Agricultural Systems 215 (2024) 103846

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

Assessing the impact on crop modelling of multi- and uni-variate climate model bias adjustments

S. Galmarini [a,*], E. Solazzo [z,1], R. Ferrise [b], A. Kumar Srivastava [f], M. Ahmed [k,l,m], S. Asseng [c], A.J. Cannon [p], F. Dentener [a], G. De Sanctis [g], T. Gaiser [f], Y. Gao [d], S. Gayler [e], J.M. Gutierrez [q], G. Hoogenboom [d], M. Iturbide [q], M. Jury [r,aa], S. Lange [s], H. Loukos [t], D. Maraun [r], M. Moriondo [i], S. McGinnis [u], C. Nendel [x,y], G. Padovan [b], A. Riccio [v], D. Ripoche [h], C.O. Stockle [m], I. Supit [o], S. Thao [w], G. Trombi [b], M. Vrac [w], T.K.D. Weber [n], C. Zhao [j]

[a] *European Commission-Joint Research Center, Ispra, Italy*
[b] *Dep. of AGRIculture, food, environment and forestry, University of Florence, Florence, Italy*
[c] *Technical University of Munich, Dep. of Life Science Eng., Digital Agriculture, HEF World Agricultural Systems Center, 85354 Freising, Germany*
[d] *Dep. of Agricultural and Biological Engineering, University of Florida, Gainesville, FL, USA*
[e] *Institute of Soil Science and Land Evaluation, University of Hohenheim, D-70599 Stuttgart, Germany*
[f] *Institute of Crop Science and Resource Conservation, University of Bonn, Bonn 53115, Germany,*
[g] *European Food Safety Authority, Parma, Italy*
[h] *National Research Institute for Agriculture, Food and Environment, France*
[i] *Institute for Bioeconomy, National Research Council (CNR), Italy*
[j] *College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China*
[k] *Dep. of Agronomy, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi 46300, Pakistan*
[l] *Dep. of Agricultural Res. for Northern Sweden, Swedish Un. of Agricultural Sc., 90183 Umeå, Sweden*
[m] *Dep.of Biological Systems Eng, Washington State University, Pullman, WA 99164-6120, USA*
[n] *Faculty of Organic Agricultural Sciences, University of Kassel, Nordbahnhofstraße 1a, D-37213 Witzenhausen, Germany*
[o] *Water Systems and Global Change, Wageningen University, Wageningen, the Netherlands.*
[p] *Climate Research Division, Environment and Climate Change Canada, Canada*
[q] *Instituto de Física de Cantabria. CSIC - University of Cantabria, Santander, Spain*
[r] *Wegener Center for Climate and Global Change, University of Graz, Austria*
[s] *Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 60 12 03, 14412 Potsdam, Germany*
[t] *The Climate Data Factory, Paris, France*
[u] *National Center for Atmospheric Research (NCAR), USA*
[v] *University of Naples "Parthenope", Italy*
[w] *Laboratoire des Science du Climat et de l'Environnement (LSCE/IPSL), France*
[x] *Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany*
[y] *Institute of Biochemistry and Biology, University of Potsdam, Germany*
[z] *UniSystems Information Technologies, Milan, Italy*
[aa] *Earth Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain*

* Corresponding author at: Joint Research Center, European Commission, Via E. Fermi, 2749, I-21027 Ispra, VA, Italy.
E-mail address: Stefano.galmarini@ec.europa.eu (S. Galmarini).
[1] Now at: Methodology and Scientific Support Unit, European Food Safety Authority, Parma, Italy.

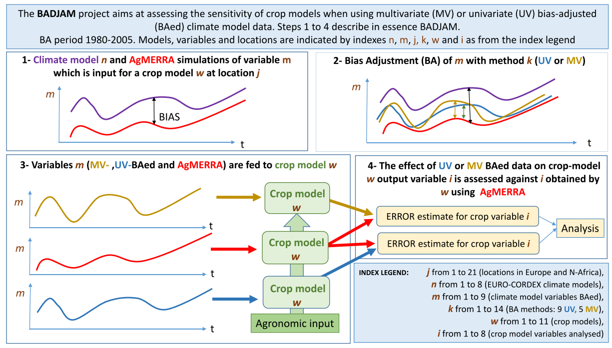
https://doi.org/10.1016/j.agsy.2023.103846

Available online 29 January 2024

HIGHLIGHTS

- The impact of climate data multivariate bias-adjustment methods versus univariate on crop model results was estimated.
- Crop model results improved when input data was treated using multivariate methods compared to univariate methods.
- Multivariate methods maintain the variables correlation as required by crop models.
- This result is attributed to the parameterized nature of the participating crop-model formulations.
- Climate, bias adjustment and the crop modelling communities joined efforts in this unprecedented collaboration.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Val Snow

Keywords:
Bias-adjustment
Crop modelling
Climate models
Impact models
Multivariate and univariate methods
BADJAM

ABSTRACT

CONTEXT: Crop models are essential tools for assessing the impact of climate change on national or regional agricultural production. Starting from meteorology, soil and crop management, fertilization and irrigation practices, they predict the yield of specific crop varieties. For long term assessments, climate models are the source of primary information. To make climate model results usable in a specific time frame context, bias adjustment (BA) is required. In fact, climate models tend to deviate from day-to-day values of the physical parameters while conserving the climate variability signal. BA brings the climatic signal to the actual values observed in a specific location and period, and to be representative of a specific period in absolute terms. BA techniques come in different flavours. The broadest categorization is univariate and multivariate methods. Multivariate methods adjust the variables considering possible cross-correlations while univariate methods treat the variables one by one without accounting for possible dependence on one another.
OBJECTIVE: The hypothesis tested in this paper is that since crop models require as input climate variables that are in most of the cases cross-correlated, the multi-variate bias adjustment of the latter is likely to improve performance compared to univariate bias adjusted climate model results.
METHODS: To verify this hypothesis, 14 BA methods were applied to 9 variables from 8 climate models at 21 locations across Europe and Northern Africa for a period of 5 years. Twelve crop models, from the AgMIP Wheat community, were run using the climate model results. All crop models, except one, were restarted at every growing season. The crop models were also run using the AgMERRA re-analysis. The latter were used as reference to compare the results when using the other climate models treated with the various sets of bias-adjustment methods.
RESULTS AND CONCLUSIONS: The results show that multivariate BA treatment should be preferred to univariate ones. The error obtained by comparing crop simulation obtained with AgMERRA with those obtained with multivariate bias-adjusted climate prediction is systematically lower. The error reduction varies as a function of the variable, the location, the crop model, and the climate model though the tendency is for smaller errors when multivariate methods are used to treat the latter. The results are attributed to the nature of crop models and the fact that multivariate methods consider more adequately the correlation existing between the meteorological variables.
SIGNIFICANCE: The study shows the importance of considering the nature of a model and the selection of input data that best suited to the former. In this case the improvements produced when using multivariate data appears to be significant especially in the light of the variety of crop models used and the similar response obtained and it is therefore recommended.

1. Introduction

Starting from meteorological data and agricultural parameters like soil characteristics and agronomic practices, crop models determine the different stages of growth and yield estimates for a large variety of crops. Crop models can use historical, predicted, as well as projected climate simulations to assess the effect of climate variability on crops.

Climate models reproduce the climate scale signal and its variability depending on external radiative forcing. The latter is the consequence of greenhouse gas (GHG) and, aerosol concentrations present in a specific period of time (paleo, pre-industrial, present or future etc.). Although they produce daily estimates of meteorological variables, climate models are not directly suitable for the prediction of the weather that drives crops on a seasonal basis. In fact, they are not initialized with initial conditions referring to a specific date in time, as would be the case of a weather prediction system, but only with the forcing pertaining to a generic period of time. This aspect introduces a *bias* in the climate simulation. The latter is defined as the overall systematic (or time-independent) deviation of the climate prediction from observed data (e.g. Maraun et al., 2017; Cannon et al., 2020).

When climate model results are to be used by other model such as crop models, and therefore over a well-defined period of time, the bias stemming from the transient character of the simulation has to be adjusted so that the variables adhere to the real conditions of the period considered for crop modelling. Several techniques exist (Maraun and Widmann, 2018) to perform what is commonly referred to as Bias-Adjustment (BA).

In Galmarini et al. (2019) the specific use of bias-adjusted climate data by crop models was addressed. In particular, the use of univariate (UV) or multivariate (MV) bias-adjustments methods. UV methods adjust climate variables independently from on another, therefore, the adjustment pertains exclusively to the relationship between the so-called *source* (reference climate variable toward which the adjustment occurs) and the *target* (climate variable to be adjusted). In the case of MV methods, BA is simultaneously performed on all the cross-correlated variables with the scope of maintaining any existing relation. The cross-correlation could be due to a direct physical connection or dependency (when *source* are monitoring data) or due to its representation within coupled conservation equations, (when the *source* is another climate model).

Galmarini et al. (2019) launched the Bias ADJustment for Agricultural Models (BADJAM) activity aimed at bringing together a community of climate modelers, and crop modelers to assess the sensitivity of crop models to the two BA techniques. The results of BADJAM are presented in this paper.

2. Material and methods

2.1. Domain and locations

The starting point of BADJAM is the selection of a domain over which the crop models will be applied and the reference data against which climate model results are BAed. Specific locations were selected within the selected region where crop models taking part in BADJAM share common calibration datasets so to reduce possible sources of uncertainty. Twenty-one locations were identified within a domain that encompasses European, North-African and Middle East regions (Fig. 1). The locations are part of the Agricultural Model Intercomparison and Improvement Project (AgMIP, URL1, 2023). At the AgMIP sites, crop management, growing conditions, plant physiology, and growing stages as well as yields and yield quality are monitored and they have been central to several studies (Asseng et al., 2013, 2014, 2019; Martre et al., 2014). The 21 locations selected are listed in **Table S1** (in the supplementary material, SM from now on), and mapped in Fig. 1 together with their elevation above sea level.

