Masterarbeit

Zur Erlangung des akademischen Grades: Master of Science (M.Sc.)

Thema der Arbeit

Simulating maize yields in Brandenburg with the Scalable Crop Yield Mapper

eingereicht von:

Gutachter/innen: Dr. Daniel Müller
Prof. Dr. Patrick Hostert

Eingereicht am Geographischen Institut der Humboldt-Universität zu Berlin am:

05.09.2019

Contents

List of figures

List of tables

List of maps

List of abbreviations

AfSBB	*Amt für Statistik Berlin Brandenburg (Office of Statistics Berlin Brandenburg)*
BGR	*Bundesanstalt für Geowissenschaften und Rohstoffe (Federal Institute for Geosciences and Natural Resources)*
BKG	*Bundesamt für Kartographie and Geodäsie (Federal Agency for Cartography and Geodesy)*
BOA	*Bottom of atmosphere*
CAP	*Common Agricultural Policy*
CI	*Chlorophyll Index*
CVI	*Combined Vegetation Index*
DOY	*Day of year (Julian Day)*
DWD	*Deutsche Wetterdienst (German weather service)*
EC	*European Commission*
ESA	*European Space Agency*
EU	*European Union*
FAO	*Food and Agriculture Organization of the United Nations*
GEE	*Google Earth Engine*
IACS	*Integrated Administration and Control System*
InVeKoS	*Integrierten Verwaltungs- und Kontrollsystem (Integrated Administration and Control System)*
LAI	*Leaf Area Index*
LGB	*Landesvermessung und Geobasisinformation Brandenburg (Land Surveying and Geo Information Brandenburg)*
MERIS	*Medium-spectral resolution, imaging spectrometer*
MODIS	*Moderate Resolution Imaging Spectroradiometer*
MONICA	*Model for Nitrogen and Carbon in Agro-ecosystems*
NDVI	*Normalized difference vegetation index*
NIR	*Near infrared*
NUTS	*Nomenclature of territorial units for statistics*
SCYM	*Scalable Crop Yield Mapper*
TOA	*Top of atmosphere*
VI	*Vegetation index/indices*
ZALF	*Leibniz-Zentrum für Agrarlandschaftsforschung e. V. (Leibniz Centre for Agricultural Landscape Research)*

Abstract

Knowledge of field scale crop yields allows us to understand small scale changes to agricultural intensification and how different management practices and environmental conditions affect production. Calculating yields with remote sensing offers the opportunity to analyse yields without the requirement for farmers to provide data. Here we use the Scalable Crop Yield Mapper (SCYM) to simulate 2017 silage maize yield at pixel and field scale in Brandenburg, Germany. We incorporated crop management, climate and soil variables to produce daily leaf area index and yield outputs from crop model simulations using the MONICA crop model and combined these with seasonal weather variables to train daily multiple regression models. We then simulated yields within known silage maize fields within the Google Earth Engine platform by estimating the leaf area index and extracting the capture date of each pixel image within a Sentinel-2 growing season composite and combining this with the associated multiple regression equation for the day the image was captured. We used district scale data from the Office of Statistics to validate results and found that SCYM explained 8% of the variability of silage maize yields in Brandenburg in 2017 with a weak non-significant positive correlation between the two datasets ($r = 0.28$, $p = 0.32$). We found that SCYM overestimated yield in 12 out of 14 districts by a factor of up to 1.29. We conclude that whilst the basic elements required for SCYM estimated yields are available for Brandenburg and that estimating yields is possible, the accuracy of the validation data available for the region is questionable and is a hindrance to assessing SCYM's performance in the region. Further research utilising highly accurate field scale validation data is recommended. Nevertheless, the results have helped us to understand the feasibility of estimating yield with SCYM in Brandenburg and offer a good grounding for further research on the subject.

1. Introduction

The World's population is predicted to reach 10 billion by 2050 leading to a global demand for food crops 50% higher than in 2013. Simultaneously, the bioenergy market continues to grow globally with demand expected to be 50% higher than 2013 by 2020. These two factors have resulted in an escalation of global demand for agriculture crops leading to increased competition for land, which is in turn driving deforestation and increasing agricultural contributions to greenhouse gasses (Food and Agricultural Organisation of the United Nations (FAO), 2017). If we are to continue to meet this demand in the future without expanding agricultural land, advancements in agricultural technology, management practices and distribution are required.

Despite this need, our knowledge on the distribution and temporal changes to agricultural intensity are limited (Kuemmerle et al., 2013). Agricultural intensification is defined as increasing production whilst maintaining inputs, or by maintaining production whilst decreasing inputs (FAO n.d. a). The required increases in agricultural intensification comes with a caveat in that intensification has both advantages and disadvantages, particularly with regard to environmental concerns (FAO, 2017). These include an increase in greenhouse gas emissions, eutrophication caused by excessive nitrates and phosphates in fertiliser, loss of biodiversity through the use of chemical substances used to remove unwanted weeds and pests, and alterations to the water cycle through irrigation (Foley, 2005; Tilman et al., 2001). The gains from intensification can also be short-lived, with increased soil erosion and salinisation becoming an issue in some regions (Foley, 2005). Recently however, research has begun to focus on sustainable or ecological intensification, which intends to decrease the yield gap by employing management practices which improve yield without subsequent environmental degradation. These techniques include increasing soil organic matter and maintaining biodiversity by managing landscapes to ensure sufficient habitats are available (Bommarco et al., 2013; Foley, 2005). However it remains to be seen whether these practices will be universally adopted. As such, any increase in agricultural production, following the adoption of new or improved management practices, must be weighed against any resulting increases in environmental degradation.

The negative consequences of agricultural intensification can be mitigated with the use of precision agriculture. In precision agriculture, farmers are able to target areas of a field where a certain management practice has been identified as beneficial. In essence, a field can be divided into an unlimited number of smaller plots, each one with its own specific management inputs dependent on its location, geomorphology, soil type and nutrient availability. Targeting smaller areas in this way leads not only to increased production but also to an increase in soil quality, a reduction in input costs and waste, and it is less damaging to the environment (Mulla, 2012). It is expected that in the near future precision agriculture will be commonplace, with autonomous instruments able to manage and maintain crops on an individual plant basis (Mulla, 2012).

If precision agriculture is to be used efficiently, it is paramount that the investment is directed at the areas where the potential increase will be most effective. It is therefore essential to establish where production gaps exist. Crop yield is defined as crop harvested (mass) divided by cropland (area) (Azzari et al., 2017; Carletto et al., 2015). Yield can be divided into three categories: 1) theoretical yield is the yield that could be achieved for a particular crop cultivar given perfect growing conditions; 2) attainable yield is the yield that could be achieved for a particular crop cultivar given the available solar radiation and temperature at the specific location; 3) actual yield is the measured yield produced. The difference between the latter two is the attainable yield gap, which is dependent on the availability of water and nutrients as well as losses due to weed and pest control (Bommarco et al., 2013; Lobell et al., 2009). Knowledge of yield gaps is one of the essential system metrics used for mapping agricultural intensity (Kuemmerle et al., 2013). Moreover, it has been suggested that by reaching 100% of attainable yield globally, production increases of between 40 – 70% can be achieved (Mueller et al., 2012). However, despite increases in agricultural investments, yield growth has slowed in recent years and in many places has stagnated (FAO, 2017; Ray et al., 2012).

Knowledge of yields provides lots of uses across spatial scales including how management, climate and soil factors affect crop growth (Lobell etal., 2015). For administrative purposes, yields are rarely reported at field level and are usually reported for administrative units, often covering a large geographical area and encompassing numerous fields (Lobell, 2013; Lobell et al., 2009). Whilst researching yields on a regional scale can be advantageous, blanket adoption of management practices can result in a waste of resources,

as areas which have a low potential of increasing yield are also targeted. Thus, it is essential to establish exactly where these resources are required at the finest possible scale. Knowledge of yields at field scale allows a detailed understanding of how crop growth responds directly to different management practices and environmental factors and can help with insurance or land rental decisions (Lobell, 2013; Lobell et al., 2015). Thus identifying yield gaps at field level or smaller is essential if we are to efficiently increase production without the expansion of agricultural land, wasted resources or further environmental degradation. However there are few options for measuring yields at a fine scale, over a large geographical area and over time.

It has long been established that estimating yields is possible through remote sensing. However, most research has aimed at calculating yield at regional scale (Becker-Reshef et al., 2010; Lobell et al., 2015; Mulla, 2012; Rembold et al., 2013) although remote sensing can provide data on crops efficiently and at a variety of temporal and spatial scales (Atzberger, 2013; Rembold et al., 2013). Different approaches to calculating yield at regional scale using remote sensing techniques, include using linear regression models to relate crop yield to vegetation indices (VI) (See section 2.1.5.2.2 for an explanation of VI) and adding meteorological data into multiple regressions (Lobell, 2013; Rembold et al., 2013; Shanahan et al., 2001). Other approaches include integrating crop models (Clevers, 1997; Lobell et al., 2003; Moulin et al., 1998; Sehgal et al., 2005; Sibley et al., 2014) but these can be resource intensive and site and date specific. Methods have also been established that capture over 50% of the field scale yield variation using remote sensing techniques (Clevers, 1997; Lobell et al., 2005; Moulin et al., 1998). Estimating field scale yields remains difficult due to cloud cover, atmospheric interference and geolocation issues (Mulla, 2012). In addition it remains uncommon as a result of the relatively high cost, lack of availability of appropriately scaled satellite data, a lack of validation data and a lack of requirement by administrations (Lobell et al., 2015). This requirement is now changing as administrations look to improve production and due to the reduction in costs of obtaining more detailed data. The costs have declined as a result of the open availability of satellite data from a number of sources (Lobell et al., 2015). This has led to further research taking advantage of these changes, in order to establish efficient and accurate methods for estimating yield at field level using remote sensing data (Azzari et al., 2017; Jin et al., 2017; Jin et al., 2019; Lobell & Azzari, 2017; Lobell et al., 2015).

The Scalable Crop Yield Mapper (SCYM) is one such method. It was developed by Lobell et al. (2015). The aim was to produce a method for estimating yield from remotely sensed images which can be adapted to any location, at any scale and with little computational resources. Lobell and colleagues updated the crop model and empirical model technique to enable it to be used on multiple dates and included season specific weather data within the regression model. Thus SCYM is neither spatially nor temporarily specific. The method can be implemented in Google Earth Engine (GEE) online meaning that no software is required, significantly reducing cost and resource requirements. The speed and efficiency with which yields can be estimated provides the opportunity for detailed analysis of management practices at site specific locations (Lobell et al., 2015). SCYM has been tested in a number of locations, including the Midwestern USA, India and Zambia and has proven to function across varying agricultural landscapes (Azzari et al., 2017; Burke & Lobell, 2017; Jin et al., 2017, 2019; Lobell & Azzari, 2017; Lobell et al., 2015). Furthermore, SCYM has been found to produce results as accurate as field based measures and as such has the potential to eliminate the need for validation data (Burke & Lobell, 2017). Thus SCYM is a good option for estimating yields at a large scale and over large geographical areas.

Utilising GEE to implement SCYM provides a number of advantages. GEE was developed by Gorelick et al. (2017) and provides a cloud based platform for acquiring, analysing and displaying satellite imagery. The computational power required to process large geospatial data is distributed across a network of computers, thus reducing the time required to process large datasets locally. GEE includes an extensive catalogue of geospatial datasets including the Landsat archive and Sentinel images as well as aerial derived images and environmental datasets. As the use of remote sensing continues to expand, it is anticipated that higher resolution images from alternative providers will become publicly available, increasing the potential of geospatial analysis within GEE and further improving yield estimation. The user interface utilises the Javascript programming language to access, analyse and display images. GEE is free to access and whilst prior knowledge of remote sensing is advantageous, it is not essential. GEE has been utilised in research across a number of disciplines including calculating growing stock volume in forest management (Sánchez-ruiz, et al., 2019), mapping wetlands (Wu et al., 2019), detecting vegetation changes (Xie, Z et al., 2019) and mapping irrigation (Xie, Y et al., 2019). As SCYM is designed using freely available

data and software, it is feasible to imagine farmers carrying out their own yield estimates using SCYM and adapting their management practices appropriately.

Whilst SCYM has been tested in various locations, it has not to our knowledge been tested in Europe. The adaptation of SCYM for use in Europe offers great potential particularly across the 28 member states of the European Union (EU-28). The EU has in place a common agricultural policy (CAP) through which member states report their annual yields of different crops. This information is used by the EU in relation to trade and development. The EU also runs a crop yield forecasting system (The MARS Crop Yield Forecasting System) which utilises satellite observations to predict future crop risks. Historical data that SCYM requires for crop modelling simulations and multiple regression training is likely to be available for all EU countries. In addition, crop masks can be easily created for each year and location thanks to the Integrated Administration and Control System (IACS), which provides georeferenced data on the crops grown in agricultural fields across the EU (EC, n.d., a). However the reporting procedure varies between regions (Amt für Statistik Berlin-Brandenburg (Office of Statistics Berlin Brandenburg) (AfSBB), 2017).

In 2016 almost 40% of the total land in the EU was classed as agricultural land, incorporating 173 m ha (Eurostat, 2018). With a total crop agricultural output of €205.6 billion in 2018, it is an important agricultural region for the World's food supply (European Commission (EC), 2019a). Over 60% of farms are less than 5 ha and 96% of farms are classed as family farms. These are expected to play an important role in providing additional income and food to rural areas in the future (Eurostat, 2018). The number of farms greater than 100 ha and with an output over €500,000 is increasing (EC, 2019a). Despite this the area of agricultural land remains stable, indicating a switch to a greater proportion of large scale farming operations in recent years (Eurostat, 2018). These changes are likely to bring in new investments and management practices. With this in mind it is important to understand how these changes may affect the distribution of yield at field level in order to manage resources and restrict negative environmental consequences.

Germany is the fourth largest state in the EU and contributes around 12% to the total agricultural output of the EU (EC, 2019a) with agricultural land covering about 16.7 m ha in 2016 (Eurostat, 2018). It is one of four EU states where growth in the number of farms is restricted to those farms greater than 100 ha (Eurostat, 2018). Between 2010 and 2016 the number of farms greater than 100 ha increased from 11.2% to 13.3% (EC, 2019b). Germany

has a history of bioenergy production and experienced a biogas boom after 2000 and especially following the incentivisation of biogas crops in 2003 (Grundmann et al., 2012; Pfeiffer & Thra, 2018). Between 2000 and 2010 the cultivated area for biogas substrate production, including silage maize, increased from 20,000 to 650,000 hectares (Pfeiffer & Thra, 2018).

Brandenburg is the 5th largest state in Germany with an area of approximately 2.95m ha (Map 1) a large proportion of which is used for arable farming (See Table 1 for a selection of facts and figures on arable farming in Brandenburg referred to in this paragraph). The number of large farms in Brandenburg is considerably higher than the German average and has been continually rising in recent years (EC, n.d., b; Landesvermessung und Geobasisinformation Brandenburg (Land Surveying and Geo Information Brandenburg) (LGB), 2017) coinciding with an overall reduction in the total number of farms (LGB, 2017). Around 75% of farms are owned by corporate bodies and partnerships (Grundmann et al., 2012; Gutzler et al., 2015). The difference between the mean size of a silage maize field and the median size of a silage maize field within the state (Table 1) suggests that a small number of very large maize fields are present within the state (LGB, 2017). The planting of silage maize crops has also been steadily increasing and silage maize recently overtook Rye to become the most planted crop by area in 2016. A total of 154715 ha was used for silage-maize in 2017 corresponding to 12% of all arable farming land in Brandenburg (LGB, 2017). Thus it is an extremely important crop for the state (LGB, 2017). Furthermore, under the Rural Development Programme for Berlin and Brandenburg (EC, n.d., b), farmers in Brandenburg are being incentivised to adopt climate and environmentally friendly management practices in order to improve biodiversity, water and soil management.

Despite these changes to the agricultural landscape in Brandenburg, very little data exists on agricultural intensity. For example, the use of fertilisers and pesticides is not officially recorded and data on yields is limited. Brandenburg is a large state with rapidly intensifying agriculture and thus is a good match for SCYM's ability to estimate yields at a large scale and over large geographical areas. Furthermore, to the best of our knowledge, the existing research on SCYM has been provided by the research group involved in the development of the method. It has not as yet been tested by researchers outside of this group. If SCYM is to be accepted amongst the research community as a durable, efficient and accurate method for estimating crop yields at any location across the World, it is necessary for other

researchers to test SCYM and report on its adaptability. As such, the aim of this research is to provide a case study on the ability of SCYM to simulate silage maize yields in the state of Brandenburg, Germany in 2017.

Map 1: Location of the state of Brandenburg within Germany

Table 1: A selection of facts and figures on arable farming in Brandenburg.

Fact	Statistic	Source
Land used for arable farming	40% (1.3m ha in 2017)	LGB, 2017
No. Farms in 2017	5611	LGB, 2017
No. Farms in 2016	5641	LGB, 2017
No. Farms in 2015	5647	LGB, 2017
Farms larger than 100 ha in 2017	40%	EC, n.d., b; LGB, 2017
Farms larger than 1000 ha in 2017	6%	LGB, 2017
Mean size of a silage maize field in 2017	13.47 ha	LGB, 2017
Median size of a silage maize field in 2017	8.39 ha	LGB, 2017
No. silage maize fields in 2017	11487	LGB, 2017
Silage maize fields < 10 ha in 2017	56%	LGB, 2017
Silage maize fields < 50 ha in 2017	97%	LGB, 2017
Silage maize field > 100 ha in 2017	0.2% (311 fields)	LGB, 2017

2. Methods

2.1 SCYM implementation

2.1.1 SCYM workflow

The research was conducted following the SCYM method developed by Lobell et al. (2015). The method is divided into four distinct steps (Table 2) and was initially developed for maize and soybean yields in the Midwestern USA, using satellite images from Landsat 5 and 7 (Lobell & Azzari, 2017; Lobell et al., 2015). It has also been tested on wheat in India and maize in Zambia using images obtained from a variety of Landsat sensors as well as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Azzari et al., 2017). Figure 1 provides an overview of the SCYM workflow we adopted for this study which is described in detail in the following sub-sections.

Table 2: The four step process of SCYM (Source: Lobell et al. (2015))

Step	Name	Description	Requirements
1	Crop model simulations	Simulate daily crop growth and yields for multiple sites, years, and management (sow dates, cultivars, plant density, etc.)	Well-tested crop model (e.g., APSIM, Hybrid-Maize, DSSAT) Daily weather and soil data for some sites
2	Pseudo-observations	Translate daily crop model output to daily observations for relevant sensor(s)	Literature-based equation to relate crop model outputs and remotely sensed variables (e.g., LAI vs. WDVI, LAI vs. c-band backscatter, water supply:demand ratio vs. evaporative stress index)
3	Regression calibration	Train regression to predict yields from pseudo-observations (e.g. VI) and weather	N/A
4	Yield estimation	Per-pixel yield predictions by applying regression to satellite observations and gridded weather	Satellite images, gridded monthly weather data, and crop type maps

