Harper Adams University

Proceedings of the 4th Symposium on Agri-Tech Economics for Sustainable Futures

20th – 21st September 2021, Harper Adams University, Newport, United Kingdom.

Global Institute for Agri-Tech Economics,

Food, Land and Agribusiness Management Department,

Harper Adams University

Proceedings of the 4th Symposium on Agri-Tech Economics for Sustainable Futures

Global Institute for Agri-Tech Economics (GIATE)
Food, Land & Agribusiness Management Department
Harper Adams University
Newport, Shropshire, United Kingdom TF10 8NB

Harper Adams University

Website: https://www.harper-adams.ac.uk/research/giate/

Symposium Website: https://www.agritechecon.co.uk/

ISBN: 978-1-7398183-2-6

Edited by K. Behrendt and D. Paparas

Published by **HAU Publications (Ebooks)**

Cover Image: Hands Free Hectare, Harper Adams University

Citation: [Authors, 2021. Title.] In: K. Behrendt and D. Paparas (eds.) *Proceedings of the 4th Symposium on Agri-Tech Economics for Sustainable Futures*. Global Institute for Agri-Tech Economics, Food, Land & Agribusiness Management Department, Harper Adams University. HAU Publications, Newport, United Kingdom, 20-21 September 2021, [pp].

All Full Papers published in this volume have been subject to single-blind peer-review administered by the proceeding's editors. Reviews have been conducted by expert referees, who have been requested to provide unbiased and constructive comments aimed, whenever possible, at improving the work. The proceeding's editors have taken all reasonable steps to ensure the quality of the materials they publish and their decision to accept or reject a paper for publication has been based only on the merits of the work and its relevance to the symposium.

Estimation of the weather-yield nexus with Artificial Neural Networks

Lorenz Schmidt, Martin Odening and Matthias Ritter

Humboldt-Universität zu Berlin, Faculty of Life Sciences, Department of Agricultural Economics, Berlin, Germany

Abstract

Weather is a pivotal factor for crop production as it is highly volatile and can hardly be controlled by farm management practices. Since there is a tendency towards increased weather extremes in the future, understanding the weather-related yield factors becomes increasingly important not only for yield prediction, but also for the design of insurance products that mitigate financial losses for farmers, but suffer from considerable basis risk. In this study, an artificial neural network is set up and calibrated to a rich set of farm-level yield data in Germany covering the period from 2003 to 2018. A nonlinear regression model, which uses rainfall, temperature, and soil moisture as explanatory variables for yield deviations, serves as a benchmark. The empirical application reveals that the gain in forecasting precision by using machine learning techniques compared with traditional estimation approaches is substantial and that the use of regionalized models and disaggregated high-resolution weather data improve the performance of artificial neural networks.

Keywords

Yield Prediction, Machine Learning, Weather Risk, Index Insurance, Basis Risk.

Presenter Profile

Lorenz Schmidt holds a master's degree in Agricultural Economics from Humboldt-Universität zu Berlin. Since August 2021, he has been a research assistant at the Department of Agricultural Economics at Humboldt-Universität zu Berlin, Germany. His research focus lies in the field of applied econometrics and machine learning.

Introduction

Understanding yield variability is essential for agricultural risk management at the sectoral as well as farm level. Crop yields depend on a variety of factors including soil and weather conditions, fertilizer, and pest control. Among these factors, weather is pivotal because, in contrast to other production factors, it is highly volatile and can hardly be controlled by farm management practices. Extreme weather events lead to harvest failures and thus threaten food security all over the world (Wheeler and Braun, 2013). Since there is a tendency towards increased weather extremes in the future, understanding the weather-related yield factors will become increasingly important not only for yield prediction, but also for the design of insurance products that mitigate financial losses for farmers. Indeed, weather-based insurance products, such as index insurance and weather derivatives, have been propagated as a promising alternative to classical crop insurance (Barnett and Mahul, 2007). Elabed *et al.* (2013) and Jensen, Mude and Barrett (2018) find that the uptake of weather index insurance products depends to a great extent on the inherent basis risk, i.e., the discrepancy between the insurant's losses and the indemnity payment which is derived from the weather index (Elabed *et al.*, 2013; Woodard and Garcia, 2008). This discrepancy can evolve from weather differences between the insurant's location and the reference station of the weather index (geographical basis risk, see for example Ritter, Mußhoff and Odening (2014)) or an imperfect correlation between crop yields and the weather index (production basis risk or design risk). The relationship between weather and crop yield, however, is complex and brings challenges to the design of appropriate weather indices for various reasons. Firstly, several weather variables must be considered simultaneously, particularly precipitation and temperature. Secondly, these variables interact in a highly nonlinear way (Schlenker and Roberts, 2009). Finally, not only the aggregated level but also the temporal distribution of weather variables affects crop yields (Musshoff, Odening and Xu, 2011).

Two general approaches have been used for modelling the weather-yield nexus. The first is crop growth models that rest on biological and physical relations and simulate the dynamics of water, nitrogen, carbon, and other yield determinants in a specific soil context considering phenological stages and plant requirements (e.g. Asseng, 2004). The second approach consists of statistical methods, particularly regressions models, which have been employed to estimate crop yields as a function of weather variables (see Section 2 for a detailed literature review). These methods are mainly data driven and do not strive for an identification of causal relations. In this paper, we focus on statistical approaches, as they are most common in the context of weather insurance. Musshoff, Odening and Xu (2011) show that a trade-off exists between the regression model's simplicity and the yield variation that cannot be explained by weather variables, i.e. basis risk. Several directions have been suggested to improve the fit of statistical yield models, including nonlinear regression or quantile regression (Conradt, Finger and Bokusheva, 2015). More recently, machine learning techniques have been applied to yield modelling (e.g. Khaki and Wang, 2019). The strength of this approach compared with traditional statistical methods arises from its flexibility in capturing complex functional relations and its capability of handling large data sets. This is particularly useful because it allows the consideration of weather variables with high temporal resolution, such as daily precipitation and temperature.

Against this backdrop, the objective of our paper is to explore the potential of machine learning for estimating the relationship between crop yield and weather conditions on a farm level and to use it as a tool for reducing basis risk in index insurance applications. More specifically, we want to investigate three hypotheses: First, we conjecture that machine learning allows a better fit to yield data compared with traditional regression models due to its flexibility. Second, we hypothesize that disaggregated weather data contain more information compared with aggregated weather variables, which allow for improving the estimation of crop yields. Third, we expect that the definition of small and homogeneous production regions eases the design of tailored weather indices and thus reduces the level of basis risk. We test these hypotheses for a large set of farm-level yields. Our data set contains 68,944 observations for winter wheat and 14,624 observations for rapeseed and in total covers many production regions in Germany over an observation period of 16 years. The use of individual farm yields avoids the underestimation of yield volatility that arises from the use of aggregated data, such as county yields (Popp, Rudstrom and Manning, 2005). To answer the aforementioned research questions, we specify an Artificial Neural Networks (ANN) and measure its performance relative to a nonlinear regression model (Hypothesis 1). Firstly, we focus on Germany as a whole and investigate the model performance for different aggregation levels of weather data, namely using monthly and daily weather data (Hypothesis 2). Subsequently, we repeat the analysis for selected homogeneous soil-climate regions within Germany (Hypothesis 3). We trace estimation errors back to particular time periods and regions. Moreover, we distinguish the viewpoint of insurers and the insured when analysing deviations between actual and predicted farm yields.