2.2. The climate models and the bias-adjustment methods

Given the geographical extension of the domain, EURO-CORDEX (Gobiet and Jacob, 2011) was designated as regional climate data provider and as source of climate simulations. The regional climate data with daily time resolution from EURO-CORDEX (Table 1) were interpolated at the 21 locations from the nearest grid model nodes. A 25-year period of interest that goes from 1980 through 2005 was selected. In addition to the CORDEX models the AgMERRA model (Ruane et al., 2015) is considered as *source* dataset for the BA. For the sake of reducing the sources of uncertainty and having a common reference dataset, AgMERRA is used as a surrogate of the observational data against which to perform the BA. In fact, AgMERRA provides daily, high-resolution, continuous, meteorological time series over the 1980–2010 period. AgMERRA data could be considered *analyzed* since they combine daily resolution data from retrospective analyses (the Modern-Era Retrospective Analysis for Research and Applications, MERRA, and the Climate Forecast System Reanalysis, CFSR) using observational data for temperature, precipitation, and solar radiation. AgMERRA shows substantially improved representations of daily precipitation distributions and extremes. Further to that, it couples humidity to the maximum daily

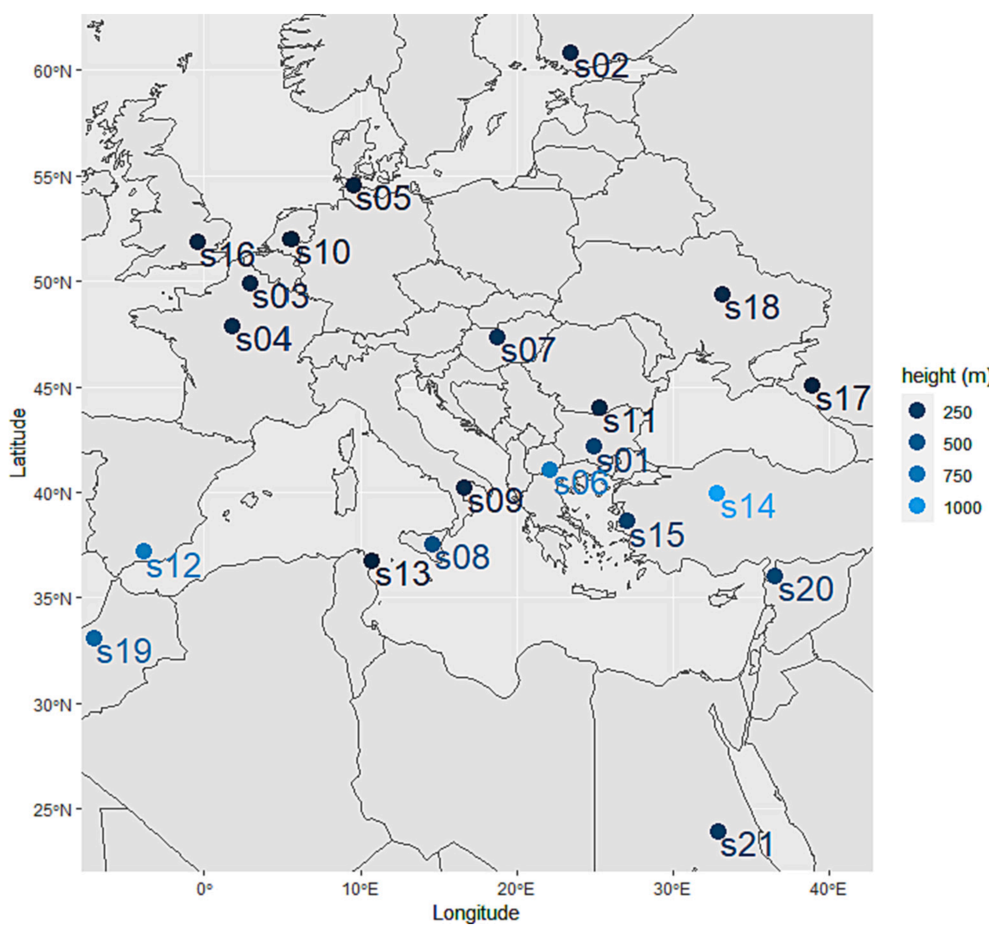


Fig. 1. Map of AgMIP stations listed in **Table S1** and their elevation above sea level in [m].

Table 1
EURO-CORDEX climate models and list of bias adjustment methods used in BADJAM. The names of climate models reflect the combination of global and (/) regional climate models.

Climate model	Institution	Acronym in the Figures
ICHEC-EC-EARTH/RCA4	Irish Centre for High-End Computing	ICHEC
MPI-M-MPI-ESM-LR/CCLM4-8-17	Max Planck institute	MPICCLM
MPI-M-MPI-ESM-LR/RCA4	Max Planck institute	MPIRCA4
MPI-M-MPI-ESM-LR/REMO2009	Max Planck institute	MPIREMO09
CSIRO-QCCCE-CSIRO-Mk3-6-0/RCA4	CSIRO/Queensland Climate Change Centre of Excellence	CSIRO
NCC-NorESM1-M/RCA4	Norwegian Climate Center	NCC
IPSL-IPSL-CM5A-MR/RCA4	Institute Philippe Simon Laplace	IPSL
NOAA-GFDL-GFDL-ESM2M/RCA4	National Oceanic and Atmospheric Administration	NOAA
AgMERRA (global)	NASA Goddard Institute for Space Studies	

Method	**UV or MV**	**Institution/ Acronym**	**References**
MBCn (*)	MV	Environment Canada and Climate Change, Canada/ECCC	Cannon *et al.*, 2015
MBCp*	MV		Cannon, 2016
MBCr*	MV		Cannon, 2018
QDM	UV		
EQM	UV	Univ. of Cantabria, Spain/UC	Dosio and Paruolo, 2011
EQMs	UV		
CDFt	UV	LSCE/IPSL and The Climate Data Factory, France	Vrac, 2018
R2D2	MV		
KDDM	UV	National Center for Atmospheric Research, USA/NCAR	McGinnis *et al.*, 2015
QM	UV	University of Graz, Austria/UG	Dosio and Paruolo, 2011
SDM	UV		Maraun and Widmann, 2018 Switanek et al., 2017
CDFt	UV	Universita' Partenope, Italy/UP	Vrac, 2018
ISIMP3 (PIK_ISIMIP in figures)	UV	Potsdam Institute for Climate Impact Research, Germany/PIK	Lange, 2019
ISIMIP3BASD (*)	MV		

temperature, thus better representing the diurnal cycle of near-surface moisture in agricultural models. AgMERRA is used in a number of model inter-comparisons taking place under the auspices of AgMIP. The use of AgMERRA results as *source* dataset for the BA is justified by the fact that in spite of not being the same as observations, they are what comes closest to the latter and are continuous, with no gaps in the time series. Furthermore, if affected by systematic errors, all crop models participating in BADJAM will be subject to the same kind of uncertainty. Such condition is preferable at this stage of the investigation to using observations from multiple sites as *source*. In fact the latter would be collected with independent instruments and could be characterized by independent and less controllable kinds and levels of uncertainties. AgMERRA is a climate model that includes climate signals like the others selected and therefore becomes a natural selection for both the bias correction and the evaluation of the crop models.

The BADJAM time-window has been split into a training sub-period (1985–2005) and an evaluation one (1980–1984). As from normal practice in BA, all BA methods were trained over 20-years and evaluated over the 5-year period. The participating BA methods, described in Galmarini et al. (2019), are also listed in Table 1. The table shows the methods that are MV or UV. UV methods are for the vast majority based on quantile-quantile mapping. It is the simplest BA approach, which aims at adjusting the distribution of the values of each climate variable independently. MV methods are original developments and drastically different approaches from one another. An example in this respect is MBCn (Cannon et al., 2015; Cannon, 2016; Cannon, 2018) which is based on an image processing algorithm. The scope of BA is to reduce the differences existing between the climate model results (*target*) and *source* data by preserving the ballpark statistics, the trend of the time series, and any low frequency component such as a climate signal. Other than that, MV methods preserve also the correlations of variables.

2.3. *The crop models*

The crop models participating in BADJAM are listed in Table 2. They are part of the AgMIP-Wheat community, a subgroup of the above-

Table 2
List of the crop models taking part in BADJAM. DE02 is the only model of the list that carries information of one variable from one season to the next.

BADJAM	Model Name	Institution	References
CN02	WHEATGROW	Beijing Normal University, China	
DE02	SIMPLACE<LINTUL-5+> PM-SlimWater-SoilCN	Univ. of Bonn, Germany	Gaiser *et al.*, 2013; Wolf 2012
DE04	Expert-N Spass	Hohenheim University, Kassel Univ.	Priesack 2006, Wang and Engel 2000; Gayler *et al.* 2002
DE06	SIMPLACE<LINTUL-2>	Univ. of Bonn, Germany	Gaiser et al., 2013; Wolf 2012
DE08	MONICA	Leibniz-Zentrum für Agrarlandschaftsforschung, Germany	Nendel et al., 2011 URL2
INT02	STICS	National Research Institute for Agriculture, Food and Environment, France	Brisson et al. 2003 Brisson *et al.* 2009
IT01	CropSyst Ver.3	CNR IBE, Italy	Stöckle et al. (2003)
IT02	SSM-Wheat	Univ. of Florence, Italy	Soltani et al. (2013)
NL01	LINTUL4	Univ. of Wageningen, The Netherlands	Shibu et al. (2010)
PK01	CROPSYST	Dep of Agronomy, Pir Mehr Ali Shah Arid Agriculture University, Pakistan Dep of Agricultural Research for Northern Sweden Dep of Biological Systems Engineering, Washington State University Pullman	Stöckle et al., 2003
US01	DSSAT-Nwheat	TU Munich, Germany	Kassie *et al.*, 2016
US04	DSSAT CSM-CERES-Wheat	Univ. of Florida, USA	Jones *et al.* (2003); Hoogenboom *et al.* (2019a, 2019b)

mentioned AgMIP project that deals specifically with crop models that treat wheat varieties. The crop models are described in detail in the peer-reviewed literature mentioned in the table. They have been also part of several model evaluation or sensitivity studies (Asseng et al., 2013; Asseng et al., 2014). We can grossly divide them into two groups that are relevant to the current analysis. The first group treats every growing season as an independent event. This is due to the absence in the model construct of processes description that relate to process occurring in the inter-seasonal period, or the way they were run. For these models at every season, new input data are provided and no memory exists of the past season conditions (Asseng et al., 2013). The second group on the contrary, to which only one model of BADJAM belongs, follows the evolution of one variable throughout the years and at the beginning of the new growing seasons uses the values evolved from the previous one. It is worth mentioning that no indication was prescribed by BADJAM on whether to re-run the models at every season or continuously throughout the period considered, leaving the choice to the modelling groups.