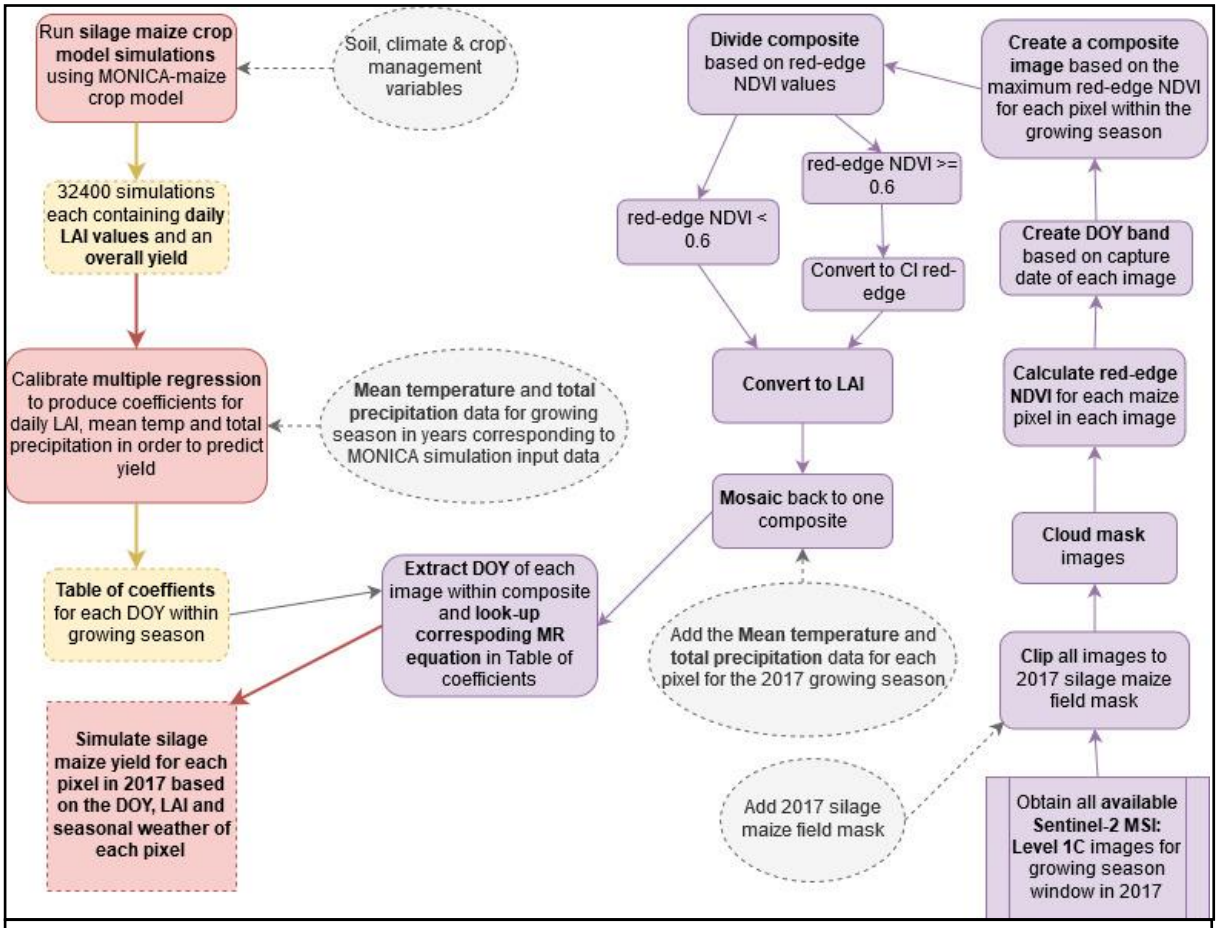


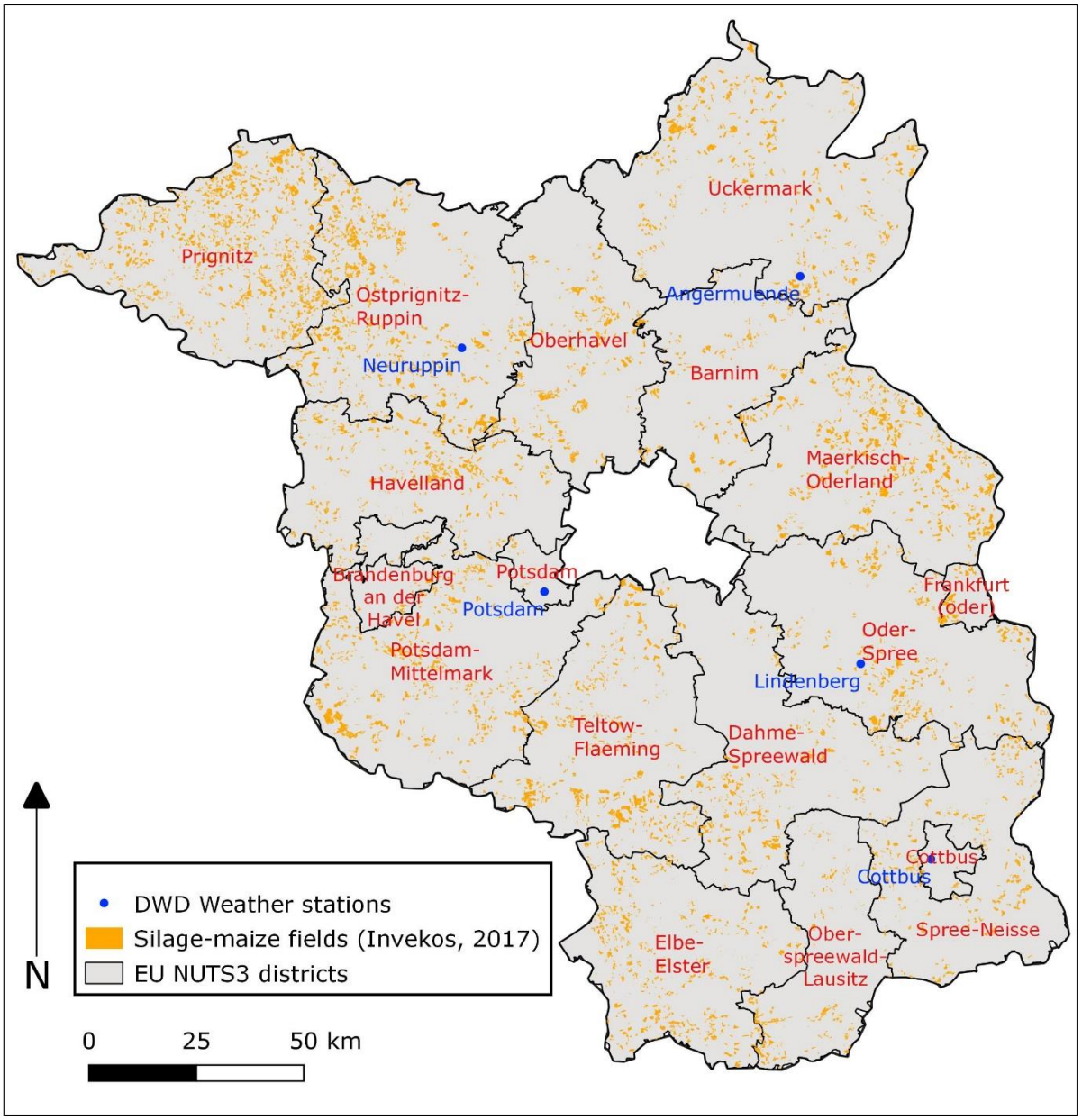
Figure 1: SCYM workflow. pink = the major steps within the process., purple = steps carried out in GEE, grey = external inputs, yellow = outputs

2.1.2 Step 1: Crop model simulations

SCYM requires a number of crop model simulations from a well-tested crop model which can produce daily outputs from a variety of input data (Lobell et al., 2015). Crop simulations are used in crop research and management in order to help improve our understanding of the factors directly influencing crop growth. They integrate crop, soil, and climate data in order to predict crop growth depending on input parameters (van Keulen, 2013). We simulated daily Leaf Area Index (LAI) values and annual yields for silage maize growth in Brandenburg using the Model for Nitrogen and Carbon in Agro-ecosystems (MONICA) (Nendel et al., 2014). LAI is a dimensionless quantity used to describe the extent of canopy foliage of plants. It is defined by the area of one half of leaf tissue within a given area at ground level and is a measurement often used in plant research (Bréda, 2008). LAI can be measured indirectly based on an empirical relationship with radio-metric measurements, such as VI (Fang and Liang, 2008).

MONICA was developed by the Leibniz-Zentrum für Agrarlandschaftsforschung e. V. (Leibniz Centre for Agricultural Landscape Research) (ZALF) to predict how climate change affects crops at the plot scale in smaller Central European regions. It was originally designed to simulate crop growth as well as water and nitrogen uptake, and soil matter dynamics (Nendel, 2010; Nendel, 2014). It has been calibrated on a number of crops typically found in central Europe, including maize (Nendel, 2014). We considered MONICA a suitable crop model for this study, as it is similar to the APSIM model used by Lobell et al. (2015) but was developed in Brandenburg and has been calibrated to silage maize under growing conditions found within the Brandenburg area. It is nevertheless suitable for simulating crop growth across a range of different climates and soils (For further information on MONICA and related research, see Appendix 1).The input data utilised for the simulations can be calibrated to represent the growing conditions found within the study area (Table 3). We followed closely those inputs utilised by Lobell et al. (2015) and added soil type which was considered important as the soil in Brandenburg is generally poor with a large proportion of sandy soils with little water holding capacity (Grundmann et al., 2012; Gutzler et al., 2015). All combinations of the input parameters were used for the SCYM model for this study, including all years and weather stations, resulting in 32400 simulations (Table 3). These were then used to generate variability in the multiple regression model.

In order to establish the values for each parameter, we analysed a number of datasets acquired from the Deutscher Wetterdienst (German weather service) (DWD) (climate (weather stations), years, sowing dates, initial soils moisture) (DWD, 2019a - i) and from the Bundesanstalt für Geowissenschaften und Rohstoffe (Federal Institute for Geosciences and Natural Resources) (BGR) (soil type) (BGR, 2019a). Each of the DWD datasets were acquired for five DWD weather stations within Brandenburg, which we selected due to the distribution of their geographical location and temporal availability of the data (Map 2 and Table 4). The climate in Brandenburg is described as humid continental (Dfb) in the Köppen climate classification with a mean annual temperature of 9.3 °C and total annual precipitation of around 580 mm (MLUL, 2016). The climate data summary statistics are outlined in Table 4. For fertiliser rates and irrigation, we reviewed reports and existing research to establish values (For detailed information of the methods used to establish parameter values, see Appendix 2).



Map 2: NUTS3 level districts, the 5 DWD weather stations selected for this study and silage maize fields in 2017 (LGB, 2017).

Table 3: Summary of MONICA input parameters used to calibrate SCYM for this study (top row) and those used to calibrate SCYM by Lobell et al (2015) (Maize only). All combinations of the input parameters were used for the SCYM model for this study, including all years and weather stations, resulting in 32400 simulations. These were then used to generate variability in the multiple regression model (adapted from Lobell et al., 2015)

Region	Crop	# weather stations	Years	Sowing dates	Sowing densities	Fertiliser rates (kg N ha^{-1})	Initial soil moistures (% capacity)	Cultivar (s)	Irrigation	Soil types
Brandenburg	Silage maize	5	1992-2018	15-Apr 1-May 15-May	N/A	150, 170	58, 85	ZALF	N/A	20 types, See Appendix 5
Central U.S. (Lobell et al., 2015)	Maize	3	1980-2015	24-Apr 1-May 15-May	4.5, 7.75, 9.0	200, 300	60, 100	Pioneer_3394, Pioneer_long, Pioneer_short	N/A	N/A
Southern Zambia (Lobell et al., 2015)	Maize	1	1998-2012	20-Nov 1-Dec 15-Dec	3.0, 3.5, 4.0, 4.5, 5.0	0, 25, 50	60, 80, 100	sc501	N/A	N/A

Table 4: Annual climate data for the 5 weather stations for the period 1992-2018

Station name	**Mean annual air temperature (°C)**	**Mean annual wind speed ($m\ s^{-1}$)**	**Mean total annual sunshine hours (h)**	**Mean total annual precipitation (mm)**	**Mean annual relative humidity (%)**
Angermuende	16.32	3.83	1728.42	225.77	79.80
Cottbus	17.06	2.89	1758.58	241.89	75.86
Lindenberg	16.81	3.49	1767.06	238.87	77.80
Neuruppin	16.47	2.98	1759.75	236.61	79.62
Potsdam	16.72	4.27	1787.96	243.90	78.94
All station mean	**16.68**	**3.49**	**1760.35**	**237.41**	**78.40**

2.1.3 Step 2: Pseudo-observations

SCYM allows the conversion of daily crop model outputs (LAI) to pseudo observations for the relevant sensors (VI). However it is also possible to convert VI to LAI (Jin et al., 2017). In this case we converted satellite observations into red-edge normalized difference vegetation index (NDVI) and subsequently into LAI. A full description of this method is outlined in section 2.1.5.2.2.

2.1.4 Step 3: Regression calibration

One advantage of SCYM is that it estimates yields using both crop model outputs and seasonal weather data. Weather data is included to incorporate climatic changes which occur after an observation image is captured and which may affect crop growth. For example, an image captured early in the growing season with a relatively high LAI may not necessarily result in a high yield if drought occurs following image capture. Including the seasonal weather data allows for a change in the weather after a particular observation date (Lobell et al., 2015). We calculated mean temperature and total precipitation for each weather station and year for the mid-late growing season (Day of Year (DOY) 166 to 288) from the DWD weather data to include as seasonal weather variables in the regression calibration. These weather variables were chosen as they are expected to influence crop growth. We then added this data to the corresponding years and weather stations of the MONICA crop model simulation outputs. We converted this data into a table with each row representing one daily (DOY) LAI output with additional columns for seasonal mean temperature (Tp_av) and seasonal total precipitation (Psum). To reduce computation requirements, we kept only data with a DOY corresponding to the DOY of observed images for use in the multiple regression analysis (See Step 4) (A sample of the data structure is included as Appendix 7). Duplicated rows and rows with an LAI of 0 were removed from the data. A number of rows were found to differ only in yield, with all other values remaining the same. These rows were aggregated to a mean yield in R (Appendix 6).

We performed multiple regression analysis in R using the lm() function, after first checking for collinearity between variables (Appendix 6) (R Core Team, 2018). The SCYM generic multiple regression equation is (Lobell et al., 2015):

$$\boldsymbol{Yield} = \boldsymbol{\beta_{0,d}} + \boldsymbol{\beta_{1,d}} * \boldsymbol{W} + \boldsymbol{\beta_{2,d}} * \boldsymbol{RM_d} + \boldsymbol{\beta_{3,d}} * \boldsymbol{W} * \boldsymbol{RM_d} \quad (1)$$

where W is a vector of seasonal weather attributes, RM is a vector of remote sensing based measures on specific dates (d), and all coefficients (β) are specific to the particular dates. For this study, the multiple regression equation is therefore:

$$\boldsymbol{Yield} = \boldsymbol{\beta_{0,d}} + \boldsymbol{\beta_{1,d}} * \boldsymbol{Tp_{av}} + \boldsymbol{\beta_{2,d}} * \boldsymbol{Psum} + \boldsymbol{\beta_{3,d}} * \boldsymbol{LAI_d} + \boldsymbol{\beta_{4,d}} * \boldsymbol{Tp\_av} * \boldsymbol{LAI_d} + \boldsymbol{\beta_{5,d}} * \boldsymbol{Psum} * \boldsymbol{LAI_d} \quad (2)$$

where W is replaced by the seasonal weather variables (mean temperature and total precipitation) and LAI is the remote sensing based measure. This model includes interaction terms between mean temperature and LAI, and total precipitation and LAI. The model was therefore tested against simpler versions to establish how well the simpler models performed (Appendix 8). We found that including the interaction terms produced a better model. The model is able to explain 53% of the variability in silage-maize yields. A table of coefficients was created with each line representing each DOY found within the input dataset (Appendix 9).

2.1.5 Step 4: Yield estimation

2.1.5.1 Silage maize field mask

It is a requirement of SCYM that fields containing the relevant crop are known in advance so they can be masked in order to calculate yield at the field level. We masked silage maize fields in Brandenburg in 2017 using data from the Integrierten Verwaltungs- und Kontrollsystem (InVeKoS) (Integrated Administration and Control System (IACS)) (LGB, 2017). The IACS was introduced by the EC in 1992 to improve efficiency in distributing subsidies to farmers under the CAP. Under the system, farmers are obliged to annually report the crops planted on each parcel of agricultural land under their control. Since 2005, the reporting has been carried out online using GIS software. Each parcel of agricultural land is therefore documented, including size, geometry and crops. Each member state is responsible for implementing their own system in order to ensure accurate reporting (FAO, n.d. b; EC, n.d., b). In Germany, the reporting system is known as InVeKoS (Bundesamts für Justiz, 2015). Each state within Germany is responsible for administering its own system. In Brandenburg, this is carried out through the Ministerium für Ländliche Entwicklung, Umwelt und Landwirtschaft

(Ministry for Rural Development, Environment and Agriculture) and is available online through the LGB geoportal (LGB, 2019). Within the InVeKoS dataset for 2017, four types of maize are listed. Two of these, 'Mais (Biogas)' (Maize for Biogas) and 'Silomais (als Hauptfutter)' (Silage maize as animal feed) relate to silage maize. We extracted fields containing these two crops from the InVeKoS dataset to create a silage maize field mask consisting only of fields where silage maize was grown in Brandenburg in 2017. Any fields located within the Berlin border were removed. The summary statistics from the resulting data can be found in table 5.

Table 5: Summary statistics for silage maize fields in Brandenburg in 2017 (LGB, 2017)

No. of fields	Total area (ha)	Min area (ha)	Max area (ha)	Mean area (ha)	Median area (ha)	Standard deviation	1st quartile area (ha)	3rd quartile area (ha)
11488	154734.2628	0.0005	228.7700	13.4692	8.3935	15.3062	3.5665	17.8330

2.1.5.2 Google Earth Engine implementation (See Appendix 10 for GEE code)

2.1.5.2.1 Importing data into Google Earth Engine

The SCYM method employs the GEE platform in order to obtain satellite images, to calculate LAI on a per pixel basis and to estimate yields. To achieve this we imported the table of coefficients resulting from the multiple regression analysis in GEE through a google fusion table (Appendix 9). We imported the silage maize field mask into GEE as a GEE table.

SCYM requires a selection of images with a high temporal resolution in order to obtain a sufficient number of LAI values across the growing season. The Sentinel-2 mission's twin satellite constellation provides a combined temporal resolution of 5 days and a spatial resolution of 10m, 20m and 60m depending on the spectral band (European Space Agency (ESA), 2015). We obtained all available 'Sentinel-2 MSI: MultiSpectral Instrument, Level-1C' images covering the study area for the growing season period from DOY 135 to 288 for 2017 and clipped this image collection to the silage maize field mask. The image collection was then mapped to a cloud mask to create a cloud free collection (Google Developers, 2019a). Level-1C imagery includes radiometric and geometric corrections with top of atmosphere (TOA) reflectance (ESA, 2015).

2.1.5.2.2 Calculating vegetation indices and LAI

To convert per pixel VI to LAI, one first has to calculate an appropriate VI from the image spectral bands. VI describe the vegetation present at a given location by maximising the sensitivity to vegetation and minimising other factors such as soil or optical effects. A multitude of VI have been used in research. VI take the form of ratios and can be calculated from the spectral bands of remotely sensed optical images, usually incorporating red and near infrared (NIR) reflectance (Fang and Liang, 2008).

Lobell et al. (2015) converted LAI to Green Chlorophyll Vegetation Index using a regression equation developed by Nguy-robertson et al. (2012). Nguy-Robertson et al. (2012) tested twelve VI for estimating LAI in maize and calculated regression equations for estimating LAI from VI. For high resolution images containing red-edge and NIR spectral bands, they suggest using a Combined Vegetation Index (CVI) in the form of $CVI\{red-edge\ NDVI, CIred-edge\}$. Using this CVI offers maximum sensitivity to LAI across the range of variability.

In order to calculate red-edge NDVI, Nguy-Robertson et al. (2012) use the equation $(NIR - red\ edge)\ /\ (NIR + red\ edge)$ from Gitelson and Merzlyak (1994) and suggest that this equates to $(MERIS\ 12 - MERIS\ 9)\ /\ (MERIS\ 12 + MERIS\ 9)$ or $(778.75nm - 708.75)\ /\ (778.75nm + 708.75)$ when adapted to the medium-spectral resolution imaging spectrometer (MERIS) satellite (ESA, 2012). Sentinel-2 data contains one NIR band (band 8) and three vegetation red edge bands (bands 5, 6 and 7). However band 8 has central wavelengths of 832.8nm (Sentinel-2A) and 832.9nm (Sentinel-2B) placing it spectrally distant from the MERIS NIR band suggested by Nguy-Robertson et al. (2012). The nearest equivalent spectral bands in Sentinel-2 images are band 5 (Red Edge 1 = 703.9nm (2A) and 703.8nm (2B)) and band 7 (Red Edge 3 = 782.5nm (2A) and 779.7nm (2B)) (ESA, 2015). Onyia & Balzter (2018) use $(750nm - 705nm)\ /\ (750nm + 705nm)$ for calculating red-edge NDVI. Arroyo-mora et al. (2018) use Sentinel bands 5 and 6 (703.9nm and 703.8nm; 740.5nm and 739.1nm). We elected to use Sentinel-2 bands 5 and 7 to calculate red-edge NDVI, in order to correspond with the MERIS spectral bands used by Nguy-Robertson et al. (2012) when developing the appropriate equation. Thus we added an additional red-edge NDVI band to the Sentinel-2 image collection using the equation:

$$\boldsymbol{red-edge\ NDVI} = (\boldsymbol{band\ 7} - \boldsymbol{band\ 5})\ /\ (\boldsymbol{band\ 7} + \boldsymbol{band\ 5}) \quad (3)$$

We added a DOY band to the image collection by extracting the capture date from the metadata for each image. We then created an image composite based on the maximum red-edge NDVI for each pixel across the image collection using the embedded GEE function '.qualityMosaic'. This resulted in a composite image where each pixel's data relates to the DOY within the growing season on which the maximum red-edge NDVI was recorded and for which a cloud free image was available.

The resulting composite is comprised of images captured throughout the growing season (Figure 2, Map 3 (a) and Map 4 (b)). However the majority of images originated from four acquisition dates (DOY: 211, 218, 221 and 241). There are a number of outliers in the early and late season with dates ranging from 151 to 281.

