The multiple linear regression modelling algorithm ABSOLUT v1.0 for weather-based crop yield prediction and its application to Germany at district level

Tobias Conradt[1]

[1]Potsdam Institute for Climate Impact Research, Potsdam, Germany

Correspondence: Tobias Conradt (conradt@pik-potsdam.de)

Abstract. ABSOLUT v1.0 is an adaptive algorithm that uses correlations between time-aggregated weather data and crop yields for yield prediction. At its core, locally (i.e. district-) specific multiple linear regressions are used to predict the annual crop yield based on four weather aggregates and a linear trend in time. In contrast to other statistical yield prediction methods, the input weather features are not predefined or based on a limited number of observed correlations but they are exhaustively

5   tested for maximum explanatory power across all of their possible combinations in all districts of the modelling domain. Principal weather variables (such as temperature, precipitation, or sunshine duration) are aggregated over two to six consecutive months from the 12 months preceding the harvest. This gives 45 potential input features per original weather variable. In a first step, this zoo of possible input features is subset to those very probably holding explanatory power for observed yields. The second, computationally demanding step is making out-of-sample predictions for all districts with all possible combinations

10   of the remaining features. Step three selects the seven combinations of four different weather features that have the highest explanatory power averaged over the districts. Finally, the district-specific best performing regression among these seven is used for district predictions, and the results can be spatially aggregated. To evaluate the new approach, ABSOLUT v1.0 is applied to predict the yields of ten major crops at the district level in Germany based on two decades of yield and weather data from about 300 districts. When aggregated to the national level, the predictions explain 70–90 % of the observed variance

15   between years depending on crop type and time frame considered. District-level performance maps for winter wheat and silage maize show areas with >40 % variance explanation covering about two thirds of the country.

1 Introduction

Weather-based crop yield prediciton have a long history; correlations between weather variables and agricultural yields had

20   already been studied in the first quarter of the 20th century (Meinardus, 1901; Hooker, 1907; Fisher, 1924), and estimating regional yields by multiple linear regressions from time-aggregated weather data has been around for decades. The entirety of crop and landscape specific weather effects is however hardly captured by the existing models. The ABSOLUT algorithm

presented in this article attempts to minimize this gap by finding and adapting the most informative correlations between weather variables and yields, finally aiming for reliable extrapolations in climate change scenario assessments.

25   At its core, the ABSOLUT algorithm applies a multiple linear regression of the form

$$y(t) = \beta_0 + \beta_1 t + \beta_2 w_{1,t} + \beta_3 w_{2,t} + \beta_4 w_{3,t} + \beta_5 w_{4,t} + \varepsilon \tag{1}$$

for each spatial subunit ("district") of the target domain. Herein $y(t)$ is the yield, i. e. the harvested mass per area in dt ha^{-1}, of a certain crop in the year t; the β_i are regression parameters; the $w_{j,t}$ are aggregated weather variables with specific time windows associated to t (for instance the precipitation sum of December, January, and February preceding the harvest in summer), these

30   are henceforth called "weather features" to avoid confusion with weather variables like temperature or precipitation in general; and ε is the estimation error to minimize. This is finally demonstrated here for Germany and its district-level administrative subunits (*Kreise*).

But what is so special about a couple of linear regressions with six coefficients that justifies calling the approach a new algorithm and naming it ABSOLUT? The challenge of weather-based yield regressions is hardly in demonstrating the gen-

35   eral approach; it usually works and explains a significant part of the observed yield variations. Numerous studies have just demonstrated that for different crops in different locations around the world (e. g. Ceglar et al., 2016; Nemoto et al., 2016; Schauberger et al., 2017b). It's the details which matter and finally decide to what extent the unavoidable prediction errors can be reduced; relevant questions include but are not limited to the following:

1. Shall a linear trend over time be assumed or not?

40   2. Shall absolute yields or interannual variations be estimated?

3. Which are the most relevant meteorological variables?

4. Should the input be augmented with non-meteorological data?

5. Shall input variables be transformed before they are used?

6. How many input variables should be used?

45   7. What are the optimal time windows for their aggregation?

8. Is it better to fit regression models separately for each subunit or to make panel models sharing the same parameters for groups of subunits?

9. Are there other options to make the estimations more robust?

From these questions alone it should be obvious that even the decision to resort to simple linear regression models leaves

50   the modeller with further decisions galore, frequently too easily taken based on personal beliefs but in fact defining the true challenge of research and development with this class of models:

Questions 1 and 2 are rather decisive if the method is to be used for evaluating the impact of long-term climate shifts. Gornott and Wechsung (2016) elegantly circumvented dealing with long-term yield trends by estimating year-on-year yield differences from the respective year-on-year changes in weather variables, thus implicitly assuming a linear trend component through the
55   intercept in the regression equations. The approach chosen here – estimating absolut values and an explicit linear trend for weather-independent long-term yield changes – makes isolating purely climate change induced yield trends easier in scenario studies by freezing the time variable in the scenario estimations. Given the stagnations after several decades of technological progress with ever increasing yields reported many times for different crops and world regions (e. g. Chen, 2018; Schauberger et al., 2018; Mehrabi and Ramankutty, 2019) the assumption of a linear time trend might be oversimplistic (cf. the stochastic
60   trend separation by Agnolucci and De Lipsis, 2020) and eventually be replaced by a better approach.

Questions 3, and 6 about the selection of predictor variables define the focus of this contribution. The "oldschool approach" still taken by many researchers is largely an expert choice. It might have been guided by selective correlation analyses for pre-selected candidate variables (González-Fernández et al., 2020; Ji et al., 2019), stepwise regression (Kern et al., 2018; Salehnia et al., 2020), or consideration of crop growth stages for suitable time windows (question 7 Butts-Wilmsmeyer et al., 2019;
65   Zhang et al., 2017). In some cases the selection effort is critically flattened, even if continental climate impact assesments are at stake: Moore and Lobell (2014, 2015) used temperature and precipitation averages of the growing season as sole meteorological basis for that purpose. If non-meteorological predictor variables were to be considered (question 4) the specific data availability, e. g. from a certain remote sensing product, often determines the input variables. Reluctance to ponder on suitable input variables however paves the way to machine learning. As an example, Gómez et al. (2019) had potato yields automati-
70   cally fitted to 54 spectral bands and indices of Sentinel 2 satellite images by a bunch of methods (including generalized linear model, quantile regression, support vector machines, and neural networks). While a massive, automated search for the best predictor features and methodology evades modeller subjectivity, explanations for a certain system behaviour can hardly be given, and an eventual gain in predictive performance remains uncertain.

By no means the method presented here – "Assessing Best-predictive Sets fOr multiple Linear regressions throUgh exhaus-
75   tive Testing" (ABSOLUT) – can be claimed to be the last word on the subject; it also contains a number of arbitrary settings, however it systematically optimizes the often disregarded input variable selection process while maintaining the principal approach of multivariate linear regressions. The approach can be counted as a kind of brute force machine learning and as such it seems to be pioneering the field of crop yield prediction: Among the 362 crop forecasting studies published in the years 2004–2019 that were evaluated by Schauberger et al. (2020) there are 258 utilizing regression, and a few dozen additionally or
80   alternatively implementing established machine learning approaches, mostly automated neural networks (28 cases) followed by random forests (12). This study should therefore serve as proof of concept proposing another building block for more accurate predictions in similar setups, e. g. with panel (question 8) or nonlinear regression models.

2 Materials

2.1 General requirements

2.1.1 Hard- and software

85 As ABSOLUT builds on exhaustive searches for optimal feature combinations in linear regressions, parallel execution of the code on a cluster computer using several dozen cores is advisable. A state-of-the-art single PC or notebook would probably need 2–3 days for the Germany example given all CPU cores (usually four) are hooked up.

The R software (https://www.r-project.org/) is required in version 3.5.1 or newer, and the following R packages must be in-
90 stalled to run the code as provided (version numbers indicate the versions actually used by the author): `leaps` (v.3.0, exhaustive search for regression subset selection) and `doMPI` (v.0.2.2, for parallel computing via MPI). If the hardware does not support MPI, the `foreach` package (v.1.4.4, automatically loaded by `doMPI`) can alternatively be combined with `doParallel`. Finally, the `ncdf4` package (v.1.16.1) can be useful for preprocessing gridded weather data in NetCDF-format within R, it however requires a NetCDF system library.

95 Further recommendations for preprocessing and output visualization are the open source programs GDAL (v.3.0.4, https://gdal.org/index.html), GRASS GIS (v.7.8.3, https://grass.osgeo.org/), and GMT (Generic Mapping Tools, v.6.0.0, https://www.generic-mapping-tools.org/).

2.1.2 Input data

Any ABSOLUT application needs a domain divided into many spatial subunits, henceforth called districts, for which there
100 are individual crop yield time series and monthly weather data available. Longer yield time series are also required, and a minimum requirement can be set to automatically exclude counties lacking enough data points from the calculation. Large numbers of districts (ideally more than 100) and years with yield data (preferably more than 20) are required for the selection of valid regression feature combinations (cf. sections 3.1.2 and 3.3).

Monthly weather data are needed for each district and should spatially correspond to the agricultural areas within the them.
105 The weather data should start at least one year before the yield records start and end not earlier than with the growing season of the final year covered by the yield data, otherwise not all yield information can be considered. In detail, the modeller fixes the last calendar month whose weather data are to be considered for the growing seasons, and the model will evaluate weather–yield correlations in the twelve-month periods ending with this month. For example, if there are yield data for the years 1996–2015 and this "cut-off month" is set to May, the weather data must cover the period from June 1995 to May 2015
110 to utilize the complete yield data.

As it is not necessary to include the very end of the growing season (the actual harvest dates shift between years anyway), a timely cut-off and weather data reaching far enough into the current year can be used for pre-harvest yield forecasts (cf. Schauberger et al., 2017b, regarding the effects of shortened weather input). Eventual additional input of seasonal weather forecasts or climate scenario data may serve as basis for extended forward guidance or climate change impact studies.

The actual input files and directories to be provided (green elements in Fig. 1) consist of:

- `absolutcontrol.dat`, a short control file containing human-readable, UTF-8 encoded text to be edited before program execution.

- `yield-indat.csv`, a CSV file with the lines holding unique district–year combinations and the associated yield observations from a number of different crops. It does not matter if cities without crop yield data are missing from this list, and the program can also deal with data gaps (single missing numbers indicated by NA-strings).

- Another CSV file named `crop-areas.csv` after its contents, the per-district crop areas for the same crop species.

- A directory (here called `DistrictWeather`) filled with the monthly weather data for the agricultural areas of all district level units. For each unit there is an ASCII file whose name contains the district ID. In each of these the months correspond to lines with the meteorological variables in columns.

The file `absolutcontrol.dat` is the "user interface". By editing it the user can set the paths to weather data directories, weather variables to be considered (a subset of the variables given in the district weather files), the crop species to be modelled, time limits for optionally excluding recent years from the input, the last month of weather data to be considered before harvest (cut-off), and the minimum requirement of years of historic yield data per district.

2.2 Specifics of the example application

The spatial basis of the example are Germany's 401 district-level subdivisions (including the city states of Hamburg and Berlin as single units) as they existed on 1st January 2018. The time frame was principally determined by the crop data covering the years 1999–2019, yield predictions for 2020 were however possible through more recent weather data.

2.2.1 Germany as test bed for agricultural modelling

According to the online database of the Federal Statistical Office of Germany (https://www-genesis.destatis.de, last access December 2020) an area of $117\,630\,\text{km}^2$ was used as cropland in 2016, 32.9 % of the total country area of $357\,578\,\text{km}^2$. With annual production rates of approximately 20–25 million tons of winter wheat and 80–100 million tons of silage maize Germany ranks among the top staple crop producers of the European Union. The croplands are dispersed across all parts of Germany only seriously intermitted by mountainous areas and urban agglomerations. Regional yield time series are therefore abundantly available, for some crops from even more than 300 of the 401 districhts.

The moderate climate allows autumn sowing for many cultures, nowadays average January temperatures range above 0°C in most parts of the country, and snow accumulation is usually limited to the higher elevated regions ($> 500\,\text{m amsl}$). Typical summer temperature averages are just below 20°C, and early heat periods can be harmful for cereals. A gradient of continentality is associated with more pronounced temperature extremes in the eastern and southern parts of the country, but elevation is more important for the local climate, especially for precipitation: The national long-term average of $789\,\text{mm year}^{-1}$ (DWD 2020) is geographically differentiated between the lower parts of Eastern Germany receiving less than $600\,\text{mm year}^{-1}$ and

coastal and higher elevated regions in the west where $1000\,\mathrm{mm\,year^{-1}}$ and more are not an exception. Extreme rainfall events and hailstorms can affect agriculture in all parts of the country but have been observed more frequently in the south.