The climate variables expected as input by the crop models are: Solar Radiation, Minimum and Maximum Temperatures, Average Daily Temperatures, Mean Wind Speed, Daily Accumulated Precipitation and Air Moisture and were provided by the climate models after BA. All other site-specific parameters are identical for all participants and are provided by AgMIP (Asseng et al., 2013; Rosenzweig et al., 2013).

The crop model results are divided into physical variables (Leaf Area Index (LAI), Crop Water Use (CWU) [mm]); phenological variables (Emergence, Heading, Flowering, Maturity dates [Julian day]), and bio-physical ones (YLD [Tons/ha], Biomass above ground [Tons/ha]). All abovementioned variables are sensitive to the meteorological conditions and describe the wheat's growth stages and its efficiency.

2.4. Birdseyeview of the dataset and data analysis strategy

Within BADJAM every crop model was run using AgMERRA and all BAed sets of climate results. The difference between each crop model result based on BAed climate data and those based on AgMERRA, will allow us to measure the effect of the various BA methods on the crop models variables. Fig. 2 provides a schematic representation of the data flow and the degrees of freedom expressed as the various combinations of 8 climate models at 21 locations, bias-adjusted with 14 bias correction methods, and subsequently used by 12 crop models to produce 8 crop model variables. In principle and as maximum estimate, for every input variable to a crop model at one location, we have 9856 combinations of results. In fact, all crop-model output variables can be affected in principle by the 1344 combination of the input. For every crop model variable one has 28,224 results when all locations are considered at the same time and over 220,000 variable values if all crop models are considered at the same time as in what we will call *global analysis*. Such an abundance of combinations provides a good basis for a potentially meaningful statistical analysis.

Given the high number of degrees of freedom available, a data analysis strategy was devised and the results presented accordingly. Namely:

a. Identification of a suitable metric apt to identify the differences between the crop model results based on MV and UV BAed climate data and those based on the reference case (AgMERRA).
b. Performing a *global analysis* (all crop model results, variables, locations, climate data sets used to generated them and BA methods used to BA the climate data): This allows us to quantify the ballpark difference in crop model results as a function of the UV or MV BA classification.
c. Increase the granularity of the analysis by considering the effect of individual BAs.
d. Determine the sensitivity of crop models on the various climate model results.
e. Analysis for specific climate models/BAs methods at the location level.
f. Analysis for the selected climate model methods at the variable level.

3. Results

3.1. Determining the best metrics for the evaluation

The aim of this analysis is to identify out of many existing, the best statistical metric to measure the differences among crop model results based on climate data treated with the two BA methodologies and those based on AgMERRA. Given the size of the data space under consideration, we cannot adopt too many indicators therefore, we want the few that properly address our research necessity. Determining whether a difference exists between data based on MV and UV BA methods is a dichotomous problem solvable through a machine learning (ML) method like the Random Forest (RF) (e.g. Ho, 1995, Speiser et al., 2019).

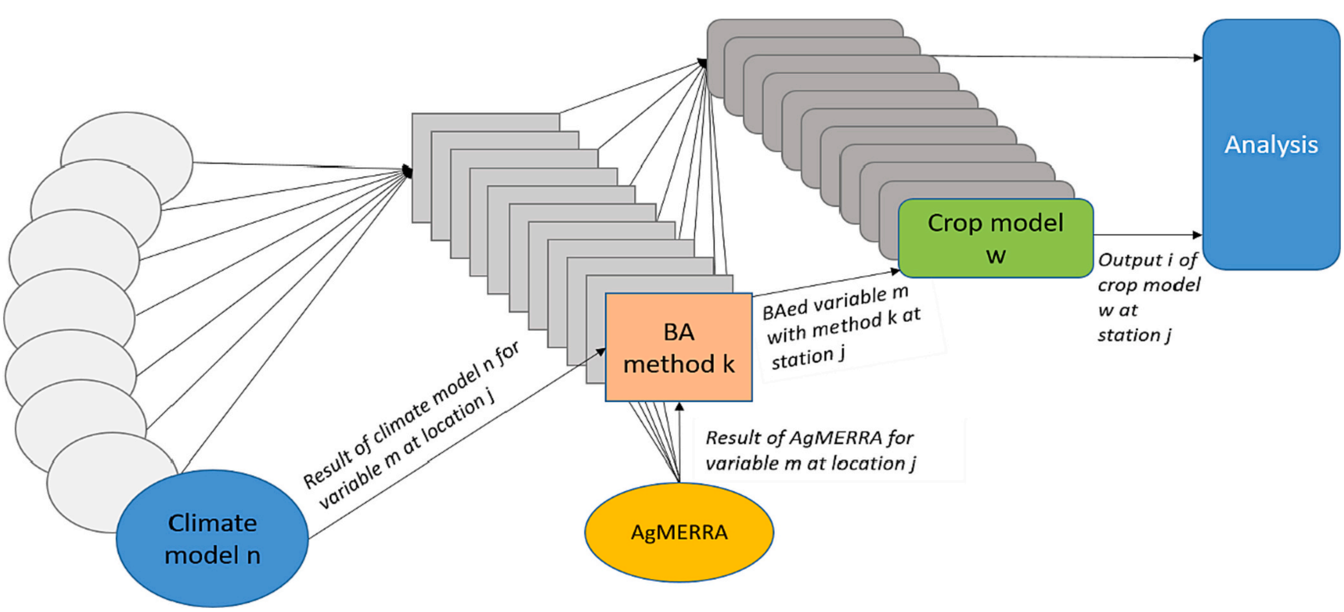


Fig. 2. Schematic sketch of the flow of data analyzed in this paper.

By building decision trees from different data samples, RF is able to define the overall classification or regression of the sample and determine which of the options is more viable. To apply the RF method to our datasets, the results for all variables of each crop model were paired in time (year) and space (station) with the corresponding AgMERRA model output. We identify the most adequate metric as that able to discriminate between MV and UV methods. For each station and variable, a RF classification was produced using several error metrics as explanatory variables: Root Mean Square Error (RMSE), Mean Bias (MB), Mean Gross Error (MGE), Index of Agreement (IoA), Pearson Correlation Coefficient (PCC), Average Mean Bias (AMB). These metrics are not independent and may scan similar error spaces though with some differences. When trained on 80% of available data, the xgbTree method (Friedman, 2001) correctly classified UV and MV of the remaining 20% of the instances with an average score of ~75%, depending on the variable and station. RMSE gave the highest predictive score for the set of physical and biophysical variables and MB for the phenological variables. The fact that the average predictive accuracy reached 75% indicates that:

- UV or MV methods share error features that can still be distinguishable with an accuracy larger than pure chance;
- the remaining 25% reveals that some methods might not fall under the UV or MV categories and that either, a finer classification would be advantageous, or that some crop models and/or climate models have peculiar features.

3.2. Five-year versus 25-year analysis

As stated above, BA climate model results encompasses 25 years of which 20 years have been used to train the BA methodologies and 5 years (1980–1985) have been used as an evaluation period. In light of that and based on considerations that will follow shortly, we limit the analysis of crop model performances to the 5 year evaluation period.

To make sure that the analyses of 5 or 25 years did not bring any relevant difference, IT02 and DE02 simulated the full 25-year period. IT02 works on a seasonal basis (as the other 6 crop models taking part in BADJAM, Soltani et al., 2013; Asseng et al., 2013, 2014, 2019; Martre et al., 2014; Rosenzweig et al., 2013) and considers every season in isolation from the previous ones. DE02 (Gaiser et al., 2013; Wolf, 2012) in contrast follows the time evolution of soil water content from one season to the next. If an impact of the climate variability is to be felt by one of the two models, DE02 will be the one that shows it and therefore we should expect a different behavior between the 5-year and the 25-year analysis. **Figs. S1a** and **b** show respectively the analysis of the results for the two models. The latter are presented in the form of error ranking diagram for all variables at all stations, and climate models BAed with 10 methods. The RMSE are ranked in ten classes (Rank#1 = smallest errors through Rank#10 = largest error) and the frequency of occurrence or the error for every method is presented for every Rank. In the majority of cases IT02 (**Fig. S1a**) confirms what we expected. The RMSE-ranking of the 5 and the 25 year-long simulations are hardly distinguishable. **Fig. S1c** shows for IT02 (blue curve) and DE02 (red curve), the difference in occupancy of rankings 1–5 between of 5-year and 25-year simulations. The average difference over the 5 ranking classes is 0 and − 0.9 for IT02 and DE02, respectively. Apart from 4 cases of the high-ranks (large errors) and a couple of low ranks (small errors) ones, the model appears to be insensitive to the time scale of climatology. In the case of DE02 (**Fig. S1b**) we notice a few more cases where a deviation between 5- and the 25-year results appears but yet it looks marginal thus indicating:

1. a marginal sensitivity of the soil water content to climate variation, at least during the considered period,
2. or an insensitiveness of the crop model to a possible variation,
3. or the absence of a noticeable climate variation that affected soil water content in the time period and the locations considered.