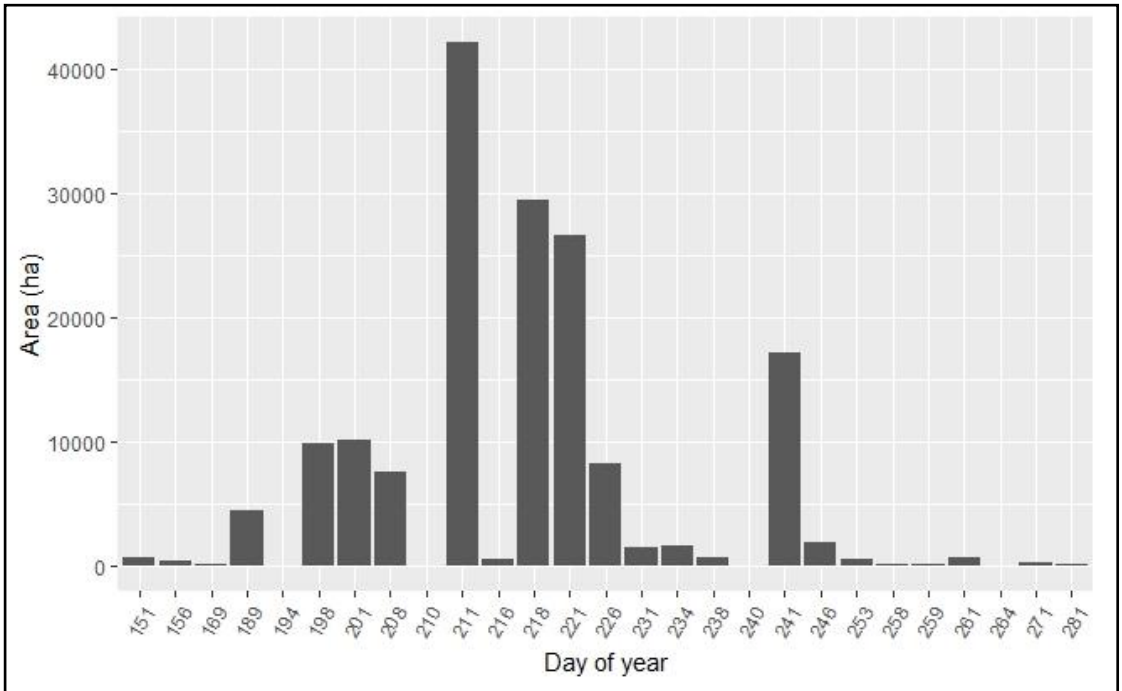


Figure 2: Per pixel distribution of the day of year of maximum red-edge NDVI from the composite satellite image

To calculate LAI we divided the composite image into two parts; pixels where red-edge NDVI was less than 0.6 and pixels where red-edge NDVI was greater than or equal to 0.6 (Nguy-Robertson et al., 2012). We then calculated the LAI for each pixel within these images using the equations outlined in Nguy-Robertson et al. (2012). For red-edge NDVI less than 0.6:

$$\boldsymbol{red-edge} < \mathbf{0.6}: (\boldsymbol{red-edge\ NDVI} - \mathbf{0.13})/\mathbf{0.14} \qquad (4)$$

For red-edge NDVI greater than or equal to 0.6, we first converted the values to Chlorophyll Index (CI) red-edge. Nguy-Robertson et al. (2012) define CI red-edge as $((NIR\ /\ red\ edge) - 1)$ or $((MERIS\ 12\ /\ MERIS\ 9) - 1)$ or $((778.75nm\ /\ 708.75)\ - 1)$. Zhang, et al. (2018) define CI red-edge as $((783nm\ /\ 705nm)\ - 1)$, which they relate to bands 7 and 5 in Sentinel-2 data respectively. Thus we calculated CI red-edge using the equation:

$$\boldsymbol{CI\ red-edge} = ((\boldsymbol{band\ 7}\ /\ \boldsymbol{band\ 5})\ - \mathbf{1}) \qquad (5)$$

We then calculated the LAI for CI red-edge pixels using the equation from Nguy-Robertson et al. (2012):

$$((\boldsymbol{CI\ red-edge} - \mathbf{0.63})\ /\ \mathbf{0.95}) \qquad (6)$$

Finally we mosaicked the two images back into one composite image using the embedded GEE function '.mosaic'.

2.1.5.2.3 Adding seasonal weather variables

We added seasonal weather variables for each pixel to correspond with the seasonal weather variables input into the multiple regression analysis. To achieve this, the 'GLDAS-2.1' (Beaudoing, 2016; Rodell et al., 2004) and 'PERSIANN-CDR' (Ashouri et al., 2014; Sorooshian et al., 2014) datasets available within GEE were utilised. GLDAS provides 3-hourly air temperature readings (K) at 2m above the ground surface with a spatial resolution of 0.25° by integrating ground based and satellite data. PERSIANN-CDR provides daily precipitation (mm) readings with a spatial resolution of 0.25°. For both variables, we created an image collection consisting of data for the growing season DOY 166 to 288. We then created two composite images based on the mean temperature (after conversion to °C) and total precipitation for each pixel within the growing season. We then added these two image bands to the LAI image composite. Each pixel within the composite thus consisted of four bands: LAI, DOY, mean temperature (Tp_av) and total precipitation (Psum).

2.1.5.2.4 Simulating silage maize yield

At this stage, each pixel contains the required data corresponding to the multiple regression equations. Thus, by first extracting the DOY for each pixel from the DOY band, then

looking up the corresponding regression coefficients in the coefficients table, we were able to simulate yield for each pixel. Finally, we added the resulting yield values as a band to the composite image and exported the image from GEE as a GeoTIFF raster. GEE divides rasters into smaller tiles for export. Thus, after export, we mosaicked the exported raster files in R using mosaic:raster (Hijmans, 2018). This final raster file was then analysed for results in QGIS (QGIS Development Team, 2019).

2.2 Validation

Silage maize yields are not officially recorded at field level. We obtained validation data from the AfSBB at the Nomenclature of territorial units for statistics (NUTS) 2 and 3 level. In Germany, NUTS2 regions consist of Bundesländer (states) and NUTS3 regions consist of Kreisfreie Städte and Landkreise (independent towns and districts) (Eurostat, 2018). In Brandenburg, there are 14 districts and 4 independent towns. We excluded independent towns from validation, as official yield data does not exist for these areas. A NUTS3 shapefile was obtained from the Bundesamt für Kartographie and Geodäsie (Federal Agency for Cartography and Geodesy) (BKG) (BKG, 2019).

To validate the field level data we first obtained the mean SCYM simulated yield (kg ha^{-1}) for each silage maize field using the zonal statistics tool in QGIS before re-calculating to produce mean yield for each field in dt ha^{-1} to match the unit of the validation data. The validation data provided by AfSBB is the fresh matter weight. The MONICA crop model yield output is dry matter content, which consists of 35% of the fresh matter (AfSBB, 2017; Gottschalk et al., 2018) and approximately 10% of the crop remains in the field as stubble after harvesting. In order to compensate for this, we first scaled the SCYM simulated yield from 35 to 100% to produce a fresh matter yield before subtracting 10% to produce an overall yield to match the validation data.

To aggregate this to NUTS3 level, we used the Aggregate tool in QGIS. We first summed the simulated total yield (kg) to produce a total simulated yield (kg) for each NUTS3 district before recalculating to dt ha^{-1} based on the total area of silage maize fields in each district. The same procedure was used to simulate mean yield in dt ha^{-1} at the NUTS2 level.

3. Results

3.1 MONICA crop model output

The crop model output produced 32400 simulations. The simulation results include a time series of crop growth recorded as daily LAI (Figure 3) together with an overall yield. The results show that in a number of simulations crop growth was limited with a lowest maximum LAI of 0.03 (Table 6). This resulted in a right skew in the maximum LAI data (Figure 4) with the 1st quantile result of 2.51 being only slightly less than the mean of 2.71. However, almost two thirds of the simulations produced an LAI greater than or equal to 2.71. The DOY on which the maximum LAI was produced ranges from the 135 to 287 with the days normally distributed around the mean of 207 (Figure 5). The distribution of the yield output follows a similar pattern to the maximum LAI in that it is skewed to the right (Figure 6). The Pearson correlation coefficient of R = 0.73 (p = < 0.01) and scatter plot (Figure 7) shows that maximum LAI and yield are strongly significantly positively correlated. Figure 9 suggests that yield increases as the day on which maximum LAI occurs becomes later until around the mean of the DOY of the maximum LAI (207), after which the yield tends to decrease with later days.

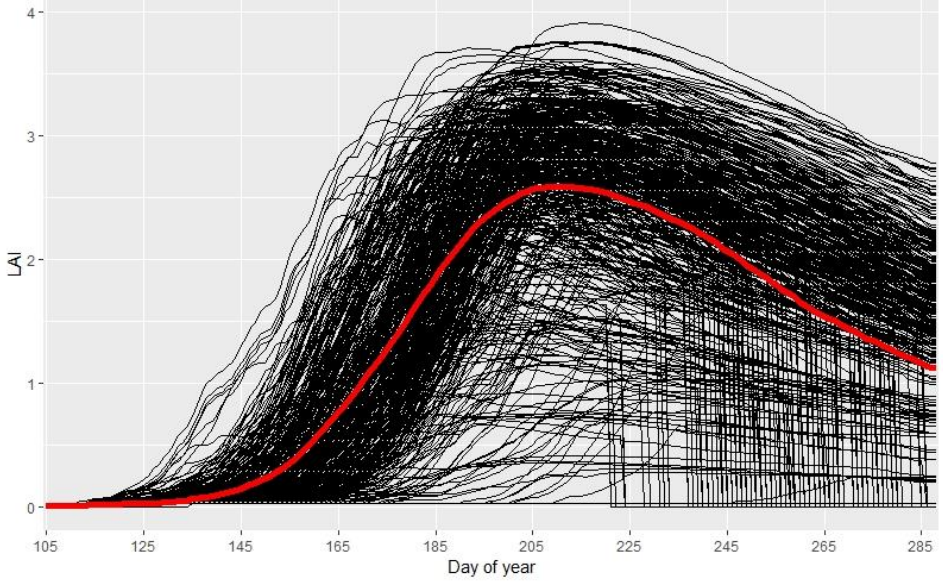


Figure 3: The LAI of a random sample of 500 MONICA crop model simulations throughout the growing season. Each black line represents one simulation. The red line represents the mean. Vertical lines show where harvesting has occurred.

Table 6: Summary statistics of the MONICA crop model simulations output

	Min	Max	Mean	Median	Mode	SD	1stQuantile	3rdQuantile
maxLAI	0.03	4.47	2.71	2.87	NA	0.70	2.51	3.17
DOY of maxLAI	135	287	206.9	207	209	13.37	200	214
Yield (dry matter) (kg ha^{-1})	10.8	17806.2	10749.1	11393.5	NA	3678.8	8213.8	13627.5

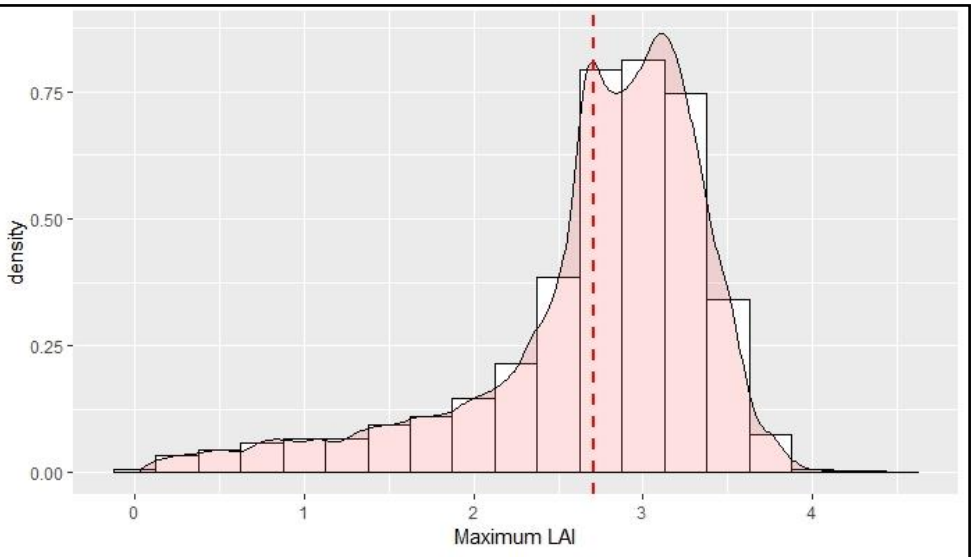


Figure 4: The density distribution of maximum LAI from the MONICA crop model simulations. The dashed red line represents the mean.

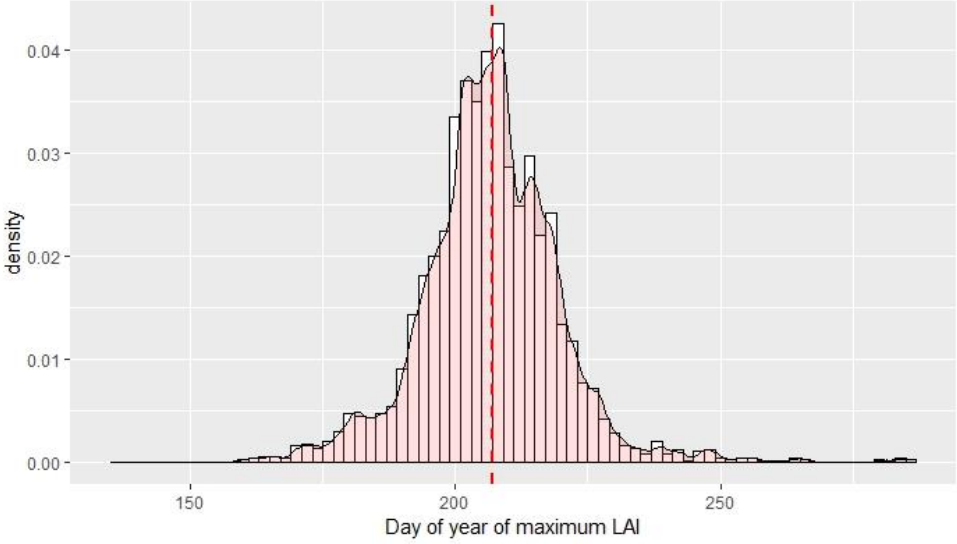


Figure 5: The density distribution of the day of the maximum LAI from the MONICA crop model simulations. The dashed red line represents the mean.

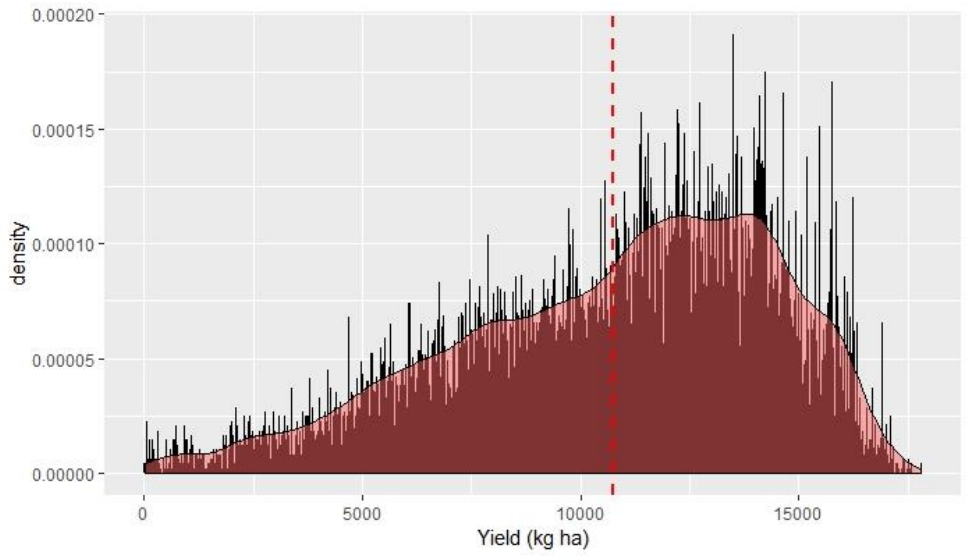


Figure 6: The density distribution of the yield from the MONICA crop model simulations. The dashed red line represents the mean.

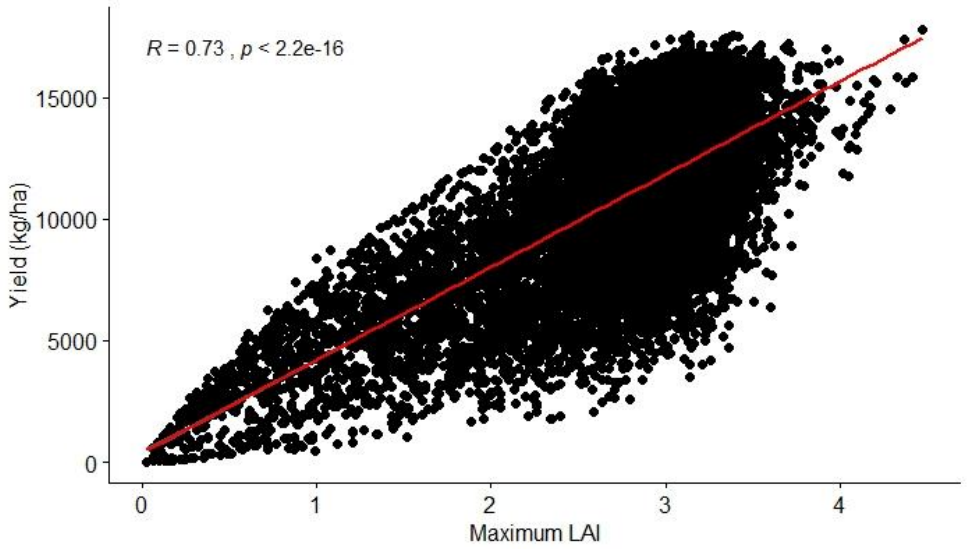


Figure 7: Scatterplot showing the correlation between yield and maximum LAI from the MONICA crop model simulations. The red line represents the regression line and shows a linear relationship between these two outputs variables.

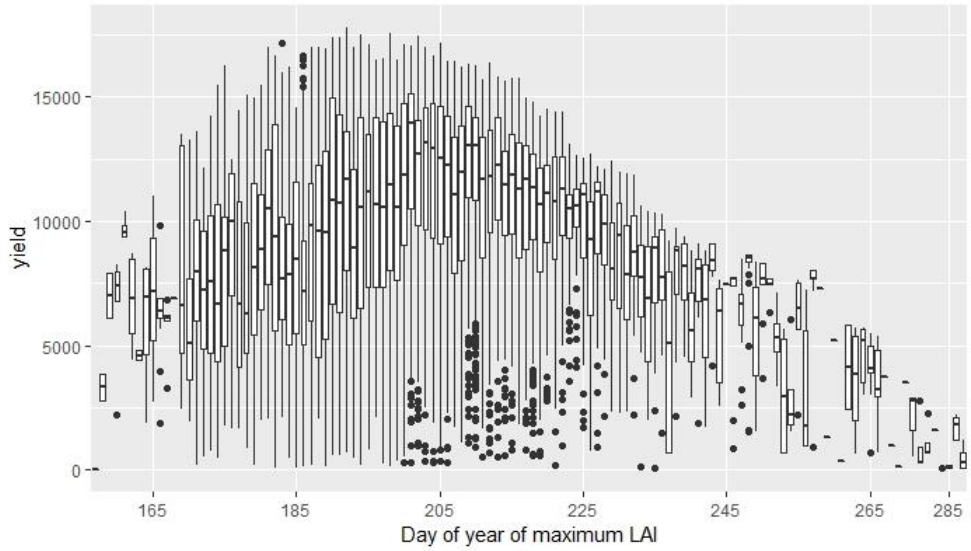


Figure 8: A comparison of yield (kg ha^{-1}) and the day of the maximum LAI from the MONICA crop model simulations

3.2 Multiple regression

The multiple regression produced an overall model with an adjusted R^2 of 0.53 ($p = < 0.001$) suggesting that the model explains around 53% of overall yield. Daily model output was also considered, as the model varies depending on the DOY. This is an important parameter as SCYM estimates yields based on the inputs on a particular day. The adjusted R^2 values for each DOY model ranged from 0.26 to 0.74 (Figure 9). Models corresponding to DOY greater than or equal to 189 provide a good model fit with R^2 values ranging from 0.63 to 0.74. Maximum R^2 was produced by the model for DOY 226. However for DOY less than 189 the model does not provide a good fit (For detailed results of the amongst DOY model variability see Appendix 11 and 12).

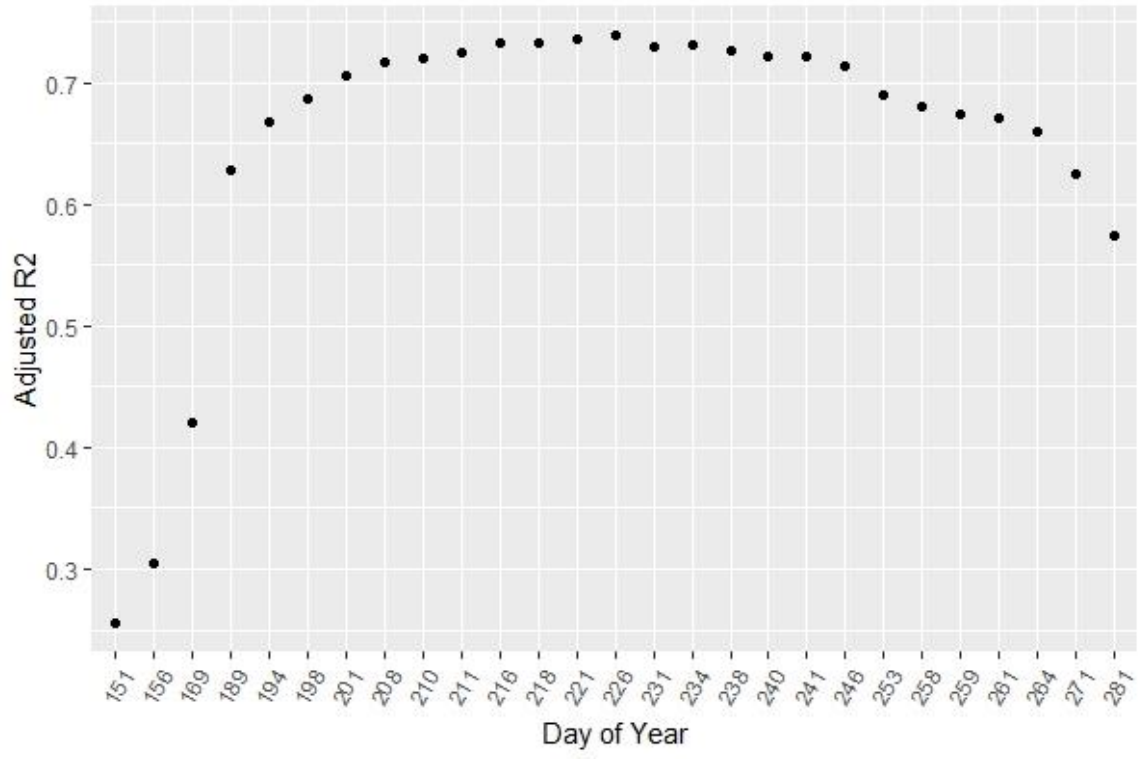


Figure 9: The adjusted R^2 for each multiple regression DOY model for days on which an image exists in the Sentinel-2 composite.