The most important factor for different cropping conditions is however a highly differentiated soil landscape: region-specific orogenetic and erosive processes including a number of glacial formations in Northern Germany and near the Alps provided

150 the full spectrum of substrates from clays to sands and peat soils (cf. European Soil Bureau Network, 2005; Richter et al., 2007; Hennings, 2013). The challenging variety of regional cropping conditions and a decent availability of weather and yield observations qualify Germany as test environment for any kind of weather-related crop yield modelling.

2.2.2 Primary data from external sources

The district geometries were taken from the official 1 : 1 million digital map of administrative areas, status 1st January 2018,

155 provided by the Federal Agency for Cartography and Geodesy (https://daten.gdz.bkg.bund.de/produkte/vg/vg1000_ebenen_ 0101/2018/vg1000_01-01.lamgw.shape.ebenen.zip, last access in August 2020).

Crop yield data were provided by the Statistical Offices of Germany through regional statistics table 41241-01-03-4 *Erträge ausgewählter landwirtschaftlicher Feldfrüchte* (yields of selected crops) accessible via https://www.regionalstatistik.de. A download made in May 2020 contained national, state, and district level data for ten crop species in the harvest years 1999–2019.

160 Additionally, table 41141-02-02-4 *Anbau auf dem Ackerland ... nach Fruchtarten* (crop growing areas ... by crop type) was obtained to use the areas reported for 2016 as weighting factors in aggregating the estimated district yields to state and national averages.

Monthly weather data were provided by the German Weather Service (DWD) through their CDC OpenData portal. Point of departure were the 1-km grid products available via http://opendata.dwd.de/climate_environment/CDC/grids_germany/

165 monthly/. Last access was in September 2020 to complete continuous time series from January 1991 to August 2020.

The spatial distribution of agricultural areas within the districts has to be taken into account for determining the locally relevant weather conditions, especially for districts including both an agricultural lowland and a mountainous part (Conradt et al., 2016). The 2012 CORINE Land Cover data were used for this purpose, namely the 100-m raster product at version v20 (release date 1 May 2019) obtained from the Copernicus Land Monitoring Service via https://land.copernicus.eu/pan-european/

170 corine-land-cover/clc-2012?tab=download.

2.2.3 Preprocessing and actual input data

A special preprocessing action was required due to a history of district reorganisations (major ones took place in Eastern Germany in 2007, 2008, and 2011). For some federal states the original yield table contained data still based on historical district geometries while pre-reorganisation data from other states had already been resampled to the current geometries. A

175 spatial homogenisation adjusted everything to the 2018 administrative geometries. Special care was taken where old districts merged into new ones differed in their agricultural area shares or the delineation of new districts was only partly based on former boundaries.

Spatial resampling of weather data to the agricultural areas within the districts is a preprocessing effort usually required for any application of ABSOLUT. One possibility would have been interpolating weather station data to area centroids (cf. Conradt et al., 2016), but here weighted averages of the DWD grid cells overlapping the croplands were used. First, 401 binary raster maps (one per district) were made with cell values of one indicating cropland (CLC code 211 = non-irrigated arable land) located within the district and zero elsewhere. These maps were then resampled by averaging the cell values to the coarser 1-km grid of the DWD weather data in another map projection. For that operation the `gdalwarp` utility of GDAL was used with the `-r average` switch. The output grids were interpreted by an R script as weight matrices for resampling the gridded weather data. This worked for practically all districts, even cities with small shares of agriculture. Only five districts in the vicinity of the Alps, the city of Suhl, and the forest-dominated mountainous domain of Siegen-Wittgenstein contained less than 1 km^2 assigned to CLC land use code 211. Class 231 (pastures) was used instead in these cases assuming a uniform interspersal with patches of cropland in the real landscape.

As these preprocessing steps are not part of the algoritm as such a directory with district weather data, the two CSV files, and a preconfigured `absolutcontrol.dat` are readily provided along the program code (Conradt, 2021b). It should be noted that it is not necessary to reproduce the very weather variables present in the example input to run the model; the example application resorts to mean air temperature (tav), precipitation (pr), and sunshine duration (sund) as selected in `absolutcontrol.dat` while any alternative combination of monthly variables would be technically possible.

3   Methods

The R code of ABSOLUT consists of five programs that have to be run in their naming order: from `100_absolut.R` to `500_absolut.R` in steps of hundreds (Conradt, 2021a). The sectional workflow, delineated in Fig. 1, is partly owing to different stages of code development and was kept to allow for checks into the intermediate output/input files. The following subsections describe the purposes and main features of these programs; for implementation details the source code should be examined as such.

3.1   Program 100 – "the prospector"

Program 100 is principally blunt exhaustive input variable testing for obtaining the "best" multiple linear regressions. Although the results are far from optimal and cannot be used for predictive purposes directly (see section 4.1.1 with Fig. 3) it contributes to narrowing the search window for regressions of higher predictive capability.

3.1.1   Initialization, preparation of weather input features

It starts with loading the libraries and reading a control file named `absolutcontrol.dat` whose contents are transferred into a couple of control variables. This information includes paths to a directory of input weather data (one file per district) and an empty one for time-aggregated weather variables to be derived and used in the regressions. The names of the original weather input files are registered, and the yield data in `yield-indat.csv` is read into memory. The official yield table may
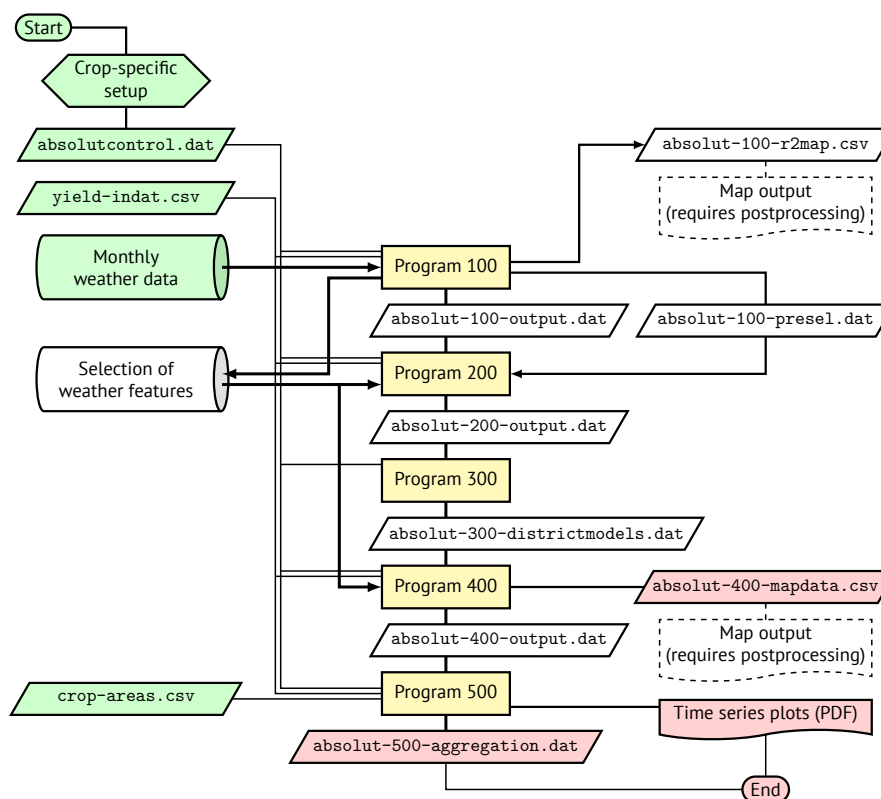
Figure 1. Flowchart of the ABSOLUT programs (yellow boxes) with input, intermediate, and output data. Coloured in green are the inputs which have to be provided to run the programs, any other data will be calculated. Principal outputs are tinted in red.

hold data for different crop species but only one of these selected via the control file is to be considered; modelling several
210   species requires several runs of the full program sequence.

The program loops over all districts for which both climate and yield data are present. Districts are identified through five-digit IDs, a case-specific setting that might need readjustment for other established key formats. The number of considered districts will usually be further reduced by the minimum requirement of yield data for the currently selected crop. Setting this minimum too low will include unreliable regressions while a high setting will unneccessarily throw out a lot of otherwise
215   useful districts; it usually requires balancing between these downsides based on the actual data availability: Longer time series allow for higher thresholds.

The "last month of weather data to be considered" to be set in the control file should be a calender month towards the end of the average growing season of the crop to be considered. Within each 12-months period ending in this month time aggregates of the chosen weather variables are calculated for periods of two to six months, these are called *weather features*. Using
220   all possible start months this means 11 different two-month features per variable, ten three-month features, nine four-month features and so on, in total 45 features per weather variable, each with one value per year. To exclude single-month features or

aggregations over more than six months was a somehow arbitrary (and possibly sub-optimal) design decision taken for limiting the number of features.

The weather features are named by combining the meteorological variable acronym with two-digit numbers of the start and
225 end months, thus pr1203 accordingly identifies precipitation in the four months from December to March. For each district considered a file with a wide table of all weather features is stored in the designated directory.

3.1.2 Naïve exhaustive search and feature selection

For each district all possible combinations of three different weather features chosen from the pool generated in the previous step are used as input to multiple linear regressions targeting the available yield data. The three best-fitting feature combinations
230 (determined by Pearson's r) are found by exhaustive search. In addition, out-of-sample predictions are evaluated: Yield data of single years are subsequently removed from the input and the search for the best regressions is repeated based on the remaining weather–yield time series. Again, the best-performing feature combinations are recorded inclusive their yield predictions for the years with censored observations.

The output from this excersize is collected in two files: `absolut-100-output.dat` containing virtually everything
235 including the weather aggregates and the coefficients of determination for both the full data regressions (R^2) and the out-of-sample predictions (R^2_{val}), and `absolut-100-r2map` containing part of the results conveniently formatted for GIS import and mapping.

As already mentioned, the best-fitting feature combinations found here cannot be directly used for reliable yield predictions, because the per-district information is not sufficient to both select a valid combination of regressors and at the same time
240 estimate their coefficients; the textbook calculation of degrees of freedom and standard error only applies if there is no extra freedom in choosing the regressors. This is why the exhaustive search for optimal regressions could be called naïve in the first place.

The better purpose of program 100 is however to subset the pool of all possible regression features to those very probably containing predictive power, and this is based on counting the number of occurrences of weather features among the "winning"
245 combinations. So why exactly three features? Why always the three best-fitting regressions on the complete time series? Once more, these are somewhat subjective choices. It is assumed that the preferred features will be similarly preferred for any multiple regression with 3–5 input features while regressions with only one or two features would suppress potential combination effects and unnecessarily limit the number of samples.

Accordingly all weather features used in the best-fitting regressions are sorted along their frequencies of use, and a relevance
250 cutoff is determined based on binomial probabilities – features need to have been selected more often then they would have occurred by pure chance with 99.9 % probability; the number of features above this threshold is henceforth denoted q. Details of the calculation are explained for the case study example (see section 4.1.1). Another small output file listing the q features, `absolut-100-preselection.dat`, is produced. Finally the district files containing the wide tables of weather features are rewritten, now constrained to these features as only those will be used in the remaining working steps.

3.2 Program 200 – "the workhorse"

Program 200 effects the highest computational burden, using parallelisation (as implemented for MPI or via the `doParallel` package) is stongly recommended. Again, complete sets of possible input feature combinations are investigated for their predictive power in multiple linear regressions, but in contrast to program 100

- there are now four weather features in addition to the time feature (calendar years) selected as input variables,

- these are chosen from the preselected set of significant features as identified in program 100,

- only predictions for single years whose yield data are censored from the regression equations (leave-one-out) are calculated, and

- the `leaps` package is skipped in favour of an explicit loop over the possible input variable combinations,

because here not only a selection of probably informative features is demanded but an evaluation of the predictive skills of the multiple linear regressions a couple of which will be finally used in the model. Therefore the regressions applied here follow the template of Eq. (1) in section 1 with four weather features. Fixing the number of weather features to $d = 4$ might not always be the best choice (see section 4.3), thus future versions of ABSOLUT will probably offer more flexibility.

