Received: 16 November 2021 | Accepted: 30 March 2022

DOI: 10.1002/vzj2.20202

Vadose Zone Journal

ORIGINAL RESEARCH ARTICLE

Same soil, different climate: Crop model intercomparison on translocated lysimeters

Jannis Groh[1,2,3] | **Efstathios Diamantopoulos**[4] | **Xiaohong Duan**[5] | **Frank Ewert**[1,3] | **Florian Heinlein**[5] | **Michael Herbst**[2] | **Maja Holbak**[4] | **Bahareh Kamali**[1,3] | **Kurt-Christian Kersebaum**[1,6,7] | **Matthias Kuhnert**[8] | **Claas Nendel**[1,6,9] | **Eckart Priesack**[5,10] | **Jörg Steidl**[1] | **Michael Sommer**[1] | **Thomas Pütz**[2] | **Jan Vanderborght**[2] | **Harry Vereecken**[2] | **Evelyn Wallor**[1,11] | **Tobias K. D. Weber**[10] | **Martin Wegehenkel**[1] | **Lutz Weihermüller**[2] | **Horst H. Gerke**[1]

[1]Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany

[2]Forschungszentrum Jülich GmbH, Agrosphere, Institute of Bio- and Geoscience IBG-3, Jülich, Germany

[3]Institute of Crop Science and Resource Conservations (INRES), Univ. of Bonn, Bonn, Germany

[4]Dep. of Plant and Environmental Science, Univ. of Copenhagen, Copenhagen, Denmark

[5]Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany

[6]Global Change Research Institute, The Czech Academy of Sciences, Brno, Czech Republic

[7]Institute for Tropical Plant Production and Agricultural Systems Modelling, Georg-August Univ., Göttingen, Germany

[8]Institute of Biological and Environmental Science, Univ. of Aberdeen, Aberdeen, UK

[9]Institute of Biochemistry and Biology, Univ. of Potsdam, Potsdam, Germany

[10]Institute of Soil Science and Land Evaluation, Univ. of Hohenheim, Stuttgart, Germany

[11]Faculty of Forest and Environment, Univ. for Sustainable Development Eberswalde, Eberswalde, Germany

Correspondence
Jannis Groh, Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany.
Email: j.groh@fz-juelich.de

Assigned to Associate Editor Asim Biswas.

Abstract

Crop model intercomparison studies have mostly focused on the assessment of predictive capabilities for crop development using weather and basic soil data from the same location. Still challenging is the model performance when considering complex interrelations between soil and crop dynamics under a changing climate. The objective of this study was to test the agronomic crop and environmental flux-related performance of a set of crop models. The aim was to predict weighing lysimeter-based crop (i.e., agronomic) and water-related flux or state data (i.e., environmental) obtained for the same soil monoliths that were taken from their original environment

Abbreviations: AC, AgroC; BL, Bad Lauchstädt; CE, Expert-N coupled to CERES; DC, DailyDayCent; Dd, Dedelow; D-W, region with drier and warmer climatic conditions; DY, Daisy; GE, Expert-N coupled to GECROS; HE, HERMES; HG, HydroGeoSphere; MM, multi-model mean; MO, MONICA; Se, Selhausen; SP, Expert-N coupled to SPASS; SU, Expert-N coupled to SUCROS; TERENO, TERrestrial ENvironmental Observatories; TH, THESEUS; W-W, region with wetter and warmer climatic conditions.

and translocated to regions with different climatic conditions, after model calibration at the original site. Eleven models were deployed in the study. The lysimeter data (2014–2018) were from the Dedelow (Dd), Bad Lauchstädt (BL), and Selhausen (Se) sites of the TERENO (TERrestrial ENvironmental Observatories) SOILCan network. Soil monoliths from Dd were transferred to the drier and warmer BL site and the wetter and warmer Se site, which allowed a comparison of similar soil and crop under varying climatic conditions. The model parameters were calibrated using an identical set of crop- and soil-related data from Dd. Environmental fluxes and crop growth of Dd soil were predicted for conditions at BL and Se sites using the calibrated models. The comparison of predicted and measured data of Dd lysimeters at BL and Se revealed differences among models. At site BL, the crop models predicted agronomic and environmental components similarly well. Model performance values indicate that the environmental components at site Se were better predicted than agronomic ones. The multi-model mean was for most observations the better predictor compared with those of individual models. For Se site conditions, crop models failed to predict site-specific crop development indicating that climatic conditions (i.e., heat stress) were outside the range of variation in the data sets considered for model calibration. For improving predictive ability of crop models (i.e., productivity and fluxes), more attention should be paid to soil-related data (i.e., water fluxes and system states) when simulating soil–crop–climate interrelations in changing climatic conditions.

