Exploring the heterogeneous effects of weather on productivity using generalized random forests

Christian Stetter*, Johannes Sauer

Technical University of Munich, Germany,
Department of Production and Resource Economics

Contributed Paper prepared for presentation at the 95[h] Annual Conference of the Agricultural Economics Society (online)

29 - 30 March 2021

*Corresponding author: Alte Akademie 14, 85354 Freising, Germany.
christian.stetter@tum.de

Abstract

With the advancement of climate change, temperature and precipitation patterns have been changing over the last decades. These changes continue to have substantial impacts on agricultural production activities. The question to which extent weather patterns affect the productivity of farm businesses has gained more and more attention in recent years. This study seeks to give a nuanced picture on the complex relationship between weather, farm management, biophysical conditions and total factor productivity. To this end, we utilize generalized random forests, a causal machine learning algorithm, to assess the complex weather-productivity nexus in European arable farming. We find considerable weather impacts in 27 EU member states at the NUTS3 level between 2005 and 2016. We find both positive and negative effects of aggregated weather events (mean temperature and precipitaion sum), and predominantly negative effects of drought spells, heat and heavy precipitation on productivity. Using model-agnostic Shapley values, important interactions between weather, farm management, technology, and the biophysical environment are found. Given the fact that climate change continues to change weather patterns, our results provide interesting insights as to how farm managers and legislators could locally respond to offset negative climate change impacts, e.g. by adjusting farm management characteristics or improving soil quality.

Keywords Total factor productivity, weather, machine learning, generalized random forest, arable farming

JEL code Q540, D240, C55

1 Introduction

According to the Food and Agricultural Organization of the United Nations (FAO), one of the major challenges for agriculture is to satisfy the steadily rising global need for food, fiber and bioenergy (FAOSTAT, 2017). Given this challenge, international organizations, legislators and scientists call for persistent efforts to further increase agricultural productivity. However, it seems that agricultural productivity growth has considerably slowed down in recent decades (e.g Thirtle et al., 2004; Alston, Beddow, and Pardey, 2009; Ball et al., 2010; Baráth and Fertő, 2017; Chambers and Pieralli, 2020). In the meantime, farmers around the globe are facing the threat of climate change, which might drastically modify the natural conditions under which they produce (OECD, 2019; Njuki, Bravo-Ureta, and O'Donnell, 2018). As farming activities directly depend upon climatic conditions, changing weather patterns are likely to have a direct effect on agricultural productivity, and a growing body of evidence suggests that this effect might be negative. Understanding the complex weather-productivity-nexus in the agricultural sector is therefore vitally important for meeting the rising demand of crop and livestock products.

In this paper, we look at the question if there exist heterogeneous effects of different weather patterns on total factor productivity in arable farming in the context of the European Union (EU). To do so, we exploit recent advances in the causal machine learning (ML) literature and insights from microeconomic production theory to estimate the impact of five weather indicators relevant for agriculture, namely average temperature and precipitation as well as drought spells, heat and heavy rains. We then use a model-agnostic interpretation approach to gain further insights into the nature of this relationship.

Several studies have found a significant relationship between weather-related phenomena and total factor productivity (TFP) measures (Mendelsohn, Nordhaus, and Shaw, 1994; Schlenker, Hanemann, and Fisher, 2005; Burke and Emerick, 2015; Njuki, Bravo-Ureta, and O'Donnell, 2018; Njuki, Bravo-Ureta, and Cabrera, 2020; Chambers and Pieralli, 2020; Chambers, Pieralli, and Sheng, 2020; Ortiz-Bobea et al., 2020; Ortiz-

Bobea, Knippenberg, and Chambers, 2018), profits (e.g. Deschênes and Greenstone, 2007, 2012) and partial productivity such as crop yields (e.g. Vogel et al., 2019; Webber et al., 2020; Schlenker and Roberts, 2009). However, many studies fail to recognize that the effect of weather on TFP interacts with important biophysical and management-related contextual covariates of farming activities (Hsiang, 2016). This might be especially true for EU agriculture, where the weather-TFP nexus has not been studied extensively yet. While some studies account for the potential heterogeneity of the weather-production relationship (Burke, Hsiang, and Miguel, 2015; Schlenker and Roberts, 2009; Keane and Neal, 2020; Cui, 2020), many fail to comprehensively capture its complexity. The effect of weather events on productivity are likely to vary both spatially and temporally due to different farming settings. These include the current state of technology, environmental conditions such as topography as well as characteristics closely related to the farms, e.g. size and intensity (Njuki, Bravo-Ureta, and O'Donnell, 2018). Failing to account for important interactions between weather and farming context might lead to biased estimates of the impact of weather on productivity. What is more, it is also important to address the possibility of compound weather effects, i.e. interactions among weather phenomena such as droughts and heat waves, which are likely to further go up in the future (Zscheischler et al., 2018; Aghakouchak et al., 2020).

One reason why such effects have not been studied extensively yet might be that traditional statistical and econometric techniques are quite limited in capturing complex relationships because they usually rely on rather inflexible functional form assumptions and are not well-suited for high-dimensional and highly nonlinear relationships (Storm, Baylis, and Heckelei, 2019). In an attempt to overcome these problems, several authors suggested a number of novel estimation techniques that combine predictive machine learning methods and inferential statistics (Athey and Imbens, 2016; Wager and Athey, 2018; Athey and Wager, 2019; Künzel et al., 2019; Shi, Blei, and Veitch, 2019; Chen et al., 2019).

Our article contributes to the literature in at least three ways. First, we use a novel machine learning technique, namely the generalized random forest (Athey, Tibshirani, and

Wager, 2019), to estimate the effect of weather on total factor productivity, which enables us to account for important interactions and nonlinearities within this relationship. Second, our empirical case is based on data for the entire EU at a granular (county-level or NUTS-3[1]) scale. Thus, we are able to draw conclusions for a very large territory comprising several million farms. In contrast to that, most previous studies in this context focused either on the United States or were based on case studies with a rather small regional extent. Third, we show that model-agnostic Shapley values can be used to interpret the results of our analyses, which allows us to identify and evaluate key drivers of the complex weather–TFP relationship.

The article proceeds as follows. In Section 2, we introduce the theoretical model of the study that is largely based on production economics and index number theory. Section 3 describes the data and provides summary statistics of the underlying variables. In Section 4, we describe our ML-based estimation strategy. Then, in Section 5 we present our empirical results, which is followed by a discussion (Sec. 6) and conclusion (Sec. 7)of the paper.

2 Economic Framework

2.1 Production technology

The starting point for our analysis is the period-and-environment-specific production possibilities set (O'Donnell, 2018) reflecting all feasible input-output combinations (x, q) using a given technology set in a given period (t) in a given environment (z) considering firm characteristics (c):

$$T^t(z, c) = \{(x, q) \in \mathfrak{R}_+^{M+N} : x \text{ can produce } q \text{ in environment } (z, c) \text{ in period } t\} \quad (1)$$

where the usual regularity conditions apply (Chambers, 1988). We can reformulate

[1]NUTS is a geographical system, according to which the territory of the European Union is divided into hierarchical levels, where NUTS-3 is the most granular resolution describing small regions for specific diagnoses. NUTS-3 regions generally have a population of 150,000 to 800,000 inhabitants.

expression (1) as an average production function using the most general, multiplicative form, which allows for nonlinearities and interactions in a potentially high-dimensional covariate space (Hastie, Tibshirani, and Friedman, 2009):

$$f_q(q_{it}) = f_a(a_{it}) \times f_x(x_{it}) \times f_z(z_{it}) \times f_c(c_{it}). \tag{2}$$

In this formulation, each component of (2) is a nonlinear function ($f_{\{q,a,x,z,c\}}$), where q is a vector of m outputs, a a vector of j covariates describing the state of technology, x a vector of n inputs, z a vector of k covariates describing the biophysical environment and c a vector of l farm characteristics of farm i at time t.