The remainder of this paper is structured as follows: Section 2 provides a literature review of standard statistical as well as machine learning approaches to estimate the weather-yield relationship; Section 3 presents details on the neural network applied in this study and introduces a regression model that is used as a benchmark; Section 4 contains the empirical application to German farm-level data; and Section 5 concludes with implications for the design of weather index insurance.

Literature Review

The estimation of the weather-yield relation by means of statistical approaches has a long tradition. Teigen and Thomas (1995) studied the relationships for US state-level yield for the period 1950–1994 and find that weather can explain 90 % of yield variation in most cases. This high percentage, however, can mostly be traced back to the time trend and not to the weather variables themselves (Vedenov and Barnett, 2004). For the application of weather derivatives to agriculture, Turvey (2001) estimates the linear dependency of county yields of corn, soybean, and hay on cumulative rainfall and cumulated degree days in Oxford County, Ontario, for the period 1935–1996, with a best fit $R^2$of 0.33. Also, in the context of weather derivatives, Vedenov and Barnett (2004) apply more complex non-linear models to estimate the relation between U.S. district-level yields in 1972–2001 and temperature and precipitation. With data-driven combinations of the weather variables and derived indices, they achieve an R^2 between 35 % and 87 %. Vroege *et al.* (2021) assess the potential of drought risk management with soil moisture data from satellites and weather stations for 89 farms in Eastern Germany. They applied quantile regression and found that the risk exposure of farmers could be reduced significantly with new insurance products based on soil moisture. Besides weather risk management, another purpose of the statistical modelling of the yield-weather relationship is the prediction of climate change impacts. Seminal papers in this context are Schlenker and Roberts (2006, 2009), who combine a county-level data set for U.S. maize yield with daily temperature observations and observe non-linear weather effects on yields; and Schlenker and Lobell (2010), who apply different specifications of the weather variables (linear, quadratic, and piece-wise linear) and find robust negative effects of climate change on agriculture in Africa. At the country-level, Lobell, Schlenker and Costa-Roberts (2011) regress yield outcomes on linear and squared monthly temperature and precipitation. It turns out that the largest share of the explained variation comes from the country-specific intercepts and the quadratic time trend rather than the weather variables. To detect spatiotemporal patterns in the yield-weather relation, Trnka *et al.* (2016) used data for ten countries and two regions in Europe over the period 1901–2012 for wheat and barley. In addition to the classical weather variables, they applied drought indicators, frost days, potential evapotranspiration, and water vapor pressure deficit, and achieved adjusted R^2 for wheat of between 0.00 and 0.71, and a normalized RMSE between 65 % and 130 % also when looking at subperiods. Nevertheless, they found an increasing influence of climatic variables in the more recent years. Bucheli, Dalhaus and Finger (2021) apply different weather indexes on a farm level yield data set in Eastern Germany and show that a tailored farm-specific drought index leads to the greatest reduction of basis risk and that no single universally best underlying drought index exists.

All of these studies show how difficult it is to explain the yield-weather relation using classical statistical approaches. Hence, a lot of hope is put in the use of machine learning and the increased computational power, which allows a more sophisticated analysis of the relationships. Van Klompenburg, Kassahun and Catal (2020) conducted a systematic literature review and identified 50 studies since 2008 that used machine learning for crop yield modelling. Explanatory variables are mostly related to weather, but also other features such as field management or nutrients. For example, Matsumara *et al.* (2015) predicted the maize yield in Chilin province, China, based on weather variables and fertilizer usage, using a multi-layer perceptron with one hidden layer and compared the results with those of a linear regression model. The artificial neural network clearly outperformed the linear regression model, and the predictive performance could mainly be traced back to fertiliser use and not to weather variables. Jeong *et al.* (2016) applied random forests to global wheat yield grid data from 2000, U.S. county-level maize grain yield 1984–2013, and potato tuber and maize silage yield data from over 1,000 points in the Northeastern U.S. in selected years. They achieved an RMSE between 6 % and 14 %, which clearly outperformed a multiple regression model (RMSE between 14 % and 49 %). Also, with random forests, Everingham *et al.* (2016) aimed to predict regional sugarcane yields at Tully, Australia, at different time points up to a year before harvest to optimise fertiliser usage. The shorter the forecast horizon, the more important variables such as rainfall and temperature range became, and up to 79 % of the variability can be explained. Using a semiparametric version of a deep neural network, Crane-Droesch (2018) model county-level yield in the U.S. Midwest from 1979 to 2016 using daily weather variables such as precipitation, temperature, humidity, wind speed, and radiation. It turns out that while the semiparametric model performs the best (with the largest effect being a time variable), the fully nonparametric neural network performed much worse than OLS regression. In a crop modelling challenge, Khaki and Wang (2019), as one of the winning teams, achieved an RMSE of 12 % with a deep neural network when predicting the yield performance of maize hybrids at over 2,000 locations in the U.S. They find considerable effects of solar radiation, temperature, and precipitation. Some studies also use remote sensing data and derived indices such as the Normalized Difference Vegetation Index (NDVI) or Enhanced Vegetation Index (EVI) (Fernandes, Ebecken and Esquerdo, 2017; Johnson *et al.*, 2016; e.g. Pantazi *et al.*, 2016; Sun *et al.*, 2019; Wolanin *et al.*, 2020).

Applications of machine learning methods for the estimation of the yield-weather relation in Germany, however, are rare. Paudel *et al.* (2021) designed a workflow for large-scale crop yield forecasting at different steps between planting and harvesting and applied it to the Netherlands, Germany, and France. For all of Germany, they achieved a normalized RMSE of between 7 % and 17 % at the end of season, which is much larger than the corresponding values from predictions by the European Commission's MARS Crop Yield Forecasting System (MCYFS). At the county-level, Webber *et al.* (2020) combine support vector machines and process-based modelling using data on weather, soil, and crop phenology to explain yield failures. Their model, however, was not able to capture the losses in 2018, an exceptionally dry year in Central Europe (Toreti *et al.*, 2019).

It can be concluded that the use of machine learning in crop yield forecasting has continuously gained more attention in recent years and that it has the potential to reduce basis risk. However, until now, it is still not clear what kind of data and what geographical aggregation form of the region is beneficial for the use of machine learning.

Methods

This study aims to explore the weather-yield relationship with different models. Even though weather is only one of the factors explaining yield deviations, the application of weather data for estimating yield variation is advantageous in comparison to other farm related information, especially when it comes to developing risk management tools such as index-based insurances. First, weather data are available at a high-resolution and independent from farmers' specific participation in the data collection process. Therefore, with weather data in general, it is possible to provide a continuous data stream. Another major advantage is the fast availability of weather data. This is especially important if it comes to an ad-hoc projection of the current expected yield. Other data such as fertiliser use, genomic information, used capital and labour as considered in Albers, Gornott and Hüttel (2017) or Khaki and Wang (2019), cannot be used for this purpose due to its lagged availability. Finally, weather data are reported by independent weather services and cannot be influenced by the insurance holder or provider. To reduce the influence of non-weather-related factors in our yield data, we do not consider in this study the yield itself, but rather the deviation of the yield from the farm yield average. By subtracting the farm-specific mean, constant location, or farmer-specific factors influencing the yield are removed to reduce the risk of omitted variable.

For a realistic insurance application, an out-of-sample evaluation is essential. Therefore, we split the data into three subsets: training data, validation data, and test data. The training data set is used to adjust the weights and to train the models. The validation set is used to evaluate the different settings of the models and to choose the optimal hyperparameters. In the end, the out-of-sample performance of the model is evaluated based on the test set. For a more realistic scenario, the split is not done randomly but by complete years. Even if this split is not necessary for the regression model, we apply this process to ensure comparability across the models. To guarantee the independence of the data sets, the aforementioned farm yield averages are calculated based only on the training data.