The result presented in **Figs. S1a** and **b** can be explained by considering the nature of the crop models taking part to BADJAM in comparison to climate models. Such crop models are not based on conservation equations in space and time as is a numerical weather prediction or climate model, but rather on direct linear or non-linear functional relationships (most of the time parameterized) among the variables governing a specific process. Therefore, while a climate model can account for the time evolution of a variable and the presence of several timescale, the crop models participating to BADJAM have generally no-capacity to process any information about timescales that extended beyond the duration of the season. Even if the initial conditions (water soil content and soil composition) at the beginning of the season are updated, they cannot capture a long-term climate signal. It should be clear that such crop model can certainly be influenced by the climate variability present in the input data or initial conditions, but only as far as the “here and now” is concerned. The “here and now” extends from the daily resolution of the data to the growing season. During that period the model formulation can only rely on the variability carried by the input data. Other models like DE02 that account for the continuous evolution of some variable beyond the growing season (i.a. water content, soil composition) and onto the next one with continuity may carry a climate signal provided, however, that a time evolving conservation equation controls those variables. As shown if **Fig. S1**, for the specific period considered in this study, DE02 is not affected either on the 25-years period possibly as a consequence of one of the points (1)–(3) presented above. It is important to stress again that this does not imply that a crop model would not react to a changing climate, it will react as long as the daily data include the variation caused by climate change at every other season.

In light of the above, for all crop models like those that took part to BADJAM (Asseng et al., 2013, 2014, 2019; Martre et al., 2014; Rosenzweig et al., 2013) in particular and in general for any model that relies on climate data in a similar way, the correspondence of input variables to “reality” and consistency with each other are of primary importance. BA is therefore a very important step in the preparation of the input of crop models for climate applications, thus emphasizing the relevance of this study.

Another consideration needs to be made that generalizes point (3) presented above. Let us assume as a hypothesis that over 2 of the considered 5 years, AgMERRA data (or conversely some or all the BA climate models) would show a prolonged widespread drought, whereas none of the BAed climate models does (conversely AgMERRA does not). This would constitute an anomaly that in all cases would not compromise the analysis that we want to perform. In fact, all BA methods will try to adjust the climate model results to AgMERRA as they will all be affected by large errors. Such a situation would constitute a further challenge for the BA methods within BADJAM. Although from **Figs. S1** it does not appear that such a condition is present, we do have to contemplate this as a possibility to make sure that the 5-year analysis can still be representative. The aspect of BA of extreme events (droughts, frost, hydrometeors, flooding, wind storms, etc.) is still a very debated and open issue in climate modelling. Cannon et al. (2020) consider the difficulty as due to two main issues. Firstly the low frequency in the occurrence of the extremes in relation to the time span of a simulation and secondly their duration, which can be either too short to be correctly recorded in the target record or too long, therefore extending outside the training window. The impact of BA on extreme modelling will require a specific analysis in the future given its relevance in particular for the agricultural sector.

For the sake of completeness it should be mentioned that the analysis over 5 years will relate always to the average of the data over 5 years. Furthermore, the granular level analyses will be also presented for IT02 and DE02 for 25 years to confirm the results presented in **Figs. S1**.

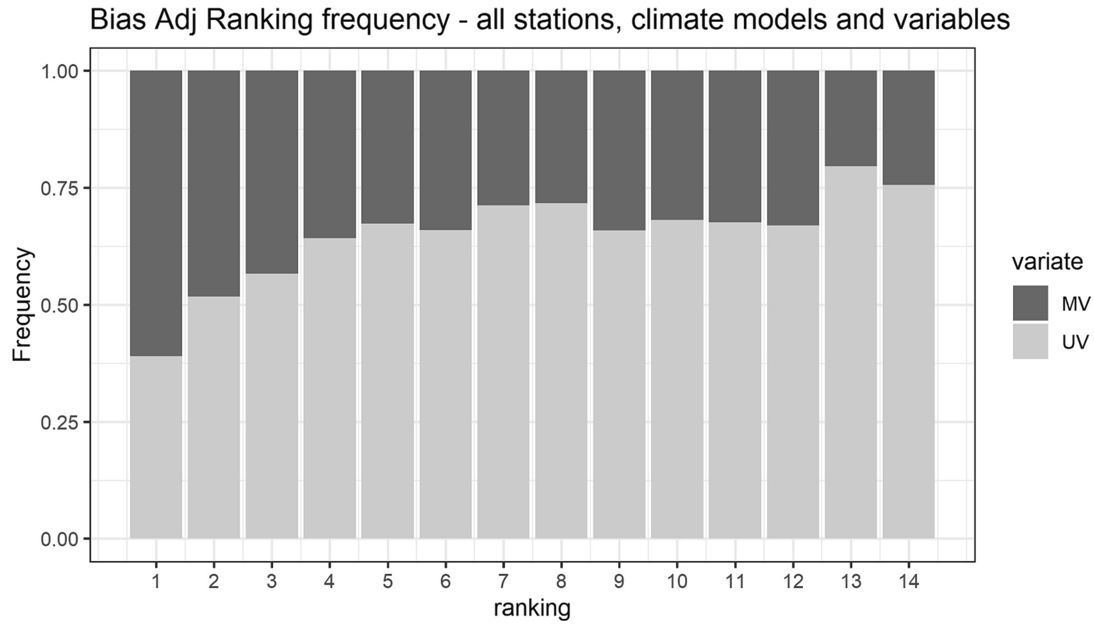


Fig. 3. RMSE-ranking frequency from global analysis. RMSE refers to the difference between crop model results based on BAed climate model data and those based on AgMERRA. The RMSE relating MV and UV BA method are showed in the stacked histogram. Ranking = 1 corresponds to lowest RMSEs, ranking =14 corresponds to maximum RMSEs.

3.3. Global analysis of all crop model results

Let us start with the global analysis, meaning the analysis of all data grouped only by the BA method underlining their generation (UV or MV).

After computing the RMSE of the crop model results obtained with BA climate data against AgMERRA reference data, the values have been organized in ranks. On the x-axis of Fig. 3, high performance corresponds to low-ranking values (small errors) while low performance corresponds to high-ranking values (big errors). Despite the proportion of 9 UV to 5 MV methods, MV methods disproportionally occupy the top ranking positions. MV-based model results decreases in percentage from almost 60 to approximately 25% between Rank#1 to Rank#14, whereas, conversely, UV-based model results increase from 40% in Rank#1 to 75% in Rank#-14. Overall, the ranking of Fig. 3 suggests a robust outscoring of UV by MV methods. One could conclude that globally, MV methods have a higher probability of a better performance than UV.

When breaking down the results of Fig. 3 in individual BA methods (keeping stations, crop variables, climate models undistinguishable) we obtain the ranking frequency of Table 3. The table is constructed as follows: once the errors are calculated for all sets of crop model results pertaining to the 14 methods, they are ranked in order of magnitude in 14 classes and the frequencies of occurrence for every class is obtained for every method. For the sake of synthesis, Table 3 presents the frequency of occurrence for the three top (Rank# 1,2 and 3, referring to smallest errors), middle (rank 8, 9 and 10, middle size errors) and bottom (Rank# 12, 13 and 14, largest errors) rankings. A full ranking diagram is presented in the SM **Fig. S2**. For every ranking, the first five highest frequencies are highlighted in the table.

As it clearly appears 4, 3 and 2 out of 5 MV methods score the five highest frequencies at ranking position #1, #2 and #3, respectively. As

Table 3
Ranking positions 1 through 3, 8–10 and 12–14 for the individual bias adjustment methods. For every rank, the percentage of cases in which the error occupies the rank is shown. The colored cells show the 5 highest frequency values.

BA method	BA type	Error ranking (Rank 1 => smallest errors, Rank 14=> largest errors)										
		#1	#2	#3	[...]	#8	#9	#10	[...]	#12	#13	#14
ECCC_MBCn	MV	7.84	6.13	6.83		8.58	7.87	8.25		5.49	3.60	2.50
ECCC_MBCp	MV	4.65	13.69	9.97		7.35	7.71	7.63		5.02	2.96	1.94
ECCC_MBCr	MV	8.96	11.12	13.14		5.13	5.09	6.74		7.97	4.55	4.69
IPSL_R2D2	MV	31.78	7.12	6.64		12.64	10.38	8.69		3.54	1.53	0.93
PIK_ISIMIP3_multi	MV	7.14	10.19	6.82		11.67	8.18	8.08		7.95	5.97	4.90
ECCC_QDM	UV	1.04	2.76	5.67		2.68	3.45	3.86		7.28	3.98	8.24
IPSL_CDFt	UV	3.42	4.02	6.37		4.53	9.97	5.42		7.31	5.33	7.04
NCAR_KDDM	UV	4.69	7.52	9.58		10.04	9.40	8.78		4.80	2.29	1.40
PIK_ISIMIP3	UV	3.85	6.11	6.52		8.01	7.79	6.47		5.25	2.81	2.66
UC_EQM	UV	5.65	8.86	6.89		4.57	5.51	11.90		9.56	9.06	8.07
UC_EQMs	UV	2.29	4.25	6.07		10.55	9.96	8.55		4.37	2.87	2.21
UG_QM	UV	2.96	4.42	3.77		4.88	5.24	4.78		12.01	20.78	19.32
UG_SDM	UV	2.86	5.18	4.63		4.57	4.44	4.52		8.05	25.18	21.05
UP_CDFt	UV	12.87	8.63	7.10		4.80	5.02	6.34		11.39	9.10	15.05

for the middle ranking, we can notice that the balance shifts toward the UV method. In the middle ranking, up to three of the highest frequencies are in fact from UV methods. In the last three ranking classes (#12, #13, #14), the predominance of high frequencies is almost totally shifted to UV methods.

The method UP_CDFt represents an interesting case. As it appears from Table 3 it presents a noticeable performance for the low rank classes (small errors) as well as however for the high ranks (large errors). The mixed performance by UP_CDFt will require a special attention in another context.

The results of Table 3 indicate that using MV methods is likely to result in smaller crop model errors when compared with those based on the *source* data (AgMERRA). This is particularly true if we consider the ratio 5:9 of MV methods compared to UV. One should also consider that almost all UV are based on the same technique (quantile-quantile mapping) and if that is underperforming compared to a MV method, in general all methods would. In this respect noticeable exceptions among the UV group are NCAR_KDDM and UC_EQM as can been seen in Rank#1 column of Table 3.

3.4. Increasing the granularity of the analysis

We will now proceed to increase the detail of the analysis as a function of the various degrees of freedom we have identified in Section 1.5.