1 | INTRODUCTION

Process-based crop growth models are an essential and invaluable tool to assess the impact of climate change on future crop yields (Adam et al., 2020; Challinor et al., 2018). The crop models used in these climate change impact assessments studies generally assume an ecosystem response to changing climatic conditions. For instance, the response to rising temperature and CO_2 in plant productivity can be described with similar functional relationships and parameterization as under current climatic conditions (Ainsworth & Rogers, 2007; Asseng et al., 2015; Lenka et al., 2020). In addition, Boote et al. (2010) pointed out that in the past, developers of crop models have often assumed that these functional relationships are correct, and that insufficient effort has been made to improve these relationships or to test them in light of the latest scientific findings.

Future crop growth assessments based on process-based crop model intercomparison studies have mostly been carried out using model parameters calibrated for soils under original climatic conditions in their original regions (Mulla et al., 2020; Ruane et al., 2017; F. Tao et al., 2020; Webber, Hoffmann, & Rezaei, 2018). All these studies were based on soils developed at a site under a specific past climate and management, assuming that the soil–crop–atmosphere system is adapted to the atmospheric and lower boundary conditions. The problem with this approach is that modelers calibrate and validate their specific model in a range of environments (i.e., soil and climate conditions) and use the parameterized models with given structures to predict crop development under environments not included in the calibration, as pointed out by Wallach, Palosuo, Thorburn, Gourdain, et al. (2021) and Wallach, Palosuo, Thorburn, Hochman, Andrianasolo, et al. (2021). Changes in climatic conditions (e.g., induced by future climate change) will affect not only crop growth but also the soils (e.g., Robinson et al., 2016, 2019). Thus, the ability of individual crop models to predict effects of changing climatic conditions on soil ecosystems remains untested for the range beyond the measured site-specific variability of environmental conditions.

Furthermore, the analyses of crop model simulations mostly focus on the physiological processes such as growth in biomass, yield, and phenological development during the vegetation period (e.g., Martre et al., 2015; Wallach, Palosuo, Thorburn, Gourdain, et al., 2021; Wallach, Palosuo, Thorburn, Hochman, Andrianasolo, et al., 2021; Yin et al., 2017). However, for the simultaneous prediction of agronomic (crop production) and environmental (boundary and internal fluxes and states) variables under changing climatic conditions, the soil–plant–atmosphere system must be considered as a whole

(i.e., adequate attention should be paid to crops and soil water fluxes and states). This includes continuous simulations, including crop-free periods and lead times (e.g., Groh, Diamantopoulos, et al., 2020). Analyzing water and element fluxes, as well as soil water storage, becomes increasingly important when trying to predict crop development under changing climatic and environmental conditions because of the feedbacks of those fluxes and states with the crop development.

A common strategy to develop site-specific soil use and management strategies and predict agricultural productivity in the same soil under changing climatic conditions is to carry out crop model simulations for climate scenarios, generated based on global atmospheric circulation models (e.g., Abd-Elmabod et al., 2020; Chisanga et al., 2020; Zydelis et al., 2021). Typically, crop models have first been calibrated and validated based on multiannual observations for a specific crop (or crop rotation) at various sites (e.g., Kollas et al., 2015; Nendel et al., 2011) to represent a wide range of environmental conditions in these studies. In a second step, the calibrated models have been applied to predict crop productivity under future climatic conditions (e.g., Schleussner et al., 2018). However, the ideal situation regarding data quality for calibration and validation (Kersebaum et al., 2015) has been realized only in a few studies (Archontoulis et al., 2020). Even more frequently, the models are often calibrated at a single site only, for example, due to other experimental constraints. Crop model intercomparison studies as performed by the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Ruane et al., 2017) or by the Modelling Agriculture with Climate Change for Food Security (MACSUR) (Asseng et al., 2013; Bassu et al., 2014) among other initiatives, explained differences in predictions between the crop models at the same site by a differential response to increased temperature and atmospheric CO_2 concentrations in leaf area development due to different model formulations (e.g., F. Tao et al., 2020) or by the limited representation of differences in soil properties at heterogeneous landscapes (Groh, Diamantopoulos, et al., 2020). Well balanced, high-quality datasets as demanded by Kersebaum et al. (2015) for model calibration and validation are rare, and in various cases, datasets have been used in which the crop has not entirely been free of stress. For the model calibration procedure, this means that modelers calibrate plant-specific parameters, while they should have also accounted for soil-related parameters (e.g., Asseng et al., 2013; Bassu et al., 2014). The resulting parameter set then often delivers a successful simulation in the application case, but for the wrong reasons (Wallach, Palosuo, Thorburn, Hochman, Gourdain, et al., 2021)