Different measures are applied in this study to assess the performance of the models and their potential to reduce the basis risk of an index insurance. The main tool is the root mean squared error (RMSE), which can be used to assess the average deviation between predicted and observed values. This measure, however, is an absolute value. Thus, a comparison across different regions and crop types is only possible to a limited extent due to the different yield levels. Because of this, we use the normalized root mean squared error (nRMSE) as a second measure. This puts the RMSE in relation to the respective average yield level in the region. A drawback of both indicators is that overestimates as well as underestimates are weighted equally, although they have different implications for both the insurance holder and insurance provider. Hence, the level of basis risk is not reflected properly. From the perspective of an insurance holder such as a farmer, basis risk is defined as the probability of having a loss but not receiving compensation: $P(\text{no indemnity} \mid \text{loss})$ (Elabed *et al.*, 2013). This is the case when a negative value is observed, but a positive value is predicted. From the perspective of an insurance provider, however, the opposite is considered as basis risk: an indemnity payment despite no actual loss, $P(\text{indemnity} \mid \text{no loss})$. This is the case if a negative value is predicted, but a positive value is observed. Please note that this definition of basis risk only focuses on the presence, but not on the severity. Complementing the RMSE and the nRMSE, we use both categories of basis risk (of the insurance holder and the insurance provider) as additional metrics in the model comparison and evaluate the shares of misclassified observations as realizations of the related basis risk. By studying different ways of exploiting and aggregating the weather data, we focus on production basis risk or design risk.

Regression Model

Our regression model, which serves as a benchmark for the neural network model, is a multiple regression model. As the dependent variable for both the regression model and the ANN, we use the previously described deviation of the yield from the farm yield average in the training data measured in dt/ha. Following Vedenov and Barnett (2004) and Vroege *et al.* (2021), we use the average temperature, total precipitation, and average soil moisture as independent variables. All weather variables are calculated as monthly values for April, May, and June, which represent the growing period for winter wheat and rapeseed. As in Vedenov and Barnett (2004), we additionally apply squares and same-month interactions of these variables to allow for a non-linear relation. The regression model can be defined as follows:

$$\boldsymbol{\Delta y_{it}} = \boldsymbol{\beta_0} + \sum_{k=\text{April, May, June}} \beta_{1k}T_{kit} + \beta_{2k}P_{kit} + \beta_{3k}M_{kit} + \beta_{4k}T_{kit}^2 + \beta_{5k}P_{kit}^2 + \beta_{6k}M_{kit}^2 \\ + \boldsymbol{\beta_{7k}T_{kit}P_{kit}} + \boldsymbol{\beta_{8k}T_{kit}M_{kit}} + \boldsymbol{\beta_{9k}P_{kit}M_{kit}} + \boldsymbol{\beta_{10k}T_{kit}P_{kit}M_{kit}} + \boldsymbol{\epsilon_{it}} \quad (1)$$

where Δy_{it} denotes the yield deviation for farm i in year t and T_{kit}, P_{kit}, and M_{kit} the values of the weather variables temperature, precipitation, and soil moisture, respectively, at farm i in month k (April, May, June) of year t. The βs denote the coefficients to be estimated and ϵ_{it} the error term. To estimate the model parameters, we use the ordinary least square (OLS) method.

Artificial Neural Network

Second, we apply an artificial neural network (ANN) to estimate the weather-yield relationship based on the same dependent variable as in the regression model ANNs with at least two hidden layers are able to recreate any form of mathematical model, which is in line with the non-linear relationship between weather and crop yields (Sharma, Sharma and Athaiya, 2020). In this study, we use an ANN with one input layer, two hidden layers, and one output layer. Since we are facing a regression problem, we have one neuron in the output layer. The used layers are all fully connected layers, which means that all neurons in the previous layer are connected to all neurons in the latter one. While setting up and training an ANN, hyperparameter tuning is essential. In our study, we develop an ANN of two hidden layers and perform grid search on a search space (Table A1) with Tune as platform (Liaw *et al.*, 2018) for hyperparameter tuning. The application of grid search, as opposed to other methods such as random search, allows us to use a reproducible approach of hyperparameter tuning. This is important since we apply different machine learning models with a separate grid search for each model. To decide for the best setting of hyperparameters, the lowest RMSE on the validation set is used. This is also known as cross validation. The search space for the grid search included learning rate, batch size, and the number of neurons per hidden layer as hyperparameters. For training the model, we use stochastic gradient descent and the Adam optimizer (Kingma and Ba, 2014). To account for the non-linear relationship between weather and crop yields, we opt in line with Sharma, Sharma and Athaiya (2020) for a non-linear activation function and use the ReLU function (rectifier linear unit) $g(x) = \max(0, x)$. The activation function is used for all neurons in the layers except for the output layer. The ANN was implemented in Python using the PyTorch library and trained on a Linux engine (Paszke *et al.*, 2019). Before training the ANN, the input variables were normalized. With this it was tried to counteract overfitting and to enhance the performance of the model (Ioffe and Szegedy, 2015). This also accounts for the different dimension in the input variables.

Empirical Application

Study Region and Data

In the empirical application, we use annual yield data for winter wheat and rapeseed of German farms. Germany is a convenient study region for the effects of drought on yield since only 2.7 % of the agricultural area in Germany is irrigated (Schimmelpfennig, Anter and Heidecke, 2018). Moreover, the conditions of farmland vary largely across Germany, which allow us to study the effect of different spatial aggregation levels (Hypothesis 3). Germany is subdivided into 50 regions with comparable soil and weather conditions, so-called soil-climate-regions (SCRs), by the chambers of agriculture of the federal states and the Federal Biological Research Centre for Agriculture and Forestry. A clustering procedure was used to combine municipalities with similar characteristics in terms of soil quality, temperature, and precipitation into larger areas, which have relatively homogeneous conditions for agricultural production (Roßberg *et al.*, 2007). In addition to Germany as a whole, we will later estimate regionalized models for five selected SCRs.

Our data set consists of annual winter wheat yields from 4,309 farms and annual rapeseed yields from 914 farms in 2003–2018, measured in deciton/hectare (dt/ha). In total, the data set consists of 68,944 observations for winter wheat yields and 14,624 observations for rapeseed yield. The data were provided by a financial accounting firm and an insurance company who collected the data via a farm survey about planting areas and harvest quantity for various crops. The farms are spread across Germany with a higher density in Southern Germany. Their exact locations have been deleted for confidentiality reasons, but the municipalities in which they are located are available in the data set. To correct for outliers from inaccuracies in the data collecting process, we identify farms within the 1st percentile and the 99th percentile of yearly yield per hectare in the years from 2003–2018 and delete those farms from the data set. This is done for Germany and the SCRs individually. For both crops, the complete data sets are split by years into training data (2003–2012), validation data (2013–2015), and testing data (2016–2018).

In line with our research aim and Hypothesis 3, we first use the entire data set (Germany) and then turn to regionalized models for the three SCRs with the largest number of farms in our data set (SCR South 1, SCR South 2, and SCR South 3) as well as one SCR in north-western Germany (SCR Northwest) and one in eastern Germany (SCR East). Figure 1 shows the location of these SCRs. The descriptive statistics for the cleaned yield dataset for all of Germany and the selected SCRs are depicted in Table 1.