3.4.1. Break down by BA method and crop model

As next step, we determine the role of crop models in the aggregated performance of BA methods discussed so far. For every crop model, Table 4 shows the RMSE frequency obtained from the AgMERRA-based results and BAed-based one. As for Table 3**,** the ranking goes from 1 to 14 (the number of BA methods), where Rank#1 comprises the frequencies of smallest errors and Rank#14 those of the largest ones. For the sake of synthesis, only the results pertaining to the first and last three ranks are presented. Furthermore, since for every of the 11 crop models there are 14 frequencies (one for every BA method), we grouped them into those relating to MV and UV methods and the average frequency values for each group are presented in Table 4. In the SM the full set of results are presented by means of a rank diagram (**Fig. S3**).

From Table 4 we can see that when crop models use MV-based BA data they outperform the results of the same models based on UV-BAed data. In fact MV compared to UV based results show systematically higher frequencies in Rank#1 and #3, in some cases by large margins. As before, this indicates that when using MV BAed climate data, crop models produce more frequently smaller errors than UV BAed data. In Rank#2 UV shows cases in which the frequency is higher though the margins are not as high as those showed by MV methods in Rank#1 and #3. In Rank#2, MV frequencies are higher than those of UV in 6 cases out of 11. Conversely looking at Ranks #12, #13, #14 (large errors) the abundance of cases in which UV is higher appears clearly. In particular, at Rank#13 and #14, UV-BAed based crop model results have systematically higher frequencies than MV based data for all models. One should notice that in order to obtain 100% of the frequencies for every ranking class one has to sum all the frequencies corresponding to all crop models times all BA methods, typically those shown in **Fig. S3**.

Among the UV methods (see **Fig. S4** for details), UP_CDFt and UG_SDM and UC_EQM are appreciably better than the others as they occupy distinctively the first ranking positions with frequencies noticeably higher than the other UV methods. However, as pointed out earlier in the case shown in Table 3, UP_CDFt also shows a noticeable frequencies in the lower ranks.

Having determined the MV BA methods are the ones that produce most frequently the smallest errors, we can further break down the results to a finer scale. In fact it is important to verify whether all crop models react in a similar way to the individual MV BA methods and whether any BA-method as a prevailing best performance. In Table 5 the average results presented in Table 4 for Rank#1 are expanded for all methods and crop models and ranked from the highest to the smallest frequency value. From the table we can deduce that there are two methods in particular that produce the best results among the MVs, namely ECCC_MBCn and IPSL_R2D2. At the same time in the first 12 positions, 11 crop models (DE04 results not included in this analysis) are present with the majority of them producing the best results when using data BAed by IPSL_R2D2.

3.4.2. Break down by climate model

As anticipated above, the same analysis has been performed for climate models. This was a precautionary action in case some would stand out more than others in producing improved crop model results. Not unexpectedly all climate models except NOAA's, do not co-vary with any BA method as shown in **Fig. S4a and b**. NOAA is the only model that produces errors in all crop models with a frequency that is systematically different from the others and for all BA methods. A statistical analysis was conducted to see how the two groups of methods (MV and UV) where performing with respect to the climate models BA. As we can see from Table 6 for every other climate model, at mean, median and percentile levels crop model results based on MV BAed data show for the majority of cases lower errors than those obtained using UV BAed climate model results.

From the results presented in this section, we can therefore select almost any climate model to further increase the granularity of the analysis at the crop model variable and station level sure, that it would

Table 4
Ranking positions 1 through 3 and 12–14 for the individual crop models. For every rank the percentage of cases in which the error occupies the rank is shown. The results are presented as average of the frequencies obtained when running the crop models on 5 MV and 9 UV BAed climate model results.

Crop model	Error ranking Average frequency (%), across BA methods, of errors (Rank# 1 ≥ smallest errors, Rank# 14≥ largest errors)												
	#1		#2		#3		[…]	#12		#13		#14	
	MV	UV	MV	UV	MV	UV		MV	UV	MV	UV	MV	UV
CN02	0.82	0.48	0.97	0.47	1.10	0.40		0.53	0.71	0.30	0.84	0.44	0.89
DE02	0.77	0.70	0.67	0.64	0.78	0.57		0.76	0.59	0.43	0.77	0.69	0.76
DE06	0.74	0.73	0.54	0.77	0.79	0.66		0.59	0.75	0.68	0.71		
DE08	0.73	0.53	0.60	0.68	0.72	0.61		0.76	0.59	0.53	0.72	0.78	0.71
INT02	1.05	0.47	1.05	0.43	1.10	0.40		0.48	0.74	0.30	0.84	0.37	0.94
IT01	0.95	0.58	0.81	0.43	0.90	0.51		0.70	0.62	0.41	0.78	0.49	0.87
IT02	0.74	0.67	0.81	0.56	0.78	0.58		0.69	0.63	0.50	0.73	0.63	0.79
NL01	0.64	0.58	0.65	0.65	0.72	0.61		0.67	0.64	0.54	0.71	0.65	0.78
PK01	0.89	0.75	0.57	0.70	0.79	0.57		0.52	0.72	0.50	0.73	0.56	0.54
US01	0.71	0.54	0.61	0.67	0.80	0.53		0.71	0.61	0.44	0.77	0.68	0.76
US04	0.67	0.56	0.64	0.65	0.81	0.56		0.73	0.61	0.51	0.73	0.76	0.72

Table 5
Break down of the results of Rank#1 MV column of Table 4. Every crop model was run on 5 MV BAed climate datasets. The RMSE for every combination of MV and crop model is shown. The frequency of occurrence of the errors are presented from the largest to the smallest.

#	MV BA method	Crop model	Rank#1 Table 4	#	MV BA method	Crop model	Rank#1 Table 4
1	ECCC_MBCn	CN02	2.38	28	PIK_ISIMIP3_multi	DE08	0.53
2	ECCC_MBCn	INT02	2.33	29	ECCC_MBCn	US01	0.53
3	IPSL_R2D2	DE06	1.94	30	ECCC_MBCp	DE02	0.51
4	IPSL_R2D2	INT02	1.74	31	PIK_ISIMIP3_multi	IT02	0.51
5	IPSL_R2D2	PK01	1.73	32	PIK_ISIMIP3_multi	CN02	0.45
6	IPSL_R2D2	DE08	1.64	33	PIK_ISIMIP3_multi	NL01	0.45
7	IPSL_R2D2	IT01	1.60	34	ECCC_MBCn	IT01	0.44
8	IPSL_R2D2	NL01	1.60	35	PIK_ISIMIP3_multi	IT01	0.43
9	IPSL_R2D2	US04	1.59	36	ECCC_MBCn	IT02	0.41
10	IPSL_R2D2	IT02	1.47	37	ECCC_MBCr	PK01	0.41
11	IPSL_R2D2	DE02	1.37	38	ECCC_MBCr	US01	0.40
12	IPSL_R2D2	US01	1.33	39	ECCC_MBCp	NL01	0.36
13	PIK_ISIMIP3_multi	PK01	1.15	40	ECCC_MBCp	US01	0.35
14	PIK_ISIMIP3_multi	DE02	1.04	41	ECCC_MBCn	DE02	0.33
15	ECCC_MBCr	IT01	1.03	42	ECCC_MBCp	CN02	0.31
16	PIK_ISIMIP3_multi	US01	0.95	43	ECCC_MBCr	CN02	0.31
17	ECCC_MBCr	INT02	0.81	44	ECCC_MBCp	DE08	0.31
18	ECCC_MBCr	IT02	0.80	45	ECCC_MBCp	IT02	0.28
19	ECCC_MBCr	US04	0.73	46	ECCC_MBCp	PK01	0.26
20	ECCC_MBCp	IT01	0.71	47	ECCC_MBCn	US04	0.25
21	ECCC_MBCr	NL01	0.68	48	PIK_ISIMIP3_multi	INT02	0.24
22	IPSL_R2D2	CN02	0.65	49	ECCC_MBCp	US04	0.21
23	ECCC_MBCr	DE06	0.64	50	ECCC_MBCn	DE06	0.19
24	ECCC_MBCr	DE08	0.64	51	ECCC_MBCn	PK01	0.19
25	ECCC_MBCr	DE02	0.63	52	ECCC_MBCp	DE06	0.18
26	PIK_ISIMIP3_multi	US04	0.56	53	ECCC_MBCp	INT02	0.14
27	ECCC_MBCn	DE08	0.54	54	ECCC_MBCn	NL01	0.13

Table 6
RMSE statistics of crop model results based on MV and UV BAed sets of climate data with respect to the climate models BAed.

Climate model	Median (%)		Mean (%)		Percentile (%)	
	MV < UV	UV > MV	MV < UV	UV > MV	75^{th}MV < 75^{th}UV	25^{th}MV < 25^{th}UV
ICHEC-EC-EARTH/RCA4	*70*	30	*87*	*13*	82	*67*
MPI-M-MPI-ESM-LR/CCLM4–8-17	*42*	58	*76*	*24*	61	*31*
MPI-M-MPI-ESM-LR/RCA4	*49*	51	*75*	*25*	68	*47*
MPI-M-MPI-ESM-LR/REMO2009	*61*	*39*	*76*	*24*	69	*65*
CSIRO-QCCCE-CSIRO-Mk3–6-0/RCA4	*76*	*24*	*86*	*14*	70	*37*
NCC-NorESM1-M/RCA4	*60*	*40*	*83*	*17*	77	*57*
IPSL-IPSL-CM5A-MR/RCA4	*84*	*16*	*93*	*7*	87	*79*
NOAA-GFDL-GFDL-ESM2M/RCA4	*52*	*47*	*76*	*24*	68	*56*

represent the totality minus one of the climate models.