Dividing both sides of the equation by $f_x(x_{it})$, we obtain the following expression:

$$\frac{f_q(q_{it})}{f_x(x_{it})} = TFP_{it} = f_a(a_{it}) \times f_z(z_{it}) \times f_c(c_{it}) \tag{3}$$

where the ratio of all outputs over all inputs is defined as total factor productivity (Chambers, 1988). Thus, TFP is a function of the state of technology, the farming environment comprising factors that naturally influence the production process such as weather, other biophysical factors or topography, as well as farm (management) characteristics. Finally, we are interested in analyzing the partial effect (θ_{it}) of some weather pattern z_w on TFP, which is equivalent to the first derivative of equation (3) w.r.t. z_w:

$$\theta_{it} = \frac{\partial TFP}{\partial z_w} \tag{4}$$

2.2 Lowe TFP index

In order to define the TFP index, we follow O'Donnell (2012c) and assume that $f_q(q_{it})$ and $f_x(x_{it})$ are linear aggregator functions, which can be expressed as $f_q(q_{it}) = \mathbf{p_0'q_{it}}$ and $f_x(x_{it}) = \mathbf{w_0'x_{it}}$, where $\mathbf{p_0'}$ and $\mathbf{w_0'}$ are reference output and input prices, respectively. To allow for comparisons across different observations and time periods, the following quantity indices are constructed

$$QI_{hsit} = \frac{\mathbf{p}_0'\mathbf{q}_{\mathbf{it}}}{\mathbf{p}_0'\mathbf{q}_{\mathbf{hs}}} \quad \text{and} \quad XI_{hsit} = \frac{\mathbf{w}_0'\mathbf{x}_{\mathbf{it}}}{\mathbf{w}_0'\mathbf{x}_{\mathbf{hs}}} \tag{5}$$

such that the input and output quantities of region i in year t are compared with the input and output quantities of region k in year s. Provided expression (5), we can define the Lowe TFP index (O'Donnell, 2012b):

$$TFPI_{hsit} = \frac{QI_{hsit}}{XI_{hsit}} = \frac{\mathbf{p}_0'\mathbf{q}_{\mathbf{it}}}{\mathbf{p}_0'\mathbf{q}_{\mathbf{hs}}} \times \frac{\mathbf{w}_0'\mathbf{x}_{\mathbf{hs}}}{\mathbf{w}_0'\mathbf{x}_{\mathbf{it}}} \,. \tag{6}$$

In contrast to other popular indexes such as the Paasche, Laspeyres, or Fisher index the Lowe TFP index satisfies all relevant axioms from index number theory by using time and observation invariant reference prices (O'Donnell, 2016).

3 Data

We use farm accountancy data on specialized crop farms in the EU-28 obtained from the EU Farm Accounting Data Network (FADN) and aggregate these at the NUTS-3 (county) level. The data cover the years 2005–2016 and originally consist of a total of 185,984 observations, which reduce to 9,693 after aggregation (taking the median at the NUTS-3 level). We only included farm-level observations with a revenue share of specialized products such as wine, olives, vegetables and fruits of at most 33%. All EU-28 countries are represented including the United Kingdom but excluding overseas departments and Croatia, which only joined the EU in 2013 and for which no weather record was available. Furthermore, 0.7% of the observations with highly implausible values were deleted from the dataset. Spatial data on the chemical and physical properties of soil were retrieved from the LUCAS 2009/2012 topsoil database (Panagos et al., 2012). Weather variables are based on 0.1 degree gridded daily data obtained from the European Climate Assessment & Dataset (ECA&D) project (Cornes et al., 2018). The panel dataset is unbalanced.[2]

The Lowe productivity index is calculated using six output variables, namely cereals, protein crops, root and tuber crops, oilseeds, and other output (e.g. vegetables, fruits) as well as five input variables: crop-specific inputs (seed, fertilizers, pesticides), materials (fuel, electricity, contract work, insurance and other farming overheads), capital, labor, and land. Labor is measured in annual working hours, including both family and hired labor. Land is expressed in hectares. For these labor and land, farm-level prices are available. All other variables are expressed as costs and are deflated using agricultural price indices from EUROSTAT to the year 2015 to obtain implicit quantities (European Commission, 2020). Capital is proxied by deflated depreciation costs. The transitivity of the Lowe TFP index relies on a fixed representative price vector (O'Donnell, 2012a). The median value of real prices (2015=100) is chosen for this purpose. Summary statistics of input and output variables can be found in Table 1.

The right hand side contextual variables from eq. (3) describing the state of technology, biophysical conditions, and farm characteristics are summarized in Table 2. Weather patterns, which are of primary interest in this study, are described by five indicators. Annual mean temperature (°C) and annual precipitation sum (mm) reflect mean yearly conditions, which are expected to keep being affected by climate change. Furthermore, beside changing annual mean values, climate change also affects specific weather patterns such as the number of heavy rain events, droughts and heat (Lüttger and Feike, 2018; Westra et al., 2014). Therefore, we include the number of consecutive dry days (precipitation $<$ 1mm), heavy rain days during the agricultural season (days with rainfall $>$ 20mm from March–October) and hot days (days with max. temperature $>$ 30°C) in our analysis. Furthermore, technology is approximated by location (longitude and latitude) and a time trend. A total of 22 variables comprising weather (see above), physical and chemical soil properties as well as elevation reflect the biophysical conditions of the

[2] This is partly because there were several territorial and nomenclature reforms concerning NUTS-3 regions during the observed time period, which caused that multiple farm-level observations could not unambiguously be attributed to a NUTS-3 region (e.g. Denmark). For several regions and years, there is no information available on specialized crop farms. Finally, there were several missing values on weather for some regions.

Variable	Mean	SD	Min.	Max.
Cereals output (implicit quantity)	916.25	1298.86	0	36780.63
Protein crop output (implicit quantity)	21.35	70.93	0	1844.98
Root/tubers output (implicit quantity)	427.96	820.76	0	12020.1
Oilseeds output (implicit quantity)	63.27	177.84	0	8606
Forage crops output (implicit quantity)	44.35	112.18	0	3374.96
Other crops output (implicit quantity)	41.82	286.56	0	17129.24
Crop specific input (implicit quantity)	659	855.89	0	22154.64
Materials input (implicit quantity)	630.33	2313.28	1.78	214582.62
Capital input (implicit quantity)	319.26	401.71	0	10192.09
Labor input (annual working hours)	5281.14	7211.16	36	382804
Land input (hectares)	174.79	263.25	0.79	12047.5
Cereals output (reference price, dimensionless)	100		100	100
Protein crop output (reference price, dimensionless)	93.69		93.69	93.69
Root/tubers output (reference price, dimensionless)	99.75		99.75	99.75
Oilseeds output (reference price, dimensionless)	100		100	100
Forage crops output (reference price, dimensionless)	100		100	100
Other crops output (reference price, dimensionless)	99.49		99.49	99.49
Crop specific input (reference price, dimensionless)	98.95		98.95	98.95
Materials input (reference price, dimensionless)	100.16		100.16	100.16
Capital input (reference price, dimensionless)	100		100	100
Labor input (reference price, dimensionless)	90.08		90.08	90.08
Land input (reference price, dimensionless)	151.34		151.34	151.34

Table 1: Descriptive Statistics of farm inputs and outputs (N = 9,693).

observations. Farm and management characteristics are summarized by 17 variables, carefully derived from the production economics literature, ranging from average farm size to crop diversity to share of irrigated land.

4 Empirical approach

The empirical strategy to estimate $\frac{\partial TFP}{\partial z_w}$ is based on generalized random forests (GRF Athey, Tibshirani, and Wager, 2019), which fundamentally rest upon the random forest algorithm introduced by Breiman (2001). It belongs to the class of tree-based prediction methods (Breiman et al., 1984). Random forests are basically an ensemble of regression or classification trees (CART), which are grown based on recursive partitioning such that the covariate space is divided into binary subregions (aka nodes) according to an optimality criterion, e.g. minimizing the in-sample prediction error of one node (Breiman