The validation of crop model predictions has furthermore been limited by the availability of data from climate change experiments, especially from long-term monitoring studies (Peng et al., 2020) that also include water and matter fluxes in the soil. Although future climatic conditions may be mimicked (e.g., Deltedesco et al., 2019; Köhler et al., 2019), it is widely unclear how to consider future soil conditions. So far, the coevolution of soil–crop–ecosystems under changing climatic conditions as a longer term process cannot be described yet, but models were developed that could be used for predictions. However, this is crucial, as the effect of soils on climate can lead to large-scale climatic changes. One step to test and develop simulation models or model components could be to compare simulated crop developments and fluxes obtained with calibrated crop model with experimental data from sites with different climates. However, this approach does not allow to disentangle the effect of soil from that of climate because it neglects that soil development is site-specific, depending on climate, vegetation, and parent material, and soils may reach another state with changing boundary conditions (Maurer & Gerke, 2016; Robinson et al., 2016; Zeng et al., 2004).

Testing the ability of crop models for different climatic conditions but for the same soil characteristics requires the translocation of soils into climate zones where they did not develop. Soil monoliths in cropped weighing lysimeters have been proposed for this purpose (Pütz et al., 2016). Thereby, the original soil structure of the soil profile during extraction is preserved. Furthermore, water balance components such as precipitation and evapotranspiration including non-rainfall water (i.e., dew and fog) can be determined more accurately with high-precision weighable lysimeters compared with other typical methods and instruments (Alfieri et al., 2012; Gebler et al., 2015; Groh et al., 2019; Haselow et al., 2019). In general, precise data help to improve model parameterization (Groh, Stumpp, et al., 2018). These lysimeters can be either installed in climate chambers as larger-scale laboratory systems (e.g., Roy et al., 2021) or moved to other regions exhibiting different climatic conditions (e.g., Pütz et al., 2016) to observe and compare the soil–ecosystem response under different climatic conditions, which perhaps includes

Core Ideas

- We demonstrate the use of high precision weighable lysimeter for full model calibration and validation.
- Lysimeter data from translocated soils represent effects of changing climatic conditions.
- We compare calibration with blind forward simulations (fixed soil and calibrated crop parameter).
- We compare individual crop model predictions with multi-model mean.
- We test the predictive ability of crop models and multi-model mean.

the transition period towards arriving at a new state. In the TERENO SOILCan lysimeter project (TERrestrial ENvironmental Observatories; https://www.tereno.net), the transfer of intact soil monoliths between the lysimeter stations followed a space-for-time substitution approach, and sites are representative of agricultural regions in Central Europe with the highest vulnerability to expected climate change effects (Zacharias et al., 2011). The transfer of soils along environmental gradients therefore mimics expected climate changes effects.

Recent field experiments on ecosystem responses to climate change focused mainly on plant- or crop-related processes (e.g., Drought-Net) (Hilton et al., 2019; Loik et al., 2019) and environmental aspects such as soil water content; water and element fluxes received far less attention. Agronomic variables such as grain yield and aboveground biomass are inherently connected with the water and plant nutrient status of the soil (e.g., Kersebaum, 2007). Therefore, in addition to variables controlling crop growth, the soil status controlling the environmental fluxes, as well as the knowledge about the lower boundary conditions (i.e., groundwater level), are needed for a holistic crop model calibration and validation (Groh, Diamantopoulos, et al., 2020; Groh et al., 2016). So far, lysimeter data from the same site (climate) but for different soil profiles have been used for crop model predictions (Groh, Diamantopoulos, et al., 2020). Applying these same data from the SOILCan lysimeter network, especially from those translocated to regions with different climatic conditions, bears the opportunity to test whether the structure of the models is adequate to simultaneously predict crop development and environmental fluxes for same soil under different climatic conditions.