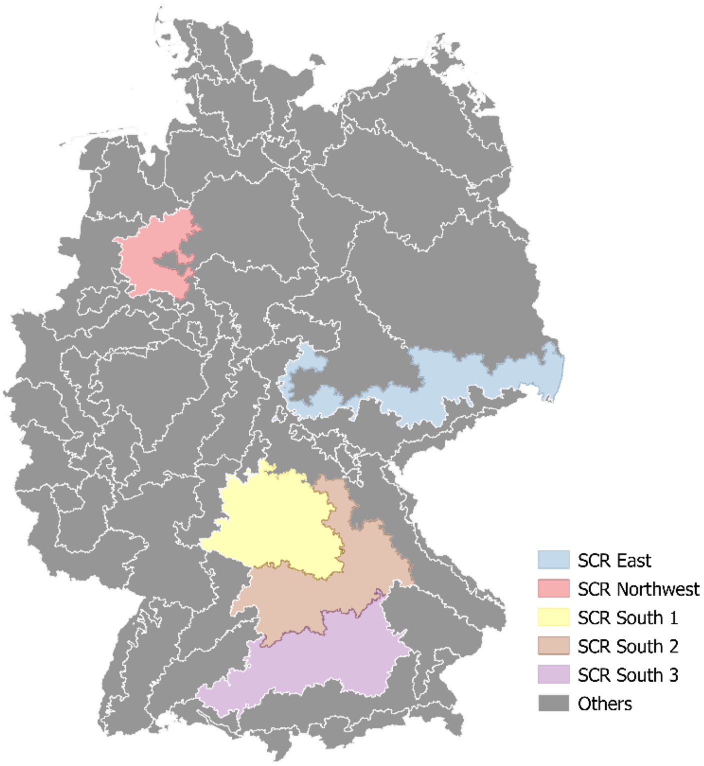

Figure 1: Soil-climate-regions (SCR) considered in this study

Table 1: Descriptive statistics of the yield data (dt/ha) for whole Germany and the considered soil-climate-regions (SCRs)

	# farms	# obs.	Mean	St. Dev.	Min.	25 %	50 %	75 %	Max.
Winter wheat									
Germany	3,344	53,504	74.20	13.57	26.59	65.86	75.09	83.02	113.58
SCR South 1	373	5,968	67.57	14.27	21.54	59.48	69.65	78.21	101.69
SCR South 2	482	7,712	73.96	12.17	32.21	66.84	74.99	81.21	114.25
SCR South 3	394	6,304	77.64	12.06	31,26	71.11	78.84	85.00	113.44
SCR Northwest	97	1,552	80.11	12.83	37.65	72.63	80.28	89.01	136.16
SCR East	7	112	79.33	13.50	51.48	70.19	79.91	89.84	106.81
Rapeseed									
Germany	698	11,168	37.79	8.71	10.44	32.72	38.47	43.49	63.59
SCR South 1	72	1,152	35.95	9.58	10.66	30.91	37.35	42.27	63.27
SCR South 2	104	1,664	37.67	8.77	10.67	32.48	38.44	43.72	63.30
SCR South 3	64	1,024	40.82	8.36	10.48	36.15	41.48	46.21	62.50
SCR Northwest	26	416	40.56	6.96	12.79	36.57	40.60	45.02	62.00
SCR East	6	96	40.90	10.95	11.17	34.45	42.55	49.98	62.47

Note: Due to the separate outlier removals in all data sets, minima/maxima of the SCRs can be smaller/larger than the ones for Germany.

The five considered SCRs comprise about 40 % of all farms in the dataset, but the number of farms per SCR varies largely with a minimum of less than ten farms for SCR East, which was included to obtain a larger regional variation. The average winter wheat yield is 74.20 dt/ha for all farms and varies between 67.57 dt/ha (SCR South 1) and 80.11 dt/ha (SCR Northwest) for the selected SCRs. Naturally, the average yield of rapeseed is 37.79 dt/ha lower than the average yield for winter wheat and there is a smaller range among the SCRs.

Weather data are provided by the Climate Data Center of *Deutscher Wetterdienst* (DWD) and contain information on daily precipitation, daily temperature, and daily soil moisture spanning the same period as the yield data (2003–2018). This detailed information allows us to feed the ANN with daily or monthly data and hence to study the effect of different temporal aggregation levels according to Hypothesis 2. Daily temperature is an average of 24-hourly values and is measured in Celsius two meters above the surface. The amount of precipitation is measured in mm. Soil moisture data are estimated by the water balanced model AMBAV (agrometeorological model to calculate the current evaporation) (Löpmeier, 1994). Since we do not know the exact locations of the farms, we connect the yield data with the weather data via the respective municipality. The DWD interpolates temperature and precipitation data coming from around 300 weather stations to a 1 km x 1 km grid based on the interpolation method by Frei (2014). Descriptive statistics for the monthly aggregated weather variables (April–June) for Germany and the selected SCRs are depicted in Figure 2. It can be observed for Germany that the conditions in 2018 were more extreme compared to 2016 and 2017. While the temperature was generally higher in 2018, median soil moisture and precipitation were lower. This development is also reflected in the selected SCRs.

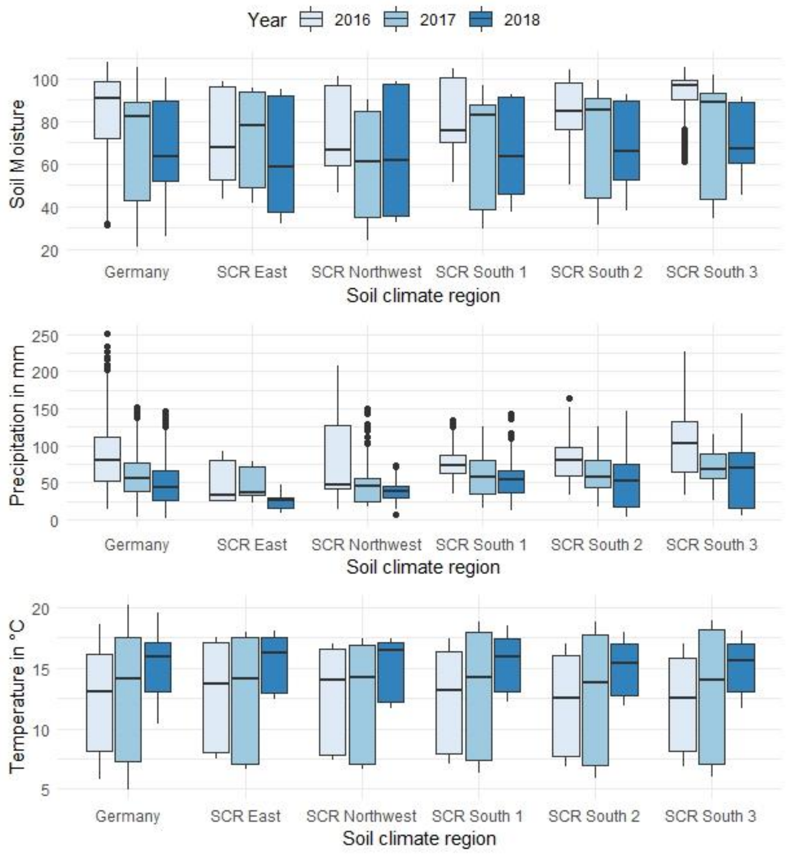

Figure 2: Monthly weather values for all farm locations (Germany) as well as for selected soil-climate-regions (SCRs)

Results

First, we consider the models for all of Germany before moving to the regionalized models. A separate grid search was performed for each model. While in some models only marginal improvements could be achieved, performance could be increased by about 40 % in other models through grid search. The best performing hyperparameter configurations are shown in Table A2. All ANNs in this study are trained with 100 iterations each and during the training process no overfitting occurred.