Variable	Mean	SD	Min.	Max.
State of technology (A)				
Time trend	5.82	3.41	0	11
Longitude (WGS 84 coordinates)	10.61	8.65	-9.18	33.23
Latitude (WGS 84 coordinates)	48.89	5.2	35.05	67.73
Biophysical environment (Z)				
Annual mean temperature (°Celcius)	10.57	2.61	-1.84	21.87
Annual precipitation sum (mm/year)	718.29	223.62	9.28	2600.7
Hot days (Days with temperature > 30 °Celcius)	22.76	24.01	0	141.07
Longest period of consecutive dry days (Days with precipition < 1mm)	27.75	17.4	9.08	187
Heavy rain days (Days with precipition > 20mm)	2.96	2.51	0	23.6
Calcium carbonates (CACO3, mg/kg)	40.24	48.13	0.33	333.26
Cation Exchange Capacity (CEC, cmol/kg)	16.79	5.64	6	40.22
C:N ratio	12.73	2.46	9.31	25.53
Potassium (K, g/kg)	206.22	72.68	64.44	663
Nitrogen (N, g/kg)	2.15	0.54	1.12	5.24
Phosphorus (P, g/kg)	32.61	12.87	4.77	71.1
pH in CaCl2 solution	5.72	0.7	3.63	7.32
pH in water (H2O)	6.29	0.69	4.39	7.91
pH in H2O,CaCl	0.58	0.1	0.28	1.05
Available Water Capacity (AWC) for the topsoil fine earth fraction	0.09	0.02	0.05	0.13
Bulk density	1.23	0.11	0.94	1.48
Clay content (%) in topsoil (0-20cm)	19.99	7.09	3.6	40.6
Coarse fragements (%) content in topsoil	13.33	4.85	4.63	30.8
Sand content (%) in topsoil	41.08	17.23	10.84	88.65
Silt content (%) in topsoil	38.93	11.96	7.75	69
USDA soil textural classes derived from clay, silt and sand maps	9.12	1.92	3	12
Elevation (10m)	68.19	42.27	0.79	177.13
Farm (management) characteristics (C)				
Environmental constraints Area	0.74	0.49	0	2
Economic size class (1=very small,..., 14=very large)	7	1.52	2	14
Share family labor	0.81	0.22	0	1
Share rented land	0.56	0.27	0	1
Cashflow to capital ratio	0.14	1.07	-2.91	88.75
Share of total subsidies in total farm income	0.52	0.21	0	1
Share of decoupled subsidies in total subsidies	0.77	0.21	0	1
Share of environmental subsidies in total subsidies	0.05	0.08	0	0.95
Insurance intensity - share of insurance in total cost	0.03	0.02	0	0.25
Crop diversity (Shannon index)	1.34	0.29	0	2.16
Capital intensity (€/ha)	2.9	17.88	0	1739.27
Labor intensity (€/ha)	79.56	148.93	0.27	4692.37
Fertilizer intensity (€100/ha)	1.72	0.89	0	14.8
Chemicals intensity (€100/ha)	1.26	0.97	0	16.99
Energy intensity (€100/ha)	1.35	1.29	0	66.18
Contract Work Intensity (€100/ha)	1	1.06	0	41.99
Share irrigated land	0.06	0.18	0	1

Table 2: Descriptive Statistics of the set of contextual covariates comprising technology, biophysical environment, and farm (management) characteristics (N = 9,693).

et al., 1984) until the final nodes (aka leaves) contain a number of observations greater than a given minimum. In regression trees, the mean outcome of such a leaf is then the prediction for the observations contained in that leaf. Random forests make predictions in the form of an average of the predictions of a large ensemble of $b = 1, \dots B$ such trees [3], each of which is grown on a random subsample of the data. One of the key attractions of tree-based methods is that they can take an extremely complex, non-linear problem, with a wide range of covariates (Tiffin, 2019).

4.1 Generalized random forests

Instead of making a precise prediction of the outcome Y_i itself (here TFP), we are primarily interested in an accurate prediction of the effect of a specific covariate z_w (weather indicators 1–5) on TFP (Eq. 4). Athey and Imbens (2019) demonstrate how partial or treatment effects, respectively, can be computed based on regression trees by means of an adjusted splitting rule. They regard random forests not as an ensemble method (aka averaging the results of multiple trees) but as an adaptive kernel method, e.g. outcome Y_i could be predicted by means of $\widehat{Y_i} = \sum_{i=1}^{n} \alpha_i(x) Y_i$, where $\alpha_i(x)$ is a data-adaptive-kernel measuring how often the i-th observation falls in the same leaf as a test point x. However, we are not primarily interested in obtaining an exact prediction of Y_i but rather of the partial effect $\theta(x)$ on Y_i, which can be expressed by the following local moment condition:[4]

$$\hat{\theta}(x) = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \alpha_i(x) \times (Y_i - \theta Z_i)^2 \tag{7}$$

From (7) it becomes apparent that the heterogeneity of the partial effect stems from the weights $\alpha_i(x)$.[5] Given the local kernel framework of equation (7) and the high-dimensional setting of the research problem, Athey and Imbens (2019) suggest using random forests for estimating kernel weights $\alpha_i(x)$, whose splitting rule has to be adjusted to the fact that not a precise prediction of the outcome is required but an esti-

[3]These trees base each split on a randomly selected subsample of covariates.

[4]The condition derives from the simple linear regression problem: $Y_i = c + \theta Z_i + \epsilon$. We omit the constant c in the above equation for simplicity reasons.

[5]Basically, eq. 7 could also be estimated using traditional k-NN estimates. However, k-NN is limited in the sense that it does not distinguish with respect to variable importance. Hence, in high dimensional cases with many covariates, where the signal is concentrated along very few covariates, k-NN does not yield precise weights. As random forests are data-adaptive and thus prioritize high-signal contextual variables, it is better-suited to yield precise weights in a high-dimensional covariate space (Wager and Athey, 2018).

mation of the effect of weather on the outcome. Wager and Athey (2018) show that minimizing the squared-error loss in regression trees (as used in Breimann regression trees) is equivalent to maximizing the heterogeneity of the estimate of interest across subregions. Athey and Imbens (2019) make use of this finding and define the generally valid splitting rule

$$\max_{S_1,S_2} = \frac{n_{S_1}\, n_{S_2}}{n_P^2}\left(\hat{\theta}_{S_1} - \hat{\theta}_{S_2}\right)^2. \tag{8}$$

Equation (8) states that a parent node is split into two children nodes (S_1, S_2) such that the balanced difference in $\hat{\theta}$ is maximized between the two children nodes. Within this framework, Athey, Tibshirani, and Wager (2019) define an optimality criterion $\Delta(S_1, S_2)$ as to how an individual tree of a subsample s^{tr} split the covariate space $\mathbf{X_i}$ of the parent node P into binary regions (S_1, S_2) to greedily [6] maximize the heterogeneity of $\hat{\theta}$ across (S_1, S_2). ${}^{N_{S_{1,2}}}/_{N_P}$ is the fraction of training examples $i : X_i \in s^{tr}$ belonging to the two children nodes obtained from the parent P. Our parameter of interest $\theta_{S_j}(s^{tr})$ is identified by locally estimating a simple equation of the form (i.e. in a subregion of the predictor space):

$$\hat{\theta}_{S_j}(s^{tr}) \in \operatorname*{argmin}_{\theta} \sum_{i}^{N} \left(Y_i - \theta T_i\right)^2. \tag{9}$$

In order to find the optimal split, eq. 9 is solved for multiple random splits of $\mathbf{X_i}$, where the split is selected that maximizes the optimality criterion (8), i.e. maximum heterogeneity across nodes. This procedure may now be repeated until a certain stopping criterion is reached, e.g., when a minimum size of observations per node is left. [7] Given the local estimating equation 9, we can train a random forest based on trees that greedily optimize for partial effect heterogeneity (9), from which we can derive similarity weights $a_i(x)$. These data-adaptive kernels measure how often the *i*-th individual falls into the

[6] *Greedy* means that an optimal choice is made at each step rather than considering the entire tree when trying to find the optimal split.

[7] For computational reasons, the algorithm applied is based on an approximation of the described procedure. Technical details can be found in Athey and Wager (2019).

same leaf as a test point q:

$$a_i(x) = \frac{1}{B} \sum_{b=1}^{B} \frac{\mathbf{1}\left(\{X_i \in L_b(q), i \in T_b\}\right)}{|\ \{i : X_i \in L_b(q), i \in T_b\}} \tag{10}$$

where $L_b(q)$ is the leaf of the b-th tree that contains the test point q and T_b denotes the subsample used to grow the b-th tree.

Using only a random subsample of observations for estimating each tree is an effective way to avoid overfitting (Hastie, Tibshirani, and Friedman, 2009). Furthermore, Athey and Imbens (2016) and Wager and Athey (2018) formally establish asymptotic normality for regression trees and random forests through *honest splitting* of trees, i.e. the training sample is split into two parts, one part is used to train the tree and the other part is used to predict the outcome of interest. Athey, Tibshirani, and Wager (2019) show that valid confidence intervals for generalized random forest estimates can be obtained by means of the 'bootstrap of little bags method', where basically small groups of trees are trained and their predictions are then compared within and across groups to estimate the variance. For a more technical description of the method, see Sexton and Laake (2009).

So far, we have not accounted for the panel structure in our data, i.e. the correlation between two observations of one individual (i.e. county) at different time points. To adjust the predictions and standard errors for this circumstance, we use clustered sampling (at the NUTS-3 level) when training the the b-th tree of the random forest, i.e. instead of drawing a random subsample from the full data directly, we add an extra step. We first draw a random set of clusters (e.g. half of the NUTS-3 regions), and then sample these selected clusters for training and honest estimation (Athey and Wager, 2019; Tibshirani, Athey, and Wager, 2020).

4.2 Model-agnostic Shapley values

To explain the individual partial effects at the NUTS-3 level, we make use of Shapley values (Shapley, 1953), a model-agnostic concept stemming from cooperative game theory, which is well-suited for complex prediction models (Molnar, 2019; Lundberg and Lee, 2017; Tiffin, 2019).