The output of program 200 is a big table, its lines correspond to all possible input feature combinations, so their number is

$$\binom{q}{d} = A \tag{2}$$

with q indicating the number of preselected weather features as provided by program 100. The first column contains the line numbers, and the columns of the table body correspond to the districts which could be considered according to data availability, their IDs are listed in an additional header line. The table body is filled with the Pearson correlations (r values) between reported yields and the out-of-sample yield predictions from the regressions (applications of all feature combinations to all districts). A former development version had the table filled with squared correlations, but these "R^2 values" erroneously indicate predictive power when there are in fact stronger negative correlations.

3.3 Program 300 – "the gold pan"

The logically following task addressed by this program is determining the optimal regression model for each spatial subunit. Simply choosing the feature combinations producing the highest correlations per district ("local heroes") would however be misleading because there is not enough validating information in a single yield time series; any feature combination needs to be cross-validated by above-average performances in many districts.

The solution currently implemented is a compromise between resorting to a single "generalist" combination with the best average performance and a zoo of isolated regressions: The seven on-average best-performing combinations are determined globally, and in each district the locally best-performing regression out of these is implemented.

There are in fact a number of alternative selection methods implemented in the code, however deactivated in the distributed version. Among the tested methods whose results are reported for the example application (Tables 2 and 3) are:

Significantly elevated r: Inclusion of any combination with a significant shift towards higher-than-average r. This is determined by a rank-based approach; combinations which are significantly overrepresented in a the upper halves, thirds, quarters, and so on (down to five per cent) of the district rankings are globally selected. This may yield several thousand "allowed" feature combinations.

Optimal rank sum of seven: Out of the combinations validated in the previous approach seven are chosen such that the rank sum of their best per-district correlations is minimized.

Global and local heroes: The 42 globally best performing combinations with the highest average r values are merged with those working exceptionally well in smaller subsets of the districts (down to 5 % of the districts). The idea behind is to account for special conditions in certain landscapes.

Best global: Finally the set of combinations is simply confined to a certain number of global top performers, this is what is currently implemented for a number of seven.

Whatever confinement is used, the general principle is always subsetting the A possible combinations to a subset of C "allowed" combinations. As for each district the locally best-performing feature combination belonging to this subset is used, only c combinaions will be finally used, and it holds $A > C \geq c$. And $C = 7$ of the example appication is chosen somewhat arbitrarily – why not use the 3, 21, or 42 best combinations? Or an alltogether different selection approach?

3.4 Programs 400 and 500 – "crucible and mould"

What remains is fitting the selected regression equations to the complete time series of observed yield data and predicting the yields of the current harvest season using recent weather data or calculating yield scenarios from climate scenario data. This is done by program 400. Leave-one-out estimations for observed yields are also repeated, because the complete results for all possible feature combinations are not recorded by program 200.

Program 500 reads the yield predictions on district level and the observed yields of the past plus the file `crop-areas.csv` containing the crop areas per district. Currently static areas according to the agricultural survey of a certain year are applied, changes of cropping intensities over time might be considered in future versions. This information is used to aggregate the district yield predictions through weighted averaging to predictions for the full modelling domain – the actual implementaion makes intermediate aggregations for federal states and derives national level predictions in a subsequent aggregation step. Among the outputs are time series plots of observations and predictions for the spatial aggregations, Fig. 2 shows an example.

4 Application to Germany: setup, results and discussion

The district-level administrative regions of Germany are identified by a five-digit key (*Regionalschlüssel*, "regional key", abbreviated RS) which was adopted for the application, e. g. in the file names for weather variables and features. Three principal weather variables were selected: average temperature, precipitation, and sunshine duration. The minimum of yield data per
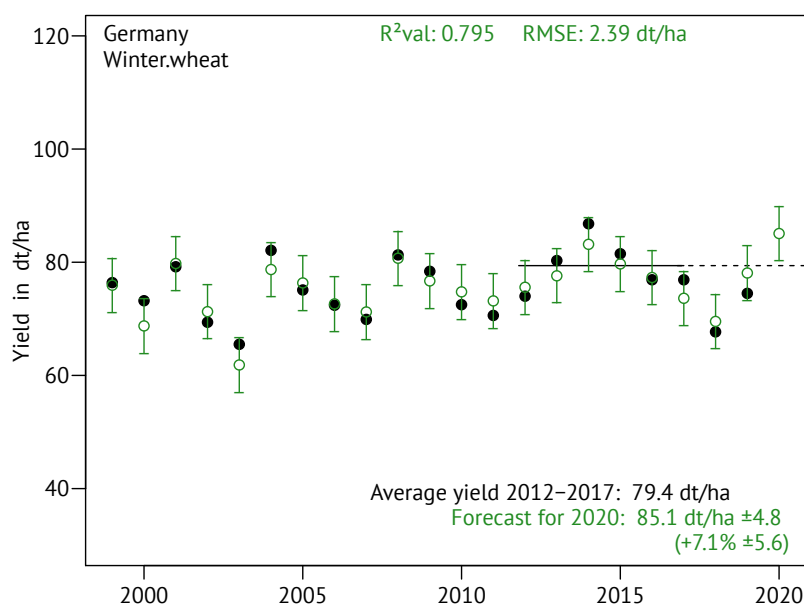
Figure 2. Average winter wheat yields in Germany according to the official statistics (black) and their predictions (green) with confidence intervals. For the year 2020 the official value is not yet available.

district had been set to 17 because the example dataset was limited to 21 years (1999–2019) and a higher requirement would have excluded many districts with incomplete observations.

The principal test crop was winter wheat, the last month for weather input before each year's harvest (typically in July or August) was set to June. For the secondary test crop silage maize the weather input season was set to end by August; maize harvest may occur late in the year but growth stagnates in autumn.

4.1 Observations along the workflow for winter wheat

4.1.1 Running program 100

With the three weather variables $3 \cdot 45 = 135$ weather features were generated (cf. section 3.1.1). This meant $\binom{135}{3} = 400\,995$ different regressions per district to test which required a couple minutes using 24 CPUs in parallel.

Figure 3 shows maps of the goodness-of-fit for winter wheat, visualizations of the CSV output of program 100. There is a stark difference between the performances of the chosen regression models depending on wether they are applied to the same data that was used for selecting them or if the observed yield value of any single year to predict is censored from the input (out-of-sample validation). The average R^2 values are 0.784 and 0.124, respectively; this exemplifies that the regressions selected in this step cannot be immediately used for predictions.

Which weather features do however appear significantly often in these regressions so that the subsequent assessment can be confined to them? What is their number, q? The algorithm requires that the number of occurrences exceed a frequency
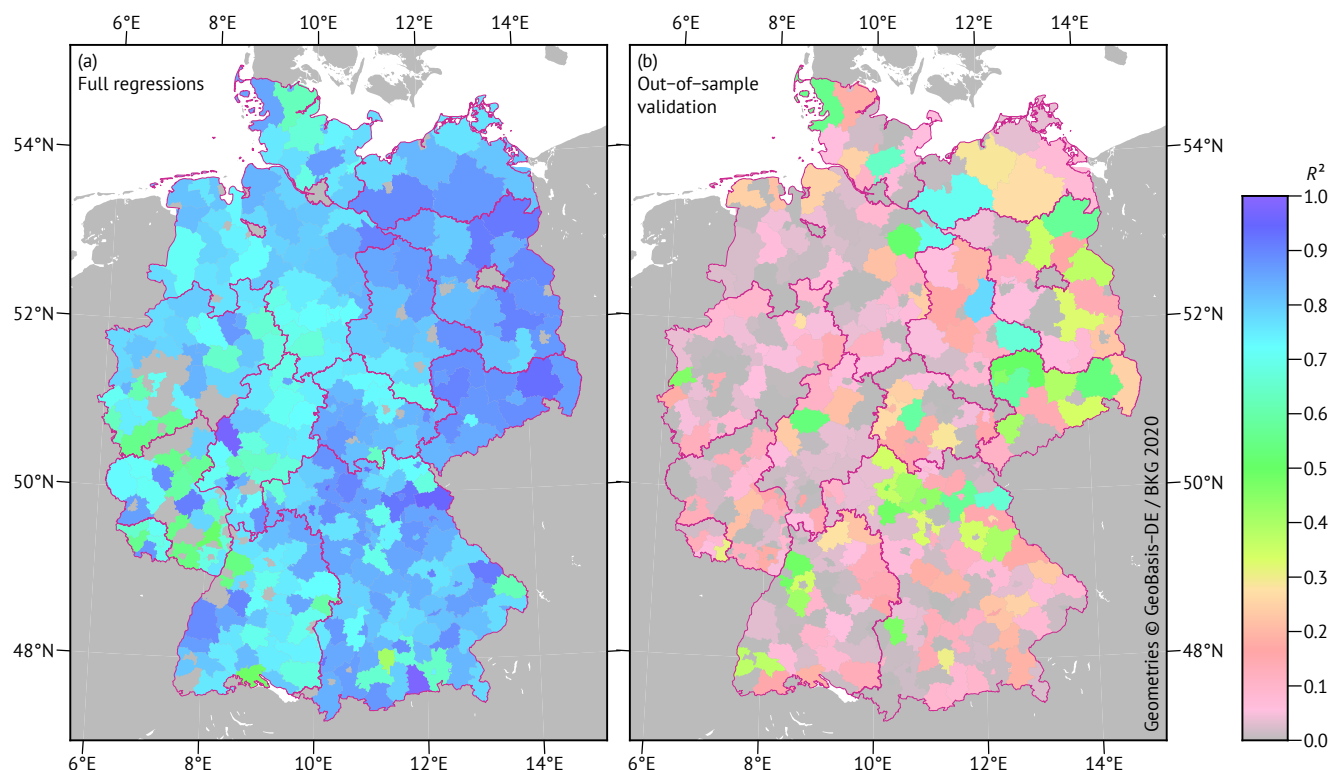
Figure 3. Spatial distributions of coefficients of determination for "optimal" winter wheat yield regressions (based on year and three weather aggregates) across Germany for the years 1999–2019. (a) Fitted to the entire observational data. (b) Squared Pearson correlations between predictions and observations; predictions made without the actual observed value used in parameter estimation (leave-one-out validation).

expected by pure chance with 99.9 % confidence. If there were pure noise in the data, each of the 135 features originally provided would turn up with a constant probability of $p = \frac{3}{135} = \frac{1}{45}$ per regression sample. With a finite number of samples their frequencies follow a binomial distribution. In the winter wheat case there were $n = 20\,892$ samples, thus the expectation value

335 for any weather feature in the noninformative case would be $E(x) = np = 464.267$ occurrences with an expected standard deviation of $E(\sigma) = \sqrt{np(1-p)} = 21.306$. The number of occurrences not to be exceeded in 99.9 % of the cases would be $P_{999}(n, p) = 531$; consequently 37 weather features called more frequently in the winter wheat case were selected, their frequencies are shown in Fig. 4

The average temperature towards the end of the growing season (temperature aggregates for May and June and March to

340 June) stick out clearly; this is in full agreement to the often described temperature sensitivity of wheat during and after anthesis (Akter and Islam, 2017; Farooq et al., 2011; Schauberger et al., 2017a). Consequently more than half of the aggregates shown in Fig. 4 are temperature averages while less than one third are precipitation depths and only five of them are sunshine durations.
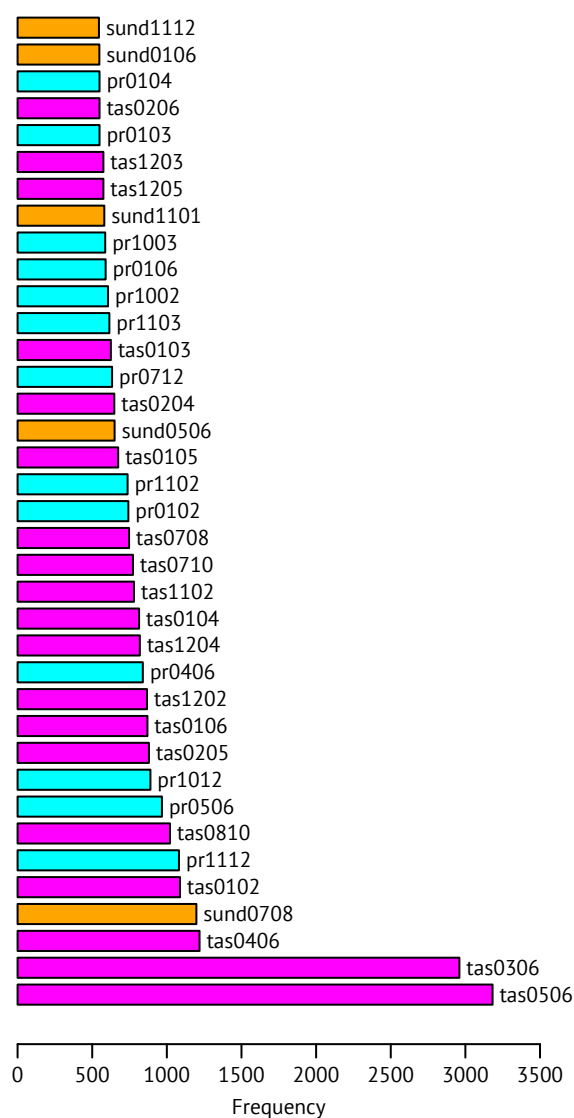
Figure 4. The 37 weather features used most frequently in the regression equations optimally describing observed winter wheat yields in the district-level administrative units of Germany. Their names consist of variable abbreviations (tas = temperature, pr = precipitation, and sund = sunshine duration) and two-digit numbers of the start nd end months of their time aggregation. The length of the bars (Frequency) indicates the number of the best-fit regressions containing the respective variable, the colours correspond to the meteorological variables.