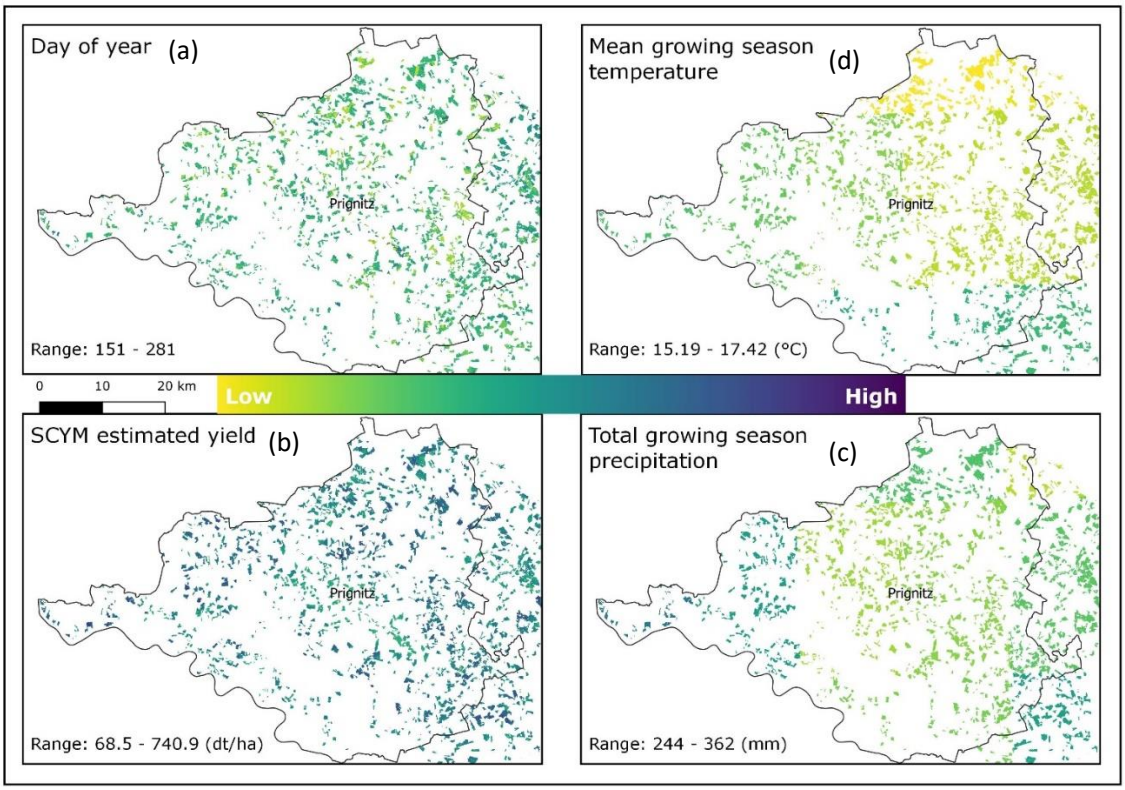
3.3 SCYM silage maize yield simulation

3.3.1 Pixel level

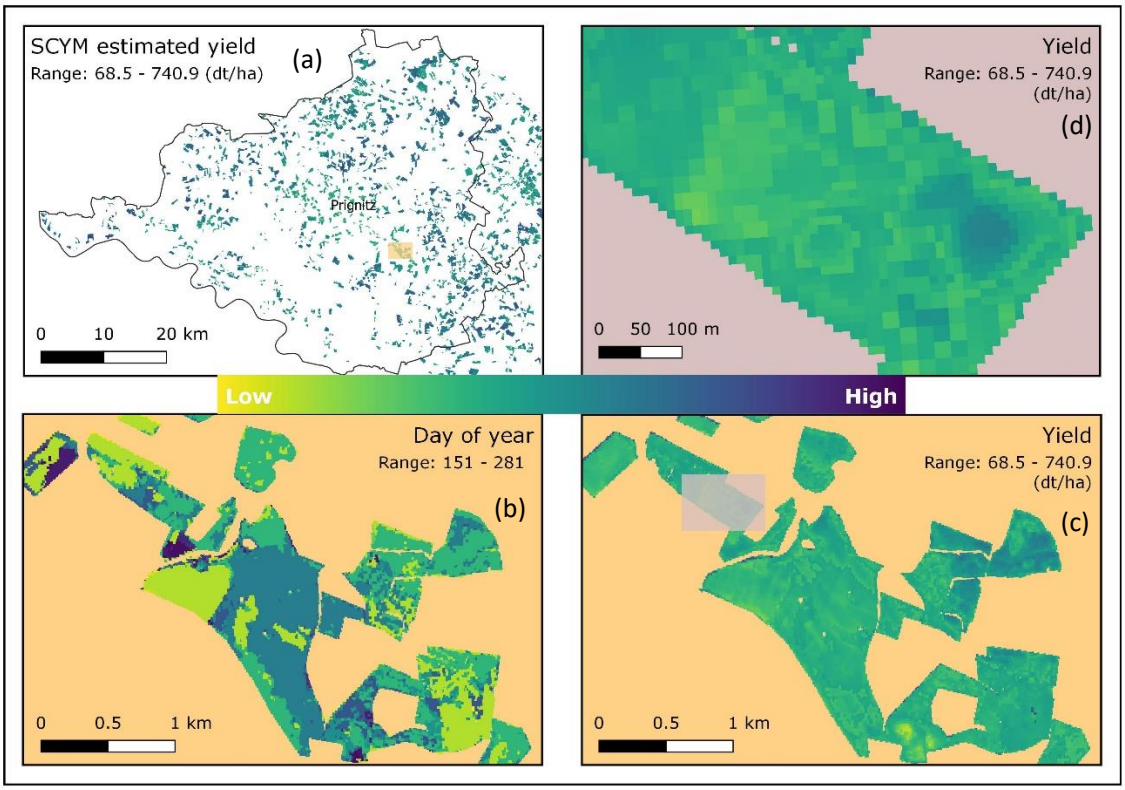
At pixel level, SCYM simulated yield ranged from 69 to 741 dt ha^{-1} (Map 3 (b), and Map 4 (a), (c) and (d)). Map 3 provides an indication of the distribution of DOY, simulated yield and the weather variables input into the multiple regression across the Prignitz (NUTS3) district. The distribution of the weather variables follows that of the spatial scale, revealing the low variance within the data (Map 3 (c) and (d)). The spatial distribution of DOY is fragmented (Map 4 (b)). In contrast, the simulated yield displays a fluid spatial distribution pattern (Map 4 (c) and (d)).

3.3.2 Field level

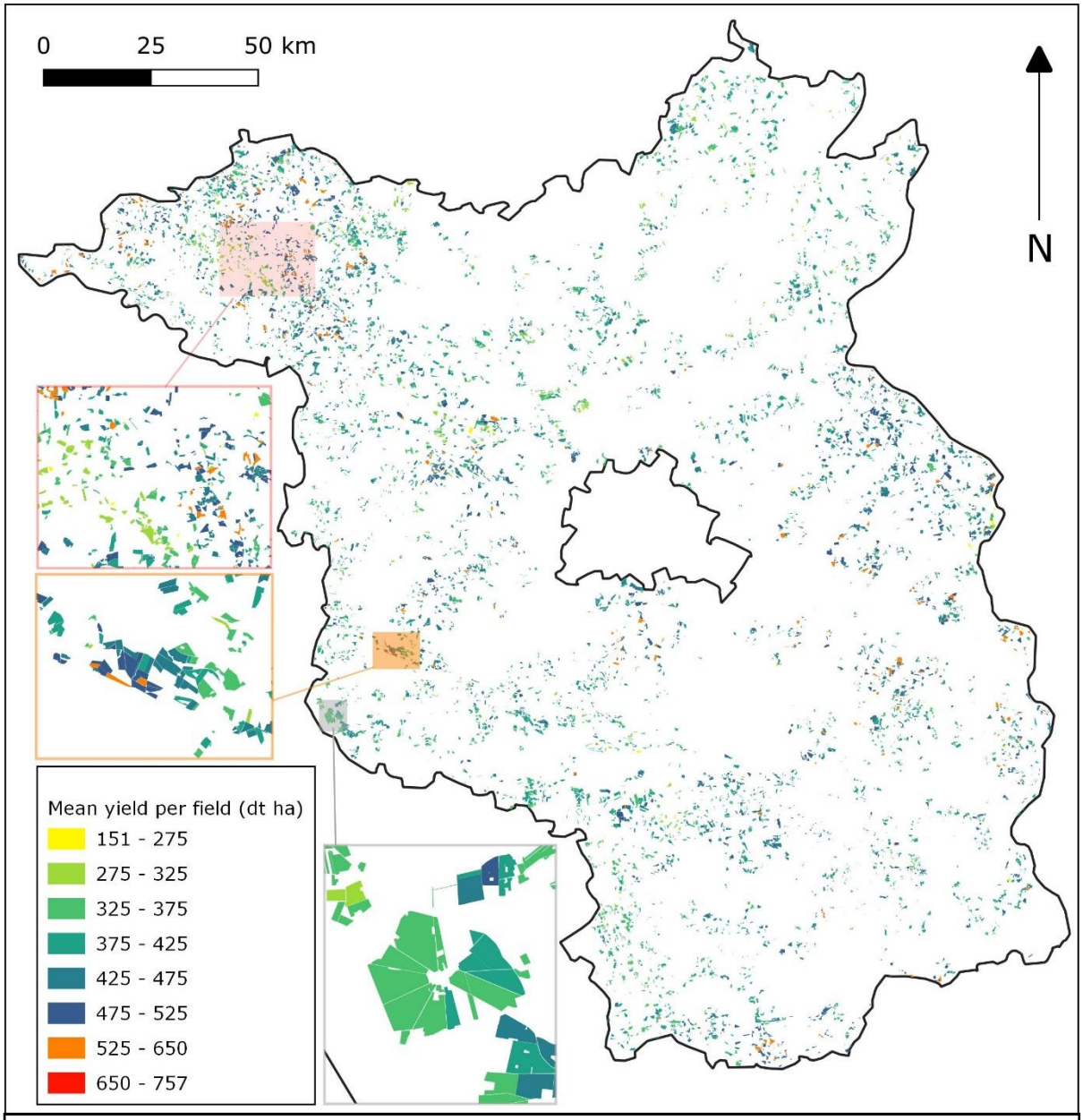
At field level, simulated yield ranges from 151 – 757 dt ha^{-1} with a mean of 409.1 dt ha^{-1} (Map 5 and Table 7). Yields display a normal distribution around the mean with a small number of outliers at the extremes (Figure 10).



Map 3: Close-up view of the distribution of (a) Maximum red-edge day of year composite pixels, (b) SCYM simulated silage maize yield, (c) Total growing season precipitation and (d) Mean growing season temperature



Map 4: Detailed view of the distribution of SCYM simulated maize yield in Prignitz (a, c and d) and the composite day of year (b). The background colours correspond to the inset areas shown on the coarser scale maps.



Map 5: SCYM simulated maize yields in 2017 at field level (dt ha^{-1}). Colour scale is designed to highlight the lowest and highest values whilst providing a scale for the majority of values distributed around the mean.

Table 7: Summary statistics of simulated maize yields for Brandenburg in 2017 following SCYM method at field and NUTS3 level

	Min	Max	Mean	Median	SD	1stQuantile	3rdQuantile
Field	150.7	757.3	409.1	405.8	60.27	364.7	449.6
NUTS3	386.4	481	428.2	430.6	25.62	411.4	442.8

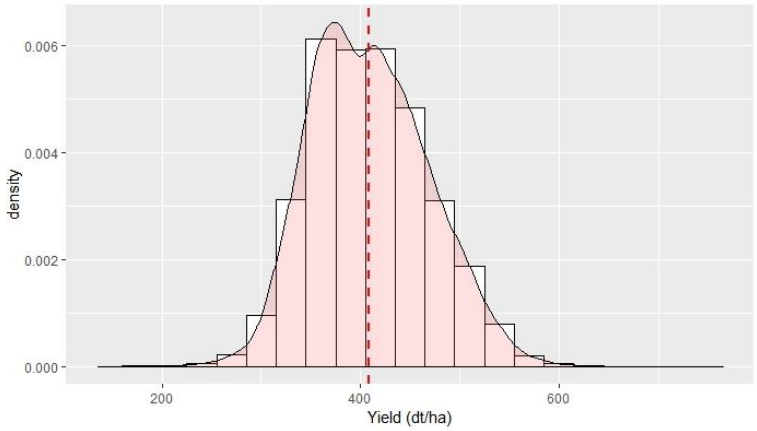
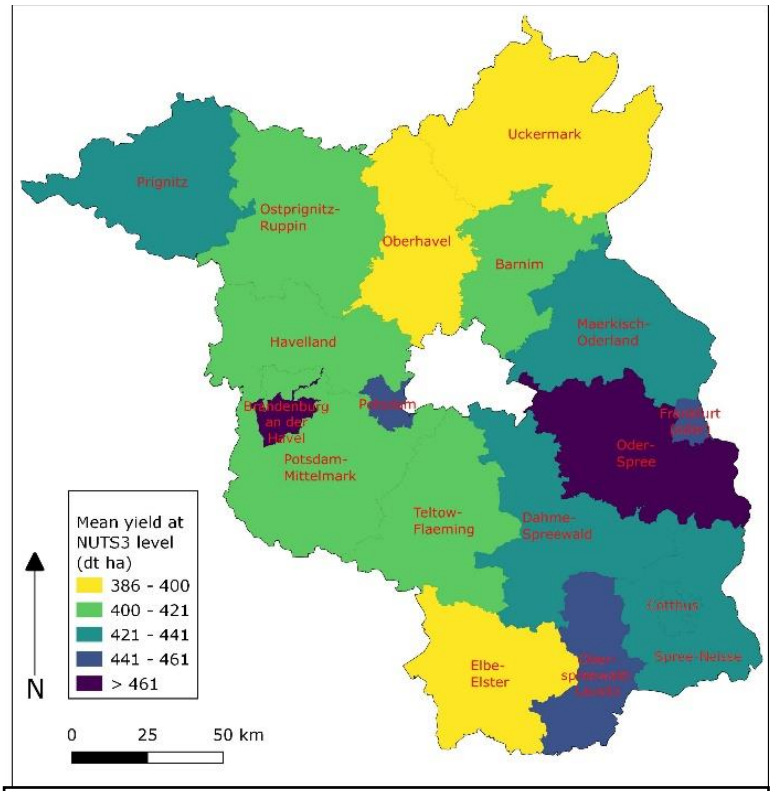


Figure 10: Distribution of simulated yields across maize fields in Brandenburg following SCYM methods. Red dashed line represents the mean.

3.3.3 NUTS3 level

At NUTS3 level, simulated yield ranged from 386 (Uckermark) to 481 (Brandenburg an der Havel) dt ha^{-1} (Table 7 and Figure 11). The yield variation is well distributed across the state, with some areas appearing to show patterns of clustering with neighbouring states (Ostprignitz-Ruppin, Havelland, Potsdam-Mittelmark, Teltow Flaeming; Dahme-Spreewald, Cottbus, Spree-Neisse) (Map 6).



Map 6: Simulated yields at NUTS3 level following SCYM method

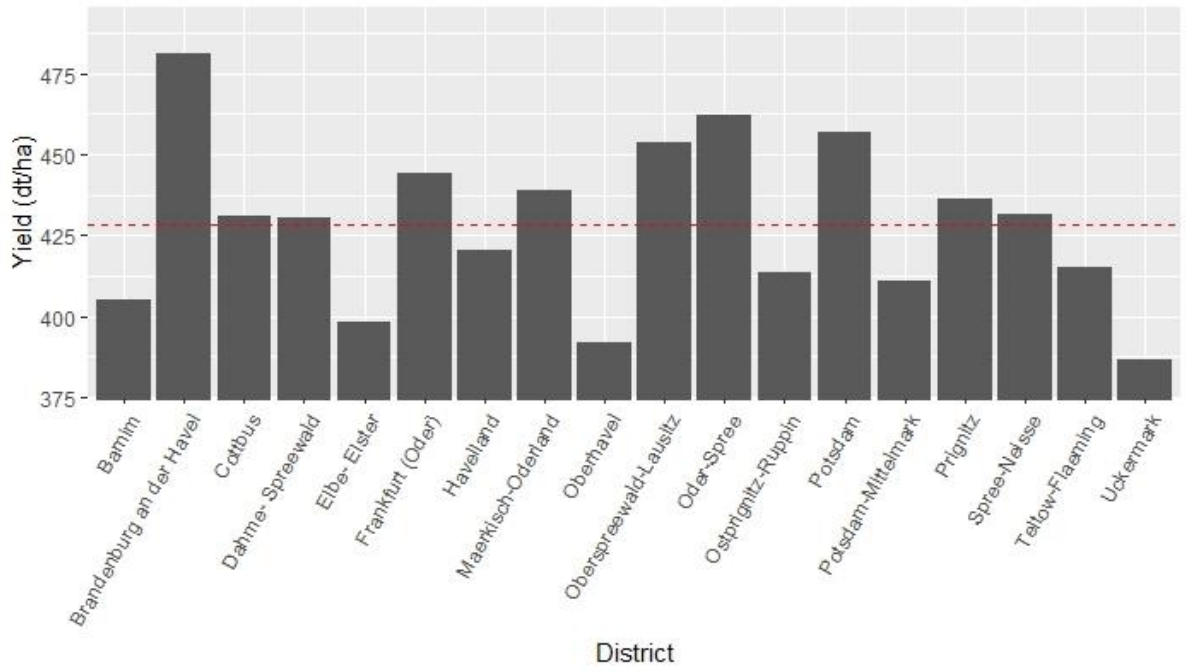
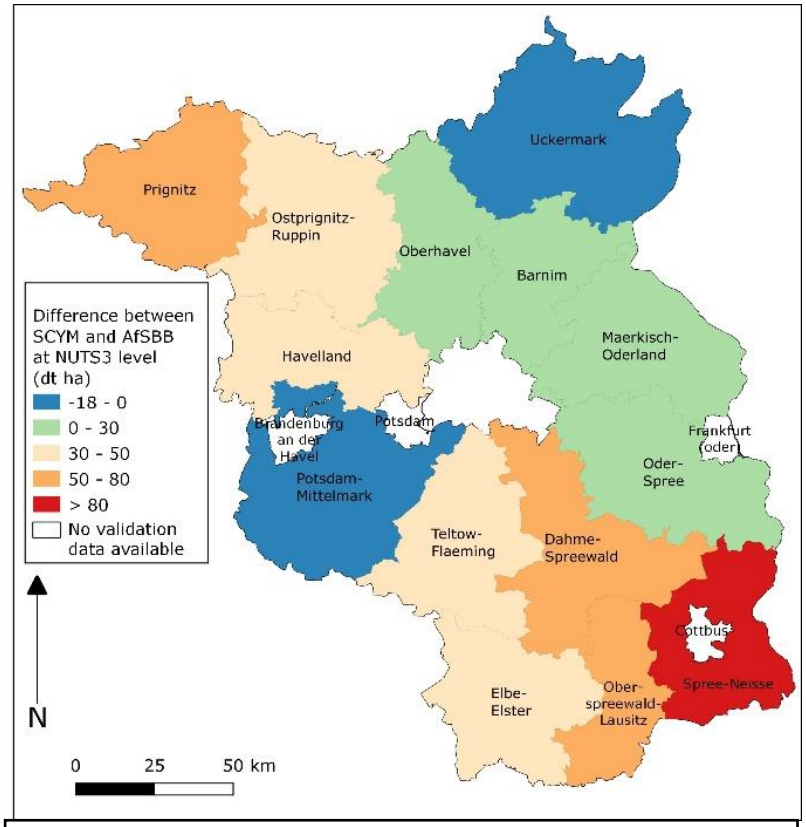


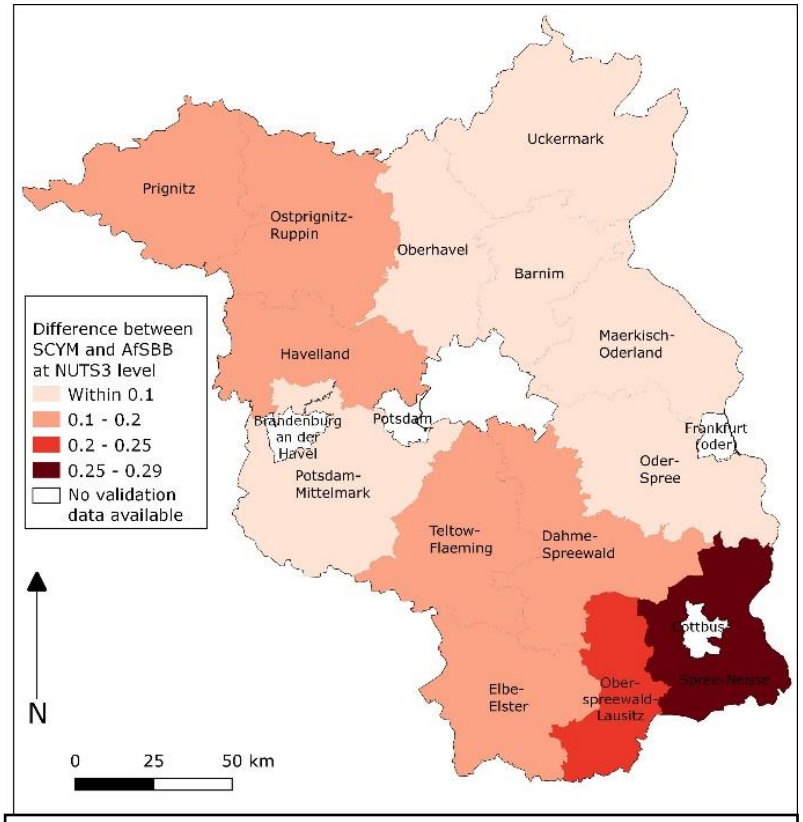
Figure 11: Simulated maize yields for each district in Brandenburg in 2017 following SCYM method. Dashed red line represents the mean.