Shapley values capture the contribution of each covariate to the difference between the actual farm-level estimation and the sample mean estimation. Hence, Shapley values reflect each contextual variable's relative contribution to the predicted outcome and can be seen as a special case of marginal effects, where interactions and redundancies between covariates are taken into account (Štrumbelj and Kononenko, 2014). Shapley values represent estimates of variable importance (size of the contribution) and direction (sign) in explaining an outcome. To get a better intuition as to what Shapley values are, Fig. 1 presents stylized examples of their use. Assume, we estimated a GRF for a sample using three contextual covariates. The sample mean prediction of the weather effect would be 7 units. Variable 1 takes on a high value for region A, which contributes negatively to the predicted effect (Shapley value=–2). For region B it is the other way around. Variable 1 is low, which contributes positively to the effect prediction (+1), i.e. it increases the mean prediction. We find similar patterns for Variables 2 and 3. Finally, after summing over all Shapley values, we end up with the individual prediction for each observation. A more detailed description and further discussions on the method can be found in Molnar (2019).

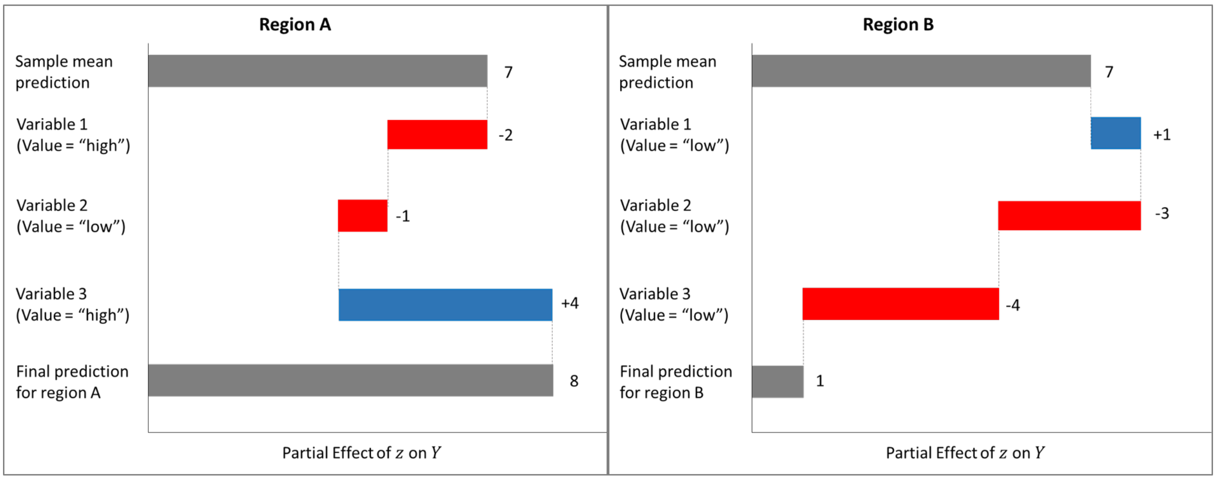

Figure 1: Stylized graph of two fictitious NUTS-3 regions, for which partial effects and Shapley values were computed. Source: Own compilation based on Tiffin (2019) and Molnar (2019).

5 Empirical Results

First, we calculated our outcome variable, the Lowe TFP index as described in Section 2. We then trained a generalized random forest for each weather indicator separately. For a better interpretation of the result we log-transformed the outcome variable (TFP), which allows us to interpret the results as semi-elasticities. To be stringent in terms of retrieving reliable estimates and standard errors, we train 40,000 trees for each weather indicator. By doing this, we make sure that the excess error – measuring the stability of our estimates if we repeated the estimation – is negligibly small (Wager, Hastie, and Efron, 2014). Furthermore, we tested different versions of hyperparameter tuning via cross-validation (James et al., 2013) on the minimum number of observations in each tree leaf, the fraction of the data used for the subsample to build each tree, the number of variables tried for each split, as well as split balance parameters. No significant performance differences could be detected, which is why we adhered to the base settings of the algorithm as determined by Tibshirani, Athey, and Wager (2020). For the calculation of the Shapley values, we used approximations of the above mentioned GRFs based on 2,000 trees to keep computational burden manageable.

5.1 Total factor productivity estimates

Figure 2 illustrates the spatial and temporal dynamics of total factor productivity in the EU between 2005 and 2016. The TFP of each region in each year is compared to the TFP of the NUTS-3 region Stockholm County (SE010) in 2005. Panel 2A describes the spatial distribution of the mean TFP level in the sample period. Based on the United Nations geoscheme (UNSD, 2021), we compare crop farm productivities across four European subregions (East, North, West, South). We can see, that crop farming in the Western subregion, and especially in Northeast France, Belgium, the Netherlands and parts of Germany show the highest TFP levels, together with the Southeast of the UK, as well South Sweden, East Denmark (Northern region). Lower TFP levels are found for most parts of Southern and Eastern Europe. These finding are largely confirmed by Panel B of Fig. 2. Crop farming in Western Europe was most productive followed by Northern Europe, which however shows a greater TFP growth than Western Europe. Western European crop production appeared to have been twice as productive as Eastern and Southern Europe. Generally, we can observe a positive development for all regions. Two very strong aggregation effects can be found in 2007. First, Danish regions join the sample for Northern Europe, which strongly increased mean productivity.[8] Also, Romania and Bulgaria joined the EU in 2007, which seem to have strongly affected the mean TFP for Eastern Europe. Our results are in parts very similar to previous findings for European crop farming, e.g. Bokusheva and Čechura (2017) compares the TFP development of arable farming in selected EU countries at the farm-level using stochastic frontier analysis. They find an average yearly productivity increase of 1.8% for England between 2003 and 2015, for which we find a 1.7% TFP growth. However, in general, our results appear to be slightly more fluctuating. Our results also correspond well with Martinho (2017), who studies TFP growth at the NUTS-2 level using data envelopment analysis. We are not aware of any EU-wide crop farming study that uses a transitive TFP index.

5.2 Heterogeneous weather effects on total factor productivity

Table 3 summarizes the GRF estimation results for all five weather indicators. Regarding the average weather patterns temperature and precipitation sum, we find mixed results. On average, a temperature increase by one °Celcius leads to a TFP increase of 4.56%. However, this effect varies considerably. Turning to the first column of Table 3, we find that 75% percent of observations react positively to an average temperature increase, while in 25% of the cases TFP decreases. In total, we find in 60% of the observations a significant impact of average temperature on TFP (at the 95% significance level). Those regions that experience a negative effect, suffer a 5.1% loss in TFP due to a one degree temperature increase, while the majority of cases (>50%) gain a 8.35% average increase

[8] We had to omit Danish regions before 2007, since farms were spatially not unambiguously attributable to the respective NUTS-3-regions after a reform of the NUTS classification in Denmark.

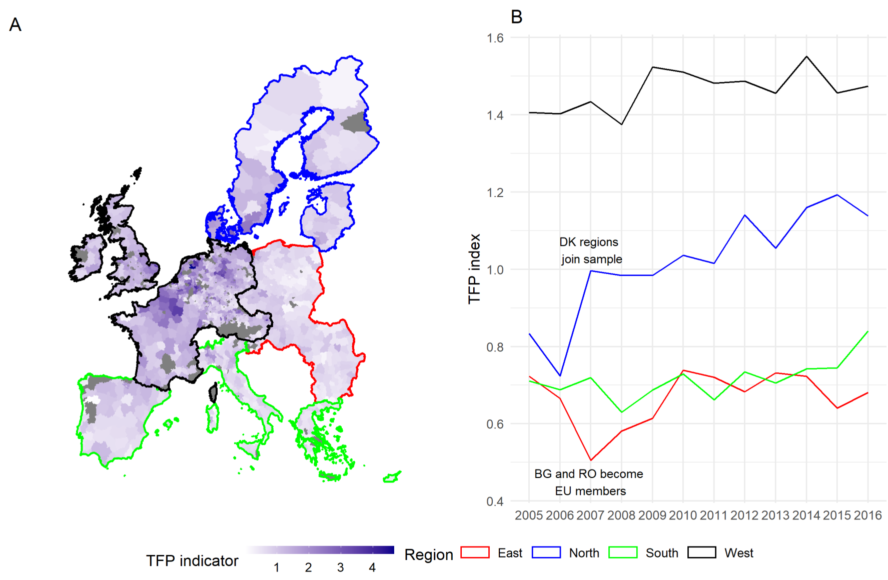

Figure 2: Summary of the total factor productivities for crop farms across the EU at the NUTS-3 level. regions are aggregated according to the UN geoschemes for Europe. Panel A: Spatial distribution of mean TFP values between 2005 and 2016. Panel B: Mean development of TFP index.

in productivity. As described in the second column of Table 3, the response to an increase in the yearly precipitation sum appears to be rather small. We could not find significant effects for more than 50% of the observations. However, the 33% of the observations with a semi-elasticity significantly smaller than zero, 100mm additional rainfall would signify a TFP decrease of 4% on average.