Table 1. Combinations of input weather aggregates performing best across all districts and the average Pearson correlation of their out-of-sample predictions in the winter wheat example for Germany. Output of program 300.

Input weather aggregates				$\bar{r}$
tas0506	sund0506	tas1202	tas0710	0.479
sund0506	tas0104	tas0306	pr1103	0.472
tas0710	pr1102	tas0104	tas0306	0.468
sund0506	tas0204	tas0710	tas0306	0.463
tas0102	tas0506	sund0506	tas0710	0.463
sund0506	tas1102	tas0306	tas1204	0.461
tas0506	sund0506	tas1202	pr0103	0.458

4.1.2 Running programs 200 and 300

In the winter wheat example the big output table of program 200 has $A = \binom{37}{4} = 66\,045$ numbered lines (below a header) for

345 all possible input variable combinations, and 325 columns representing the districts for which enough yield data were available. Negative correlation coefficients could be found in no less than 28 % of the table cells. The negative extreme in the example is a near perfect anticorrelation of $r_{\min} = -0.964$.

Given the number of approximately 20 out-of-sample regression estimates behind every correlation coefficient, the computational demand is significant. The winter wheat example took about 2 hours using 112 CPU cores in parallel.

350 The repeated estimation of parameters for the $q = 37$ weather features (in varying combinations) per district is still rather noisy given only approximately 20 data points from each observed yield time series. Consequently 304 different combinations make up the best-performing regressions in the 325 districts of the winter wheat case, and despite an average correlation of $\bar{r} = 0.788$ they perform better than the regressions from the initial "naïve" search (Fig. 3b) but still very poorly when used for predictions outside the training period.

355 The input feature combinations leading to the highest average correlations over all spatial subunits as determined by program 300 are shown in Table 1. The individual district rankings among these finally decide which one is applied where (and if all of them are used at all – like in this example with $C = c = 7$). For winter wheat 13 different weather features are finally included, and the average correlation of the so determined district models is $\bar{r} = 0.597$.

4.1.3 Running programs 400 and 500

360 Program 400 – utilizing the district-specific regressions determined in the previous step for the quasi-out-of-sample yield estimations and target-year predictions – runs within seconds through all districts. Their spatial time-series aggregation towards Germany's national winter wheat yields, computed by program 500, is shown in Fig. 2.

Figure 2 includes the R^2_{val} and root mean square error (RMSE) for the predictions: 0.795 and 2.39 dt ha^{-1}, respectively. The RMSE is used as basis for the 95 % confidence interval indicated by the green error bars. Assuming a normal distribution of the residuals, the interval should be ±1.96 RMSE, but here a factor of 2.00 was used to account for the fact that all data have been used to determine the weather aggregate combinations in the regressions so that not a pure out-of-sample approach had been applied despite the regression parameters being estimted without the observed values in the target years. The rough and rather intuitive confidence correction serves the needs of this example framework but should eventually be refined.

4.2 How restrictive should the final selection be?

Tables 2 and 3 show the results for winter wheat and silage maize, both for the target years 2018 and 2019 whose yield predictions could already be compared to the final results of the official statistics. Only the yield observations preceding the target years were used for feature and regression selections, thus the predictions of the target years are truely out-of-sample while the leave-one-out predictions of the preceding years are only quasi out-of-sample. It must however be noted that both target years (chosen just because they were at the end of the observation time series) accidentally coincide with a historic drought over Germany; the linear extrapolations from smaller weather and yield fluctuations of the past are challenged by higher-than-average error levels.

The prognostic overconfidence from smaller residuals in the past can be traced in the RMSE columns: The more restrictive the selection of weather aggregate combinations, i. e. the smaller c, the larger the average prediction error along the time series of national yields RMSE$_{\mathrm{g}}$. While this can be expected – less combinations mean less flexibility in fitting the real-world behaviour – RMSE$_{\mathrm{p}}$, the averaged prediction errors for the federal state yields in the target year generally reaches its minimum if only few regression alternatives are used. This rule is violated in the 2018 silage maize case (Table 3), and RMSE$_{\mathrm{p}}$ shows also minima for the least restrictive "significantly elevated r" method, but it is clear that the lower RMSE$_{\mathrm{g}}$ values connected to flexible regressor choices systematically overestimate their predictive performance.

The seven best globally performing combinations have been defined as standard approach despite the suboptimal results of this method in the 2018 silage maize case. This can be justified by comparably low errors in the true out-of-sample cases and a relative parsimony of different feature combinations. Parsimony of different weather aggregates included adds to the latter, because the seven combinations of the best global approach require only about 13 different weather features while the seven combinations of the optimal rank sum approach regularly contain 20 or more.

However, this currently adopted standard might be revised in the future, cf. the minimal errors achieved for the target year forecasts of winter wheat yields using more than 200 district-specific combinations in the "significantly elevated r" approach (Table 2). The solution chosen here might not be optimal but is a viable selector for demonstrating the power of the exhaustive search approach.

Table 2. Prediction performances for different approaches in the global preselection of weather aggregate combinations – Winter wheat from 1999 to 2018 and 2019. The number of preselected combinations is C while c is the number of different combinations finally used. The performance indicators R^2_{val} and $RMSE_g$ refer to the time series of national quasi out-of-sample predictions while $RMSE_p$ is the true out-of-sample RMSE for the federal states and Germany in the respective final year.

Preselection approach	Winter wheat 2018					Winter wheat 2019				
	C	c	R^2_{val}	$RMSE_g$	$RMSE_p$	C	c	R^2_{val}	$RMSE_g$	$RMSE_p$
Significantly elevated r	7656	293	0.924	1.58	7.61	13216	291	0.873	1.88	5.89
Optimal rank sum of seven	7656	7	0.842	2.18	7.74	13216	7	0.784	2.43	6.19
Global and local heroes	109	94	0.882	1.88	7.88	97	83	0.837	2.12	6.43
Best global: Top 96	96	85	0.886	1.85	8.30	96	82	0.831	2.16	6.74
Best global: Top 42	42	40	0.872	1.96	7.98	42	40	0.819	2.23	6.87
Best global: Top 21	21	21	0.835	2.20	8.05	21	21	0.820	2.23	6.64
Best global: Top 7	7	7	0.807	2.38	7.70	7	7	0.800	2.35	6.54
Best global: Top 3	3	3	0.784	2.51	8.02	3	3	0.710	2.85	6.75
Best global: Number 1	1	1	0.701	2.93	8.33	1	1	0.665	3.10	7.30

Table 3. Prediction performances for different approaches in the global preselection of weather aggregate combinations – Silage maize from 1999 to 2018 and 2019. Column headings as in Table 2.

Preselection approach	Silage maize 2018					Silage maize 2019				
	C	c	R^2_{val}	$RMSE_g$	$RMSE_p$	C	c	R^2_{val}	$RMSE_g$	$RMSE_p$
Significantly elevated r	8056	277	0.921	10.23	49.68	7579	278	0.935	9.12	43.12
Optimal rank sum of seven	8056	7	0.857	13.52	56.84	7579	7	0.916	10.32	38.47
Global and local heroes	123	95	0.899	11.44	50.09	138	97	0.922	9.87	41.30
Best global: Top 96	96	80	0.903	11.23	46.35	96	79	0.920	9.99	36.95
Best global: Top 42	42	41	0.888	11.99	49.53	42	41	0.917	10.16	36.05
Best global: Top 21	21	20	0.889	11.93	50.43	21	21	0.909	10.64	37.78
Best global: Top 7	7	7	0.831	14.48	61.15	7	7	0.908	10.68	34.11
Best global: Top 3	3	3	0.743	17.68	66.76	3	3	0.900	11.14	32.90
Best global: Number 1	1	1	0.656	20.44	76.58	1	1	0.896	11.34	33.01

Table 4. Dependence of quasi out-of-sample prediction fidelity on the number d of input weather features. A is the number of possible feature combinations, $\bar{r}$ the average correlation of the finally applied combinations, and R_{g}^2 and $\mathrm{RMSE_g}$ are coefficient of determination and root mean square error for the time series of spatially aggregated (national) yields.

Crop	d	A	$\bar{r}$	R_{g}^2	$\mathrm{RMSE_g}$
	2	666	0.489	0.593	3.38
	3	7770	0.583	0.800	2.37
Winter wheat	4	66045	0.597	0.795	2.39
	5	435897	0.591	0.798	2.38
	2	496	0.634	0.864	13.13
	3	4960	0.639	0.882	12.10
Silage maize	4	35960	0.659	0.910	10.55
	5	201376	0.664	0.914	10.32
	2	231	0.644	0.640	2.59
	3	1540	0.656	0.646	2.58
Winter rape	4	7315	0.650	0.646	2.59
	5	26334	0.664	0.721	2.30

4.3 The number of climate aggregate features

To fix the number of climate aggregate input features to $d = 4$ had been a fundamental design decision. Table 4 shows the
395 goodness-of-fit metrics for two to five climate input variables for winter wheat, silage maize, and winter rape; now the complete available input data (with the observed yields up to and including the year 2019) are used.

For winter wheat, three weather aggregates would have been sufficient, three show higher performance than two in all three indicators – the average correlation of predicted district yields with local observations $\bar{r}$, the coefficient of determination for the time series of national yield averages R_{g}^2, and the national forecast error level $\mathrm{RMSE_g}$. With four or five input weather
400 aggregates the results deteriorate slightly.

The silage maize results support four weather inputs. Although there are monotonous performance improvements for each additional weather aggreagate regressor, the advance with going from four to five is too small to justify the additional computational burden. Winter rape yield estimations show a somehow opposite behaviour: While there is practically no performance change from two to four regressors, introducing a fifth one causes an upward jump. Here it probably plays a role that, as only
405 22 weather aggregates are preselected, even five variables mean just 26 334 different combinations, much fewer than there are for wheat and maize with four variables.

As using four weather input variables hardly spoils the predictive power in cases where two or three weather inputs would suffice but clearly enhances the performance for silage maize while five variables often afford extreme computational demand

– for winter wheat it took 112 CPU cores running 14 hours – four weather aggregates were set as standard. Making the number
410 of weather inputs flexible remains an option worth considering for future model versions, though.

4.4 Selection of principal weather variables

As already mentioned above, mean air temperature, precipitation, and sunshine duration had been chosen as meteorological
input. Other variables have not been systematically investigated for their predictive potential due to a number of reasons.
Temperature and precipitation are fundamental climatological variables definitely governing plant growth, they should always
415 be considered. Aggregates of daily maximum or minimum temperatures are considerable alternatives for mean temperature,
but combinations of different temperature "flavours" should be avoided because of strong multicollinearity. Experimenting
with temperature differences (diurnal temperature variations) would also be an option, but these are correlated with radiation.