3.3.4 Validation

A comparison of SCYM simulated yields with AfSBB data at NUTS 3 level, shows that SCYM underpredicted in two districts (Uckermark and Potsdam-Mittelmark) (Maps 7 and 8, Figure 12 and Table 8). Some areas display patterns of clustering with neighbouring states. Oberhavel, Barnim, Maerkisch-Overland and Oder-Spree all overpredicted within 30 dt ha^{-1} (Map 7). When considered as factors, six districts were within 10% of the AfSBB data, all of which are located on the eastern side of Berlin. Six districts were within 10-20% of the AfSBB data, all of which are located on the west and south-west of Berlin (Map 8). The districts of Oberspreewald-Lausitz and Spree-Neisse, both of which are located in the far south of Brandenburg, show the largest differences between SCYM and AfSBB with a difference of 21% and 29% respectively. When the entire state is considered, SCYM overpredicts maize in Brandenburg in 2017 by 9% (Figure 12 and Table 8). Statistically, SCYM and AfSBB data have a weak non-significant positive correlation at NUTS3 level (R = 0.28, P = 0.32) (Figure 13) (Shapiro-Wilks test for normal distribution: AfSBB = 0.95, SCYM = 0.98).



Map 7: Difference between SCYM simulated yields and AfSBB yields (dt ha^{-1}).



Map 8: The difference between SCYM simulated yields and AfSBB yields as a Factor. Within 0.1 = SCYM is within 10% of AfSBB.

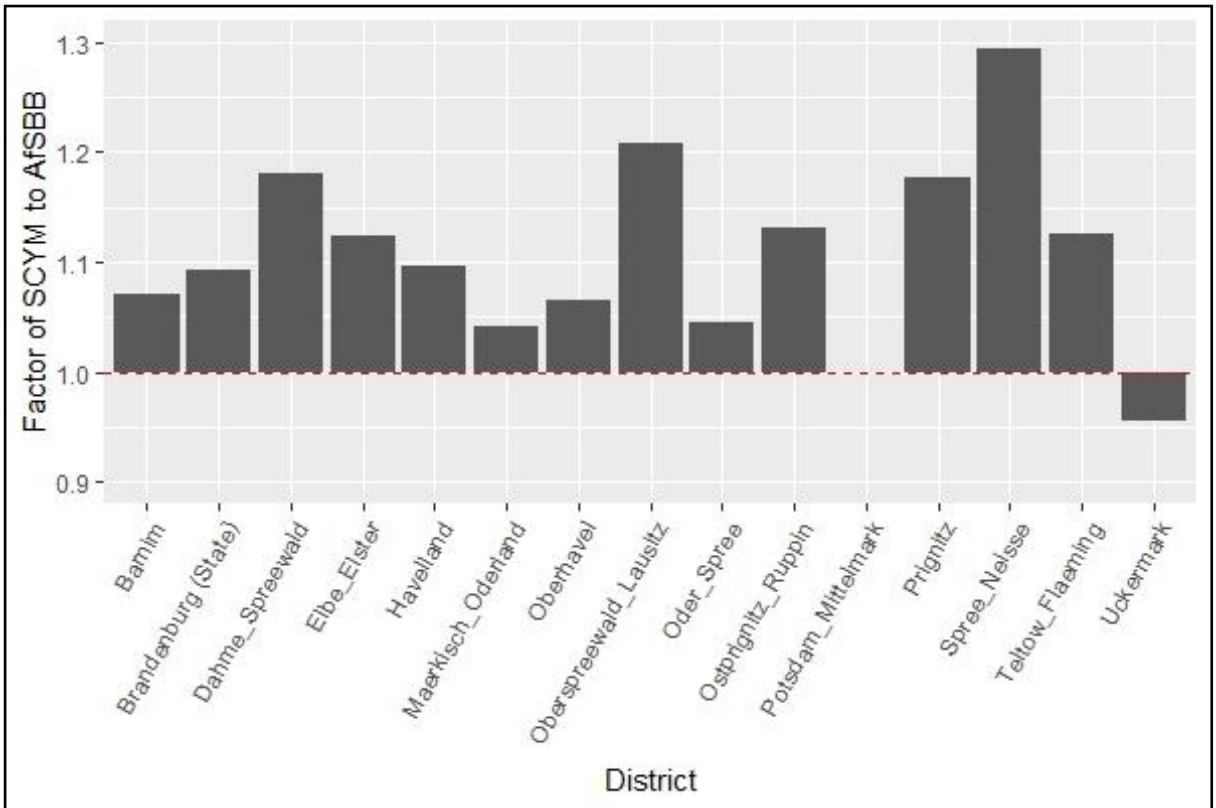


Figure 12: Difference between simulated maize yield in Brandenburg in 2017 following the SCYM method and the AfSBB results at district level as factors. The dashed red line represents a factor of 1.0. Points above the line represent overpredicted districts, below the line represents underpredicted districts.

Table 8: Comparison of simulated maize yields in Brandenburg in 2017 at NUTS3 and NUTS2 level following SCYM method with the AfSBB results.

District	SCYM	AfSBB	Difference (dt/ha)	Difference (factor)
Barnim	405.1	378.3	26.8	.07
Dahme-Spreewald	430.3	364.8	65.5	.18
Elbe-Elster	398.1	254.2	43.9	.12
Havelland	420.3	383.7	36.6	.10
Maerkisch-Oderland	439	421.3	17.7	.04
Oberhavel	391.6	367.9	23.7	.06
Oberspreewald-Lausitz	453.5	375.2	78.3	.21
Oder-Spree	462.2	442.3	19.9	.04
Ostprignitz-Ruppin	413.6	365.5	48.1	.13
Potsdam-Mittelmark	410.6	411	-0.4	0
Prignitz	436.4	370.7	65.7	.18
Spree-Neisse	431.7	333.7	98	.29
Teltow-Flaeming	415	368.5	46.5	.13
Uckermark	386.4	404.9	-18.5	-.05
Brandenburg (state)	420.9	385.4	35.5	.09

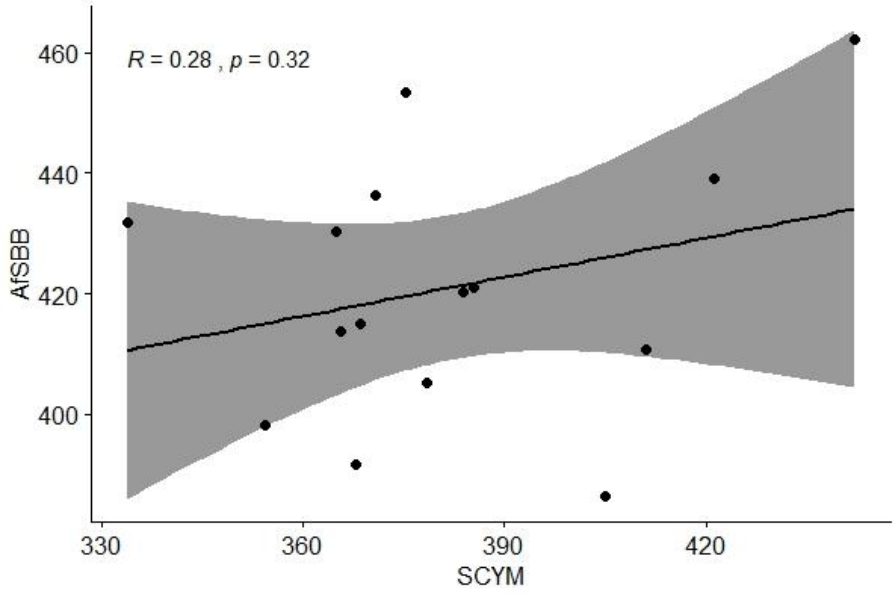


Figure 13: Scatterplot of the correlations between the simulated maize yields in Brandenburg in 2017 following SCYM methods and the AfSBB results with a Pearson correlation of R = 0.28

4. Discussion

4.1 Summary of findings

We found that when compared to official statistics provided by the AfSBB, the SCYM method showed a weak positive correlation of 0.28, although the correlation was not found to be statistically significant (p = 0.32) ((Figure 13). Therefore, around 8% of the variation in maize fields in Brandenburg is captured by the SCYM method. We found that in three districts, the SCYM method was within a factor of 0.05 and for a further three districts within a factor of 0.1. Furthermore, we found that SCYM overestimated maize yields in 12 of the 14 districts for which validation data was available.

4.2 MONICA crop model simulations

The MONICA crop model simulations produced a range of maximum LAI from 0.03 to 4.47. In their research on the interactions of water and Nitrogen on silage maize, Gheysari et al. (2009) reported maximum LAI of between 3 and 4 for silage maize, dependent on fertiliser application and irrigation levels. Nguy-robertson et al. (2012) researched the temporal dynamics of LAI in a selection of corn maize varieties in Nebraska, USA and found that maximum LAI was 4.3 in rainfed fields. Lobell et al. (2015) describe a large variability in the

maximum LAI from their simulations, spanning a factor of two to three. Whilst they do not report specifically the range of the maximum LAI their simulations produced, it can be deduced from their report that the range of maximum LAI was between 2 and 4. Thus, the maximum LAI from the MONICA simulations is in the range of what one would expect for rainfed silage maize.

The mean and median values of the maximum LAI (Table 6) suggest that a number of simulations produced low maximum LAI values. These are likely as a result of water stress in simulations caused by a combination of low rainfall and poor soils. Nevertheless approximately two thirds of the MONICA simulations produced LAI above the mean.

The MONICA crop model simulations produced a range of dry matter yields from 0.1 to 178.1 dt ha^{-1}. Lobell et al. (2015) reported yields from their crop model simulations ranging from 0 – 155 dt ha^{-1}. It is presumed (although not directly stated) that these results relate to dry matter. In Brandenburg the mean silage maize yield per year for the period 2002 to 2018 is 173 dt ha^{-1} (Ebel & Barthelmes, 2019) suggesting that the MONICA simulations are within the realms of the expected results. However, there are a number of low yields linked to the simulations which produced a low maximum LAI, resulting in a distribution of yields following a similar pattern to the distribution of maximum LAI (Figures 4 and 6). This is to be expected given the correlation identified between these outputs (Figure 7). Overall the simulations provide a wide variety of results in terms of both LAI and yield for input into the multiple regression model.

4.3 Multiple regression

The multiple regression produced an overall model with a good model fit, but the models ability to predict yield is dependent on the DOY on which a particular image was captured (Figure 9). The distribution of model fit across the growing season produced an arching effect, with model accuracy improving up until the peak of the growing season following which model accuracy begins to decline. There are a number of factors which influence this. Firstly, as the growing season progresses, model accuracy improves due to the declining variability in future crop growth. This is intuitive, as in the early stages the scale of possibilities for crop growth is large and there are many unknowns in the coming months. However as the season progresses, the number of possible outcomes declines, potentially leading to more accurate

predictions, i.e. the more established a crop is, the more accurate the prediction is likely to be. Furthermore, fields sown at a later date will show late onset of crop growth which may not be captured in the earlier images, e.g. fields sown on day 135 (the last day for which sowing was possible in the MONICA simulations) may still show zero LAI by the time the earliest images are captured on DOY 151. That said, if they show zero LAI, it is unlikely that this will be the maximum LAI image available for this location through the growing season. The declining accuracy in the latter part of the growing season, is likely due to maximum LAI having already occurred, i.e. for images taken after maximum LAI has occurred, predictions are based on LAI that is less than the LAI was at its maximum. In maize, maximum LAI is reached when silking begins, which is the beginning of the plant reproductive stage. This is followed by a decrease in LAI until zero LAI occurs at maturity (Nguy-robertson et al., 2012). The model is not able to predict yields for these images as accurately as earlier images, as it cannot know what the maximum LAI was and the variety of maximum LAI is too varied.

Although the model produced by Lobell et al. (2015) provided a better overall model fit, they also found that the model accuracy improved through the growing season resulting in an optimum window from which images could be selected in order to capitalise on the most accurate model. By first training a regression model on all of the days within the growing season, it is possible to choose a minimum DOY and maximum DOY between which satellite images could be selected. For instance, if only images on days where the regression model was able to explain 60% of variability were used, the window for images would be days 185 to 275 (Figure 9). This would have the advantage of increasing overall model accuracy but would reduce the chances of obtaining cloud free images for every location within the study area. Furthermore, an advantage of SCYM is that images obtained on any date within the growing season, can be used to predict yield (Lobell et al., 2015). As such, whether or not an advantage can be gained by utilising this narrow window approach is questionable. This could however be tested in future research to establish whether it would be advantageous for the Brandenburg region. For this study, the overall model accuracy was considered sufficient for a first look into the feasibility of simulating yield in Brandenburg using the SCYM method (For a detailed discussion about the amongst DOY model variability see Appendix 11 and 12).

4.4 SCYM Pixel level yield simulation

The red-edge NDVI composite image is comprised of images captured throughout the growing season with a core element captured in the peak season and a number of outliers in the early and late season. The early season outliers are likely a result of a lack of available cloud free images before maximum LAI is reached, as one would expect any vegetation at these locations to continue to grow unless an event occurs which prevents further growth. The late season outliers may also be as a result of a lack of cloud free images at these locations until the late season. However, two further factors may also play a role here. Firstly, late sowing at these location may have resulted in maximum LAI occurring at a later date. Secondly, mixed pixels at these locations could be influencing the timing of the maximum yield.

The fragmented distribution of DOY revealed at the pixel level (Map 4, (b)) is the result of a combination of image availability and the maximum red-edge NDVI. When compared to the yield distribution (Map 4 (c) and (d)) it is evident that yield variability is independent of the DOY. The fluid pattern of the yield displays a realistic pattern of how one would expect yield to vary within a particular field. Geomorphological elements such as depressions and rises affect parameters such as soil type, soil moisture and soil layer distribution, all of which can influence yields. This pattern indicates that despite the variation both in image dates and LAI on image dates, the SCYM method is able to nevertheless produce a realistic pattern of yields, in line with the patterns one would expect to find within a field of maize.

4.5 SCYM Field level yield simulation

Although no validation data is available at this level, the rescaling of the pixel level results to field level provides results per hectare and therefore offers an opportunity to discuss and compare the summary statistics to previous research and expected results.

The distribution of yield reveals a wide range of results with some fields producing considerably lower than expected yields and some producing considerably higher (Map 5, Table 7 and Figure 10). Whilst it is possible that crops in some parts of fields are subject to stresses which limit growth, it is unlikely that the yields at the higher end of this range are realistic, given the conditions found within Brandenburg. It may be that pixel size and field size both play a role here. When Burke & Lobell (2017) investigated the role of field size in

yield variation, they found that smaller field sizes produced less accurate results than larger fields. The minimum field area within the silage maize mask is 5.5 m^2 (LGB, 2017). This is considerably smaller than the pixel area of 400 m^2 obtained from the 20x20 m spectral bands used to estimate LAI. If a particularly small field is found within a pixel with a high LAI due to the presence of other vegetation within the pixel, this field will subsequently receive an excessively high yield, once adjustment for field size is taken into account.

However, the yields follow a normal distribution with a mean of 409 dt ha^{-1} (Table 7 and Figure 10). Given the standard deviation of 60 dt ha^{-1}, 68% of all fields are within 349 to 469 dt ha-^{1} and 95% are within 289 to 529 dt ha^{-1}. As the mean silage maize yield per year for the period 2002 to 2018 in Brandenburg is 444.8 dt ha^{-1} (90%) with a standard deviation of 71 dt ha^{-1} (Ebel & Barthelmes, 2019) and that the mean yield for Brandenburg in 2017 was reported as 385.4 dt ha^{-1} (AfSBB, 2017), one can state that the results at field level are within the realms of the expected results. Nevertheless, it is difficult to provide more concrete assertion, given the lack of validation data at this level.

4.6 SCYM NUTS3 level yield simulation and validation

Aggregation of the results to district level offers the opportunity to validate the results against those provided by AfSBB. As previously discussed, validation data is limited to the NUTS3 level, due to a lack of official data provided by the state. At this level, validation data is limited to 14 observations, each of which covers a large area, reducing the yield variation found within each of these areas into one value. As such, it is difficult to provide an accurate assessment of the accuracy of the SCYM method in Brandenburg. However, it is possible to provide a detailed overview of the results and to compare the results to the validation data and previous research involving SCYM, in order to make some assertions on the ability of SCYM to simulate yields for Brandenburg.

Examination of the correlations between the SCYM method results and AfSBB data reveals a weak positive correlation between the two datasets ($r = 0.28$, $p = 0.32$, $R^2 = 0.08$) (Figure 13) meaning that the SCYM method is able to explain 8% of the yield variability within Brandenburg. This is lower than the weakest result of Lobell et al. (2015) who found that SCYM was able to explain between 14 and 58% of yield variation within maize fields in the US Midwest, albeit at field level. Lobell & Azzari (2017) were able to capture 67% of the variation

in maize yields at county level and 55% at within county level in the US Midwest. Similarly, Azzari et al. (2017) were able to capture at least 50% of maize yield variation in the US but reported poorer results when adapting SCYM for irrigated wheat fields in India and maize in Gambia. Jin et al. (2019) reported that SCYM was able to capture between 40 and 55% of maize yield variation at district level in Kenya. Thus, when validated in this manner, SCYM performed poorly compared to previous studies.

SCYM has been proven to produce results which match the accuracy of ground based measures in the study region (Burke & Lobell, 2017) and is designed to estimate yield without the need for ground based validation data (Azzari et al., 2017; Burke & Lobell, 2017; Jin et al., 2017, 2019; Lobell & Azzari, 2017; Lobell et al., 2015). Furthermore, Azzari et al. (2017) noted a decrease in model accuracy correlated with a decrease in the quality of the field based measures used for validation. Burke & Lobell (2017) noted that agreements between SCYM and validation data were strongest when confidence in the validation data was highest and that errors in the field data were just as likely to cause discrepancies as errors in the satellite based estimates. As such, it is prudent to consider whether the ground based field data available in Brandenburg is accurate enough to validate SCYM.

The AfSBB estimate yields based on a survey completed by farmers or by reporters appointed on their behalf. The completion of the survey is not mandatory. The selection of farms chosen is not random, due to a lack of response from farmers unwilling to complete the survey. Instead a selection from a list of participants known to respond are asked to compete the survey. The estimations are based on subjective impressions and therefore vary in accuracy depending on the experience of the farmer or reporter. The location attributed to the yield estimation is based on the registered company address of the farm and not on the location of the particular field. The estimated yield per hectare for a particular region is calculated based on the estimated yield from these surveys and the total area of crop included in the surveys in that region. These results are aggregated to higher levels including state and national level. In some regions too few or even no responses are received from farmers. In these circumstances, the yield is estimated based either on neighbouring regions or weighted by the national average. Furthermore, the quality of the yield estimations is decreasing and estimations are becoming harder to obtain (AfSBB, 2017).

This form of non-probability survey sampling adopted by AfSBB is known as Convenience sampling and comes with a number of sources of bias. It is unlikely to provide

good coverage of the entire population and the reliance on volunteers with varying degrees of expertise can lead to discrepancies in the results. Furthermore, there is a high risk that survey results obtained from Convenience sampling will differ substantially from the reality (Baker et al., 2013). This method of yield estimation fails to account for the heterogeneity of maize yields found amongst farms and within individual maize fields (Lobell & Azzari, 2017). Thus, one must conclude that the sampling method adopted by the AfSBB in order to estimate maize yield at NUTS3 level, carries large uncertainties and cannot be relied upon in these circumstances.

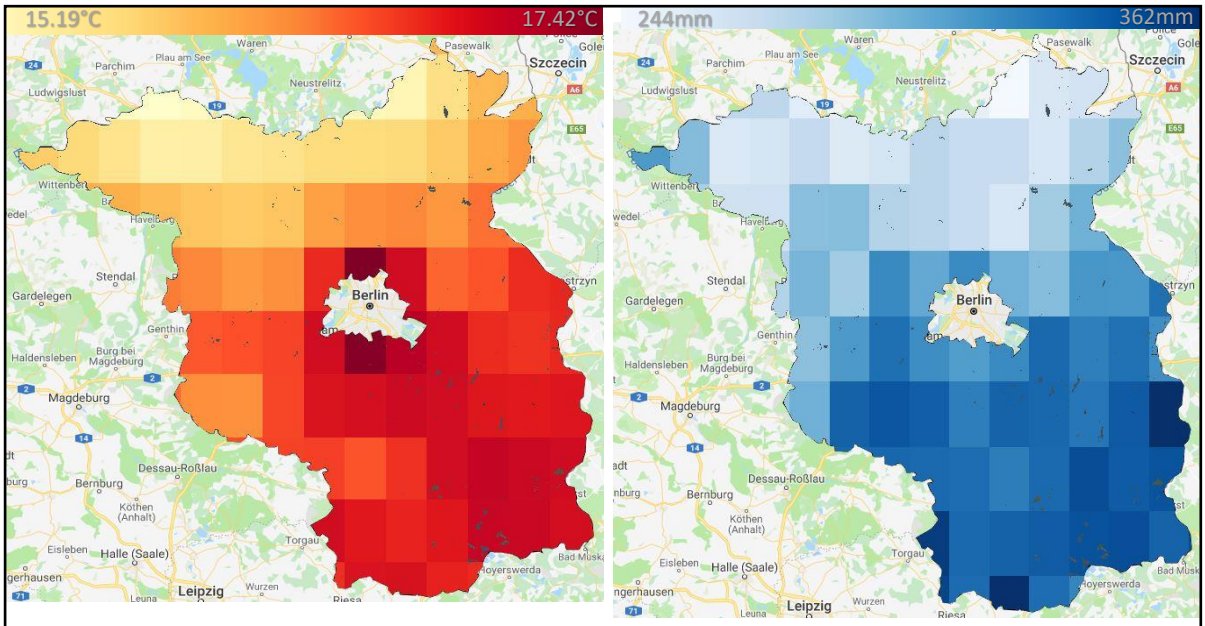
Given that SCYM model accuracy correlates with the quality of validation data, it is fair to speculate that a large proportion of the inaccuracy found when analysing the results of this research is a result of the inaccuracy of the validation data and not of the model itself. It is impossible to check the validity of this statement without repeating this research using highly accurate validation data. Burke & Lobell (2017) recommended using survey teams to precisely measure yields in fields which have been georeferenced using GPS coordinates. This method could be carried out in Brandenburg by selecting a large number of geographically well distributed silage maize fields, measuring their precise yield in a particular year and comparing this to the SCYM results for those fields in that year.