Regarding specific weather events, namely drought spells, heat and heavy rain, we find almost exclusively negative effects on TFP (Table 3, columns 3–5), e.g. an extra day with a maximum temperature of more than 30°Celcius means an average TFP reduction of 0.7% for approx. 80% of the observations. What is more, approx. 75% suffer a TFP loss of -0.8% from one extra day of the maximum rain-free period (i.e. drought spell). The effect of heavy rain events varies from -10% to 4.3%. In summary, we can say that average weather had a mixed effect on TFP in the EU between 2005 and 2016, while specific weather phenomena had a clearly negative effect on productivity.

Figure 3 gives an overview of the spatio-temporal dynamics of the weather-TFP nexus. Panel A summarizes the weather impacts on a yearly basis, from which two primary conclusions can be drawn. First, there is no clear time trend observable in the given time

Table 3: Summary of the GRF estimation results.

	Weather indicator				
	Mean temperature (°C)	Precipitation sum (mm)	Consecutive dry days	Hot days	Heavy rain days (precipitation > 20mm)
Full sample					
Mean semi-elasticity (% change in TFP)	4.56	-0.01	-0.69	-0.63	-2.83
SD semi-elasticity (% change in TFP)	5.1	0.04	0.35	0.3	2.05
Precentage of N with semi-elasticity < 0	24.4	66.1	98.7	98	90.9
Precentage of N with semi-elasticity > 0	75.6	33.9	1.3	2	9.1
Subsample (semi-elasticity θ < 0 at 95% significance level)					
N	474	3165	7332	7842	5767
Share in full sample (%)	4.9	32.7	75.6	80.9	59.5
Mean semi-elasticity (% change in TFP)	-5.1	-0.04	-0.8	-0.7	-4.07
SD semi-elasticity (% change in TFP)	1.69	0.01	0.31	0.25	1.31
Subsample (semi-elasticity θ > 0 at 95% significance level)					
N	5202	1068	0	3	54
Share in full sample (%)	53.67	11.02	<1	<1	<1
Mean semi-elasticity (% change in TFP)	8.35	0.06	-	0.29	3
SD semi-elasticity (% change in TFP)	2.34	0.03	-	0.06	0.64

period for any of the weather indicators. Second, for most indicators, the effect on TFP is rather constant or fluctuates very little, respectively. We only find somewhat more pronounced volatility for drought spells (Panel A3), where the negative impact appears to have been largest in 2005 and 2006.

Panel 3B characterizes the temporal dynamics of the share of observations whose semi-elasticities are significantly different from zero (at the 95% significance level). For average temperature (Panel A2), the number of observations with negative effects is quite constant at 4% to 6%, while the number for positive values fluctuates by roughly 10 percentage points from 41% (2005) to 51% (2015). Regarding precipitation (Panel B2), we find fluctuations especially for the significant negative relationship (28.2%–38.7%). The largest variability in terms of significant weather impacts can be found for drought spells (Panel B3), varying from roughly 85% (2005, 2006) and 66% in 2007.

Finally, weather effects also vary spatially (Panel 3C). We can see that an increase in yearly mean temperature benefits primarily Central and Eastern Europe but also Scandinavia, the UK, Belgium, the Netherlands and parts of France. Contrarily, adverse temperature effects occur in most parts of Southern Europe. As for precipitation, a East-West gradient can be observed, where Eastern Europe would mostly benefit from more rainfall. This is particularly true for most regions in Poland. Prolonged drought spells strongly affect Western Europe, especially Northwest France and adjacent regions. Similar patterns can be found for heat and heavy rain, while hot days have a relatively strong negative impact on Eastern Germany. Poland seem to profit from additional heavy rain days, while the French regions *Haut-de-France* and *Grand Est* suffer rather strongly from heavy rain events.

5.3 Exploring impact heterogeneity

After having detected a prevalent heterogeneity in the weather TFP relationship in the European Union, we want to find the major drivers behind that relationship. Further explaining the predicted weather effects should give important insights into features that can mitigate adverse and amplify positive weather impacts on total factor productivity. We computed Shapley values for all observations and covariates and present a extensive assortment in Figure 4. The figure summarizes individual Shapley values and compares them with their repsective covariate values (indicated by the color). For instance, for observations with a high value for the elevation covariate (=high altitude), we find a positive contribution on the predicted temperature effect of up to almost 10% (Panel D1). Contrary to that, low altitude seems to decrease the predicted temperature effect. In Panel D2, we find the opposite effects for the yearly precipitation sum.

With regard to farm (management) characteristics, we find mixed results for the different weather indicators (Panel 4A), e.g. fertilizer intensity seem to have rather small contributions to mean temperature and rainfall compared to droughts, heat and heavy rain. Generally, contextual variables reflecting production intensity seem to have a noticeable effect (to various degrees) on the weather-TFP relationship, e.g. high labor intensity seems to shift the impact on rainfall downward. Crop diversity, which is seen as a climate change mitigation measure (Falco et al., 2014), does not seem to have strongly affected the weather-TFP nexus. It is worth noticing that the share of decoupled subsidies

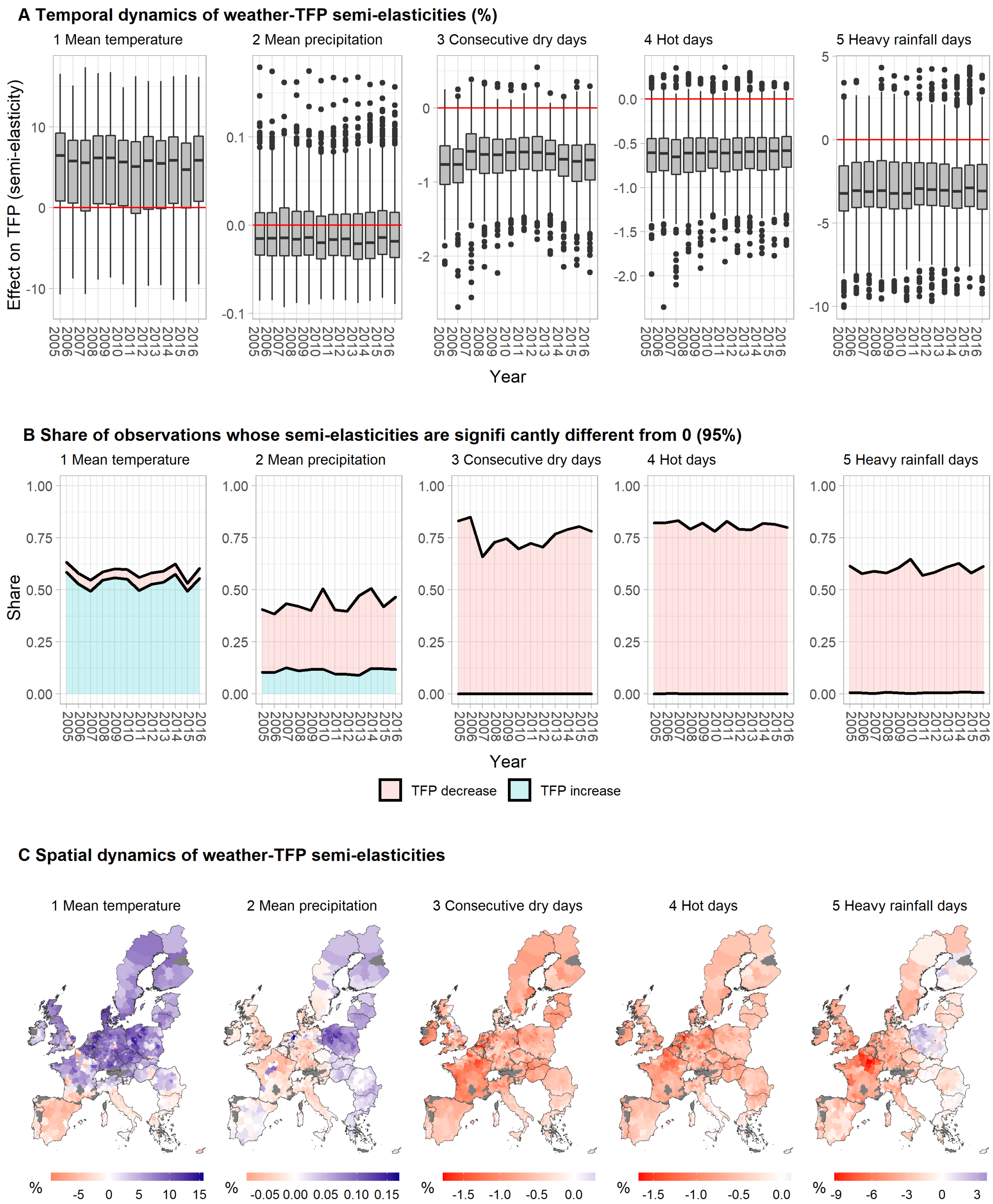

Figure 3: Overview of the spatio-temporal dynamics of the weather-TFP relationship in the EU.

in total subsidies seems to increase the TFP-robustness against all 5 weather phenomena.