3.4.3. Break down by location and physical and biophysical variable level

Since crop models are insensitive to the underlying climate models we can reduce one degree of freedom when analyzing the data at location. We therefore selected two climate models (IPSL-IPSL-CM5A-MR/RCA4 and MPI-M-MPI-ESM-LR/REMO2009) and analyzed the crop model results based on their BAed output. For the sake of synthesis results at two stations representative of specific regions of the domains were selected, namely: Aswan, the southernmost location of the domain and Jokioinen as the Northernmost location. Fig. 4 (a-b) shows the analysis at Aswan and Jokioinen for all physical variables and all crop models using BAed data from MPI-M-MPI-ESM-LR/REMO2009. The rest of the stations are presented in **Fig. S5** of the SM. The distributions of RMSE errors of every crop model against the AgMERRA based-results are presented as a box-plot that accounts for the distribution of RMSE produced when using climate model data BAed with the 5 MV methods (red box) and the 9 UV methods (black box). The first column in the figure shows the ensemble average of all crop model results for the two MV and UV groups. In Fig. 4 the sensitivity of the different crop models to the two sets of input data and how the latter also varies among the four physical variables are shown. Fig. 4 clearly show an improved performance of crop models when using climate data BAed with MV methods. By splitting the stations in three groups denominated Southern, Central and Northern stations (see Table S1), we find that, in most cases, the Ensemble average error of all 4 physical variables obtained when using MV methods is smaller 66% of the times for southern locations and 42 and 72% of the time for Central and Northern locations respectively. For the variable Yield, MV based input produces an error smaller than a UV based one 90%, 50%, 80% of the time for Southern, Central and Northern locations respectively.

The results presented in Fig. 4 are confirmed when using the 25 year runs for model IT02 and DE02(see SM **Fig. S6** and comment).

A different approach to the analysis consists in calculating the RMSE of the marginal statistics of the crop simulations once averaged over time and space without pairing the data in time and considering the individual locations. In fact climate model outputs, such as the ones used here, are provided by transient simulations, which, as explained in the introduction are free to develop their own states of internal climate variability driven only by the GHG and aerosol concentrations. The use RMSE makes sense for the evaluation of predictions for which the source and target are synchronous, but in the case of transient simulations the internal variability will be a source of noise in the RMSE calculation. In order to show that no significant noise arises from the use of RMSE, marginal-statistic has been calculated and the results are presented in Fig. 5. In the latter IPSL-IPSL-CM5A-MR/RCA4 climate model data (similarly for MPI-M-MPI-ESM-LR/REMO2009 in **Fig. S7**) are used. As it clearly appears MV methods outscores UV ones.

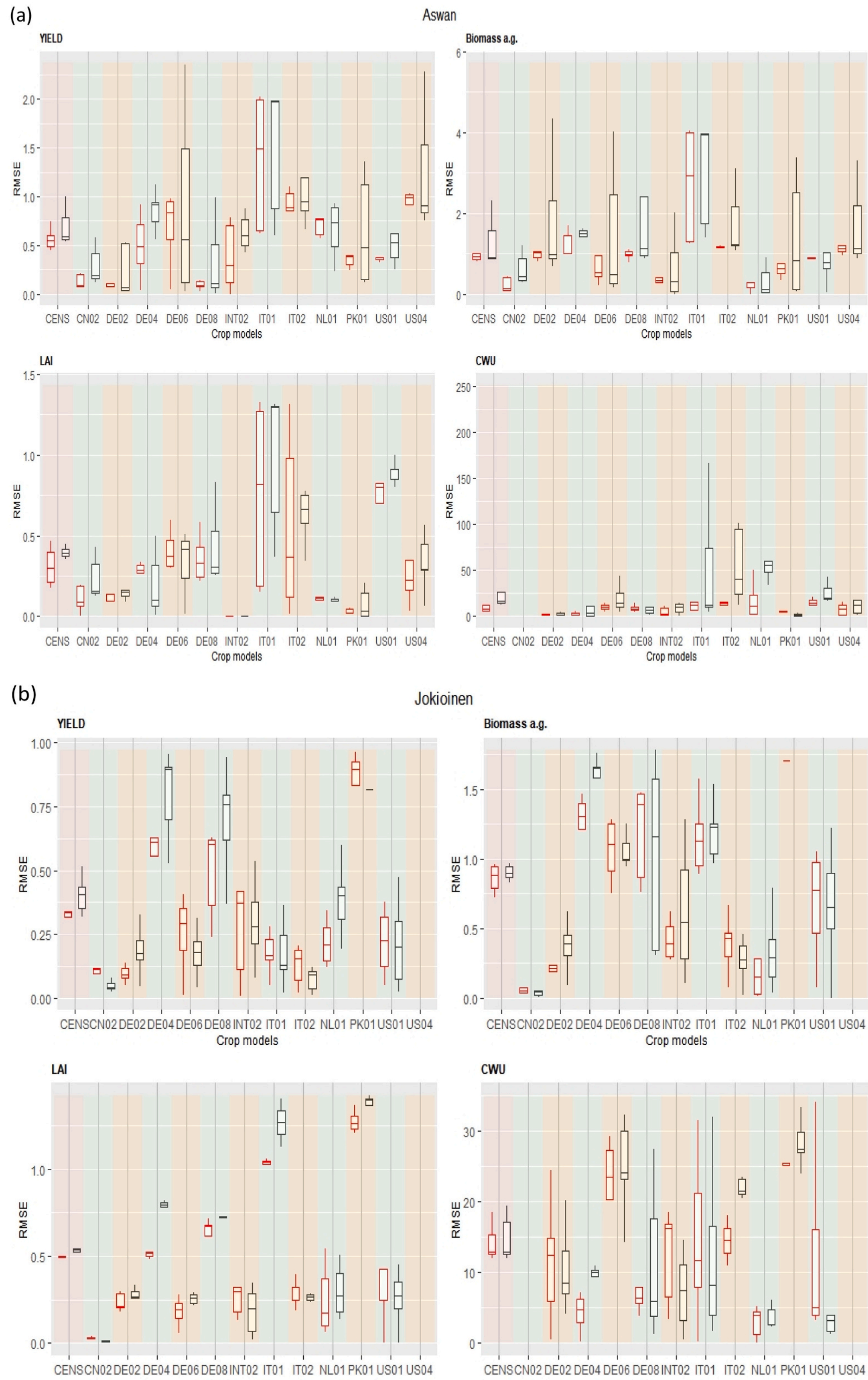


Fig. 4. (a and b): RMSE of the different crop model results at two locations: Aswan (a) and Jokioinen (b). The red box-plot represents the distribution of RMSE produced when analyzing crop model results based on MV methods. The black box-plot relates to those based on UV methods. The pink column to the left (CENS) contains the box-blot obtained when averaging all crop model results for the two groups MV (red) and UV (black). Results from MPI-M-MPI-ESM-LR/REMO2009 climate model. The units of the y-axis are respectively: yield = [Tons/ha], Biomass = [Tons/ha], LAI = dimensionless number, CWU = [mm]. CWU results for CN02 were not submitted.

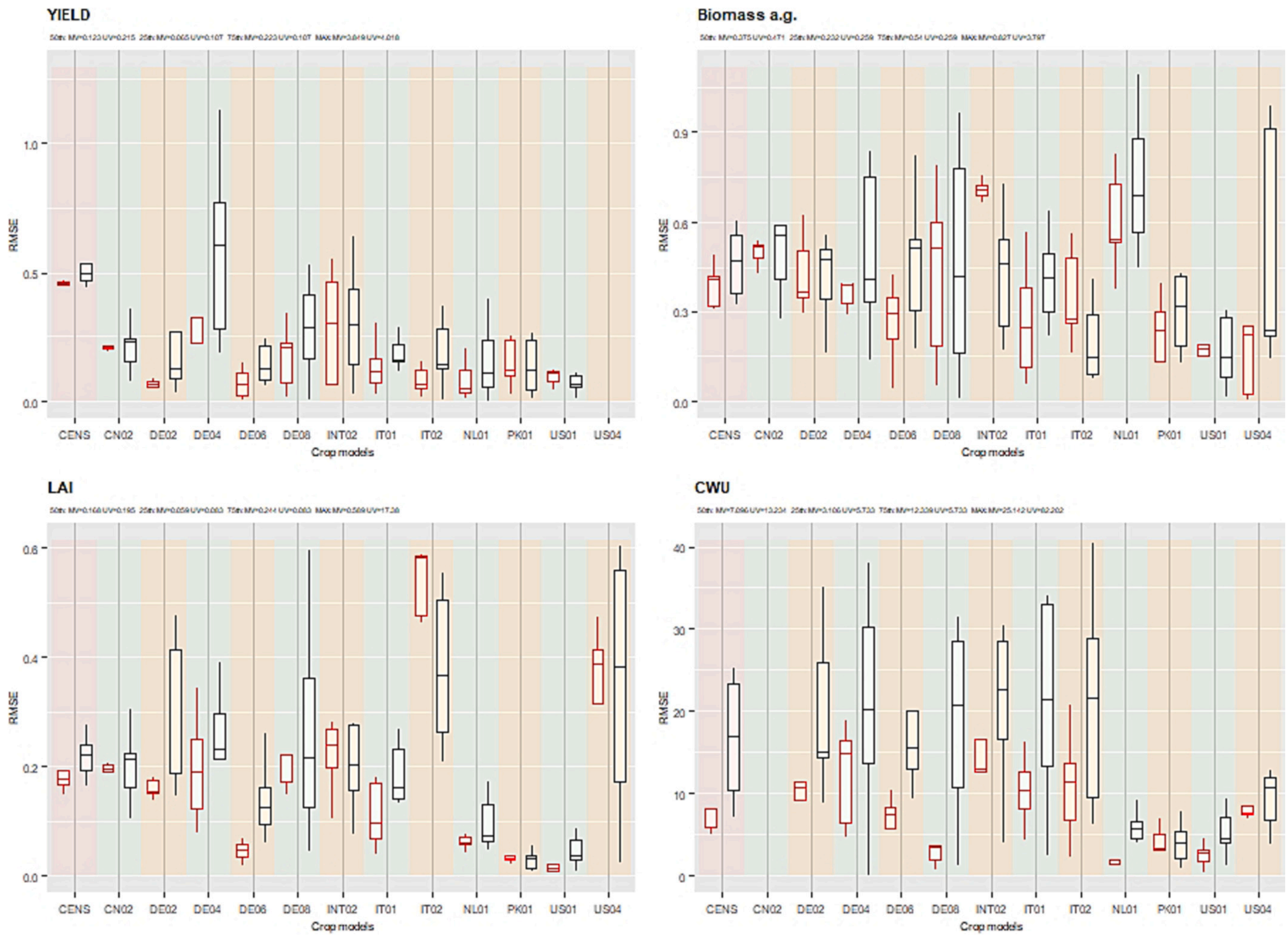


Fig. 5. Analysis of crop model physical variables resulting from IPSL-IPSL-CM5A-MR/RCA4 climate data treated with MV and UV methodologies. This time the RMSE is calculated on the marginal statistics (space and time averaging) rather than pairing the variables. The units of the y-axis are respectively: yield = [Tons/ha], Biomass = [Tons/ha], LAI = dimensionless number, CWU = [mm]. The results based on MPI-M-MPI-ESM-LR/REMO2009 are presented in the SM.