Sunshine duration, chosen as proxy for photosynthetic active radiation, could also have been replaced by direct radiation
measurements or cloud coverage, another proxy. Tests using global radiation instead of sunshine duration showed however
420 gradual performance losses which may be related to differences in data acquisition: There are only few stations recording
radiation directly compared to those recording sunshine duration. Sunshine duration was preferred for another reason, too –
the respective monthly DWD grid product becomes regularly available shortly after the turn of the month while the radiation
product takes several days longer to be released. Wind was also tested occasionally but crop yields seem to be rather insensitive
to it, the devastating effects of regional gales and local tornadoes hit too erratic to be captured by multi-month wind averages.
425 Data availability will generally determine the decision which variables to select. Climate scenario data may e. g. lack sun-
shine duration and provide downwelling short radiation instead. Among the DWD data, evapotranspiration and especially soil
wetness, estimated by the AMBAV model (Löpmeier, 2014; Friesland and Löpmeier, 2007), seem very promising for yield
forecasting. They were deliberately cold-shouldered here as the general applicability of ABSOLUT with purely meteorological
regressors should be demonstrated.

430 4.5 Regional performance

Using district regressions provides not only a basis for aggregate results but can also help identify spatial patterns despite the
higher noise in single district results. In contrast to major agricultural zones of the world like the North China Plain or the
US corn belt Germany is characterized by a high diversity of soil landscapes forming a distinct pattern of high and low yield
regions (Hennings, 2013; Kruse, 2016). Hence it makes a difference whether the ability to predict spatial yield patterns is
435 measured by absolute values or relative changes which is also demonstrated in the following.

The silage maize harvest forecast for the drought year 2018 was chosen due to extreme yield losses observed in Eastern
Germany. For 2019 yield changes of winter wheat, the most frequently grown crop in Germany, are presented. Unless stated
otherwise, quoted figures in this section refer to all of Germany.

Table 5. Statistical properties of 242 unweighted silage maize district yields predicted for 2018 and their observed counterparts according to the official statistics (observation).

	Prediction dt ha^{-1}	Observation dt ha^{-1}
Mean	406.4	372.8
Std devn	90.0	96.2
Maximum	765.4	612.2
Q3	465.2	433.2
Median	418.0	374.3
Q1	354.5	305.3
Minimum	95.0	124.8
Pearson's r	0.630	

4.5.1 Silage maize 2018 on district level

440   Silage maize became increasingly popular as fodder and energy crop over the last decades and is now grown on approximately two million hectares. An average yield of 441.3 dt ha^{-1} was observed in the six pre-drought years 2012–2017 (DESTATIS). The lowest yield levels are regularly observed in the federal state of Brandenburg where sandy soils are abundant, 340.3 dt ha^{-1} during the pre-drought period, while the most fertile regions like the Lower Rhine Bay landscape usually report yields above 550 dt ha^{-1}. In 2018 the regional variation expanded because the drought hit hardest in Saxony-Anhalt and the south of Bran-

445 denburg where district yields dropped below 200 dt ha^{-1} and many farmers were even confronted with complete failure.

    Table 5 summarizes the district yield predictions and the officially reported statistics for the same districts; except for a certain upward bias and an overestimated range the distribution of predictions looks convincingly similar to what had been observed. The pairwise correlation between predicted and observed district yields is however not very high and partly owing to the soil fertility pattern. Moving from absolute yields to relative changes (departures from the average district yields of the

450 years 2012–2017) spoils the picture even more: The correlation coefficient drops to 0.404 ($n = 223$; some districts excluded due to yield data gaps in the pre-drought period) indicating a mere 16.3 % of explained variance.

    Despite this high noise level, the maps of the predicted and observed relative yield changes in districts shown in Fig. 5 (a) and (b) show a general similarity of spatial patterns. There are regional misses along the coast to the Baltic Sea and in some western and south-eastern parts of the country, but the center of gravity of the strongest yield losses could correctly be located.

455   Panel (c) of Fig. 5 gives an impression of the regional distribution of prediction power through absolute RMSE values calculated from the complete record of out-of-sample prediction errors. In general, the method works fine in Northern Germany and some parts of the south, but has some issues in western to central areas. The actual 2018 errors (d) were much larger in

Table 6. Statistical properties of 280 unweighted winter wheat district yields predicted for 2019 and their observed counterparts according to the official statistics (observation).

	Prediction dt ha^{-1}	Observation dt ha^{-1}
Mean	78.93	74.03
Std devn	10.60	11.95
Maximum	100.69	103.80
Q3	86.51	82.85
Median	80.14	75.85
Q1	73.17	66.45
Minimum	32.16	39.00
Pearson's r	0.809	

many districts including those with comparably small RMSEs. It may be assumed that the training period (1999–2017) did not contain enough reference drought years; in fact only 2003 might have been comparable to some extent.

4.5.2 Winter wheat 2019 on district level

Winter wheat is Germany's most frequently grown crop covering approximately three million hectares of the agricultural areas. The national yield average of the years 2012–2017 was 79.5 dt ha^{-1} (DESTATIS).

The distributions of predicted and observed district yields for the year 2019, shown in Table 6, match reasonably well and there was also a certain general overprediction as with silage maize for 2018. The correlation between predicted and observed district yields is however a little higher, for the relative changes to the 2012–2017 yields Pearson's r reaches 0.518 ($n = 253$) which means 26.8 % of the variance being explained.

The spatial distributions of predicted and reported changes shown in maps (a) and (b) of Fig. 6 are also rather similar, except that the magnitude of observed yield losses was not captured by the prediction. On the other hand, the situation in the coastal and western parts matches the predictions quite well contrasting the poor performance for the 2018 silage maize. In the south-east the results suffer from slight overestimations, but a local gradient towards the southern boundary is generally matched.

4.5.3 Error compensation in spatial aggregates

By aggregating district results to regional averages some noise is filtered out by mutual error compensation and error levels diminish. The 2018 silage maize predictions for 242 districts expose an unweighted root-mean-square error (RMSE) of 88.2 dt ha^{-1}, their unweighted mean absolute error (MAE) is 71.3 dt ha^{-1}. The 13 state yield predictions, weighted aggregates
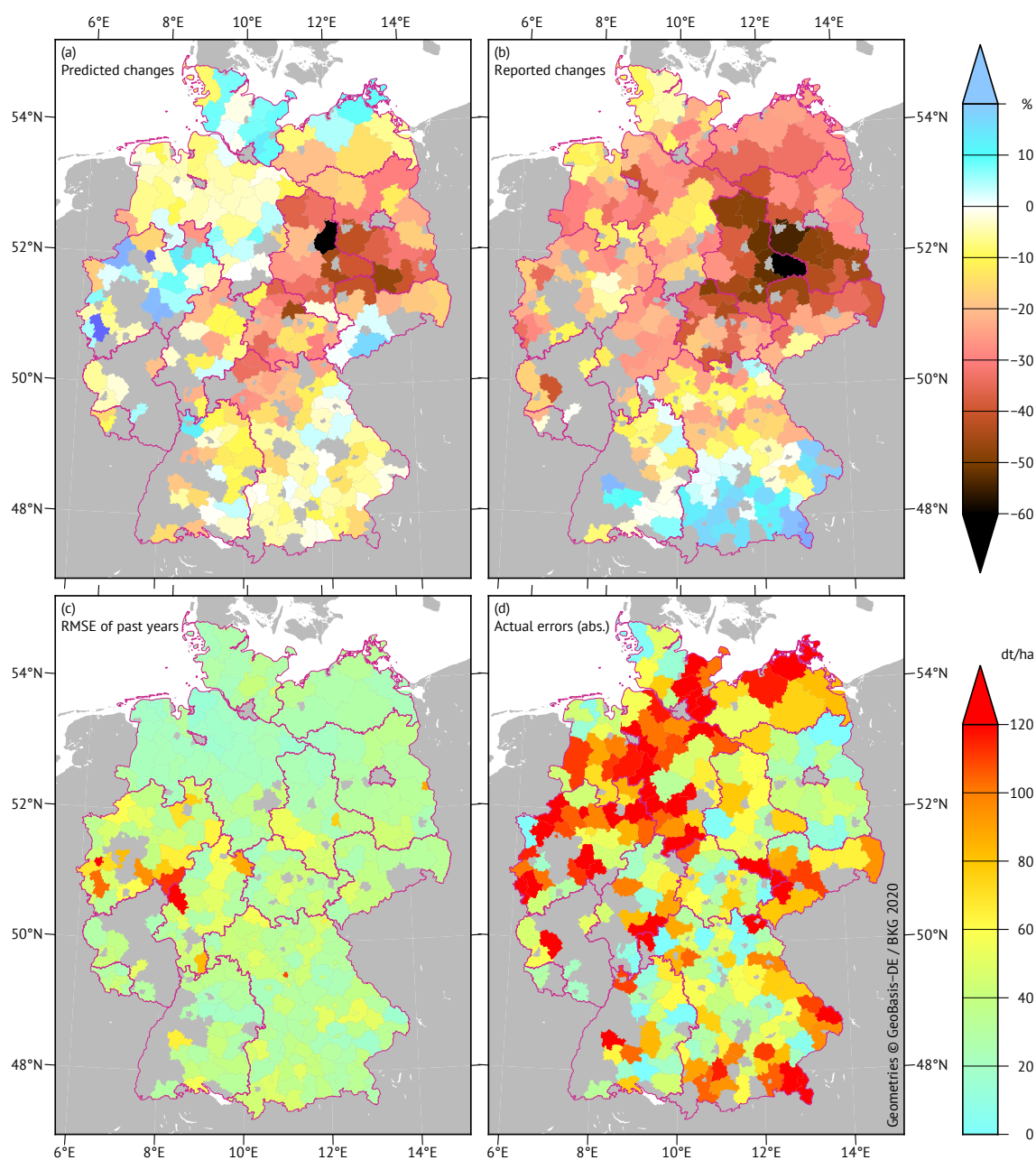
Figure 5. Spatial characteristics of the silage maize yield prediction for the drought year 2018. (a) Predicted yield changes compared to the average district yields of the years 2012–2017. (b) Observed changes according to the official statistics. (c) Root-mean-square errors (RMSE) of the out-of-sample district yield predictions for the years 1999–2017. (d) Absolute values of prediction errors for 2018. Panels (a) and (b) show relative deviations in percent (upper scale), (c) and (d) refer to absolute deviations in dt ha^{-1} (lower scale).
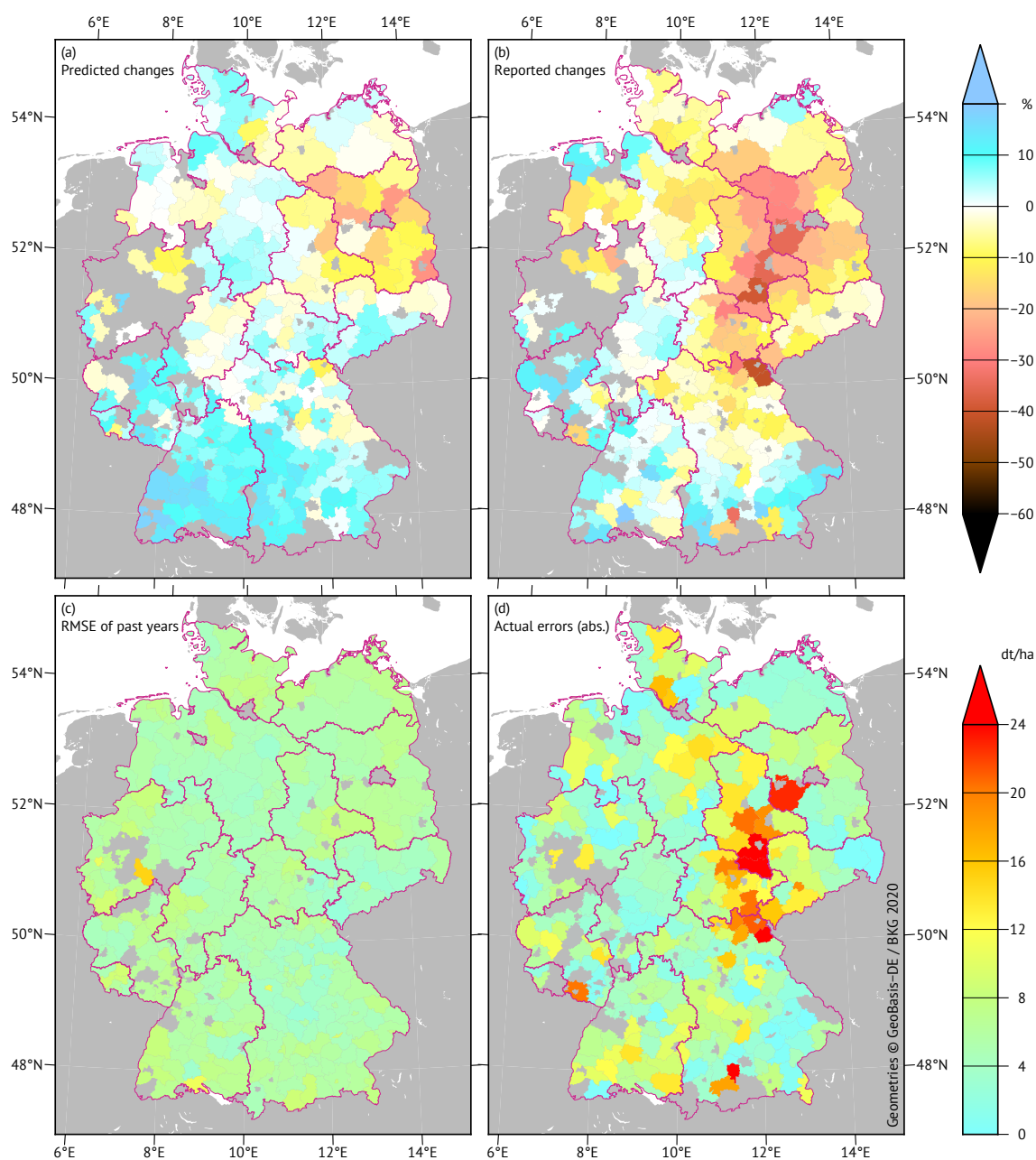
Figure 6. Spatial characteristics of the winter wheat yield prediction for the year 2019. (a) Predicted yield changes compared to the average district yields of the years 2012–2017. (b) Observed changes according to the official statistics. (c) Root-mean-square errors (RMSE) of the out-of-sample district yield predictions for the years 1999–2018. (d) Absolute values of prediction errors for 2019. Panels (a) and (b) show relative deviations in percent (upper scale), (c) and (d) refer to absolute deviations in dt ha^{-1} (lower scale).