The Shapley values for the proxies of the state of technology confirm the spatio-temporal patterns observed in the previous Section (Panel 4B), e.g. the time trend barely has an impact on productivity in comparison to other covariates, while we find a strong East-West relationship for rain-related weather patterns and a North-South gradient for temperature and rain.

The results in Panel 4C partly point toward the existence of compound weather effects, e.g. the occurrence of prolonged drought periods and hot days seem to negatively impact the effect of mean temperature on TFP, while the effects of rainfall seem to become more positive with higher temperatures. Although no clear pattern can be observed, our results also hint toward compound extreme weather effects (4C 3–5).

As for the biophysical farming environment, we find several effects (4D). For instance, a high sand content in the top soil strongly intensifies the negative impact of heat on productivity. Also, there appears to be an important interaction between weather and soil nitrogen content, where soils with high nitrogen contents appear robust against the weather in that high Shapley values are clustered around zero.

Overall, we find that for different weather events, different interactions matter most. While the geographical location influences the weather-TFP relationship quite strongly, the interaction effect is small for consecutive dry days, where as for this weather indicator as well as heat days farm (management) characteristics matter a lot. This gives farmers the possibility to actively develop strategies that mitigate negative weather effects, and hence negative climate change impacts. What is more, some covariates induce opposite effects, e.g. small economic size contributes positively to the mean temperature effect, but negatively to the drought effect.

6 Discussion

Recent advancements in the assessment of the weather-TFP index in the agricultural context largely explored weather as one component of TFP growth (Chambers and Pieralli, 2020; Njuki, Bravo-Ureta, and O'Donnell, 2018; Njuki, Bravo-Ureta, and Cabrera, 2020), which have their strengths in analyzing the interplay between weather patterns and the components of productivity growth. However, these productivity decomposi-

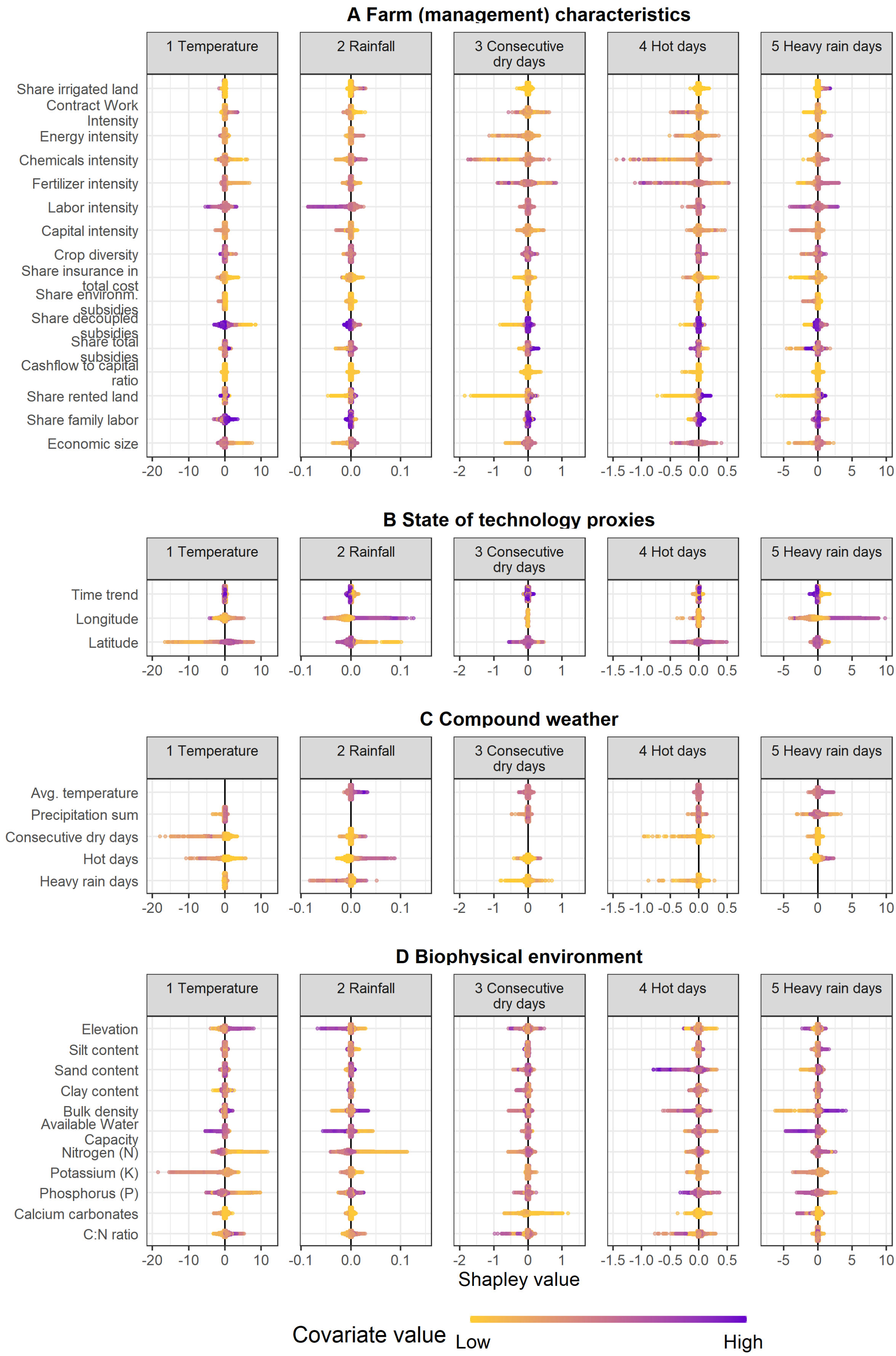

Figure 4: Summary of the temporal-spatial dynamics of the weather-TFP relationship.

tion approaches do not allow for a detailed analysis on the marginal effects of weather trends and patterns on TFP. In many circumstances, we are particularly interested in how farming systems respond to specific weather events such as heavy rain or hot days, which is necessary to derive strategies to adapt to changing weather patterns, which become increasingly prevalent in the climate change context.

The theoretical foundation of our study can most likely be compared to Ortiz-Bobea et al. (2020) and Damania, Desbureaux, and Zaveri (2020). Ortiz-Bobea et al. (2020) study the historical impact of anthropogenic climate change on global agricultural productivity using a similar production technology assumption to ours. In contrast to our results, they find a consistently negative relationship between temperature, precipitation and productivity growth globally and find noticeable regional heterogeneity. The most widespread estimation approach to the analyzed relationship is based on parametric panel data models linking productivity (growth) to weather (change), which optionally include quadratic terms (e.g. Van Passel, Massetti, and Mendelsohn, 2017; Bozzola et al., 2018; Ortiz-Bobea et al., 2020; Damania, Desbureaux, and Zaveri, 2020; Letta and Tol, 2019) or more sophisticated methods such as regression splines (Schlenker and Roberts, 2009) or quantile regression (DePaula, 2020) to account for effect heterogeneity/non-linearity. Many of the authors find that weather effects are particularly detrimental in hot and arid locations. This, however does usually not provide information regarding how to best respond to (extreme) weather events. One way to come to such a conclusion and another likely source of heterogeneity are interactions between weather patterns and the production context, which is often neglected. In an attempt to account for some of this heterogeneity source, Letta and Tol (2019) uses interaction terms of temperature with a dummy for being poor, and finds that poor countries are particularly vulnerable to temperature shocks. Our machine learning based approach goes beyond these attempts in that it flexibly accounts for nonlinearities and potentially myriad interactions in a high covariate space; thus more reliably reflects the complexity of the underlying relationship. According to Storm, Baylis, and Heckelei (2019), this is one of the prime examples, where machine learning adds value to the agricultural and applied economics literature.