Addressing our first hypothesis, we first examine errors for the regression models and the ANN models for Germany as a whole and then examine the basis risk of these models. Table 2 depicts the RMSE and nRMSE for models using all farms in the data set (Germany) for the two different crop types. For winter wheat, the regression model achieves an RMSE for the testing data of 13.06 dt/ha. Compared to an average yield of 74.20 dt/ha during the entire study period, this error appears quite substantial (17.6 %). Even for the training data, the RMSE of the regression model is substantial (10.23 dt/ha or 13.8 %), which demonstrates that the regression model cannot explain a large share of the yield deviations. This finding is also reflected by an R^2 of 0.172. The daily and monthly ANN models perform better in-sample with an RMSE of 7.99 dt/ha (10.8 %) and 8.37 dt/ha (11.3 %) on the training set, respectively. However, this superiority does not hold for the test data, as the neural network with monthly data has a higher RMSE (14.44 dt/ha) than the benchmark model. The use of daily weather variables, however, reduces the RMSE to 12.38 dt/ha (16.7 %), so that it seems beneficial not to aggregate the data. Evaluating the performance of the models for the five SCRs separately reveals that the ANN with monthly data performs the worst in all southern SCRs whereas it outperforms the regression model in SCR East and SCR Northwest (Table 2). The ANN with daily data constantly performs the best, even though only with small differences in some cases. Comparing these results with other applications of machine learning models, e.g., Khaki and Wang (2019), a similar level of the RMSE (14.96 dt/ha) in the out-of-sample data can be observed.

For rapeseed, the regression model performs worse than the machine learning models. The RMSE of the test set reduces from 9.02 dt/ha for the regression model to 7.89 dt/ha for the ANN with daily data. Compared to the average yield of 37.79 dt/ha, these errors remain substantial (23.86% and 20.9% for the regression model and ANN with daily data, respectively) and are even larger compared to the nRMSE for winter wheat. Evaluating the performance of the models for the selected SCRs shows a similar picture: Except for SCR South 1 and SCR South 2 – where the RMSE remains more or less constant across models – the use of the ANN with daily data improves the results.

These first results support Hypothesis 1 that the ANN is in general better performing in comparison to the regression model. The results also support Hypothesis 2 that the use of non-aggregated data is in general beneficial.

Table 2: RMSE and nRMSE for regression and ANN models based on all farms, evaluated for whole Germany and the five selected SCRs

	Data set	Winter Wheat						Rapeseed					
		Regression Model		ANN Monthly Data		ANN Daily Data		Regression Model		ANN Monthly Data		ANN Daily Data	
Germany	Training	10.23	*13.8 %*	7.99	*10.8 %*	8.37	*11.3 %*	6.98	*18.47%*	8.57	*16.2 %*	6.07	*16.1 %*
	Validation	12.21	*16.5 %*	12.77	*17.2 %*	12.18	*16.4 %*	9.13	*24.15%*	8.94	*21.5 %*	9.62	*25.5 %*
	Testing	13.06	*17.6 %*	14.44	*19.5 %*	12.38	*16.7 %*	9.02	*23.86%*	8.62	*22.2 %*	7.89	*20.9 %*
SCR East	Testing	16.00	*20.2 %*	14.96	*18.9 %*	14.72	*18.6 %*	9.68	*23.7 %*	7.84	*19.1 %*	7.62	*18.6 %*
SCR Northwest	Testing	13.95	*17.4 %*	12.75	*15.9 %*	12.40	*15.5 %*	9.05	*22.3 %*	8.28	*20.4 %*	7.45	*18.4 %*
SCR South 1	Testing	13.35	*19.8 %*	14.24	*21.1 %*	12.44	*18.4 %*	9.01	*25.0 %*	8.86	*24.6 %*	9.00	*25.0 %*
SCR South 2	Testing	13.15	*17.8 %*	15.58	*21.1 %*	12.59	*17.0 %*	7.83	*20.7 %*	7.67	*20.3 %*	7.80	*20.7 %*
SCR South 3	Testing	12.43	*16.0 %*	14.98	*19.3 %*	12.42	*16.0 %*	8.13	*19.9 %*	8.39	*20.5 %*	7.53	*18.4 %*

To further explore the spatial variation of the forecasting power of the ANN, the RMSE of the daily model is depicted at the municipality level for both crops in Figure 3. The maps reflect the unequal distribution of the farms over Germany, their concentration in the south and northwest of Germany, and the lower number of farms with rapeseed. The RMSE shows a large range from 0.3 dt/ha to 37.2 dt/ha for winter wheat and from 1.1 dt/ha to 21.1 dt/ha for rapeseed. It seems that there are clusters with a lower RMSE and isolated municipalities with a very high RMSE. This spread of the results underlines the conclusion that the model is not performing equally across the regions. It shows large heterogeneity in model performance, which could be due to the unequal representation of the regions in the model. This finding supports our research aim to investigate whether more homogeneous regions can improve the model and thus reduce risk and improve the performance of the model.

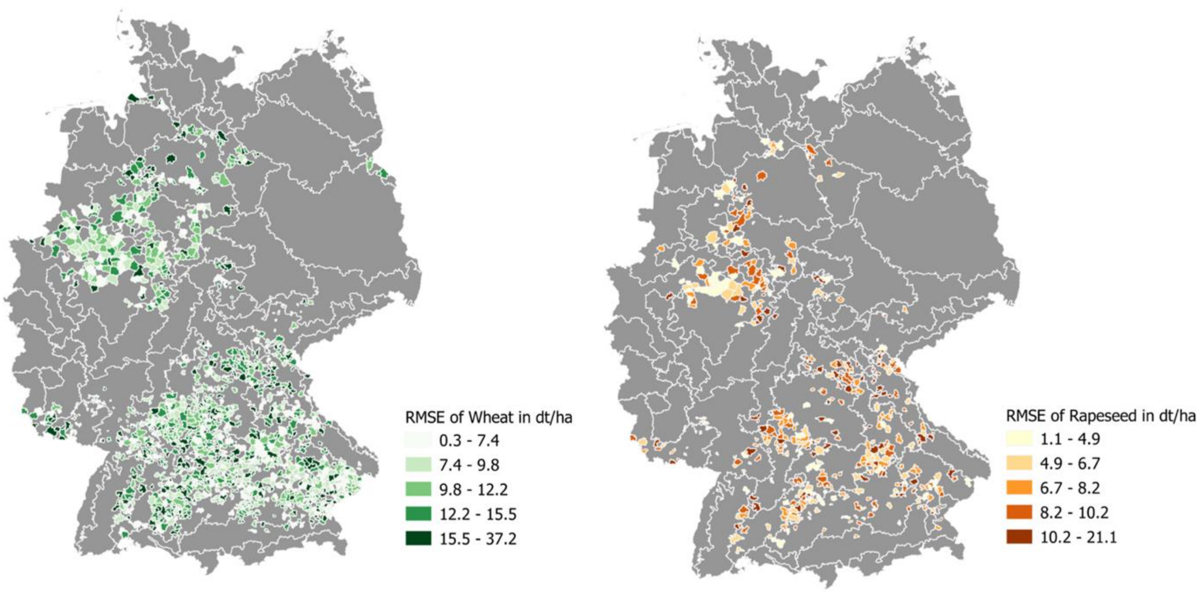

Figure 3: RMSE for test years per municipality for ANN based on daily data

To further investigate these errors, we take a closer look at the residual plots of the models for Germany (Figure 4). The variance of the predicted values is lower for the ANN, especially for the ANN with daily data, compared to the regression models. Table 3 depicts the share of observations with positive predicted but negative observed yield deviations (a disadvantage for the insurance holder) and the share of observations with negative predicted but positive observed yield deviations (a disadvantage for the insurance provider). These observations can be interpreted as realizations of the basis risk for the insurance holder and the insurance provider, respectively. The minimal share of misclassifications for the insurance holder is achieved by the ANN with monthly data for winter wheat (19.4 %) and the ANN with daily data for rapeseed (18.1 %). This supports our first hypothesis that using ANNs can improve the estimation of the weather-yield nexus.