Despite the discrepancies with the validation data, it is nevertheless prudent to discuss the distribution of results at NUTS3 level, as the SCYM method discussed here maybe more accurate than the validation data suggests. The simulated yields ranged from 386.4 to 462.2 dt ha^{-1} giving a range of 75.8 dt ha^{-1} between districts. Given the maximum of 462.2 dt ha^{-1} one can say that the yield varies by 16.4% across districts.

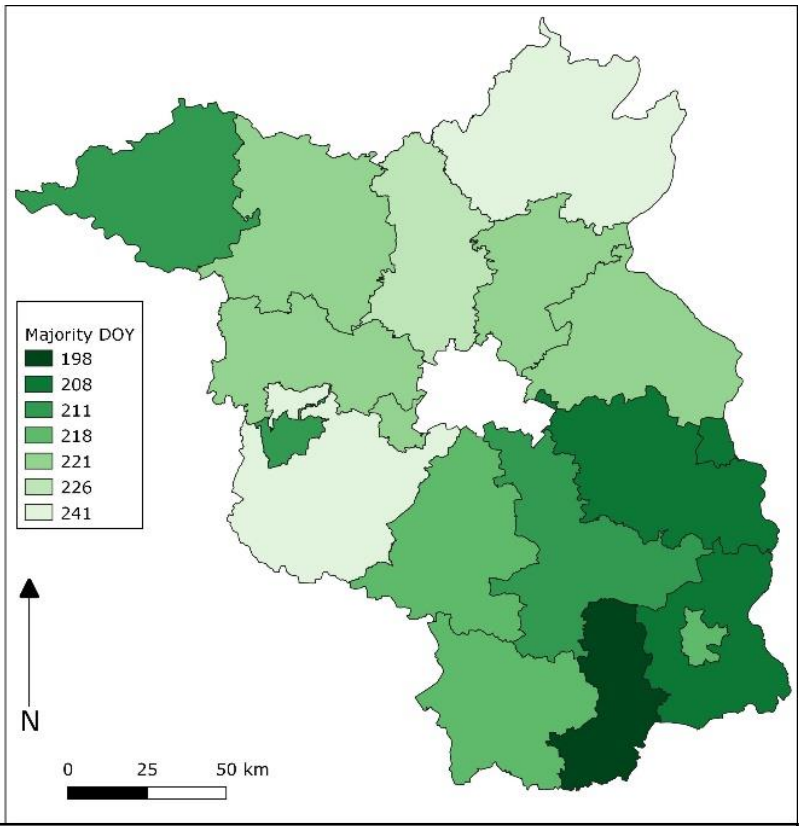
The spatial distribution of simulated yield provides some interesting results in terms of the clustering of similar yields with neighbouring districts (Map 6). Larger yields tend to be found within the south east of Brandenburg. When compared to the spatial distribution of the seasonal weather variables (Map 9) obtained from GEE and used to predict yield, one could hypothesise that the seasonal weather in the southeast, which was both warmer and wetter in 2017 than the rest of Brandenburg, could be having a greater influence on yields than other parameters within the model. Whilst, low-medium positive correlations were found between both of these variables and simulated yield (Temp: R = 0.36; Prec: R = 0.50) neither of the correlations were found to be statistically significant.

If the validation data are taken as true values then the SCYM model overestimated yield in twelve out of fourteen districts and by a factor of 1.09 on average across districts (Table 7 and Figure 13). Lobell & Azzari (2017) also found that SCYM tended to overestimate yield and attributed this to the crop model they used to produce simulations' (APSIM) inability to account for loss factors, such as disease or pests, or to underestimation of leaf area to yield production. Whilst this may of course be the case here, it is difficult to pinpoint overestimation to a particular factor, given the quality of the validation data previously alluded to.

A visual analysis of the spatial distribution of the differences between SCYM and AfSBB yields also appears to show some geographical clustering, with all districts whose SCYM yield was simulated within 10% of AfSBB yield being located in the northeast, aside from one (Map 8). The two districts with differences between SCYM predicted yield and AfSBB yield of greater than 20% (Oberspreewald-Lausitz and Spree-Neisse) are located in the southeast. This could mean that the AfSBB results for these districts are the most underestimated of all the districts, resulting in larger discrepancies between the estimates. However SCYM could be overestimating the yield in these districts due to a parameter in the multiple regression model. Analysis reveals that the SCYM-AfSBB difference shows no correlation with either seasonal mean temperature or seasonal total precipitation. However, correlations were found between both the SCYM simulated yields and DOY of the majority of pixels within each district ($R = -0.76$, $p = < 0.01$) and between the SCYM-AfSBB difference and the DOY of the majority of pixels within each district ($R = -0.83$, $p = < 0.001$). This results suggests that the clustering of both simulated yields and SCYM-AfSBB difference, is likely to be caused by the majority DOY of captured images within each district. The strong negative correlation suggests that districts within which the majority of images are captured later in the season, are likely to result in simulated yields closer to those detailed in the validation data. Map 10 shows the distribution of majority DOY at NUTS3 level. When compared with map 8, it is clear that similarities exist between the two. If the validation data was to be accepted as true, one could therefore conclude that SCYM is overestimating yield in the southeast of Brandenburg due to the relatively large proportion of images captured earlier in the season in these districts.



Map 9: The distribution of mean growing season temperature (left) (GLDAS-2.1 (Beaudoing, 2016; Rodell et al., 2004)) and total growing season precipitation (right) (PERSIANN-CDR (Ashouri et al., 2014; Sorooshian et al., 2014)) across Brandenburg in 2017. (Basemap: Gorelick et al., 2017*)*



Map 10: The distribution of majority DOY at district level across Brandenburg

4.7 Uncertainties and potential sources of error

With any research it is necessary to consider the uncertainties within the methods and data, and the limitations of the study. Table 9 lists the uncertainties and possible sources of error within this research. For a detailed discussion on the uncertainties and potential sources of error, see Appendix 13.

5. Conclusion

We assessed the feasibility and performance of utilising SCYM for estimating silage maize yields in the state of Brandenburg, Germany by simulating yields in 2017. We found that whilst it is possible to adapt SCYM for this purpose, the accuracy of the validation data available for the region is a hindrance to assessing SCYM's performance and further research is required in this regard before official yields can be declared using this method.

Accurate yield data is essential for monitoring food supplies and yield data at individual field scale allows a thorough assessment of the impact of management practices. The ability to estimate yields using remotely sensed data drastically improves the efficiency of yield estimation. By testing SCYM in the Brandenburg region, we have proven that the basic elements required for SCYM are available for this area and that the use of SCYM to estimate yields is feasible.

Bearing in the mind the sampling method of the currently available yield estimations for Brandenburg, it is essential that a new method for yield estimation is established. Notwithstanding the limitations previously addressed, our study has identified that SCYM could be a helpful tool for estimating yields in Brandenburg, once the issues identified in this research have been refined. These include the lack of validation data and the scale of the available soil map and gridded weather data in GEE.

Whilst previous research has concluded that SCYM can be adopted without validation data, it would nevertheless be prudent to carry out further research using highly accurate and georeferenced yield results for validation. This is considered especially necessary if SCYM is to be adopted for the generation of official yield statistics. This would enable a much more definitive conclusion on the question as to whether or not SCYM can be universally adopted for estimating yields in Brandenburg. If this additional research produces accuracies on par with SCYMs performance in other regions, there is no reason why it cannot be used to

estimate yields in Brandenburg and indeed throughout Germany, as it is likely to produce more accurate yield estimations than those currently available.

6. Acknowledgements

I would like to thank my thesis supervisors Dr. Daniel Müller and Prof. Dr. Patrick Hostert of the Geography department at Humboldt-Universität zu Berlin for their continued support throughout the project. I would also like to thank Dr. Claas Nendel, Michael Berg-Mohnicke and Dr. Tommaso Stella at the Leibniz Centre for Agricultural Landscape Research for their time, expertise and training in the use of the MONICA crop model. Finally I would like to express my profound gratitude to my wife Hannah for her invaluable support and encouragement and without whom I may never have embarked on this journey. Thank you.

Table 9: Uncertainties and potential sources of error in the methods and data

Stage	Source	Uncertainties	Possible solution
MONICA data inputs	DWD	Collection method	
		Instrument calibration	
		Station selection	
		Assumption that climate variation in Brandenburg is restricted to these sites	Increase number of stations
	Soil type	Collection method	
		Scale	Finer scale maps
	Cultivar	Generic cultivar	Include all cultivars
MONICA	LAI and yield outputs	Model uncertainty	Test alternatives
Multiple regression	Weather variables	Effects on crop growth	Test alternatives
		Seasonal window	Test alternatives
Maize field mask	InVeKoS	Farmer error	Alternative sources
		Lack of full coverage	
		Unrealistic field sizes	
		Split overlaps wrongly assigned	
Image collection	Sentinel-2 MSI: MultiSpectral Instrument, Level-1C	Instrument calibration	
		Mixed pixels	Procure finer scale images
		TOA reflectance	Procure BOA
Cloud mask	Google Earth Engine	Instrument calibration	
Calculating VI and LAI	VI	Choice of VI	Test alternatives
	Regression equations	Spectral band selection	Test alternatives
		Model uncertainty	Test alternatives
Yield simulation	Weather data	Collection method	
		Instrument calibration	
		Interpolation	
		Scale	Procure larger scale data

7. References

Amt für Statistik Berlin-Brandenburg. (2017). *Ernte- und Betriebsberichterstattungen (EBE): Feldfrüchte und Grünland (einschließlich Anbauflächen und Vorräte)*. Available at: https://www.statistik-berlin-brandenburg.de/Publikationen/metadaten/MD_41241_2016.pdf

Arroyo-mora, J. P., Kalacska, M., So, R., I, G., Leblanc, G., Schaaf, E. S., & Lucanus, O. (2018). *Evaluation of phenospectral dynamics with Sentinel-2A using a bottom-up approach in a northern ombrotrophic peatland*. Remote Sensing of Environment, 216(August 2017), 544–560. https://doi.org/10.1016/j.rse.2018.07.021

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., & Nelson, B. R. (2014). *Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies*. American Meteorogical Society, (January), 69–84. https://doi.org/10.1175/BAMS-D-13-00068.1

Atzberger, C. (2013). *Advances in Remote Sensing of Agriculture : Context Description , Existing Operational Monitoring Systems and Major Information Needs*. Remote Sensing, 949–981. https://doi.org/10.3390/rs5020949

Azzari, G., Jain, M., & Lobell, D. B. (2017). *Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries.* Remote Sensing of Environment, 202, 129–141. https://doi.org/10.1016/j.rse.2017.04.014

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M. P., Couper, M. P., Dever, J. A., … Tourangeau, R. (2013). *Summary Report of the AAPOR Task Force on Non- probability Sampling*. Journal of Survey Statistics and Methodology, (April 2015). https://doi.org/10.1093/jssam/smt008

Balkovic, J., Velde, M. Van Der, Schmid, E., Skalsky, R., Khabarov, N., Obersteiner, M., … Xiong, W. (2013). *Pan-European crop modelling with EPIC : Implementation , up-scaling and regional crop yield validation.* Agricultural Systems, *120*, 61–75. https://doi.org/10.1016/j.agsy.2013.05.008

Barthelmes, G., & Ebel, G. (2017). *Ergebnisse der LSV Mais sowie Sortenempfehlungen zur Aussaat 2017*. Landesamt für LändlicheEntwicklung, Landwirtschaft und Flurneuordnung.Available at: https://lelf.brandenburg.de/media_fast/4055/Landessortenversuche_Mais.pdf

Beaudoing, Hiroko and Rodell, M. NASA/GSFC/HSL (2016) *GLDAS Noah Land Surface Model L4 monthly 0.25 x 0.25 degree*, Version2.1. Greenbelt, Maryland, USA. Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [Via Google Earth Engine]

Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., Tucker, C., Anyamba, A., … Doorn, B. (2010). *Monitoring Global Croplands with Coarse Resolution Earth Observations: The Global Agriculture Monitoring (GLAM) Project*. Remote Sensing, 2(6), 1589–1609. https://doi.org/10.3390/rs2061589

Bommarco, R., Kleijn, D., & Potts, S. G. (2013). *Ecological intensification: harnessing ecosystem services for food security*. Trends in Ecology and Evolution, 28(4), 230–238. https://doi.org/10.1016/j.tree.2012.10.012

Bréda, N.J.J. (2008) *Lead Area Index.* In: Jørgensen, S.E. and Fath, B.D. (eds) Encyclopedia of Ecology. Elsevier Science: Unknown

Bundesamts für Justiz. (2015). *Verordnung über die Durchführung von Stützungsregelungen und des Integrierten Verwaltungs- und Kontrollsystems* (InVeKoS). Available at: http://www.gesetze-im-internet.de/invekosv_2015/

Bundesamt für Kartographie and Geodäsie (BKG) (2019). *Verwaltungszuordnungen 1:250 000 (VZ250)*. Data licence Germany. Version 2.0. https://www.bkg.bund.de/DE/Produkte-und-Services/Shop-und-Downloads/Digitale-Geodaten/Verwaltungsgebiete-Verwaltungsgrenzen/verwaltungsgebiete.html

Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) (2019a) *Bodenübersichtskarte der Bundesrepublik Deutschland 1:1.000.000*. Accessed: 22.03.2019. https://www.bgr.bund.de/DE/Themen/Boden/Projekte/Informationsgrundlagen-abgeschlossen/BUEK1000_und_Ableitungen/BUEK1000_und_Ableitungen.html

Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) (2019b) *Bodenübersichtskarte der Bundesrepublik Deutschland 1:200*. Accessed: 24.07.2019. https://www.bgr.bund.de/DE/Themen/Boden/Projekte/Informationsgrundlagen-laufend/BUEK200/BUEK200.html

Burke, M., & Lobell, D. B. (2017). *Satellite-based assessment of yield variation and its determinants in smallholder African systems*. Proceedings of the National Academy of Sciences, 114(9), 2189–2194. https://doi.org/10.1073/pnas.1616919114

Carletto, C., Jolliffe, D., & Banerjee, R. (2015). *From Tragedy to Renaissance : Improving Agricultural Data for Better Policies*. The Journal of Development Studies, 51(2), 133–148.
Clerc, S. (2019). *Sentinel-2 MPC L1C Data Quality Report*. European Space Agency. Available at: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2-data-quality-report-issue-03-may-2016-?redirect=https%3A%2F%2Fsentinel.esa.int%2Fweb%2Fsentinel%2Fuser-guides%2Fsentinel-2-msi%2Fdocument-library%3Fp_p_id%3D101_INSTANCE_Wk0TKajiISaR%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-1%26p_p_col_count%3D1
Clevers, J. G. P. W. (1997). *A Simplified Approach for Yield Prediction of Sugar Beet Based on Optical Remote Sensing Data*, Remote Sensing of Environment *228*, 221–228.
Conradt, T., Gornott, C., & Wechsung, F. (2016). *Extending and improving regionalized winter wheat and silage maize yield regression models for Germany : Enhancing the predictive skill by panel definition through cluster analysis*. Agricultural and Forest Meteorology, 2*16*, 68–81. https://doi.org/10.1016/j.agrformet.2015.10.003
Deutsche Wetterdienst (DWD) (2019a) Climate Data Center (CDC): *Daily station observations of mean temperature at 2 m above ground in °C for Germany*. Version v18.3 & recent, last accessed: 22.01.2019.
Deutsche Wetterdienst (DWD) (2019b) Climate Data Center (CDC): *Daily station observations of minimal temperature at 2 m above ground in °C for Germany*. Version v18.3 & recent, last accessed: 22.01.2019.
Deutsche Wetterdienst (DWD) (2019c) Climate Data Center (CDC): *Daily station observations of maximal temperature at 2 m above ground in °C for Germany*. Version v18.3 & recent, last accessed: 22.01.2019.
Deutsche Wetterdienst (DWD) (2019d) Climate Data Center (CDC): *Daily station observations of mean wind speed in m/s for Germany*. Version v18.3 & recent, last accessed: 22.01.2019.
Deutsche Wetterdienst (DWD) (2019e) Climate Data Center (CDC): *Daily station observations of sunshine duration in hours for Germany*. Version v18.3 & recent, last accessed: 22.01.2019.
Deutsche Wetterdienst (DWD) (2019f) Climate Data Center (CDC): *Daily station observations precipitation height in mm for Germany*. Version v18.3 & recent, last accessed: 22.01.2019.
Deutsche Wetterdienst (DWD) (2019g) Climate Data Center (CDC): *Hourly station observations of relative humidity in % for Germany*. Version v18.3 & recent, last accessed: 07.02.2019.
Deutsche Wetterdienst (DWD) (2019h) Climate Data Center (CDC): *Phenological observations of crops from sowing to harvest (annual reporters, recent)*. Version: recent 2018, last accessed: 12.03.19.
Deutsche Wetterdienst (DWD) (2019i) Climate Data Center (CDC): *Calculated daily values for different characteristic elements of soil and crops*. Version V0.x, 2015, last accessed: 22.03.19
Ebel, G., & Barthelmes, G. (2019). *Ergebnisse der LSV Mais sowie Sortenempfehlungen zur Aussaat 2019*. Landesamt für LändlicheEntwicklung, Landwirtschaft und Flurneuordnung. Available at: https://lelf.brandenburg.de/media_fast/4055/2018_12_06-LSV%20Mais.pdf
European Commission (EC) (n.d., a) *Integrated Administration and Control System (IACS)* [Online] Accessed: 01.07.2019 https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/income-support/controls-and-transparency/managing-payments_en
European Commission (EC) (n.d., b). *Factsheet on 2014-2020 Rural Development Programme for the German Länder Berlin and Brandenburg*. Available at: https://ec.europa.eu/agriculture/sites/agriculture/files/rural-development-2014-2020/country-files/de/factsheet-berlin-and-brandenburg_en.pdf
European Commission (2019a). *European Union Agricultural Statistical Factsheet*. Available at: https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/farming/documents/agri-statistical-factsheet-eu_en.pdf
European Commission (2019b). *Germany Agricultural Statistical Factsheet*. Available at: https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/farming/documents/agri-statistical-factsheet-de_en.pdf
European Space Agency (ESA). (2012). *MERIS Frequently Asked Questions. Available at: https://earth.esa.int/documents/700255/0/MERIS+Frequently+Asked+Questions%2C%20version+1.1/da28e55b-925b-4eff-89b1-915b308badd7*
European Space Agency (ESA). (2015). *Sentinel-2 User Handbook*.Available at: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook
European Space Agency (ESA) (2019) *Sentinel-2 MSI Technical Guide, Products and Processes, Level-1C Processing, Cloud Masks* [Online] Accessed: 26.07.2019, Available at: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks

Eurostat (2018) F*arms and farmland in the European Union – statistics* [Online] Accessed 02.08.2019. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Farms_and_farmland_in_the_European_Union_-_statistics#Farmland_in_2016

Fang, H., & Liang, S. (2008). *Leaf Area Index Models*. In: Jørgensen, S.E. and Fath, B.D. (eds) Encyclopedia of Ecology. Elsevier Science: Unknown

Foley, J. A. (2005). *Global Consequences of Land Use*. Science, 309(5734), 570–574. https://doi.org/10.1126/science.1111772

Food and Agricultural Organisation of the United Nations (FAO) (n.d. a) *Agricultural Intensification* [Online] Accessed: 03.09.19 www.fao.org/3/j0902e/j0902e03.htm

Food and Agricultural Organisation of the United Nations (FAO) (n.d. b) *The Integrated Administration and Control System* [Online] Accessed: 01.07.19 http://www.fao.org/3/A0464E09.htm

Food and Agriculture Organization of the United Nations (FAO). (2017). *The future of food and agriculture - Trends and challenges*. Available at: http://www.fao.org/3/a-i6583e.pdf

Gheysari, M., Majid, S., Bannayan, M., Homaee, M., & Hoogenboom, G. (2009). *Interaction of water and nitrogen on maize grown for silage*. Agricultural water management, 96, 809–821. https://doi.org/10.1016/j.agwat.2008.11.003

Gitelson, A.A., and Merzlyak, M.N. (1994) *Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves: Spectral features and relation to chlorophyll estimation*. J. Plant Physiol. 143:286–292. doi:10.1016/S0176-1617(11)81633-0

Google Developers (2019a) *Earth Engine Data Catalog, Sentinel-2 MSI: MultiSpectral Instrument, Level-1C* [Online] Available at: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2

Google Developers (2019b) *Earth Engine Data Catalog, Sentinel-2 MSI: MultiSpectral Instrument, Level-2A* [Online] Available at: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). *Google Earth Engine : Planetary-scale geospatial analysis for everyone*. Remote Sensing of Environment, 202, 18-27. https://doi.org/10.1016/j.rse.2017.06.031

Gottschalk, P., Lüttger, A., Huang, S., Leppelt, T., & Wechsung, F. (2018). *Evaluation of crop yield simulations of an eco-hydrological model at different scales for Germany*. Field Crops Research, 228(September 2017), 48–59. https://doi.org/10.1016/j.fcr.2018.07.013

Grundmann, P., Ehlers, M. H., & Uckert, G. (2012). *Responses of agricultural bioenergy sectors in Brandenburg (Germany) to climate, economic and legal changes*: *An application of Holling's adaptive cycle*. Energy Policy, 48, 118–129. https://doi.org/10.1016/j.enpol.2012.04.051