However, increasing model complexity usually comes at the expense of reduced interpretability (Hastie, Tibshirani, and Friedman, 2009). Using model-agnostic Shapley values, stemming from the interpretable machine learning literature, makes it possible to identify important patterns underlying such a complex model (Molnar, 2019). In contrast to many other studies, we are able to explore relevant interactions between weather events and production context and their effect on productivity. This allows us to analyze key points concerning weather vulnerability of farming systems, which eventually inform climate change adaptation strategies, e.g. increasing economic size appears to guard against the negative (and positive) effects of weather shocks in the European context. In contrast to this, Reidsma, Oude Lansink, and Ewert (2009) find that economic size amplify the negative impact of higher temperature in assorted European countries. There are also weak signals in our analysis that insurance might give a disincentive to adapt to weather shocks (compare Annan and Schlenker, 2015). Similarly, there are weak signals in our model that increased crop diversity might guard farms from productivity losses due to heavy rain events (Gaudin et al., 2015).

What is more, we can confirm the link between technology and weather sensitivity (Lipper et al., 2018; Ortiz-Bobea, Knippenberg, and Chambers, 2018). The agricultural weather impact analysis literature increasingly acknowledges the important interplay between several weather events (Zscheischler et al., 2018; Ortiz-Bobea et al., 2019; Haqiqi et al., 2021). For instance, Haqiqi et al. (2021) find that the yield response to water can be up to four times higher in hot weather. We find a less pronounced but similar result for TFP (Fig. 4, Panel C2). Higher avg. temperature as well as more hot days positively impact the effect of precipitation on productivity. Finally, we are able to analyze in what ways the biophysical environment affects the buffer potential of farms against negative weather effects.

One problem with ML approaches is that through their flexibility, they allow researchers to include a myriad of predictors in their models, which are prone to lead to bias structures in cause-effect relationships, e.g. by including bad control variables

(Cinelli, Forney, and Pearl, 2020). Thus, if researchers are interested in statistical inference rather than pure outcome prediction, it is absolutely necessary to come up with a credible identification strategy (Pearl, 2018). We do that by basing our analysis on microeconomic production theory.

Further, while the generalized random forests allows for statistical tests and inference (Athey, Tibshirani, and Wager, 2019), Shapley values as model-agnostic interpretation tool refer to the modeled relationship and not the ground truth, which is not the same (Lipton, 2018). This is a major epistemological difference from more traditional statistical methods. Hence, we rely on the assumption that our model approximates the causal mechanisms of the true relationship well (Páez, 2019). Bearing this in mind is particularly important if we were to include an excess set of covariates that are not causally related to the outcome variable, which would lead to Shapley values explaining a spurious relationship. For the assumption that the estimated model reflects the true relationship, beside having a reliable identification strategy, it is important to conduct robustness checks to see if the estimated relationship is stable across multiple configurations. At this stage, our study is lacking several important robustness checks.

7 Conclusion

In this study, we demonstrate the importance of taking into account the farming environment when assessing the effects of different weather patterns on total factor productivity. Recent advances in the causal machine learning literature allow us to explore the complex relationship between weather, technology, farm management, biophysical environment and total factor productivity. We derive important contextual variables and estimate a Lowe TFP index, consistent with index number theory. Based on five weather indicators, namely mean temperature, precipitation sum, consecutive dry days, hot days and heavy rain days, we find considerable weather impacts for crop farming in 27 EU member states at the NUTS-3 level between 2005 and 2016. We find both positive and negative effects of aggregated weather events (mean temperature and precipitaion sum), and predominantly negative effects drought spells, heat and heavy precipitation on productivity. Using model-agnostic Shapley values, we find that the farming environment plays an important role in determining the effect size of weather on TFP. Important interactions between weather, farm management, technology, and the biophysical environment are found. Our modeling approach also allows to analyze compound weather effects.

While our research approach is very flexible, we do not account for potential accumulating weather effects. Furthermore, we rely on a rather short time horizon of a total of 12 years, which might reduce the generalizability of our results to a broader climate change impact context. Given the fact, that we have not conducted important robustness checks, the results of this study should be interpreted with care. Nevertheless, given the fact that climate change continues to change weather patterns, our results provide interesting insights as to how farm managers and legislators could locally respond to offset negative climate change impacts. This might lead to more effective adaptation strategies. For instance, we find that larger farms might be less vulnerable to weather shocks in the EU, a finding that should be considered in the Common Agricultural Policy, which currently prioritizes small and medium-sized farms.

There remain several promising paths for future research. Including data on specific agricultural practices could give more nuanced recommendations in terms of reducing negative weather impacts on productivity. Furthermore, it would be interesting to see our research approach being applied at the farm-level. Finally, future research could also explore how future climate scenarios might affect the weather-TFP relationship in agriculture.

8 References

Aghakouchak, A., F. Chiang, L.S. Huning, C.A. Love, I. Mallakpour, O. Mazdiyasni, H. Moftakhari, S.M. Papalexiou, E. Ragno, and M. Sadegh. 2020. "Climate Extremes and Compound Hazards in a Warming World." *Annual Review of Earth and Planetary Sciences* 48:519–548.

Alston, J.M., J.M. Beddow, and P.G. Pardey. 2009. "Agricultural research, productivity, and food prices in the long run."

Annan, F., and W. Schlenker. 2015. "Federal crop insurance and the disincentive to adapt to extreme heat." *American Economic Review* 105:262–266.

Athey, S., and G. Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences of the United States of America* 113:7353–7360.

Athey, S., and G.W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11.

Athey, S., J. Tibshirani, and S. Wager. 2019. "Generalized random forests." *Annals of Statistics* 47:1179–1203.

Athey, S., and S. Wager. 2019. "Estimating Treatment Effects with Causal Forests: An Application.", pp. 1–15.

Ball, V.E., J.P. Butault, C.S. Juan, and R. Mora. 2010. "Productivity and international competitiveness of agriculture in the European Union and the United States." *Agricultural Economics* 41:611–627.

Baráth, L., and I. Fertő. 2017. "Productivity and Convergence in European Agriculture." *Journal of Agricultural Economics* 68:228–248.

Bokusheva, R., and L. Čechura. 2017. "Evaluating dynamics, sources and drivers of productivity growth at the farm level." Working paper No. 106, OECD, Paris.

Bozzola, M., E. Massetti, R. Mendelsohn, and F. Capitanio. 2018. "A Ricardian analysis of the impact of climate change on Italian agriculture." *European Review of Agricultural Economics* 45:1–23.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45:1–122.

Breiman, L., J. Friedman, C.J. Stone, and R.A. Olshen. 1984. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series, Taylor & Francis.

Burke, M., and K. Emerick. 2015. "Adaptation to climate change: Evidence from US agriculture." *American Economic Journal: Economic Policy* 8:106–140.

Burke, M., S.M. Hsiang, and E. Miguel. 2015. "Global non-linear effect of temperature on economic production." *Nature* 527:235–239.

Chambers, R.G. 1988. *Applied Production Analysis*. New York: Cambridge University Press.

Chambers, R.G., and S. Pieralli. 2020. "The Sources of Measured US Agricultural Productivity Growth: Weather, Technological Change, and Adaptation." *American Journal of Agricultural Economics* 102:1198–1226.

Chambers, R.G., S. Pieralli, and Y. Sheng. 2020. "The Millennium Droughts and Australian Agricultural Productivity Performance: A Nonparametric Analysis." *American Journal of Agricultural Economics* 102:1383–1403.

Chen, P., W. Dong, X. Lu, U. Kaymak, K. He, and Z. Huang. 2019. "Deep representation learning for individualized treatment effect estimation using electronic health records." *Journal of Biomedical Informatics* 100:103303.

Cinelli, C., A. Forney, and J. Pearl. 2020. "A Crash Course in Good and Bad Controls." *SSRN Electronic Journal*, pp. 1–10.

Cornes, R.C., G. van der Schrier, E.J.M. van den Besselaar, and P.D. Jones. 2018. "An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets." *Journal of Geophysical Research: Atmospheres* 123:9391–9409.

Cui, X. 2020. "Climate change and adaptation in agriculture: Evidence from US cropping patterns." *Journal of Environmental Economics and Management* 101:102306.

Damania, R., S. Desbureaux, and E. Zaveri. 2020. “Does rainfall matter for economic growth? Evidence from global sub-national data (1990–2014).” *Journal of Environmental Economics and Management* 102:102335.

DePaula, G. 2020. “The distributional effect of climate change on agriculture: Evidence from a Ricardian quantile analysis of Brazilian census data.” *Journal of Environmental Economics and Management* 104:102378.