Performance differences between years can be seen in both Figure 4 and Table 3. For the regression models, Figure 4 shows a clear separation of the years into layers. This is also confirmed by the results in Table 3, where most incorrect classifications disadvantageous to the insurance holder can be traced back to observations from 2018. In 2017, there is a small share of observations with no payout despite an observed loss that can be identified across the models and crop types (between 0 % and 12.6 %). Thus, for this year the share of misclassifications disadvantageous to the insurance holder is lowest. However, at the same time the insurance provider faces the largest share of misclassifications in 2017 (between 37.8 % and 71.1 % across models and crop types). These results demonstrate the expected asymmetric distribution of the basis risk between the insurance holder and insurance provider, which could not be seen from the (n)RMSE.

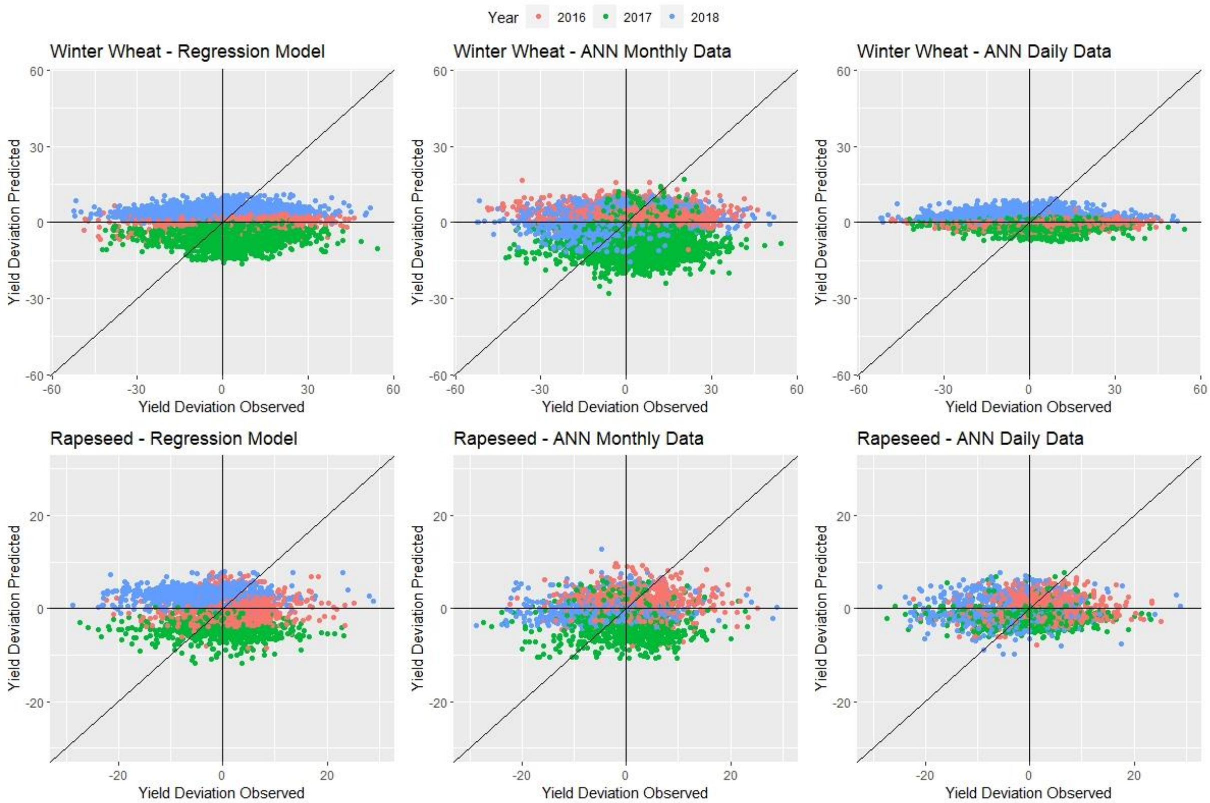

Figure 4: Residual plots for winter wheat and rapeseed for Germany

Table 3: Share of observations with (no indemnity | loss) disadvantaging the insurance holder (H) and (indemnity | no loss) disadvantaging the insurance provider (P) for winter wheat and rapeseed for Germany

	Winter Wheat						Rapeseed					
Year	Regression Model		ANN Monthly Data		ANN Daily Data		Regression Model		ANN Monthly Data		ANN Daily Data	
	H	P	H	P	H	P	H	P	H	P	H	P
2016	21.9%	22.8%	31.5%	6.6%	22.2%	28.2%	13.5%	28.2%	25.8%	19.8%	14.1%	37.8%
2017	0.0%	71.1%	1.2%	68.4%	12.6%	37.8%	0.0%	51.9%	7.5%	46.5%	7.2%	44.7%
2018	42.0%	0.6%	25.8%	19.5%	42.0%	0.9%	68.4%	0.9%	30.0%	13.8%	32.7%	14.1%
Overall	21.3%	31.6%	19.4%	31.6%	25.6%	22.3%	27.3%	27.0%	21.1%	26.7%	18.1%	32.3%

To investigate Hypothesis 3 that more homogenous regions can improve the performance of the models, we will split the data into subsets using the aforementioned SCRs and estimate separate models for each SCR. Moreover, we examine the temporal differences in the performance of the regionalized models. Due to the greater availability of yield data, we focus on winter wheat.

The results for the SCR-specific models in Table 4 strongly differ between the three southern SCRs and the other two SCRs. Regarding the regression model, the southern SCRs have an nRMSE for the test data between 15.8 % and 19.3 %. This is close to the results of the model that has been specified for the entire data set (cf. Table 2). The ANN with monthly data does not change the performance substantially, but the ANN with daily data is able to reduce the nRMSE up to 14.5 %. The latter outperforms the model based on all farms with an nRMSE between 16.0 % and 18.4 %.

On the other hand, the results for SCR East and SCR Northwest show a different picture. The nRMSE for the regression model increases to 35.4 % (SCR Northwest) and 49 % (SCR East) and for the ANN with monthly data it increases to 20.8 % and 22.3 %, respectively. These errors are much larger compared to those based on one model for all farms (between +1.9 and +28.80 percentage points). Only the ANN with daily data shows comparable results, with a clear decrease in the nRMSE for SCR East (–3.6 pp.) and a slight increase for SCR Northwest (+1.0 pp.). It turns out that estimating SCR-specific models can substantially worsen the results whereas only the NN with daily data seems to have a robust performance. By using daily weather data, the ANN has far more parameters that can be trained compared to the ANN with monthly data. Thus, the ANN with daily data can better capture certain weather events. A substantial difference between the southern SCRs and the other two is the number of farms and hence the number of observations in the data set. The southern SCRs include between 373 and 482 farms whereas the other two consist only of 97 (SCR Northwest) or even 7 farms (SCR East). Given the size of the data sets, the results may lead to the conclusion that the ANN can reduce the error by using individual models for homogeneous sub-regions (supporting Hypothesis 3), but that these regions must contain enough observations to benefit from these similarities.