In the *IPSL-IPSL-CM5A-MR/RCA4* MV BAed case, the error is always lower than UV based results for the 50th,25th and 75th percentile. In the *MPI-M-MPI-ESM-LR/REMO2009* MV BA-based case, 75%, 25% and 75% of the time the error lower for the respective percentiles values, indicating that MV outperforms UV methods most of the time at median and extreme portion of the distribution.

3.4.4. Break down by location and phenological variables

The analysis of the impact of MV-data and UV BA-data on the various stages of maturity of the crop at Aswan and Jokioinen are presented in Figs. 6 and **S8**. The same driving climate models and locations have been chosen to facilitate the cross comparison of the results with those of the physical variable. As anticipated by the random forest analysis, the mean bias is considered a better metrics to determine the different impact of the two BA methods. The crop growth stages are identified by the letters “E” for Emergence, “H” for Heading, “F” for Flowering and “M” for Maturity. The dates are calculated as distance in time from the sowing date that is common to all crop models, therefore the plot shows the difference in time of every stage as calculated by each model using all BA corrections and climate model input with the same data calculated using AgMERRA data. The two plots show that the error produced when using MV BA data is on average smaller or equivalent to that when using UV. Even in this case, the difference between the different model results are small even in cases like for example Aswan where all models show a 20-day error in all dates in the case of UV data. Only D04 shows a smaller error at this location. The differences in the calculated dates are in line with the small RMSE of the physical variables especially for what concerns Yield. In fact the duration of the growing stages is one of the controlling parameters of the crop production.

4. Conclusions

Within the BADJAM project several bias adjustment methods belonging to two distinct categories, (univariate and multivariate) and applied to climate model results were compared in terms of how they affect the performance of crop models when climate data are used. A selection of 8 EURO-CORDEX results were bias corrected against AgMERRA by using 9 univariate and 5 multivariate methods and then used by the 8 AgMIP-Wheat community crop models. The crop models were run at 21 locations across Europe, North Africa and the Middle-East. and the results were compared against those obtained using the observationally constrained AgMERRA model data. The bias correction was performed from 1980 to 2005 of which the latter 20 were used as training period and the first five as evaluation. The choice of a 5-year period for the evaluation was tested by comparing the results from a model restarted at every growing season and a model calculating the water soil content with continuity through out the period. No difference was found between the two results when the models are run for 5 years or 25 years thus supporting the choice of running the models only for the evaluation period of the BA. An explanation was provided that depends on the nature of the crop models that took part to BADJAM and the way they were run.

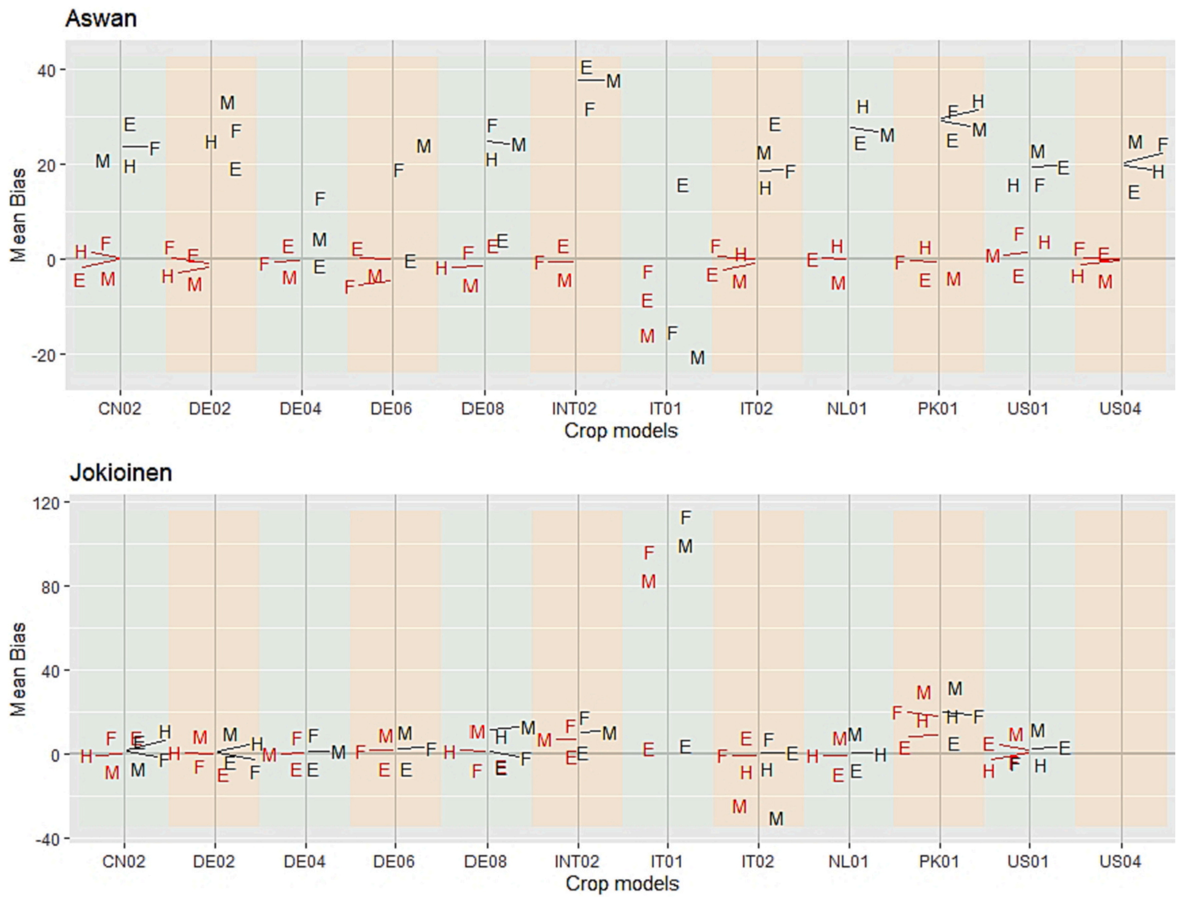


Fig. 6. Mean bias of phenological stages (E = emergence, H = heading, F = flowering, M = maturity) at some of the 21 stations. More station results are presented in Fig. S9. Data expressed as the MAB (in days) between the phenological stages calculated by each crop model when using AgMERRA and those obtained from each crop model when using MV BAed climate data (red) and UV BAed climate data (black). IPSL-IPSL-CM5A-MR/RCA4 is the BAed climate model to which these result pertain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

By means of a machine learning method we have identified RMSE and the MB as most appropriate metrics to determine the magnitude of the differences between the crop model results obtained using the bias-adjusted climate data and those obtained when using AGMERRA.

The results clearly show that crop models produce lower errors when using MV-BAed data than when using UV ones. Globally the multivariate results produce smaller errors compared to univariate most of the time. Increasing the granularity of the analysis shows that the choice of the climate model has no effect on the relative performance of the methods. Errors of crop model results based on data MV BAed for southern and Northern locations climate data are 90 and 80% of the times smaller than those based on UV BAed data. For the locations central to the domain, the percentage reduces to 50%. The reason for this result could be due to more extreme conditions present at lower and higher latitudes or more systematic differences between AgMerra and the EURO-CORDEX results, though more detailed analysis will be required.

Since crop models of the kind that took part to BADJAM strictly rely on and appear to be sensitive to the accuracy of the input data, in case of climate applications MV BAed data should be preferred. One could further generalize this result to any modelling system that directly uses BAed climate data.

Disclaimer

The views or positions expressed in this publication do not necessarily represent in legal terms the official position of the European Food Safety Authority (EFSA). EFSA assumes no responsibility or liability for any errors or inaccuracies that may appear.

Credit authorship contribution statement

S. Galmarini: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **E. Solazzo:** Conceptualization, Data curation, Formal analysis. **R. Ferrise:** Conceptualization, Data curation, Formal analysis. **A. Kumar Srivastava:** Conceptualization, Data curation, Formal analysis. **M. Ahmed:** Conceptualization, Data curation, Formal analysis. **S. Asseng:** Conceptualization, Data curation, Formal analysis. **A.J. Cannon:** Conceptualization, Data curation, Formal analysis. **F. Dentener:** Supervision. **G. De Sanctis:** Conceptualization, Data curation, Formal analysis. **T. Gaiser:** Conceptualization, Data curation, Formal analysis. **Y. Gao:** Conceptualization, Data curation, Formal analysis. **S. Gayler:** Conceptualization, Data curation, Formal analysis. **J.M. Gutierrez:** Conceptualization, Data curation, Formal analysis. **G. Hoogenboom:** Conceptualization, Data curation, Formal analysis. **M. Iturbide:** Conceptualization, Data curation, Formal analysis. **M. Jury:** Conceptualization, Data curation, Formal analysis. **S. Lange:** Conceptualization, Data curation, Formal analysis. **H. Loukos:** Conceptualization, Data curation, Formal analysis. **D. Maraun:** Conceptualization, Data curation, Formal analysis. **M. Moriondo:** Conceptualization, Data curation, Formal analysis. **S. McGinnis:** Conceptualization, Data curation, Formal analysis. **C. Nendel:** Conceptualization, Data curation, Formal analysis. **G. Padovan:** Conceptualization, Data curation, Formal analysis. **A. Riccio:** Conceptualization, Data curation, Formal analysis. **D. Ripoche:** Conceptualization, Data curation, Formal analysis. **C.O. Stockle:** Conceptualization, Data curation, Formal analysis. **I. Supit:** Conceptualization, Data curation, Formal analysis. **S. Thao:** Conceptualization, Data curation, Formal analysis. **G. Trombi:** Conceptualization, Data curation, Formal analysis. **M. Vrac:** Conceptualization, Data curation, Formal analysis. **T. K.D. Weber:** Conceptualization, Data curation, Formal analysis. **C. Zhao:** Conceptualization, Data curation, Formal analysis.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.agsy.2023.103846.