Deschênes, O., and M. Greenstone. 2007. “The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather.” *American Economic Review* 97:354–385.

—. 2012. “The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather: Reply.” *American Economic Review* 102:3761–3773.

European Commission. 2020. “Eurostat Database.”

Falco, S.D., F. Adinolfi, M. Bozzola, and F. Capitanio. 2014. “Crop Insurance as a Strategy for Adapting to Climate Change.” *Journal of Agricultural Economics* 65:485–504.

FAOSTAT. 2017. “FAO Statistics.”

Gaudin, A.C.M., T.N. Tolhurst, A.P. Ker, K. Janovicek, C. Tortora, R.C. Martin, and W. Deen. 2015. “Increasing Crop Diversity Mitigates Weather Variations and Improves Yield Stability.” *PLOS ONE* 10:e0113261.

Haqiqi, I., D.S. Grogan, T.W. Hertel, and W. Schlenker. 2021. “Quantifying the impacts of compound extremes on agriculture.” *Hydrology and Earth System Sciences* 25:551–564.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*, vol. 27 of *Springer Series in Statistics*. New York, NY: Springer New York.

Hsiang, S. 2016. “Climate econometrics.” *Annual Review of Resource Economics* 8:43–75.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An introduction to statistical learning*, vol. 112. New York: Springer.

Keane, M., and T. Neal. 2020. "Climate change and U.S. agriculture: Accounting for multidimensional slope heterogeneity in panel data." *Quantitative Economics* 11:1391–1429.

Künzel, S.R., J.S. Sekhon, P.J. Bickel, and B. Yu. 2019. "Metalearners for estimating heterogeneous treatment effects using machine learning." *Proceedings of the National Academy of Sciences of the United States of America* 116:4156–4165.

Letta, M., and R.S. Tol. 2019. "Weather, Climate and Total Factor Productivity." *Environmental and Resource Economics* 73:283–305.

Lipper, L., N. McCarthy, D. Zilberman, S. Asfaw, and G. Branca, eds. 2018. *Climate Smart Agriculture*, vol. 52 of *Natural Resource Management and Policy*. Springer International Publishing.

Lipton, Z.C. 2018. "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery." *Queue* 16:31–57.

Lundberg, S.M., and S.I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds. *Advances in Neural Information Processing Systems*. Curran Associates, Inc., vol. 30, pp. 4765–4774.

Lüttger, A.B., and T. Feike. 2018. "Development of heat and drought related extreme weather events and their effect on winter wheat yields in Germany." *Theoretical and Applied Climatology* 132:15–29.

Martinho, V.J.P.D. 2017. "Efficiency, total factor productivity and returns to scale in a sustainable perspective: An analysis in the European Union at farm and regional level." *Land Use Policy* 68:232–245.

Mendelsohn, R., W.D. Nordhaus, and D. Shaw. 1994. "American Economic Association The Impact of Global Warming on Agriculture : A Ricardian Analysis." *The American Economic Review* 84:753–771.

Molnar, C. 2019. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*.

Njuki, E., B.E. Bravo-Ureta, and V.E. Cabrera. 2020. "Climatic effects and total factor productivity: Econometric evidence for Wisconsin dairy farms." *European Review of Agricultural Economics* 47:1276–1301.

Njuki, E., B.E. Bravo-Ureta, and C.J. O'Donnell. 2018. "A new look at the decomposition of agricultural productivity growth incorporating weather effects." *PLoS ONE* 13:1–21.

O'Donnell, C.J. 2012a. "Alternative Indexes for Multiple Comparisons of Quantities and Prices." *CEPA Working Paper Series*, pp. .

—. 2012b. "An aggregate quantity framework for measuring and decomposing productivity change." *Journal of Productivity Analysis* 38:255–272.

—. 2012c. "Nonparametric estimates of the components of productivity and profitability change in U.S. agriculture." *American Journal of Agricultural Economics* 94:873–890.

—. 2018. *Productivity and efficiency analysis: an economic approach to measuring and explaining managerial performance*, 1st ed. Singapore: Springer.

—. 2016. "Using information about technologies, markets and firm behaviour to decompose a proper productivity index." *Journal of Econometrics* 190:328–340.

OECD. 2019. *Innovation, Productivity and Sustainability in Food and Agriculture*, oecd food ed. Paris: OECD Publishing.

Ortiz-Bobea, A., T.R. Ault, C.M. Carrillo, R.G. Chambers, and D.B. Lobell. 2020. "The Historical Impact of Anthropogenic Climate Change on Global Agricultural Productivity." Unpublished.

Ortiz-Bobea, A., E. Knippenberg, and R.G. Chambers. 2018. "Growing climatic sensitivity of U.S. Agriculture linked to technological change and regional specialization." *Science Advances* 4.

Ortiz-Bobea, A., H. Wang, C.M. Carrillo, and T.R. Ault. 2019. "Unpacking the climatic drivers of US agricultural yields." *Environmental Research Letters* 14.

Páez, A. 2019. "The Pragmatic Turn in Explainable Artificial Intelligence (XAI)." *Minds and Machines* 29:441–459.

Panagos, P., M. Van Liedekerke, A. Jones, and L. Montanarella. 2012. "European Soil Data Centre: Response to European policy support and public data requirements." *Land Use Policy* 29:329–338.

Pearl, J. 2018. "Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution." *arXiv*, pp. 0–9.

Reidsma, P., A. Oude Lansink, and F. Ewert. 2009. "Economic impacts of climatic variability and subsidies on European agriculture and observed adaptation strategies." *Mitigation and Adaptation Strategies for Global Change* 14:35–59.

Schlenker, W., W.M. Hanemann, and A.C. Fisher. 2005. "American Economic Association Will U . S . Agriculture Really Benefit from Global Warming ? Accounting for Irrigation in the Hedonic Approach." *The American Economic Review* 95:395–406.

Schlenker, W., and M.J. Roberts. 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* 106:15594–15598.

Sexton, J., and P. Laake. 2009. "Standard errors for bagged and random forest estimators." *Computational Statistics and Data Analysis* 53:801–811.

Shapley, L.S. 1953. *A value for n-person games.*.

Shi, C., D.M. Blei, and V. Veitch. 2019. "Adapting Neural Networks for the Estimation of Treatment Effects." Working paper, 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Storm, H., K. Baylis, and T. Heckelei. 2019. "Machine learning in agricultural and applied economics." *European Review of Agricultural Economics* 105:493.

Štrumbelj, E., and I. Kononenko. 2014. "Explaining prediction models and individual predictions with feature contributions." *Knowledge and Information Systems* 41:647–665.

Thirtle, C., L. Lin, J. Holding, L. Jenkins, and J. Piesse. 2004. "Explaining the decline in UK agricultural productivity growth." *Journal of Agricultural Economics* 55:343–366.

Tibshirani, J., S. Athey, and S. Wager. 2020. *grf: Generalized Random Forests*.

Tiffin, A. 2019. "Machine Learning and Causality: The Impact of Financial Crises on Growth." *IMF Working Papers* 19.

UNSD. 2021. "Methodology - Standard country or area codes for statistical use (M49)." Working paper, United Nations Statistical Division, New York.

Van Passel, S., E. Massetti, and R. Mendelsohn. 2017. "A Ricardian Analysis of the Impact of Climate Change on European Agriculture." *Environmental and Resource Economics* 67:725–760.

Vogel, E., M.G. Donat, L.V. Alexander, M. Meinshausen, D.K. Ray, D. Karoly, N. Meinshausen, and K. Frieler. 2019. "The effects of climate extremes on global agricultural yields." *Environmental Research Letters* 14.

Wager, S., and S. Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113:1228–1242.

Wager, S., T. Hastie, and B. Efron. 2014. "Confidence intervals for random forests: The jackknife and the infinitesimal jackknife." *Journal of Machine Learning Research* 15:1625–1651.

Webber, H., G. Lischeid, M. Sommer, R. Finger, C. Nendel, T. Gaiser, and F. Ewert. 2020. "No perfect storm for crop yield failure in Germany." *Environmental Research Letters* 15.

Westra, S., H.J. Fowler, J.P. Evans, L.V. Alexander, P. Berg, F. Johnson, E.J. Kendon, G. Lenderink, and N.M. Roberts. 2014. "Future changes to the intensity and frequency of short-duration extreme rainfall." *Reviews of Geophysics* 52:522–555.

Zscheischler, J., S. Westra, B.J. Van Den Hurk, S.I. Seneviratne, P.J. Ward, A. Pitman, A. Aghakouchak, D.N. Bresch, M. Leonard, T. Wahl, and X. Zhang. 2018. "Future climate risk from compound events." *Nature Climate Change* 8:469–477.