of the district results computed with program 500, have unweighted RMSE and MAE values of 62.0 dt ha^{-1} and 56.9 dt ha^{-1}, respectively. The national yield prediction, a weighted aggregate of the state aggregates, missed the official figure by 48.8 dt ha^{-1}.

480 In the case of the 2019 winter wheat predictions RMSE and MAE of district results ($n = 280$) are at 8.6 and 6.6 dt ha^{-1}, respectively; the state level RMSE and MAE are at 6.6 and 5.1 dt ha^{-1}, respectively; and the national aggregate prediction missed the official national yield by 5.7 dt ha^{-1}. The 2018 and 2019 predictions were generally too optimistic, thus for other years without general biases better error reductions can be expected from spatial aggregations.

Predictions from linear regressions are usually associated with confidence intervals calculated from the standard errors and correlations of the parameter estimations. Hence it should theoretically be possible to spatially aggregate these estimates

485 as well and accordingly give dynamic confidence intervals for state and national predictions of individual years. However, the covariances between the district errors would have to be known for the uncertainty propagation calculation, and this is currently unachievable. The estimated error levels of district predictions hardly correspond to the errors actually observed in single years. Each year exposes a general bias, and in addition to that the district prediction errors are locally correlated.

At the moment, the confidence intervals given for district and aggregate predictions on different levels (cf. the green bars

490 in Fig. 2) are therefore based on the RMSE of the out-of-sample errors of past year predictions for the given prediction unit. Under certain conditions (extreme weather) errors do however exceed the expected level by far, often concentrated in regional clusters as shown in Fig. 5 (d). The spatial correlation of errors should eventually be further investigated.

4.6 Comparison to existing prediction approaches

4.6.1 Official in-season yield estimations

495 There are at least two institutions regularly publishing crop yield forecasts or estimates: the German Federal Statistical Office (Statistisches Bundesamt, DESTATIS) and the Joint Research Centre (JRC) of the European Comission with their MARS (Monitoring Agricultural ResourceS) activity. The DESTATIS reports with national and federal states' estimates are based on extensive field monitoring, on-site observations during growth and harvest by farmers and travelling experts. Current issues are regularly published through their webpage (https://www.destatis.de/DE/Themen/Branchen-Unternehmen/

500 Landwirtschaft-Forstwirtschaft-Fischerei/Feldfruechte-Gruenland/_inhalt.html, last access in October 2020). The MARS forecasts utilize a number of sources and predictors but seem to largely rely on remote sensing. National aggregate predictions of MARS are released via monthly bulletins offered at https://ec.europa.eu/jrc/en/mars/bulletins (last access in October 2020).

Yield predictions for crops harvested in June or July (cereals or rape) can usually be computed in the beginning of July as soon as the monthly weather data for June are available. For silage maize or sugar beets the August weather data should

505 be completed. Both conditions were constantly observed for the examples presented. Hence the preliminary national crop yield estimations of DESTATIS, regularly published in the beginning of August and at the end of September, are used for comparison. (Only the winter barley estimates are released in between and DESTATIS does not provide early indications at all for sugar beet yields.) For MARS, the annual issues 7 (typically released at the end of July) and 9 (typically released in

Table 7. National average yields for various crops: comparison of official statistics (released in the year after harvest) and near-harvest predictions of ABSOLUT and two other sources. The winter wheat figures given for MARS have been obtained by adding 0.6 dt ha^{-1} to their original soft wheat prediction.

Year	Yield	ABSOLUT			DESTATIS			MARS		
		Date	Predn	Error	Date	Predn	Error	Date	Predn	Error
	dt ha^{-1}		dt ha^{-1}	dt ha^{-1}		dt ha^{-1}	dt ha^{-1}		dt ha^{-1}	dt ha^{-1}
Winter wheat										
2020		02 Jul	85.1		03 Aug	71.9		27 Jul	76.0	
2019	74.5	ex post	80.2	+5.7	02 Aug	73.0	−1.5	22 Jul	77.1	+2.6
2018	67.7	ex post	70.2	+2.5	02 Aug	66.4	−1.3	23 Jul	71.6	+3.9
Winter barley										
2020		04 Jul	77.2		28 Aug	67.5		27 Jul	68.8	
2019	72.2	ex post	71.6	−0.4	29 Aug	72.1	−0.1	22 Jul	71.3	−0.9
2018	60.6	ex post	52.5	−8.1	24 Aug	60.8	+0.2	23 Jul	63.5	+2.9
Rye										
2020		05 Jul	55.8		03 Aug	51.7		27 Jul	52.5	
2019	50.9	ex post	51.1	+0.6	02 Aug	51.7	+0.8	22 Jul	52.8	+1.9
2018	42.1	ex post	42.3	+0.2	02 Aug	42.7	+0.6	23 Jul	44.6	+2.5
Rape										
2020		04 Jul	40.1		03 Aug	32.8		27 Jul	32.7	
2019	33.1	ex post	35.5	+2.4	02 Aug	33.8	+0.7	22 Jul	34.7	+1.6
2018	30.0	ex post	29.5	−0.5	02 Aug	28.8	−1.2	23 Jul	30.0	±0.0
Silage maize										
2020		07 Sep	434.9		24 Sep	410.3		14 Sep	400.0	
2019	390.0	ex post	399.2	+9.2	25 Sep	383.7	−6.3	16 Sep	394.0	+4.0
2018	352.9	ex post	401.7	+48.8	26 Sep	342.7	−10.2	17 Sep	361.0	+8.1

mid-September) are the seasonal counterparts to compare with. Table 7 presents the comparisons for five crops in the years 2018–2020; the reader may fill the blanks with the 2020 yields and prediction errors when the final official figures are available (expected for February 2021).

Considering the fact that the ABSOLUT results can be obtained 2–4 weeks in advance to those of the other sources the quality of its predictions of national yield averages are in the same league with DESTATIS and MARS. There were considerable overpredictions for winter wheat and silage maize, but rye results could be anticipated remarkably exact. Winter barley figures

515   are very closely matched by DESTATIS, but they do not separate winter and summer barley before end of August when the
field surveys already converge towards their official final result.

4.6.2   Weather input of Gornott and Wechsung (2016)

Experiments with district-based crop yield prediction in Germany through multiple regression using weather aggregates had
already been presented by Gornott and Wechsung (2016). In contrast to the algorithm presented here only year-on-year (YoY)

520   changes were considered saving the explicit estimation of an underlying linear trend. The study investigated different options
to couple the coefficients of the district models (paneling), an approach further pursued with cluster analysis (Conradt et al.,
2016). The nearest equivalent to ABSOLUT are therefore the independent district regressions of Gornott and Wechsung (2016),
called there "separate time series models" (STSMs), and the most fundamental difference is that all STSMs used the same set
of input variables – predefined per crop – while ABSOLUT searches for some optimal combinations.

525   How powerful are the predefined input variables in terms of prediction accuracy compared to the combinations drawn
by ABSOLUT? To answer this question, the district regressions were charged with the weather variables originally used by
Gornott and Wechsung (2016), the base trend estimation for absolute yield outputs was however left untouched to avoid side
effects from the other methodological differences. The weather variables uniformly used in the STSMs were

– For winter wheat: potential evapotranspiration from November to April and from May to July, temperature normalized

530   radiation from May to July, and precipitation from November to April and from May to July;

– For silage maize: potential evapotranspiration from May to July and from August to October, temperature normalized
radiation from May to July, and precipitation from May to july and from August to October.

The potential evapotranspiration had been calculated daily according to Haude (1955), cf. Schrödter (1985). The same holds
for the temperature normalized radiation which is solar radiation in $J\,cm^{-2}$ divided by the average air temperature in °C elevated

535   by 20 to avoid negative denominators. All weather data were taken from meteorological stations with each district using the
data of the one closest to its centre, only in case of multiple stations located within the same district their measurements were
averaged (Gornott and Wechsung, 2016).

To exclude these peculiar differences in weather data processing from the comparison all weather data are now consistently
taken from the monthly DWD grids and preprocessed for the districts as described above in Sect. 2.2. Potential evapotran-

540   spiration is consequently calculated by AMBAV (Löpmeier, 2014; Friesland and Löpmeier, 2007), very probably closer to
reality than with Haude's formula. Only temperature normalized radiation is not originally computed on daily basis but using
the monthly grids of global radiation and average temperature. The original Gornott and Wechsung (2016) approach used also
fertilizer price and acreage of the target crop among their input variables which are simply omitted here.

The time series for national predictions of winter wheat in Germany calculated from the five predefined weather aggregates

545   is shown in Fig. 7. Compared to the result of the ABSOLUT algorithm in Fig. 2 the lacking accuracy is evident; the share of ex-
plained variance dropped from 79.5 % to 20.2 %. Figure 8 maps the spatial distribution of the out-of-sample prediction accuracy
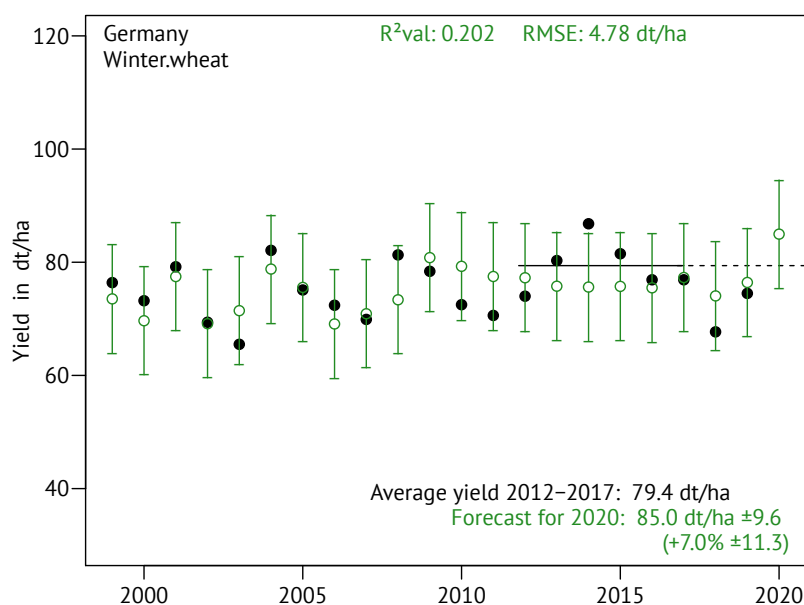
Figure 7. Average winter wheat yields in Germany according to the official statistics (black) and their predictions (green) calculated from the five weather aggregates used by Gornott and Wechsung (2016)

for both approaches. Practically everywhere the five input variables give only poor results, and the pattern of higher accuracies especially in Northern Germany observed by Gornott and Wechsung (2016) and Conradt et al. (2016) is not preserved.