Gutzler, C., Helming, K., Balla, D., Dannowski, R., Deumlich, D., Glemnitz, M., … Zander, P. (2015). *Agricultural land use changes - A scenario-based sustainability impact assessment for Brandenburg, Germany*. Ecological Indicators, 48, 505–517. https://doi.org/10.1016/j.ecolind.2014.09.004

Hijmans, R.J., (2018) *raster: Geographic Data Analysis and Modeling*. R package. Version 2.8-4. https://CRAN.R-project.org/package=raster

Jin, Z., Azzari, G., & Lobell, D. B. (2017). *Improving the accuracy of satellite-based high-resolution yield estimation : A test of multiple scalable approaches*. Agricultural and Forest Meteorology, 247(July), 207–220. https://doi.org/10.1016/j.agrformet.2017.08.001

Jin, Z., Azzari, G., You, C., Di, S., Aston, S., Burke, M., & Lobell, D. B. (2019). *Smallholder maize area and yield mapping at national scales with Google Earth Engine*. Remote Sensing of Environment, 228(September 2018), 115–128. https://doi.org/10.1016/j.rse.2019.04.016

Kuemmerle, T., Erb, K., Meyfroidt, P., Mu, D., Verburg, P. H., Estel, S., … Reenberg, A. (2013). *Challenges and opportunities in mapping land use intensity globally.* Current Opinion in Environmental Sustainability, 2013(5), 484–493. https://doi.org/10.1016/j.cosust.2013.06.002

Kuhn, T. (2017). *The revision of the German Fertiliser Ordinance in 2017* (Discussion Paper 2017:2). Institute forFood and Resource EconomicsUniversity of Bonn. Unpblished. Available at: http://www.ilr.uni-bonn.de/agpo/publ/dispap/download/dispap17_02.pdf

Landesvermessung und Geobasisinformation Brandenburg (LGB) (2017). *Agrarantragsdaten (InVeKoS)*. Available at: https://geobroker.geobasis-bb.de/gbss.php?MODE=GetProductInformation&PRODUCTID=996f8fd1-c662-4975-b680-3b611fcb5d1f

Landesvermessung und Geobasisinformation Brandenburg (LGB) (2019) *Digitales Feldblock Kataster*. [Online] https://geobroker.geobasis-bb.de/gbss.php?MODE=GetProductInformation&PRODUCTID=9e95f21f-4ecf-4682-9a44-e5f7609f6fa0

Lobell, D. B. (2013). *The use of satellite data for crop yield gap analysis*. Field Crops Research, 143, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008

Lobell, D. B., Asnera, G. P., Ortiz-Monasterio, J. I., & Benning, T. L. (2003). *Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties.* Agriculture, Ecosystems and Environment, 94, 205–220. https://doi.org/https://doi.org/10.1016/S0167-8809(02)00021-X

Lobell, D. B., & Azzari, G. (2017). *Satellite detection of rising maize yield heterogeneity in the U.S. Midwest*. Environmental Research Letters, 12(1). https://doi.org/10.1088/1748-9326/aa5371

Lobell, D. B., Cassman, K. G., & Field, C. B. (2009). *Crop Yield Gaps : Their Importance , Magnitudes , and Causes.* Annual Review of Environment and Resources. https://doi.org/10.1146/annurev.environ.041008.093740

Lobell, D. B., Ortiz-Monasterio, J. I., Asner, G. P., Naylor, R. L., & Falcon, W. P. (2005). *Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape*. Agronomy Journal, 97(1), 241–249. https://doi.org/10.2134/agronj2005.0241

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). *A scalable satellite-based crop yield mapper*. Remote Sensing of Environment, 164, 324–333. https://doi.org/10.1016/j.rse.2015.04.021

Luetter, A., Dittmann, B., & Sourell, H. (2005). *Leitfaden zur Beregnung landwirtschaftlicher Kulturen*. Landesamt für Verbraucherschutz,Landwirtschaft und Flurneuordnung. Available at: https://lelf.brandenburg.de/media_fast/4055/Leitfaden%20zu%20Beregnung.pdf

Ministerium für Ländliche Entwicklung, U. und L. des L. B. (MLUL). (2016). *Klimareport Brandenburg 2016 Das Klima von gestern , heute und in Zukunft*. Available at: https://lfu.brandenburg.de/cms/media.php/lbm1.a.3310.de/fb_150.pdf

Moulin, S., Bondeau, A., & Delecolle, R. (1998). *Combining agricultural crop models and satellite observations: From field to regional scales*. International Journal of Remote Sensing, 19(6), 1021–1036. https://doi.org/10.1080/014311698215586

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012). *Closing yield gaps through nutrient and water management.* Nature, 1–5. https://doi.org/10.1038/nature11420

Mulla, D. J. (2012). *Twenty five years of remote sensing in precision agriculture : Key advances and remaining knowledge gap5.* Biosystems Engineering, 114(4), 358–371. https://doi.org/10.1016/j.biosystemseng.2012.08.009

Nendel, C. (2010). *MONICA - Modelling effects of climate change on crop production and environment within an interactive knowledge platform*. In *Agro 2010, the XI ESA Congress, Montpellier*.

Nendel, C. (2014) *MONICA: A Simulation Model for Nitrogen and Carbon Dynamics in Agro-Ecosystems*. In: Mueller, M., Saparov, A., Lischeid, G., (Eds) Novel Measurement and Assessment Tools for Monitoring and Management of Land and Water Resources in Agricultural Landscapes of Central Asia. Springer: Switzerland

Nendel, C., Specka, Z., Berg-Mohnicke, M., Stella, T. (2014) *The Model for Nitrogen and Carbon in Agro-Ecosystems (MONICA)*. Institue of Landscape Systems Analysis, Leibniz-Zentrum für Agrarlandschaftsforschung e. V. (ZALF). Version 2.2.1 [Computer Software] Available at: https://github.com/zalf-rpm/monica

Nguy-robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., & Rundquist, D. (2012). *Green Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to Achieve Maximal Sensitivity*. Agronomy Journal, (104), 1336–1347. https://doi.org/10.2134/agronj2012.0065

Onyia, N. N., & Balzter, H. (2018). *Normalized Difference Vegetation Vigour Index : A New Remote Sensing Approach to Biodiversity Monitoring in Oil Polluted Regions*. Remote Sensing, (10). https://doi.org/10.3390/rs10060897

Pfeiffer, D., & Thra, D. (2018). *One Century of Bioenergy in Germany : Wildcard and Advanced Technology*. Chemie Ingenieur Technik, (11), 1676–1698. https://doi.org/10.1002/cite.201800154

QGIS Development Team (2019) *QGIS Geographic Information System*. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Version: 1.1.456 [Computer software] Vienna: Austria, https://www.R-project.org

Ramirez-cabral, N. Y. Z., Kumar, L., & Shabani, F. (2017). *Global alterations in areas of suitability for maize production from climate change and using a mechanistic species distribution model (CLIMEX)*. Scientific Reports, (February), 1–13. https://doi.org/10.1038/s41598-017-05804-0

Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A. (2012). *Recent patterns of crop yield growth and stagnation*. Nature Communications, 3:1293, DOI: 10.1038/ncomms2296

Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). *Using Low Resolution Satellite Imagery for Yield Prediction.* Remote Sensing, 1704–1733. https://doi.org/10.3390/rs5041704

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll (2004) *The Global Land Data Assimilation System*. Bull. Amer. Meteor. Soc., 85(3), 381-394, 2004

Sánchez-ruiz, S., Moreno-martínez, Á., Izquierdo-verdiguier, E., Chiesi, M., Maselli, F., & Amparo, M. (2019). *Growing stock volume from multi-temporal landsat imagery through google earth engine.* Int J Appl Earth Obs Geoinformation, 83(February), 101913. https://doi.org/10.1016/j.jag.2019.101913

Sehgal, V. K., Sastri, C. V. S., Kalra, N., & Dadhwal, V. K. (2005). *Farm-level yield mapping for Precision Crop Management by linking remote sensing inputs and a crop simulation model*. Journal of the Indian Society of Remote Sensing, 33(1), 131–136. https://doi.org/10.1007/BF02990002

Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., … Major, D. J. (2001). *Use of Remote-Sensing Imagery to Estimate Corn Grain Yield Agronomy* – Faculty Publications Use of Remote-Sensing Imagery to Estimate Corn Grain Yield, (May), 583–589. https://doi.org/10.2134/agronj2001.933583x

Shelestov, A., Lavreniuk, M., Kussul, N., & Novikov, A. (2017). E*xploring Google Earth Engine Platform for Big Data Processing : Classification of Multi-Temporal Satellite Imagery for Crop Mapping*. Frontiers in Earth Science, 5(February), 1–10. https://doi.org/10.3389/feart.2017.00017

Sibley, A. M., Grassini, P., Thomas, N. E., Cassman, K. G., & Lobell, D. B. (2014). *Testing remote sensing approaches for assessing yield variability among maize fields*. Agronomy Journal, 106(1), 24–32. https://doi.org/10.2134/agronj2013.0314

Sorooshian, Soroosh; Hsu, Kuolin; Braithwaite, Dan; Ashouri, Hamed; and NOAA CDR Program (2014) *NOAA Climate Data Record (CDR) of Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN-CDR)*. Version 1 Revision 1. [2.2]. NOAA National Centers for Environmental Information. doi: https://doi.org/10.7289/V51V5BWQ [02.07.2019]

Specka, X., Nendel, C., & Wieland, R. (2015). *Analysing the parameter sensitivity of the agro-ecosystem model MONICA for different crops*. European Journal of Agronomy, 71, 73–87. https://doi.org/10.1016/j.eja.2015.08.004

Stella, T. (2019) *Silage-maize parameters* [Personal email])

Tilman, D., Fargione, J., Wolff, B., Antonio, C. D., Dobson, A., Howarth, R., … Swackhamer, D. (2001). *Forecasting Agriculturally Driven Global Environmental Change*. Science, 292(5515), 281–284.

Troegel, V. T., & Schulz, C. (2018). *Ergebnisse der Agrarstrukturerhebung 2016 für das Land Brandenburg*. Zeitschrift Für Amtliche Statistik Berlin Brandenburg, 2014(571), 44–60. Available at: https://www.statistik-berlin-brandenburg.de/publikationen/aufsaetze/2018/HZ_201801-08.pdf

van Keulen H, Penning de Vries FWT, Drees EM (1982) *A summary model for crop growth*. In: Penning de Vries FWT and van Laar HH (eds) Simulation of plant growth and crop production. PUDOC, Wageningen, pp 87–97

van Keulen H. (2013). *Simulation Models as Tools for CropManagement.* In: Christou P., Savin R., Costa-Pierce B.A., Misztal I., Whitelaw C.B.A. (eds) Sustainable Food Production. Springer, New York, NY

Vuolo, F., Zółtak, M., Pipitone, C., Zappa, L., Wenng, H., Immitzer, M., … Atzberger, C. (2017). *Data Service Platform for Sentinel-2 Surface Reflectance and Value-Added Products : System Use and Examples*. Remote Sensing, (June 2015). https://doi.org/10.3390/rs8110938

Wallach, D., & Thorburn, P. J. (2017). *Estimating uncertainty in crop model predictions : Current situation and future prospects*. European Journal of Agronomy, 88, 1–7. https://doi.org/10.1016/j.eja.2017.06.001

Wu, Q., Lane, C. R., Li, X., Zhao, K., Zhou, Y., Clinton, N., … Lang, M. W. (2019). *Integrating LiDAR data and multi-temporal aerial imagery to map wetland inundation dynamics using Google Earth Engine.* Remote Sensing of Environment, 228(September 2018), 1–13. https://doi.org/10.1016/j.rse.2019.04.015

Xie, Y., Lark, T. J., Brown, J. F., & Gibbs, H. K. (2019). *Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine*. ISPRS Journal of Photogrammetry and Remote Sensing, 155(February), 136–149. https://doi.org/10.1016/j.isprsjprs.2019.07.005

Xie, Z., Phinn, S. R., Game, E. T., Pannell, D. J., Hobbs, R. J., Briggs, P. R., & Mcdonald-madden, E. (2019). *Using Landsat observations (1988 – 2017) and Google Earth Engine to detect vegetation cover changes in rangelands - A fi rst step towards identifying degraded lands for conservation.* Remote Sensing of Environment, 232(October 2018), 111317. https://doi.org/10.1016/j.rse.2019.111317

Zeileis, Achim and Grothendieck, Gabor (2005) *zoo: S3 Infrastructure for Regular and Irregular Time Series*. Journal of Statistical Software, 14(6), 1-27. doi:10.18637/jss.v014.i06

Zhang, Z., Liu, M., Liu, X., & Zhou, G. (2018). *A New Vegetation Index Based on Multitemporal Sentinel-2 Images for Discriminating Heavy Metal Stress Levels in Rice*. Sensors, 18(2172). https://doi.org/10.3390/s18072172.

8. Appendix

8.1 Appendix 1: Detailed information for MONICA crop model

MONICA has a crop growth temporal resolution of one day and is a 1D point model with a variable depth. It does not have a spatial dimension, other than that results are interpreted as an area of $1m^2$. Crop growth is based on the SUCROS model (van Keulen et al., 1982) and is limited by water and Nitrogen stress (Nendel, 2014). Inputs required for the simulation process include management data and climate data (Specka et al., 2015) (Table 2). The user is able to specify the required outputs, including input parameters, in order to compare simulation outputs derived from a variety of inputs (Nendel, 2014). MONICA has been tested and calibrated under various growing conditions and at a variety of locations in Germany and throughout Europe (Nendel, 2014). Although primarily used in Germany, where it helps support farmer's and stakeholder's decision making on management practices, it is also used in Brazil, where the results from MONICA simulations are used in modelling land cover change and farmer's behaviour (Nendel, 2014).

8.2 Appendix 2: Detailed information on value selection for the MONICA input parameters

8.2.1 Climate data

Climate data is used by MONICA to relate crop and soil management data to weather and MONICA requires specific climate variables in order to successfully run simulations (Appendix 3). We obtained climate data from five Deutscher Wetterdienst (German weather service) (DWD) (DWD, 2019a - g) weather stations (Appendix 4 and Map 2). The data was prepared in Microsoft Excel and R statistical software (R Core Team, 2018) in the required .csv format. Relative humidity data was only available in hourly format. The mean daily relative humidity values were calculated in R (Appendix 6). A small number of missing values were found within each stations data. MONICA is not able to handle missing values within climate

data. To overcome this, we used the na.approx function from the zoo package (Zeileis and Grothendieck, 2005) in R to interpolate missing values.

8.2.2 Years

We chose 01.01.1992 – 31.12.2018 as the range of dates for the climate data used in the MONICA crop model simulations, giving a total of 27 full years of climate data. We chose these dates as 01.11.1991 was the date on which the last of each of the five chosen weather stations became automatic (Appendix 4).

8.2.3 Sowing dates

The DWD keep a record of phenological dates throughout Germany for the main arable crops, including maize. This data includes an assigned weather station ID and the date upon which phenological changes take place including sowing dates. We obtained sowing date data for stations within Brandenburg and for the years 1992 – 2018 (DWD, 2019h). We established that the dates ranged from 15th April to 15th May with a mean and a median of 26th April. In view of this analysis, we chose the sowing dates of 15th April, 30th April and 15th May (DOY 105, 120 and 135 in a non-leap year) as input parameters, in order to suitably represent the range of sowing dates within the simulations.

8.2.4 Sowing densities

MONICA does not include the possibility to adjust sowing density parameters. As such we were unable to include variations in sowing densities in the MONICA crop model simulations.

8.2.5 Fertiliser rates

Fertiliser use is not officially recorded in Brandenburg. Fertiliser application is commonly based on crop demand (Conradt et al., 2016; Gottschalk et al., 2018). The Fertiliser Ordinance 2017 states that a threshold of 170 kg N ha-1 should be applied although there is no total limit. Instead farmers are expected to plan themselves how much fertiliser they require. In 2017 a compulsory fertiliser planning method was introduced. The method takes

into account the mean yield from the previous three years and the amount of additional fertiliser allowed is dependent on the amount of N remaining in the soil from the previous crop (Kuhn, 2017). Furthermore, Balkovic et al. (2013) reported that Nitrogen stress was not a limiting factor on maize yields in Germany. We therefore used two fertiliser rates as input parameters in the MONICA crop modelling simulations: The maximum threshold of 170 kg N ha-1, as it is presumed that the majority of non-organic farmers would aim for this maximum; and a slightly lower 150 kg N ha-1 to account for some variability in application.

8.2.6 Soil moisture capacities

We obtained soil moisture data from the DWD for the period 1992 to 2018 for the five weather stations for which climate data was also used (Appendix 4) (DWD, 2019i). We established that at the depth at which maize is sown (5-10 cm), soil moisture during the sowing period ranged from 27 – 62%. From 0 – 60cm, the range was 66 – 85% (DWD, 2019i). Maize requires a starting soil moisture of around 60%. It is unlikely that farmer's would risk sowing until the soil moisture is close to the 60% required. Therefore, we chose two soil moisture capacity parameters, one at 58% representing the lower end of the required starting capacity but well above the 27% DWD range, and one at 85% representing the higher end of the scale.

8.2.7 Silage maize Cultivars

The range of silage maize cultivars used in Brandenburg is extensive. A list of recommended cultivars is produced annually and is dependent on the previous year's growing conditions and crop requirements (Barthelmes & Ebel, 2017). Selecting one or two of these cultivars to use as crop model parameters was not considered to be beneficial for this study. Instead we used a simple, generic representation of a silage maize cultivar created by ZALF and covering the range of cultivars used in Brandenburg (Stella, 2019).

8.2.8 Irrigation

Irrigation of agricultural fields is rare in Brandenburg with only 2% of agricultural fields being irrigated in 2015 representing 24,400 hectares (Gutzler et al., 2015; Luetter et al., 2005;

Troegel and Schulz, 2018). Given that so little of the land is irrigated, we chose not to include irrigation as an input parameter in the MONICA crop model simulations.

8.2.9 Soil types

In addition to those parameters used by Lobell et al. (2015), we also chose a selection of soil types for inclusion in the simulations. We obtained soil data from the Bundesanstalt für Geowissenschaften und Rohstoffe (Federal Institute for Geosciences and Natural Resources) (BGR) in the form of the 'Bodenübersichtskarte der Bundesrepublik Deutschland 1:1.000.000' (Soil map of Germany) (BGR, 2019a). The data is provided in the form of a shapefile. By comparing the soil map of Germany with the silage maize fields in Brandenburg (LGB, 2017), we established that 20 different varieties of soil were present within silage maize fields in Brandenburg in 2017. We included these soil types in the MONICA crop model simulations (See Appendix 5 for a full list of the soil types included).

8.3 Appendix 3: The required daily climate data inputs per weather station for MONICA crop model simulations

	Mean air temperature 2m above ground surface	**Minimum air temperature 2m above ground surface**	**Maximum air temperature 2m above ground surface**	**Wind speed at 2m above ground surface**	**Sunshine hours**	**Total precipitation**	**Relative humidity**
Unit	°C	°C	°C	m s^{-1}	h	mm	%
Column header	tavg	tmin	tmax	wind	sunhours	precip	relhumid

8.4 Appendix 4: DWD weather stations within Brandenburg used for climate data in MONICA crop model simulations.

Station-index	**Station ID**	**Station name**	**Station height (m)**	**Latitude**	**Longitude**	**Automatic readings since**	**Readings since**
10291	164	Angermuende	54	53° 01'	13° 59'	01.11.1991	1947
10496	880	Cottbus	69	51° 47'	14° 19'	01.11.1991	1887
10393	3015	Lindenberg	112	52° 12'	14° 07'	28.11.1976	1947
10270	3552	Neuruppin	38	52° 54'	12° 48'	01.01.1976	1961
10379	3987	Potsdam	81	52° 23'	13° 03'	17.10.1978	1893

8.5 Appendix 5: List of soil types found within maize fields of Brandenburg in 2017 (LGB, 2017; BGR, 2019) and input into the MONICA crop model simulations.