Table 4: nRMSE of Winter Wheat for five SCR-specific regression and ANN models

	Data set	Regression Model	ANN Monthly Data	ANN Daily Data
SCR East	Training	16.3 %	9.5 %	10.2 %
	Validation	47.7 %	13.5 %	16.7 %
	Testing	49.0 %	22.3 %	15.3 %
SCR Northwest	Training	12.9 %	10.2 %	11.2 %
	Validation	21.8 %	18.5 %	13.8 %
	Testing	35.4 %	20.8 %	16.5 %
SCR South 1	Training	18.9 %	11.6 %	11.2 %
	Validation	21.2 %	14.3 %	14.9 %
	Testing	19.3 %	19.1 %	15.9 %
SCR South 2	Training	14.1 %	10.5 %	10.3 %
	Validation	16.2 %	15.5 %	14.6 %
	Testing	16.8 %	17.6 %	16.2 %
SCR South 3	Training	12.6 %	10.1 %	9.8 %
	Validation	16.6 %	16.1 %	13.2 %
	Testing	15.8 %	15.7 %	14.5 %

To examine the model performance over time and the influence of the drought year 2018, we compare the nRMSE for each year separately for one model for all farms (Germany) and the five SCR-specific regionalized models (Figure 5). The nRMSE for the regression model is particularly high in SCR East and SCR Northwest in 2018. From the monthly weather values in Figure 2, however, it cannot be concluded that 2018 was an exceptional year only in these regions, so that the exact reason for the high nRMSE remains unclear. The performance of the ANN based on monthly data also differs between the three years although with a smaller range. The ANN with daily data does not only lead to the smallest nRMSE, but its performance also varies little between the three years, demonstrating its robustness.

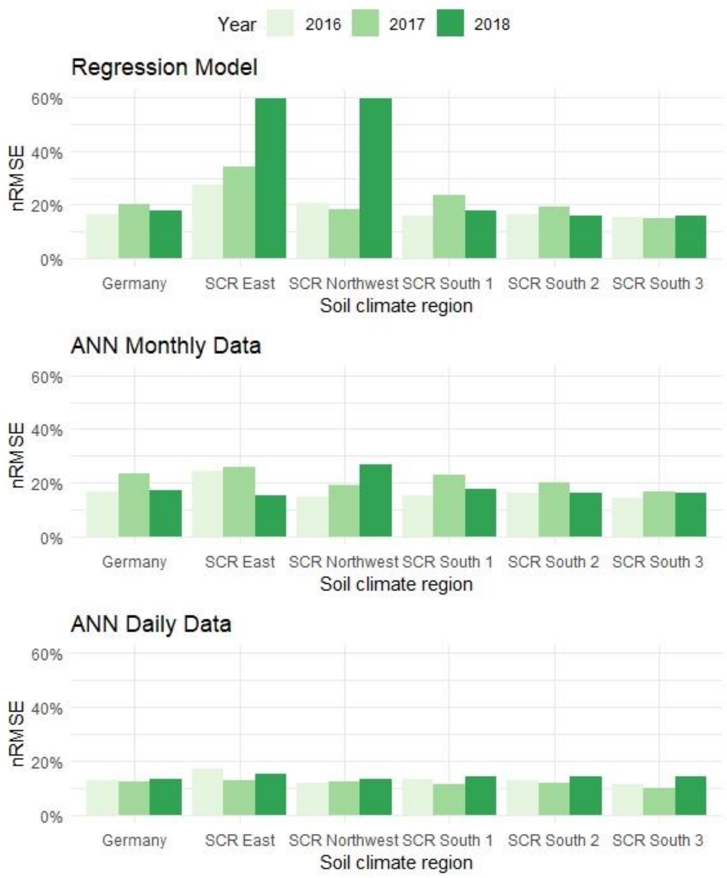

Figure 5: nRMSE by year for winter wheat of the testing set for one model for all farms (Germany) and SCR-specific regionalized models

Figure 6 depicts the share of misclassifications disadvantageous to the insurance holder and insurance provider for the regionalized models. There are two main observations. First, the total share of misclassifications is lowest for the ANN with daily data, which again seems to be more robust compared to the other models. Second, the share of misclassifications is rarely fairly distributed between the insurance holder and insurance provider – in many cases, just one side is affected. Which side is affected depends not only on the year, but also on the selected model. Compared to the results from one model for all of Germany (Table 3), it can be seen that the very high level of misclassifications disadvantageous to the insurance provider in 2017 could be reduced by using the regionalized ANN with daily data.

Figure 6: Share of misclassified observations with (no indemnity | loss) and (indemnity | no loss) for winter wheat for SCR-specific models

Conclusions

In this paper, we explore the potential of using machine learning techniques for improving the estimation of weather-induced yield losses. We specify an ANN and calibrate it to a rich set of farm-level yield data in Germany covering the period from 2003 to 2018. A nonlinear regression model, which uses rainfall, temperature, and soil moisture as explanatory variables for yield deviations, serves as a benchmark. Our empirical application reveals that the gain in estimation precision by using machine learning techniques compared with traditional estimation approaches is quite substantial. This improvement of model fit can be traced back to two sources: the flexibility inherent to ANN and the use of daily weather data instead of monthly weather data. In contrast to the common expectation that yield models can be better fitted to smaller, homogeneous regions, we find that the use of regionalized models is only beneficial if a sufficient sample size is available. From an insurance perspective, however, it is noteworthy that even for the best fitting ANN, the level of the nRMSE amounts to 14.5%. This shows that a considerable part of yield variability at the farm level cannot be captured by statistical methods which solely use "big weather data."

Our findings have important implications for the design of weather-index based insurance because they document that a rather high level of basis risk remains if insurance products are based on an estimation of the weather-yield relationship. This suggests the use of other indices, such as area yields, as an underlying index for index-based insurance. Our results, however, should be considered as a first attempt to tap the full potential of machine learning in this context. Future research should use models with flexible model structures, e.g., convolutional neural networks or locally connected layers, to better estimate the meteorological factors affecting yields. Moreover, considering basis risk explicitly in the objective function of the ANN could further improve the design of weather indices for yield insurance. Finally, we propose the application of neural networks with high-resolution data to other crops and regions to generalize the findings of our study.

References

Albers, H., Gornott, C. and Hüttel, S. (2017) 'How do inputs and weather drive wheat yield volatility? The example of Germany', *Food Policy*, 70, pp. 50–61. doi: 10.1016/j.foodpol.2017.05.001

Asseng, S. (2004) *Wheat crop systems: A simulation analysis*. Melbourne: CSIRO Publishing.

Barnett, B.J. and Mahul, O. (2007) 'Weather Index Insurance for Agriculture and Rural Areas in Lower-Income Countries', *American Journal of Agricultural Economics*, 89(5), pp. 1241–1247. doi: 10.1111/j.1467-8276.2007.01091.x

Bucheli, J., Dalhaus, T. and Finger, R. (2021) 'The optimal drought index for designing weather index insurance', *European Review of Agricultural Economics*, 48(3), pp. 573–597. doi: 10.1093/erae/jbaa014

Conradt, S., Finger, R. and Bokusheva, R. (2015) 'Tailored to the extremes: Quantile regression for index-based insurance contract design', *Agricultural Economics*, 46(4), pp. 537–547. doi: 10.1111/agec.12180

Crane-Droesch, A. (2018) 'Machine learning methods for crop yield prediction and climate change impact assessment in agriculture', *Environmental Research Letters*, 13(11), p. 114003. doi: 10.1088/1748-9326/aae159

Elabed, G. *et al.* (2013) 'Managing basis risk with multiscale index insurance', *Agricultural Economics*, 44(4-5), pp. 419–431. doi: 10.1111/agec.12025

Everingham, Y. *et al.* (2016) 'Accurate prediction of sugarcane yield using a random forest algorithm', *Agronomy for Sustainable Development*, 36(2). doi: 10.1007/s13593-016-0364-z

Fernandes, J.L., Ebecken, N.F.F. and Esquerdo, J.C.D.M. (2017) 'Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble', *International Journal of Remote Sensing*, 38(16), pp. 4631–4644. doi: 10.1080/01431161.2017.1325531

Frei, C. (2014) 'Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances', *International Journal of Climatology*, 34(5), pp. 1585–1605. doi: 10.1002/joc.3786

Ioffe, S. and Szegedy, C. (2015) *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. Available at: https://arxiv.org/pdf/1502.03167.