Regarding the loss of accuracy compared to the previous studies Fig. 7 might give a hint: While the predictions are more or less correlated to the observations until the year 2009, the ups and downs observed thereafter are not captured any more. The previous studies used yield data of the years 1992–2010 while the time interval used here is 1999–2020. Hence it might be that the once effective weather predictors have lost their explaining power about the year 2010.

Figure 9 shows the respective maps for silage maize simulated until 2019. At least here the spatial pattern produced by the five variables approach resembles the results of the former studies. The coefficient of determination for the nationally aggregated out-of-sample time series reaches at least 40.3 % while Gornott and Wechsung (2016) reported 50 % for their prediction of interannual changes. This is however clearly below the 90.8 % obtained with ABSOLUT.

5 Conclusions and outlook

Among many studies investigating the correlations between weather variables and agricultural yields and using these for yield prediction this one utilizes ordinary multiple linear regressions but massive parallelized computations for a systematic search of the most relevant input aggregates among tens of thousands of possible combinations. Test bed are about 300 district-level administrative areas in Germany for which there are comprehensive high resolution weather data and individual yield statistics for ten different crops.
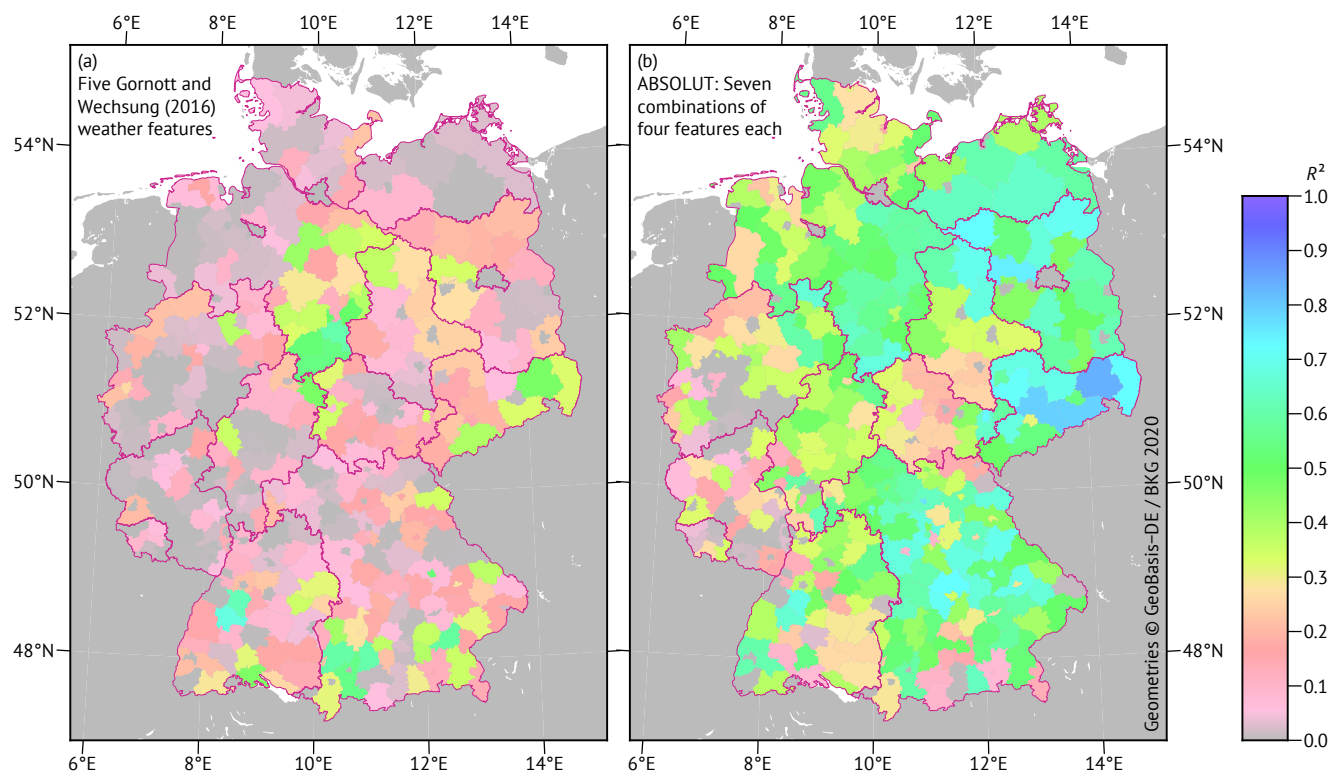
Figure 8. Shares of variance explained by out-of-sample estimations of winter wheat yields on district level: (a) based on the five weather aggregates used by Gornott and Wechsung (2016) and (b) based on seven combinations of four weather aggregates selected through ABSOLUT. The model run of (a) corresponds to the national time series in Fig. 7 and panel (b) to Fig. 2 respectively.

The first lesson learned was that the excessive testing and optimization of regressor combinations consumes degrees of freedom, thus predictive power, just like the estimation of many coefficients within the multiple regressions. As the time frame of example applications was relatively short (20–22 years from 1999 to 2018, 2019, or 2020) the solution was to not rely on the top-performing combination of each district but to require significant above-average performances in many districts for a specific combination of weather aggregates to qualify as predictor. While this approach always requires many spatial subunits with respectively distributed training data the performance of the predictions clearly exceeded the results of precursor studies (Gornott and Wechsung, 2016; Conradt et al., 2016). The prediction performance achieved by the latter could be reconstructed with more recent data to some extent for silage maize but hardly at all for winter wheat – probably the formerly observed correlations between weather and wheat yield detoriate under climate change and become replaced by others (to which the ABSOLUT algorithm will automatically adopt).

There are two critical spots of the present 1.0 version of the algorithm which are at the same time the biggest opportunities for improvements with future revisions: The first one is the assumption of a linear base trend of yields independently estimated
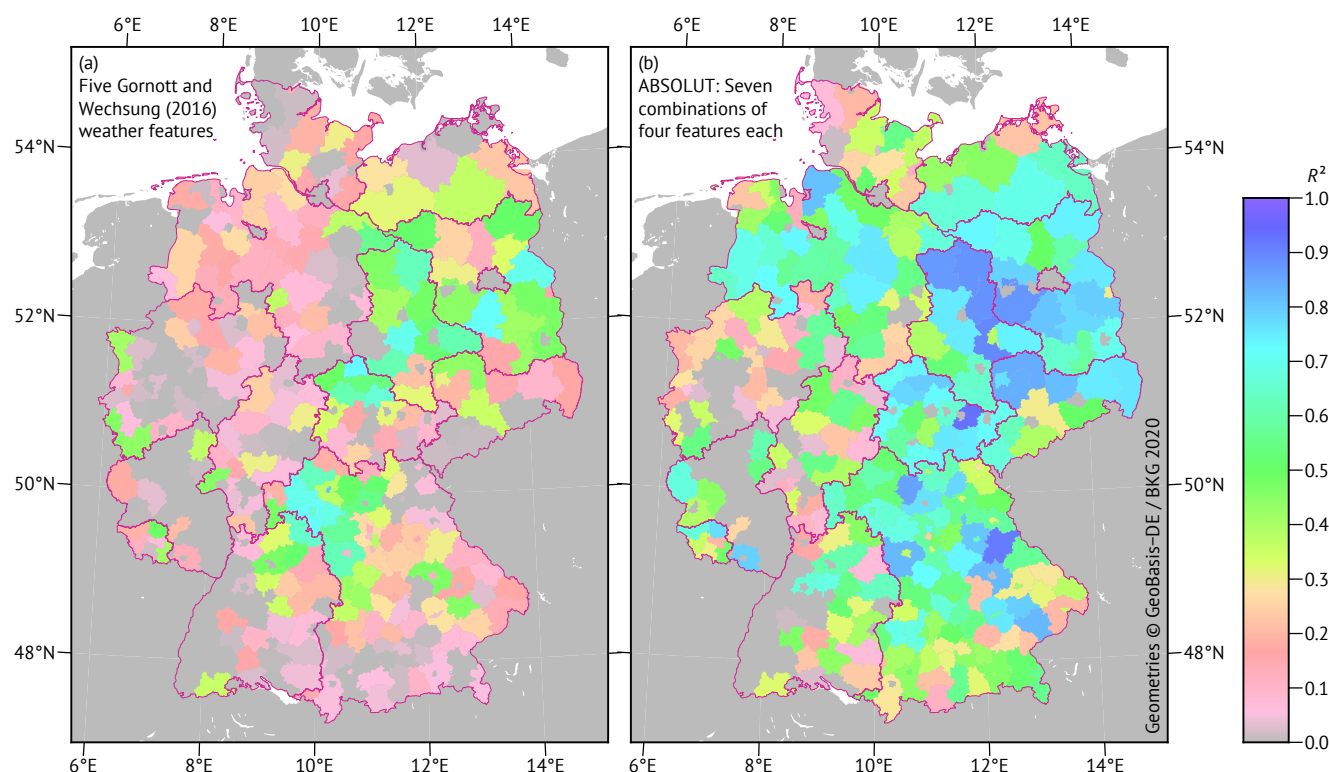
Figure 9. Shares of variance explained by out-of-sample estimations of silage maize yields on district level: (a) based on the five weather aggregates used by Gornott and Wechsung (2016) and (b) based on seven combinations of four weather aggregates selected through ABSOLUT. The target year of this simulation was 2019.

575 for each spatial subunit. This allows straightforward prediction of absolute yields instead of relative changes, but as recent observations do not support the assumption of a smooth long-term development any linear estimation basis will cause prediction biases especially if a long history of training data is applied (which is principally advantageous).

The other issue is the final selection of independent regressor variables, i.e. weather aggregates, for each spatial subunit. The variants tested in Sect. 4.1.2 are probably just building blocks for a more intelligent way to select the most reliable combi-
580 nations. The general challenge is about finding the optimum balance between the multiple, spatially distributed confirmations of predictive power and the flexibility needed to adopt to smaller regions demanding alternative combinations for more exact predictions. The assumption of the correlation of weather aggregates to yields changing over time has been confirmed by other studies (Trnka et al., 2016; Ceglar et al., 2020) which calls for additional flexibility in handling long-term time series (a luxury problem).

585 A third, less critical opportunity for improvement has been identified in Sect. 4.5.3: An adaptive estimation of confidence intervals for the spatial aggregates of yield predictions and for individual years. Some results suggest that standard errors of

district predictions are hardly connected to actual prediciton errors for state and national yield aggregates. An obviously high spatial correlation of subunit errors calls for further investigation.

590   Nevertheless it could be demonstrated that the ABSOLUT algorithm, already in its present stage of development, is capable of explaining 70–90 % of the national yield variations of major crops in Germany solely based on weather variables, a much higher share than usually assigned to weather signals alone: Global studies assessing the relative impact of weather factors on crop yield variations (Frieler et al., 2017; Schauberger et al., 2017b) give typical ranges of 50–60 % for wheat and maize yields of main producer countries, for wheat in the United States even as low as 30–40 %. A careful assessment separating the impacts of farm management and weather effects on wheat yield variations across Germany (Albers et al., 2017) found average shares

595   of only 43 % of the variations caused by the weather and 49 % owing to management. These numbers are however averaged from single farm samples and should not directly be compared to the ABSOLUT results; individual farm management effects are probably just cancelling out in the national averages of district yields. Considering also non-meteorological drivers is generally advisable in developing countries where factors like irrigation status, fertilizer price and general farming conditions are much more decisive (Assefa et al., 2020).

600   National aggregate yields of staple crops in Europe may however depend more on weather than previously assumed (cf. Agnolucci and De Lipsis, 2020). Related questions about the impact of climate change on food security consequently underline the need for further research into this field.

Code and data availability.   The model code, consisting of five R programs, is publicly available at https://doi.org/10.5281/zenodo.4468609 (Conradt, 2021a). The input data needed for the example application to Germany consist of crop yields and cropping areas in administra-

605   tive regions, district-level monthly weather data, and a control file. They are publicly available at https://doi.org/10.5281/zenodo.4468691 (Conradt, 2021b).

Author contributions.   The sole author Tobias Conradt conducted all parts of the work, including conceptualisation, data acquisition and preprocessing, software coding and testing, running the example computations, visualization, and manuscript writing.

Competing interests.   The author declares that he has no conflict of interest.