Soil number	Soil type	Soil number	Soil type	Soil number	Soil type
6	Eutric Histosols	22	Stagnic Gleysols from boulder clay with a loamy to sandy cover	32	Eutric Cambisols / Luvic Arenosols from eutrophic sand deposits
8	Fluvisols / Gleysols from loamy to clayey fluviatile sediments	24	Stagnic Chernozems from boulder clay with a loamy to sandy cover	34	Dystric Regosols from dry dystrophic sand deposits
11	Fluvisols / Gleysols from rapidly alternating sandy to clayey fluviatile sediments	26	Dystric Podzoluvisols / Luvic Arenosols / Dystric Cambisols from sandy sediments overlying boulder clay	46	Haplic Luvisols / Eutric Podzoluvisols / Eutric Cambisols from sandy loess overlying sand or loam
12	Gleysols from sandy sediments of the ice-marginal valleys and lowlands	27	Calcaric and Umbric Regosols / Luvic Arenosols from sandy to loamy end moraine deposits (alternating patches)	57	Spodic Cambisols from acid igneous and metamorphic rocks
17	Haplic Podzols / Cambic Podzols / Gleyic Podzols from sandy fluviatile sediments	28	Spodo-Stagnic Cambisols / Stagnic Podzoluvisols from loamy to sandy deposits overlying boulder clay	70	Sealed areas in larger cities (Urbic Anthrosols)
19	Haplic Luvisols / Eutric Podzoluvisols / Stagnic Luvisols from boulder clay	29	Stagnic and Spodic Gleysols from sandy deposits overlying boulder clay	71	Soils redeposited by man and large open-cast mines (Cumulic Anthrosols)
20	Luvic Chernozems / Haplic Luvisols from boulder clay or glacial-lake sediments	31	Cambic Podzols / Spodic Arenosols from dry dystrophic sand deposits		

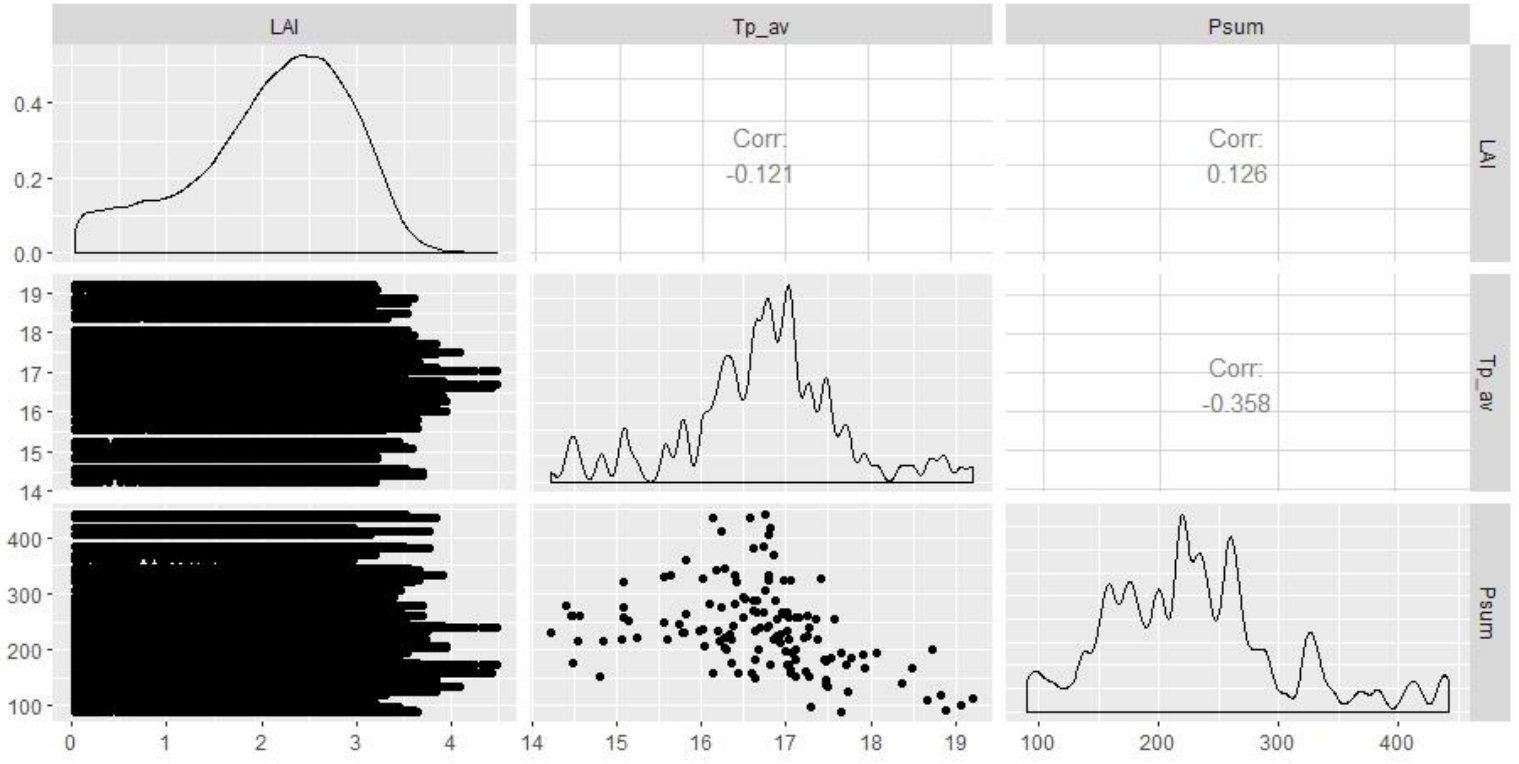
8.6 Appendix 6: R-code for data preparation and analysis

The R-code can be found on the accompanying USB drive in the folder named 'Accompanying_materials_and_data'.

8.7 Appendix 7: Example of the data structure for multiple regression training

doy	LAI	Tp_av	Psum	yield_kg_ha
221	2.84	14.49187	176.6	7376
238	0	18.66585	110.3	11960.1
210	3.18	17.1122	200.4	13345.7
151	0.31	17.06423	194.5	5035.1
241	2.79	15.23821	221.4	12367
241	2.62	17.58049	255.5	11368.3
201	3.35	16.75528	305.3	9799.6
259	0	17.47886	137.7	9289.7
210	2.64	15.98374	234.6	13520.5
208	2.7	17.4065	327.4	14343.1
234	2.51	14.80163	153.4	6990.5
234	3.29	16.38833	331.9	15376.4
238	2.1	15.81463	262.5	9561.9
226	2.84	17.36585	218.1	12601.2
151	0.13	16.13171	158.3	5732.8
241	1.86	16.63577	149.3	7714.3
218	3.33	16.81301	172.5	12766.1
253	2.83	14.49187	176.6	11887.4
189	2.78	16.79919	323.9	16392
198	0.38	17.49431	135.1	2365

8.8 Appendix 8: Overview of collinearity between variables input into the multiple regression analysis



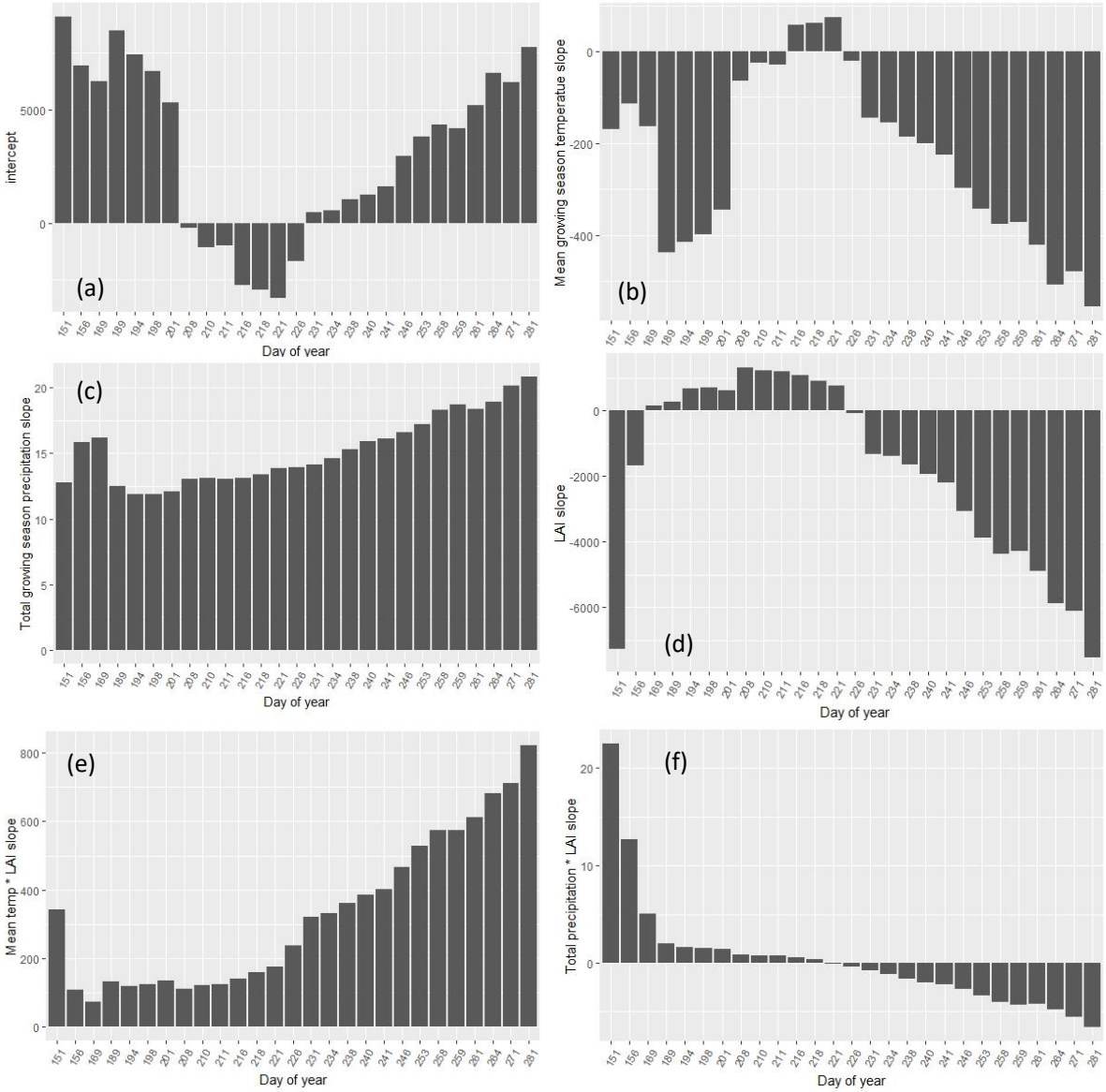
8.9 Appendix 9: The table of coefficients for each DOY

doy	intercept	Tp_av	Psum	LAI	Tp_av_LAI	Psum_LAI
151	9087.304	-168.898	12.74513	-7270.02	341.3741	22.50503
156	6925.622	-113.074	15.8403	-1682.63	108.7366	12.711
169	6227.785	-163.67	16.18621	150.0423	72.07844	5.092802
189	8466.866	-438.186	12.48869	265.477	132.0336	1.977225
194	7428.813	-415.77	11.91709	687.4397	119.0852	1.598133
198	6692.794	-399.802	11.86066	706.5794	123.91	1.508198
201	5332.754	-345.475	12.11553	628.0126	135.6106	1.402717
208	-192.761	-64.7768	13.02734	1309.499	110.8924	0.874593
210	-1057.55	-24.0444	13.09394	1236.727	120.6996	0.756451
211	-985.067	-29.8497	13.05549	1195.985	124.8229	0.732398
216	-2735.94	58.50564	13.13116	1087.442	141.2664	0.56608
218	-2940.63	62.14543	13.37215	897.852	158.4018	0.381253
221	-3296.06	74.25052	13.88409	772.1052	175.6182	-0.06765
226	-1683.66	-20.9252	13.91673	-81.1218	237.1316	-0.39855
231	476.0122	-144.755	14.10665	-1335.12	320.3559	-0.72766
234	558.1474	-154.337	14.64876	-1382.3	332.5065	-1.16214
238	1055.237	-186.993	15.31161	-1649.57	360.3186	-1.67888
240	1237.692	-201.228	15.89857	-1949.58	384.5107	-1.9906
241	1637.855	-225.827	16.1419	-2190.27	402.9807	-2.23377
246	2937.96	-297.918	16.6287	-3073.05	466.8317	-2.68829
253	3822.426	-342.294	17.2221	-3866.1	529.5404	-3.3107
258	4326.011	-376.917	18.30282	-4364.84	574.8525	-4.0303
259	4182.125	-371.882	18.74648	-4291.98	575.0189	-4.33375
261	5179.764	-422.273	18.38986	-4885.32	612.0512	-4.24843
264	6606.446	-507.647	18.92623	-5874.31	682.0363	-4.76297
271	6203.254	-479.421	20.11668	-6101.19	711.7902	-5.57136
281	7768.961	-555.974	20.82603	-7534.81	821.9727	-6.59627

8.10 Appendix 10: GEE code

The GEE-code can be found on the accompanying USB drive in the folder named 'Accompanying_materials_and_data'.

8.11 Appendix 11: The multiple regression coefficients for each DOY



8.12 Appendix 12: A detailed review of the amongst DOY coefficients

Appendix 11 shows how the coefficients differ between DOY models. Between DOY 216 and 226 higher growing season mean temperatures positively affects yield (b). Increases in temperature in the early and late growing season negatively affect yield (b). Increases in precipitation positively affect yield throughout the growing season (c). Increases in LAI at the very beginning of the season (DOY 151 and 156) negatively affect yield (d). During the peak growth period (DOY 169 to 221), increases in LAI positively affect yield (d). This reverses later in the season (DOY 226 to 281) when increases in LAI negatively affect yield (d). With regards to the variable interactions, the effect of temperature on yield increases with increasing LAI,

positively affecting yields (e). In the early season, the effect of precipitation on yield increases with increasing LAI, positively affecting yields (f). In the late season, the effect of precipitation on yield decreases with increasing LAI, negatively affecting yields.

A detailed look at the regression models for each DOY reveals that the effects of each input variable on yield varies across the growing season. We found that between DOY 216 and 221, increases in temperature positively affect yield and the effect of temperature on yield increases with increasing LAI throughout the entire growing season (b). This is to be expected as maize requires warmer temperatures and a minimum temperature above 10°C in which to grow (Ramirez-cabral et al., (2017). However the effect of increasing temperatures positively affecting yield is restricted to a very small window within the growing season. Whilst the possibility of heat stress is increased in the very early growing season, one would nevertheless expect that the window within which temperature increases positively affect yield to be larger. Whilst high temperatures do occur in Brandenburg with maximum daytime temperatures often exceeding the 25 to 33°C optimal range for maize growth (Ramirez-cabral et al., 2017), the strength of this negative impact is surprising and may require further investigation in future research.

With regards to precipitation (c), we found that an increase positively affects yield throughout the growing season. This is to be expected, as Brandenburg is a relatively dry region and maize is susceptible to drought, meaning that crops are more likely to be affected by water stress than by extreme wet conditions (Ramirez-cabral et al., 2017).

8.13 Appendix 13: A detailed explanation of the limitations and potential sources of error

The DWD data used in the MONICA crop model were collected by automatic instruments at locations across Brandenburg. The collection procedure is standardised and conforms to World Health Organisation guidelines. Nevertheless, changes in instruments, instrument height, quality control procedures, personnel and observation schedules may have affected the accuracy and comparability of the data. All data was utilised, as MONICA requires daily observation data (DWDa-I, 2019). Furthermore, the data is collected at point locations. The MONICA crop model simulations provide details on silage maize crop growth based on the climate at these locations. Whilst this method captures some of the climate

variability in Brandenburg, it does not capture all of it. This could be improved by increasing the number of stations and subsequently the variation but any improvement may be offset by additional time and computational requirements.

The soil types were selected based on the coarse scale (1:1,000,000) map produced by BGR in 1995 and by overlaying the silage maize fields. The BGR map was produced by compiling older existing maps (BGR, 2019a) but we could not establish how these existing maps were compiled. This poses a number of issues. Firstly, the accuracy of maps produced at this scale is limited. It is likely that soil type variation exist within regions declared as one particular soil type. Secondly, changes in soil type are likely to have occurred since the dates on which the original soil maps were produced. Thirdly, without knowing how the original maps were produced, it is not possible to judge their accuracy. Furthermore, when overlaying the maize fields mask in order to extract relevant soil types, scale factors produce uncertainties. At the time the modelling for this research was conducted, this was the most comprehensive soil map available for the Brandenburg region. Since then, a new soil map at a scale of 1:200 has been made available (BGR, 2019b). For future adaptations of the SCYM model in the Brandenburg region, this new soil map should be employed in order to establish soil types to be included in the MONICA crop model simulations.

The silage maize cultivar used in the MONICA crop model simulations was a general cultivar composed of average parameters found within silage maize cultivars utilised in Brandenburg (Stella, 2019). Using a generic cultivar results in the loss of detail of specific parameters within cultivars which may reduce or increase crop growth. Thus, the variability of the crop growth is reduced. Increasing the number of cultivars used may improve results, but this improvement must also be offset by additional time and computational requirements.

Whilst MONICA has been tested and calibrated on a variety of studies (Nendel, 2014), all models come with uncertainties and one can ever be certain that a model represents the true reality. Errors may exist in the structure, in the equations used within the model, in the model parameters and in the model inputs previously referred to (Wallach & Thorburn, 2017). Testing SCYM in Brandenburg by utilising alternative crop modelling software, may produce different daily LAI and yield results. Using a combination of crop models may improve results, although Jin et al. (2017) found that the benefits of this approach are minimal.

The seasonal weather variables input into the multiple regression were calculated from the DWD data. As such the same uncertainties exist in this data as those previously alluded to when discussing the DWD data. The weather variables were chosen due to their influence on crop growth, however there is no restriction on the type or number of weather variables that can be added to the multiple regression. Utilising alternative or additional weather variables or adjusting the start and end dates for calculating these variables, may result in different results. Alternative combinations could be tested in future research in order to establish whether these improve results.

The maize field mask was created from InVeKoS data. The crop data is provided by the farmers who are responsible for completing an online annual survey. The farmers are presented with a map displaying their fields as detailed in the Amtliches Liegenschaftskatasterinformationssystem (Official real estate cadastral information system) and input the crops that are present in each of their fields for the period 1st June to 15th July. They are asked to check and adjust the fields according to the reality ((Bundesamts für Justiz, 2015). Whilst systems are in place to ensure the accuracy of the information provided, the system leaves room for error, as the possibility of unintentionally inputting false information exists. Furthermore, overlaps were found to exist between fields. The overlaps were removed during pre-processing and randomly re-assigned to one of the original fields. Bearing in mind the above, it is possible that fields containing silage maize were omitted from the field mask, that fields not containing maize were included and that the area of some silage maize fields does not match the reality. An alternative method would be to create a field mask from a crop classification using remote sensing techniques, such as the random forest classification adopted by Azzari et al. (2017) and Jin et al. (2019). However it is questionable whether this approach would produce a more accurate field mask.

The collection of satellite imagery is subject to both geometric and radiometric uncertainties. The Sentinel-2 satellites are calibrated and validated to reduce uncertainties, and targets and thresholds are set internally in order to ensure that the highest possible accuracy is maintained. In terms of current levels of uncertainty, Level-1C products are measured at less than 11 m for geometric uncertainty and at less than 5% for radiometric uncertainty (Clerc, 2019). Thus, it is possible that radiometric measurements obtained at one location refer to a point up to 11m away and that measured reflectance is 5% above or below that measured by the sensor.

The Sentinel-2 bands used to calculate VI in the research have a spatial scale of 20m. Therefore each pixel covers an area of 400m^2. This introduces the problem of mixed pixels. For example, a pixel covering a field perimeter may consist of 50% crops, 30% trees and 20% road and thus produce a spectral reflectance representing this mixture of land cover. This will differ from the reflectance from a pixel representing 100% crops and thus is likely to negatively affect the correlation with validation data. This effect could be reduced by acquiring images with a higher spatial resolution or by selecting alternative bands to calculate VI (see section 2.1.5.2.2) as correlation tends to increase with higher resolution and could improve explanatory power (Burke & Lobell, 2017).

Sentinel-2 level 1C data is calibrated for top of atmosphere (TOA) reflectance (ESA, 2015). TOA reflectance is reflectance which has been corrected for variance in solar zenith angles at image acquisition and for irradiance from variance in spectral bands. However it does not account for absorption in the atmosphere or for atmospheric scattering. For this, Bottom of Atmosphere (surface) (BOA) data is required. Research suggests that using BOA data improves the accuracy of VI (Jin et al., 2019; Shelestov et al., 2017; Vuolo et al., 2017). At the time modelling was conducted for this research, BOA Sentinel-2 data (Level 2A) was not available in GEE, nor was an appropriate algorithm for converting TOA to BOA (Jin et al., 2019). However, this data recently became available in GEE (Google Developers, 2019b). Thus, future research should take advantage of this development, as it is likely to improve model accuracy.

The cloud mask algorithm utilises the existing Sentinel-2 bands to create an additional band for each pixel identifying the presence of Dense or Cirrus clouds (ESA, 2019). Despite the seemingly robust method of cloud detection, the algorithm is subject to the same geometric and radiometric uncertainties as the satellite imagery. Due to the seasonal aspect of image collection, it is highly unlikely that any snow or ice detection errors have occurred. However, errors cannot be ruled out.

The selection processes for the VI and spectral bands required to convert reflected radiance into LAI has been described in detail in section 2.2.5.2.2. The selection of the VIs and bands for this research is considered appropriate. However it is possible that selecting alternative VIs may produce more accurate results. Furthermore, there seems to be some confusion within the research community, with regard to the appropriate bands to use when calculating such VIs (See section 2.2.5.2.2). Selecting alternative bands for the VIs may

produce more accurate results, particularly if bands with a 10m spatial resolution were selected. The equations for the conversion of VI to LAI were obtained by training regression models using data acquired from experiments in the field (Nguy-robertson et al., 2012). This model was therefore subject to the same uncertainty found in all such experiments, such as instrument calibration. Whilst the model used was the best fit given the circumstances, alternative models based on alternative experiments may result in a different regression equation for the conversion of VI to LAI in maize.

The gridded weather data within GEE used to calculate seasonal weather variables is also subject to uncertainties. The GLDAS-2.1 data uses a combination of ground based and satellite observations to model temperature with data obtained from a number of sources (Beaudoing, 2016; Rodell et al., 2004). The PERSIANN-CDR data (Ashouri et al., 2014; Sorooshian et al., 2014) is derived from satellite data merged from two sources. Both of these datasets are therefore subject to errors in data collection and satellite observations previously alluded to. Furthermore, the data is provided at a relatively coarse scale with each grid cell equal to approximately 478 km^2 (0.25° = 27.82 x 17.17 km at 52° N). As such, it fails to capture more localised variations in temperature and precipitation, which may affect yield output. However, at the time of modelling, these were the finest scale weather datasets covering the study area and required dates available in GEE. If finer scale datasets become available in GEE, it would be beneficial to utilise these in future SCYM modelling in the Brandenburg region.

Finally, within all these processes there is always an element of human error, particularly when data observations are carried out manually and during the transfer of results between datasets.

ERKLÄRUNG

Ich erkläre, dass ich die vorliegende Arbeit nicht für andere Prüfungen eingereicht, selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Sämtliche fremde Quellen inklusive Internetquellen, Grafiken, Tabellen und Bilder, die ich unverändert oder abgewandelt wiedergegeben habe, habe ich als solche kenntlich gemacht. Mir ist bekannt, dass Verstöße gegen diese Grundsätze als Täuschungsversuch bzw. Täuschung geahndet werden.

Berlin, den 05.09.2019

Unterschrift