Jensen, N.D., Mude, A.G. and Barrett, C.B. (2018) 'How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern Kenya', *Food Policy*, 74, pp. 172–198. doi: 10.1016/j.foodpol.2018.01.002

Jeong, J.H. *et al.* (2016) 'Random Forests for Global and Regional Crop Yield Predictions', *PloS One*, 11(6), e0156571. doi: 10.1371/journal.pone.0156571

Johnson, M.D. *et al.* (2016) 'Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods', *Agricultural and Forest Meteorology*, 218-219, pp. 74–84. doi: 10.1016/j.agrformet.2015.11.003

Khaki, S. and Wang, L. (2019) 'Crop Yield Prediction Using Deep Neural Networks', *Frontiers in Plant Science*, 10, p. 621. doi: 10.3389/fpls.2019.00621

Kingma, D.P. and Ba, J. (2014) *Adam: A Method for Stochastic Optimization*. Available at: https://arxiv.org/pdf/1412.6980.

Liaw, R. *et al.* (2018) *Tune: A Research Platform for Distributed Model Selection and Training*. Available at: https://arxiv.org/pdf/1807.05118.

Lobell, D.B., Schlenker, W. and Costa-Roberts, J. (2011) 'Climate trends and global crop production since 1980', *Science (New York, N.Y.)*, 333(6042), pp. 616–620. doi: 10.1126/science.1204531

Löpmeier, F.-J. (1994) 'Berechnung der Bodenfeuchte und Verdunstung mittels agrarmeteorologischer Modelle', *Zeitschrift für Bewässerungswirtschaft* (29), pp. 157–167.

Matsumara, K. *et al.* (2015) 'Maize yield forecasting by linear regression and artificial neural networks in Jilin, China', *The Journal of Agricultural Science*, 153(3), pp. 399–410. doi: 10.1017/S0021859614000392

Musshoff, O., Odening, M. and Xu, W. (2011) 'Management of climate risks in agriculture–will weather derivatives permeate?' *Applied Economics*, 43(9), pp. 1067–1077. doi: 10.1080/00036840802600210

Pantazi, X.E. *et al.* (2016) 'Wheat yield prediction using machine learning and advanced sensing techniques', *Computers and Electronics in Agriculture*, 121, pp. 57–65. doi: 10.1016/j.compag.2015.11.018

Paszke, A. *et al.* (2019) *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Vancouver, Canada. Available at: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Paudel, D. *et al.* (2021) 'Machine learning for large-scale crop yield forecasting', *Agricultural Systems*, 187, p. 103016. doi: 10.1016/j.agsy.2020.103016

Popp, M., Rudstrom, M. and Manning, P. (2005) 'Spatial Yield Risk Across Region, Crop and Aggregation Method', *Canadian Journal of Agricultural Economics* (53), pp. 103–115.

Ritter, M., Mußhoff, O. and Odening, M. (2014) 'Minimizing Geographical Basis Risk of Weather Derivatives Using A Multi-Site Rainfall Model', *Computational Economics*, 44(1), pp. 67–86. doi: 10.1007/s10614-013-9410-y

Roßberg, D. *et al.* (2007) 'Definition von Boden-Klima-Räumen für die Bundesrepublik Deutschland', *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes*, Bd. 59(Nr. 7), pp. 155–161.

Schimmelpfennig, S., Anter, J. and Heidecke, C. (2018) *Bewässerung in der Landwirtschaft: Notwendigkeit der Bewässerung aus Sicht des Pflanzenbaus. Bewässerung in der Landwirtschaft*. Tagungsband zur Fachtagung am 11./12.09.2017 in Suderburg. Suderburg, 11./12.09.2017.

Schlenker, W. and Lobell, D.B. (2010) ‘Robust negative impacts of climate change on African agriculture’, *Environmental Research Letters*, 5(1), p. 14010. doi: 10.1088/1748-9326/5/1/014010

Schlenker, W. and Roberts, M.J. (2006) ‘Nonlinear Effects of Weather on Corn Yields*’, *Review of Agricultural Economics*, 28(3), pp. 391–398. doi: 10.1111/j.1467-9353.2006.00304.x

Schlenker, W. and Roberts, M.J. (2009) ‘Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(37), pp. 15594–15598. doi: 10.1073/PNAS.0906865106

Sharma, S., Sharma, S. and Athaiya, A. (2020) ‘ACTIVATION FUNCTIONS IN NEURAL NETWORKS’, *International Journal of Engineering Applied Sciences and Technology*, 04(12), pp. 310–316. doi: 10.33564/ijeast.2020.v04i12.054

Sun, J. *et al.* (2019) ‘County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model’, *Sensors (Basel, Switzerland)*, 19(20). doi: 10.3390/s19204363

Teigen, L.D. and Thomas, M. (1995) *Weather and Yield, 1950-94: Relationships, Distributions, and Data*.

Toreti, A. *et al.* (2019) ‘The Exceptional 2018 European Water Seesaw Calls for Action on Adaptation’, *Earth’s Future*, 7(6), pp. 652–663. doi: 10.1029/2019EF001170

Trnka, M. *et al.* (2016) ‘Changing regional weather-crop yield relationships across Europe between 1901 and 2012’, *Climate Research*, 70(2), pp. 195–214. doi: 10.3354/cr01426

Turvey, C.G. (2001) ‘Weather Derivatives for Specific Event Risks in Agriculture’, *Review of Agricultural Economics*, 23(2), pp. 333–351. doi: 10.1111/1467-9353.00065

van Klompenburg, T., Kassahun, A. and Catal, C. (2020) ‘Crop yield prediction using machine learning: A systematic literature review’, *Computers and Electronics in Agriculture*, 177, p. 105709. doi: 10.1016/j.compag.2020.105709

Vedenov, D.V. and Barnett, B.J. (2004) *Efficiency of Weather Derivatives as Primary Crop Insurance Instruments*.

Vroege, W. *et al.* (2021) ‘Insuring crops from space: the potential of satellite-retrieved soil moisture to reduce farmers’ drought risk exposure’, *European Review of Agricultural Economics*. doi: 10.1093/erae/jbab010

Webber, H. *et al.* (2020) ‘No perfect storm for crop yield failure in Germany’, *Environmental Research Letters*, 15(10), p. 104012. doi: 10.1088/1748-9326/aba2a4

Wheeler, T. and Braun, J. von (2013) ‘Climate change impacts on global food security’, *Science (New York, N.Y.)*, 341(6145), pp. 508–513. doi: 10.1126/science.1239402

Wolanin, A. *et al.* (2020) ‘Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt’, *Environmental Research Letters*, 15(2), p. 24019. doi: 10.1088/1748-9326/ab68ac

Woodard, J.D. and Garcia, P. (2008) ‘Basis risk and weather hedging effectiveness’, *Agricultural Finance Review*, 68(1), pp. 99–117. doi: 10.1108/00214660880001221