610   *Disclaimer.*   All parts of the ABSOLUT v1.0 program code are released under the GNU General Public License, Version 3 (https://www.gnu.org/licenses/gpl-3.0). The accompanying input data for the example application to Germany are distributed under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/).

Acknowledgements. This work was financially supported by the German Federal Office for Agriculture and Food (BLE) in the framework of the national research project OptAKlim (grant no. 281B203316) and by the German Aerospace Center (DLR) on behalf of the Federal Ministry of Education and Research (BMBF) in the frameworks of the international BiodivERsA research project SALBES (grant no. 01LC1809B) and the international AXIS research project CROSSDRO (grant no. 01LS1901A). The Federal Republic of Germany and the Federal State of Brandenburg provided PIK's high performance computing facilities though institutional funding. Special thanks goes to the workforce involved in collecting and providing meteorological and crop yield data across Germany. The Federal Agency for Cartography and Geodesy (BKG), the German Weather Service (DWD), and the German statistical offices on national and state levels are acknowledged for providing open access to all necessary input data.

References

Agnolucci, P. and De Lipsis, V.: Long-run trend in agricultural yield and climatic factors in Europe, Climatic Change, 159, 385–405, https://doi.org/10.1007/s10584-019-02622-3, 2020.

Akter, N. and Islam, R.: Heat stress effects and management in wheat. A review, Agronomy for Sustainable Development, 37, 17 pp, https://doi.org/10.1007/s13593-017-0443-9, 2017.

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

Assefa, B. T., Chamberlin, J., Reidsma, P., Silva, J. V., and van Ittersum, M. K.: Unravelling the variability and causes of smallholder maize yield gaps in Ethiopia, Food Security, 12, 83–103, https://doi.org/10.1007/s12571-019-00981-4, 2020.

Butts-Wilmsmeyer, C. J., Seebauer, J. R., Singleton, L., and Below, F. E.: Weather During Key Growth Stages Explains Grain Quality and Yield of Maize, Agronomy, 9, 15 pp, https://doi.org/10.3390/agronomy9010016, 2019.

Ceglar, A., Toreti, A., Lecerf, R., der Velde, M. V., and Dentener, F.: Impact of meteorological drivers on regional inter-annual crop yield variability in France, Agricultural and Forest Meteorology, 216, 58–67, https://doi.org/10.1016/j.agrformet.2015.10.004, 2016.

Ceglar, A., Zampieri, M., Gonzalez-Reviriego, N., Ciais, P., Schauberger, B., and van der Velde, M.: Time-varying impact of climate on maize and wheat yields in France since 1900, Environmental Research Letters, 15, 094 039, 13 pp., https://doi.org/10.1088/1748-9326/aba1be, 2020.

Chen, H.: The spatial patterns in long-term temporal trends of three major crops' yields in Japan, Plant Production Science, 21, 177–185, https://doi.org/10.1080/1343943X.2018.1459752, 2018.

Conradt, T.: ABSOLUT v.1.0 R programs, Zenodo, https://doi.org/10.5281/zenodo.4468609, 2021a.

Conradt, T.: ABSOLUT v.1.0 Input data for an example application on the districts of Germany, Zenodo, https://doi.org/10.5281/zenodo.4468691, 2021b.

Conradt, T., Gornott, C., and Wechsung, F.: Extending and improving regionalized winter wheat and silage maize yield regression models for Germany: Enhancing the predictive skill by panel definition through cluster analysis, Agricultural and Forest Meteorology, 216, 68–81, https://doi.org/10.1016/j.agrformet.2015.10.003, 2016.

DESTATIS: Wachstum und Ernte – Feldfrüchte [Growth and harvest – field crops, in German], Fachserie [Thematic report series] 3, 3.2.1, Issue 2018,16, Statistisches Bundesamt, Wiesbaden, https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00096325, last access September 2020, 2019.

DWD 2020: Nationaler Klimareport [National climate report, in German], Tech. Rep. 4. corrected edn, Deutscher Wetterdienst, Potsdam, https://www.dwd.de/DE/leistungen/nationalerklimareport/download_report_auflage-4.pdf, last access December 2020, 2020.

European Soil Bureau Network: Soil Atlas of Europe, Office for Official Publications of the European Communities, Luxembourg, https://esdac.jrc.ec.europa.eu/content/soil-atlas-europe, last accessed December 2020, 2005.

Farooq, M., Bramley, H., Palta, J. A., and Siddique, K. H. M.: Heat Stress in Wheat during Reproductive and Grain-Filling Phases, Critical Reviews in Plant Sciences, 30, 491–507, https://doi.org/10.1080/07352689.2011.615687, 2011.

Fisher, R. A.: The influence of rainfall on the yield of wheat at Rothamsted, Philosophical Transactions of the Royal Society B, 213, 89–142, https://doi.org/10.1098/rstb.1925.0003, 1924.

Frieler, K., Schauberger, B., Arneth, A., Balkovič, J., Chryssanthacopoulos, J., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., Olin, S., Pugh, T. A. M., Schaphoff, S., Schewe, J., Schmid, E., Warszawski, L., and Levermann, A.: Understanding the weather signal in national crop-yield variability, Earth's Future, 5, 605–616, https://doi.org/10.1002/2016EF000525, 2017.

Friesland, H. and Löpmeier, F.-J.: The performance of the model AMBAV for evapotranspiration and soil moisture on Müncheberg data, in: Modelling water and nutrient dynamics in soil–crop systems, edited by Kersebaum, K. C., Hecker, J. M., Mirschel, W., and Wegehenkel, M., chap. 2, p. 19–26, Springer, Dordrecht, https://doi.org/10.1007/978-1-4020-4479-3_2, 2007.

González-Fernández, E., Piña-Rey, A., Fernández-González, M., Aira, M. J., and Rodríguez-Rajo, F. J.: Prediction of Grapevine Yield Based on Reproductive Variables and the Influence of Meteorological Conditions, Agronomy, 10, 19 pp, https://doi.org/10.3390/agronomy10050714, 2020.

Gornott, C. and Wechsung, F.: Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany, Agricultural and Forest Meteorology, 217, 89–100, https://doi.org/10.1016/j.agrformet.2015.10.005, 2016.

Gómez, D., Salvador, P., Sanz, J., and Casanova, J. L.: Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data, Remote Sensing, 11, 17 pp, https://doi.org/10.3390/rs11151745, 2019.

Haude, W.: Zur Bestimmung der Verdunstung auf möglichst einfache Weise, Mitteilungen des Deutschen Wetterdienstes 11, Deutscher Wetterdienst, Bad Kissingen, 1955.

Hennings, V.: Ackerbauliches Ertragspotential der Böden in Deutschland 1 : 1 000 000 (SQR1000) [Agricultural yield potential of the soils in Germany, 1 : 1 million map, in German], Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, Germany, https://www.bgr.bund.de/DE/Themen/Boden/Ressourcenbewertung/Ertragspotential/Ertragspotential_node.html, last accessed September 2020, 2013.

Hooker, R. H.: Correlation of the weather and crops, Journal of the Royal Statistical Society, 70, 1–51, https://doi.org/10.2307/2339501, 1907.

Ji, Y., Zhou, G., Wang, L., Wang, S., and Li, Z.: Identifying climate risk causing maize (*Zea mays* L.) yield fluctuation by time-series data, Natural Hazards, 96, 1213–1222, https://doi.org/10.1007/s11069-019-03605-4, 2019.

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., and Lichtenberger, J.: Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices, Agricultural and Forest Meteorology, 260–261, 300–320, https://doi.org/10.1016/j.agrformet.2018.06.009, 2018.

Kruse, K., ed.: Bodenatlas Deutschland, Bundesanstalt für Geowissenschaften und Rohstoffe, Schweizerbart, Stuttgart, Germany, 2016.

Löpmeier, F.-J.: Agrarmeteorologisches Modell zur Berechnung der aktuellen Verdunstung (AMBAV), Online document accompanying the monthly soil moisture grid data on the CDC open data portal of the German Weather Service, probably republished from the 1983 DWD-internal report Beiträge zur Agrarmeteorologie 7/83, https://opendata.dwd.de/climate_environment/CDC/grids_germany/monthly/soil_moist/AMBAV.pdf, last accessed in September 2020, 2014.

Mehrabi, Z. and Ramankutty, N.: Synchronized failure of global crop production, Nature Ecology & Evolution, 3, 780–786, https://doi.org/10.1038/s41559-019-0862-x, 2019.

Meinardus, W.: Einige Beziehungen zwischen der Witterung und den Ernteerträgen in Nord-Deutschland, in: Verhandlungen des Siebenten Internationalen Geographen-Kongresses, Berlin, 1899, pp. II, 421–428, Sampson Low & Co., W. H. Kühl, and H. Le Sondier, London, Berlin, and Paris, https://archive.org/details/verhandlungende19unkngoog/page/n457/mode/1up, last accessed in July 2020, scan lacks reproduction of tables., 1901.

Moore, F. C. and Lobell, D. B.: Adaptation potential of European agriculture in response to climate change, Nature Climate Change, 4, 610–614, https://doi.org/10.1038/nclimate2228, 2014.

Moore, F. C. and Lobell, D. B.: The fingerprint of climate trends on European crop yields, Proceedings of the National Academy of Sciences
of the United States of America (PNAS), 112, 2670–2675, https://doi.org/10.1073/pnas.1409606112, 2015.

Nemoto, M., Hamasaki, T., Matsuba, S., Hayashi, S., and Yanagihara, S.: Estimation of Rice Yield Components with Meteorological Elements Divided According to Developmental Stages, Journal of Agricultural Meteorology, 72, 128–141, https://doi.org/10.2480/agrmet.D-15-00017, 2016.

Richter, A., Siebner, C., Schmidt, R., Adler, G. H., Stegger, U., Eckelmann, W., and Hartwich, R.: Nutzungsdifferenzierte Bodenübersichtskarte 1 : 1 000 000 (BÜK1000N) [Soil overview map, differentiated by use, 1 : 1 million map, in German], Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, Germany, https://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK1000/Nutz_BUEK/nutz_buek_node.html, last accessed December 2020, 2007.

Salehnia, N., Salehnia, N., Torshizi, A. S., and Kolsoumi, S.: Rainfed wheat (Triticum aestivum L.) yield prediction using economical, meteorological, and drought indicators through pooled panel data and statistical downscaling, Ecological Indicators, 111, https://doi.org/10.1016/j.ecolind.2019.105991, 2020.

Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C., Pugh, T. A. M., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W., and Frieler, K.: Consistent negative response of US crops to high temperatures in observations and crop models, Nature Communications, 8, 9 pp, https://doi.org/10.1038/ncomms13931, 2017a.

Schauberger, B., Gornott, C., and Wechsung, F.: Global evaluation of a semiempirical model for yield anomalies and application to within-season yield forecasting, Global Change Biology, 23, 4750–4764, https://doi.org/10.1111/gcb.13738, 2017b.

Schauberger, B., Ben-Ari, T., Makowski, D., Kato, T., Kato, H., and Ciais, P.: Yield trends, variability and stagnation analysis of major crops in France over more than a century, Scientific Reports, 8, 12 pp, https://doi.org/10.1038/s41598-018-35351-1, 2018.

Schauberger, B., Jägermeyr, J., and Gornott, C.: A systematic review of local to regional forecasting approaches and frequently used data resources, European Journal of Agronomy, 120, 15 pp, https://doi.org/10.1016/j.eja.2020.126153, 2020.

Schrödter, H.: Verdunstung – Anwendungsorientierte Meßverfahren und Bestimmungsmethoden, Springer, Heidelberg, 1985.

Trnka, M., Olesen, J. E., Kersebaum, K. C., Rötter, R. P., Brázdil, R., Eitzinger, J., Jansen, S., Skjelvåg, A. O., Peltonen-Sainio, P., Hlavinka, P., Balek, J., Eckersten, H., Gobin, A., Vučeti, V., Dalla Marta, A., Orlandini, S., Alexandrov, V., Semerádová, D., Štěpánek, P., Svobodová, E., and Rajdl, K.: Changing regional weather–crop yield relationships across Europe between 1901 and 2012, Climate Research, 70, 195–214, https://doi.org/10.3354/cr01426, 2016.

Zhang, N., Zhao, C., Quiring, S. M., and Li, J.: Winter Wheat Yield Prediction Using Normalized Difference Vegetative Index and Agro-Climatic Parameters in Oklahoma, Agronomy Journal, 109, 2700–2713, https://doi.org/10.2134/agronj2017.03.0133, 2017.