caijun Xu, and Yangmao Wen. 2018. "Sentinel-1 Observation of the 2017 Sangsefid Earthquake, Northeastern Iran: Rupture of a Blind Reserve-Slip Fault near the Eastern Kopeh Dagh." *Tectonophysics* 731–732:131–38. doi: 10.1016/j.tecto.2018.03.009.

Young, Nicholas E., Ryan S. Anderson, Stephen M. Chignell, Anthony G. Vorster, Rick Lawrence, and Paul H. Evangelista. 2017. "A Survival Guide to Landsat Preprocessing." *Ecology* 98(4):920–32. doi: 10.1002/ecy.1730.

Yu, Le, Lu Liang, Jie Wang, Yuanyuan Zhao, Qu Cheng, Luanyun Hu, Shuang Liu, Liang Yu, Xiaoyi Wang, Peng Zhu, Xueyan Li, Yue Xu, Congcong Li, Wei Fu, Xuecao Li, Wenyu Li, Caixia Liu, Na Cong, Han Zhang, Fangdi Sun, Xinfang Bi, Qinchuan Xin, Dandan Li, Donghui Yan, Zhiliang Zhu, Michael F. Goodchild, and Peng Gong. 2014. "Meta-Discoveries from a Synthesis of Satellite-Based Land-Cover Mapping Research." *International Journal of Remote Sensing* 35(13):4573–88.

Zenzo, Silvano Di, Ralph Bernstein, Harwood G. Kolsky, and Stephen D. Degloria. 1987. "Gaussian Maximum Likelihood and Contextual Classification Algorithms for Multicrop Classification." *IEEE Transactions on Geoscience and Remote Sensing* GE-25(6):805–14. doi: 10.1109/TGRS.1987.289752.

Zhang, Wei, Ping Tang, and Lijun Zhao. 2021. "Fast and Accurate Land Cover Classification on

Medium Resolution Remote Sensing Images Using Segmentation Models." *International Journal of Remote Sensing* 42(9):3277–3301. doi: 10.1080/01431161.2020.1871094.

Mario Gilcher

gilcher.mario@gmail.com
Feyener Weg 10
54294 Trier
Germany, Rhineland-Palatinate

Education

2014 - present:

PhD, University of Trier

Thesis subject: Remote Sensing Based Crop Classification of Maize

2012 - 2014:

MSc Applied Geoinformatics, University of Trier

Thesis subject: Agent Based Modeling of Migration in Western Africa

2007 - 2011:

BSc Applied Human Geography, University of Trier

Thesis subject: Vegetable Gardens in Rural Areas

Experience

2014 - present

Teaching univariate and multivariate statistics at the University of Trier

2020 - 2022

Research associate, project PANTHEON

- Precision agriculture in hazelnut orchards near Rome, Italy
- UAV based orthomosaics (RGB, multispectral, thermal)
- UGV based LiDAR and image analysis
- Fruit detection and Modeling

2018 - 2021

Project BIOVIM

- UAV data acquisition at the Institut viti-vinicole in Remich

2018 - 2019

Research associate, project SASSCAL

- Environmental monitoring in southern africa
- Large scale interactive visualization of the enhanced vegetation index

2014 - 2018

Research associate, project Dryland

- scientific preparation of the EnMAP mission
- UAV based crop status analysis of maize/wheat fields near Trier

2011 - 2014

Consultant at Taurus Eco Consulting

- Reporting and supporting regional development projects in Rhineland Palatinate
- Moderating stakeholder events in Mainz and Berlin

2008 - 2010

Students assistant in the department of spatial development, University of Trier