References

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rotter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Mueller, C., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stoeckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Chang. 3, 827–832. https://doi.org/10.1038/nclimate1916.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V. V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2014. Rising temperatures reduce global wheat production. Nat. Clim. Chang. 5, 143–147. https://doi.org/10.1038/nclimate2470.

Asseng, S., Martre, Pierre, Maiorano, Andrea, Rötter, Reimund P., O'Leary, Garry J., Fitzgerald, Glenn J., Girousse, Christine, Motzo, Rosella, Francesco Giunta, M., Babar, Ali, Reynolds, Matthew P., Kheir, Ahmed M.S., Thorburn, Peter J., Waha, Katharina, Ruane, Alex C., Aggarwal, Pramod K., Ahmed, Mukhtar, Balkovič, Juraj, Basso, Bruno, Biernath, Christian, Bindi, Marco, Cammarano, Davide, Challinor, Andrew J., De Sanctis, Giacomo, Dumont, Benjamin, Rezaei, Ehsan Eyshi, Fereres, Elias, Ferrise, Roberto, Garcia-Vila, Margarita, Gayler, Sebastian, Gao, Yujing, Horan, Heidi, Gerrit Hoogenboom, R., Izaurralde, César, Jabloun, Mohamed, Jones, Curtis D., Kassie, Belay T., Kersebaum, Kurt-Christian, Klein, Christian, Koehler, Ann-Kristin, Liu, Bing, Minoli, Sara, Martin, Manuel Montesino San, Müller, Christoph, Kumar, Soora Naresh, Nendel, Claas, Olesen, Jørgen Eivind, Palosuo, Taru, Porter, John R., Priesack, Eckart, Ripoche, Dominique, Semenov, Mikhail A., Stöckle, Claudio, Stratonovitch, Pierre, Streck, Thilo, Supit, Iwan, Tao, Fulu, Van der Velde, Marijn, Wallach, Daniel, Wang, Enli, Webber, Heidi, Wolf, Joost, Xiao, Liujun, Zhang, Zhao, Zhao, Zhigan, Zhu, Yan, Ewert, Frank, 2019. Climate change impact and adaptation for wheat protein. Glob. Chang. Biol. 25, 155–173. https://doi.org/10.1111/gcb.14481.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, G., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P., Hénault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model STICS. Eur. J. Agron. 18, 309–322.

Brisson, N., Launay, M., Mary, B., Beaudoin, N. (Eds.), 2009. Conceptual Basis, Formalisations and Parameterization of the STICS Crop Model. Editions Quae, Versailles.

Cannon, A.J., 2016. Multivariate bias correction of climate model output: matching marginal distributions and intervariable dependence structure. J. Clim. 29 (19), 7045–7064. https://doi.org/10.1175/JCLI-D-15-0679.1.

Cannon, A.J., 2018. Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. Clim. Dyn. 50, 31–49. https://doi.org/10.1007/s00382-017-3580-6.

Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? J. Clim. 28 (17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1.

Cannon, A.J., Piani, C., Sippel, S., 2020. Bias correction of climate model output for impact models. In: Climate Extremes and Their Implications for Impact and Risk Assessment. https://doi.org/10.1016/B978-0-12-814895-2.00005-7.

Dosio, A., Paruolo, P., 2011. Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: evaluation on the present climate. J. Geophys. Res. 116 (D16), 1–22.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.

Gaiser, T., Perkons, Ute, Kupper, P.M., Kautz, T., Puschmann, D.U., Ewert, F., Enders, A., Krauss, G., 2013. Modeling biopore effects on root growth and biomass production on solis with pronounced sub-soil clay accumulation, 256, pp. 6–15. https://doi.org/10.1016/j.ecolmodel.2013.02.016.

Galmarini, S., Cannon, A.J., Ceglar, A., Christensen, O.B., de Noblet-Ducoudré, N., Dentener, F., Doblas-Reyes, F.J., Dosio, A., Gutierrez, J.M., Iturbide, M., Jury, M., Lange, S., Loukos, H., Maiorano, A., Maraun, D., McGinnis, S., Nikulin, G., Riccio, A., Sanchez, E., Solazzo, E., Toreti, A., Vrac, M., Zampieri, M., 2019. Adjusting climate model bias for agricultural impact assessment: How to cut the mustard, 13, pp. 65–69. https://doi.org/10.1016/j.cliser.2019.01.004.

Gayler, S., Wang, E., Priesack, E., Schaaf, T., Maidl, F.-X., 2002. Modeling biomass growth, N-uptake and phenological development of potato crop. Geoderma 105 (3), 367–383. https://doi.org/10.1016/S0016-7061(01)00113-6.

Gobiet, A., Jacob, D., 2011. The EURO-CORDEX-initiative. In: WCRP Open Science Conference Denver, 24-28 Oct. 2011. Poster no. W108B.

Ho, Tin Kam, 1995. Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995, pp. 278–282.

Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., White, J. W., Asseng, S., Lizaso, J.I., Moreno, L.P., Pavan, W., Ogoshi, R., Hunt, L.A., Tsuji, G. Y., Jones, J.W., 2019a. The DSSAT crop modeling ecosystem. In: Boote, K.J. (Ed.), Advances in Crop Modeling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, United Kingdom, pp. 173–216. https://doi.org/10.19103/AS.2019.0061.10.

Hoogenboom, G., Porter, C.H., Shelia, V., Boote, K.J., Singh, U., White, J.W., Hunt, L.A., Ogoshi, R., Lizaso, J.I., Koo, J., Asseng, S., Singels, A., Moreno, L.P., Jones, J.W., . Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.7. www.DSSAT.net. DSSAT Foundation, Gainesville, Florida, USA.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. DSSAT cropping system model. Eur. J. Agron. 18, 235–265.

Kassie, B.T., Asseng, Senthold, Porter, Cheryl H., Royce, Frederick S., 2016. Performance of DSSAT-Nwheat across a wide range of current and future growing conditions. Eur. J. Agron. 81, 27–36. ISSN 1161-0301,. https://doi.org/10.1016/j.eja.2016.08.012.

Lange, S., 2019. Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). Geosci. Model Dev. 12, 3055–3070. https://doi.org/10.5194/gmd-12-3055-2019.

Maraun and Widmann, 2018. Statistical Downscaling and Bias Correction for Climate Research. Cambridge University Press, Cambridge.

Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L., 2017. Towards process-informed bias correction of climate change simulations. Nat. Clim. Chang. 7 https://doi.org/10.1038/nclimate3418.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., Cammarano, D., Hatfield, J.L., Rosenzweig, C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2014. Multimodel ensembles of wheat growth: many models are better than one. Glob. Chang. Biol. https://doi.org/10.1111/gcb.12768.

McGinnis, S., Nychka, D., Mearns, L.O., 2015. Machine Learning and Data Mining Approaches to Climate Science. Springer International Publishing, pp. 91–99. A new distribution mapping technique for climate model bias correction.

Nendel, C., Berg, M., Kersebaum, K.C., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K.O., Wieland, R., 2011. The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. Ecol. Model. 222 (9), 1614–1625. https://doi.org/10.1016/j.ecolmodel.2011.02.018.

Priesack, E., 2006. Expert-N Dokumentation der Modellbibliothek: Zugl.: Göttingen, Univ., Habil.-Schr., 2006. FAM-Bericht, vol. 60. Hieronymus, München.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The agricultural model Intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric. For. Meteorol. 170, 166–182. https://doi.org/10.1016/j.agrformet.2012.09.011.

Ruane, A.C., Goldberg, R., Chryssanthacopoulos, J., 2015. Climate forcing datasets for agricultural modeling: merged products for gap-filling and historical climate series estimation. Agric. Forest Meteorol. 200, 233–248. https://doi.org/10.1016/j.agrformet.2014.09.016.

Shibu, M.E., Leffelaar, P.A., Van Keulen, H., Aggarwal, P.K., 2010. LINTUL3, a simulation model for nitrogen limited situations: application to rice. Eur. J. Agron. 32, 255–271.

Soltani, A., Maddah, V., Sinclair, T., 2013. SSM-wheat: a simulation model for wheat development, growth and yield. Int. J. Plant Product. 7, 711–740.

Speiser, Jaime Lynn, Miller, Michael E., Tooze, Janet, Ip, Edward, 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst. Appl. 134, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028.

Stöckle, C.O., Donatelli, M., Nelson, R., 2003. CropSyst, a cropping systems simulation model. Eur. J. Agron. 18 (3–4), 289–307.

Switanek, M.B., Troch, P.A., Castro, C.L., Leuprecht, A., Chang, H.-I., Mukherjee, R., Demaria, E.M.C., 2017. Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes. Hydrol. Earth Syst. Sci. 21, 2649–2666. https://doi.org/10.5194/hess-21-2649-2017.

URL1, 2023. https://agmip.org/overview/ last accessed 08/08/2023.

URL2, 2023. https://github.com/zalf-rpm/monica last accessed 08/08/2023.
Vrac, M., 2018. Multivariate bias adjustment of high-dimensional climate simulations: the rank resampling for distributions and dependences (R2D2) bias correction. Hydrol. Earth Syst. Sci. 22, 3175–3196. https://doi.org/10.5194/hess-22-3175-2018.
Wang, E., Engel, T., 2000. SPASS: a generic process-oriented crop model with versatile windows interfaces. Environ. Model Softw. 15 (2), 179–188. https://doi.org/10.1016/S1364-8152(99)00033-X.
Wolf, J., 2012. User Guide for LINTUL5: Simple Generic Model for Simulation of Crop Growth Under Potential, Water Limited and Nitrogen, Phosphorus and Potassium Limited Conditions. Wageningen UR, pp. 1–63.