The overall aim of this study was to compare the ability of an ensemble of crop models to predict the ecosystem's agronomic productivity (grain yield and above ground biomass), environmental fluxes (evapotranspiration, drainage), and states (soil water content) for lysimeters containing the same soils but exposed to different climates. We hypothesize, that all the calibrated crop models when using identical soil hydraulic properties will be able to similarly well reproduce the agronomic and environmental variables at the calibration site, but model performance will diverge when predicting crop growth, yield, and soil water flux for conditions outside the range of site-specific variations. The objectives were (a) to assess the effect of changed climatic conditions on key agronomic and environmental variables and (b) to test the predictive capability of crop models to simulate this effect. The model comparison will contribute to test a widely used procedure in climate change impact assessments studies (e.g., Falconnier et al., 2020; Franke et al., 2020; Rosenzweig et al., 2014; Zydelis et al., 2021), in which calibration is based on a set of climatic conditions and prediction will be made for climatic conditions outside the calibration range.

2 | MATERIAL AND METHODS

2.1 | Site and soil descriptions

The Dedelow (Dd; Figure 1) site (53°22′2.45″ N, 13°48′10.91″ W, 50–60 m asl) is located in the region of the younger moraines in northeastern Germany, which is an intensively cultivated hummocky arable soil landscape. The Dd lysimeter station is part of the Northeast German Lowland Observatory of TERENO (Heinrich et al., 2018) and the German-wide lysimeter network SOILCan (Pütz et al., 2016). A total of nine intact eroded Luvisol soil monoliths (c.f., Groh, Diamantopoulos, et al., 2020) were extracted along an approximately 20-m transect within the experimental field in Dd in 2010 (Figure 2).

Following the concept of the TERENO SOILCan lysimeter experiment, three of the soil monoliths have been transferred from Dd to (a) a drier and warmer (D-W) region in the central (Bad Lauchstädt, BL) and three have been transferred to (b) a wetter and warmer (W-W) region in the western part of Germany (Selhausen, Se). Crop management remained similar at all sites except for regionally adopted fertilizer application, plant growth regulators, and pesticides (see Supplemental Tables S3–S6). Three soil monoliths remained at the Dd station. The experimental field station in BL (51°23′37″ N, 11°52′41″ E, 113 m asl) is located in central Germany and the station in Se (50°52′7″ N, 6°26′58″ E, 104 m asl) is located in the western part of Germany (Figure 1). The BL site is part of the Harz/Central German Lowland Observatory, whereas the Se site is part of the Eifel/Lower Rhine Valley Observatory of the TERENO network (Bogena et al., 2018; Wollschläger et al., 2016). At each site, the air temperature (T_a), global radiation (R_g), wind speed (WS), rainfall (R), relative humidity, and barometric pressure data were monitored.

The climatic conditions at BL and Se (Figure 1) differed from those of the (soil-origin) site in Dd with respect to annual mean T_a, mean annual R, grass reference evapotranspiration (ET_0), WS, and annual R_g. For 2014–2018, the soil monoliths from Dd were subjected at Se to higher annual average T_a (+1.6°C), R (+58 mm yr^{-1}), and R_g (+65.4 kWh m^{-2} yr^{-1}), but to a lower WS (−0.3 m s^{-1}), and ET_0 (−102 mm yr^{-1}). At BL, the annual averages were higher than in Dd for T_a (+1.0°C), ET_0 (+28 mm yr^{-1}), and R_g (+31.3 kWh m^{-2} yr^{-1}), but lower for WS (−0.1 m s^{-1}), and R (−77 mm yr^{-1}, see Table 1).

The D-W climatic conditions at BL are characterized by a climatic aridity index (AI = ET_0/R) of 1.47, and the W-W climate by an AI value of 0.96, as compared with the AI value of 1.23 at Dd (Table 1). The water balance components obtained from three lysimeters at each site (Supplemental Table S1) show the basic hydrological differences between

TABLE 1 Site specific mean air temperature (T_a, annual and seasonal), mean rainfall (R, annual and seasonal), potential annual evapotranspiration determined as grass reference evapotranspiration (ET_0; Allen et al., 2006), aridity index (AI) defined as the ratio between R and ET_0, annual global radiation (R_g), and daily mean wind speed (WS) during the observation period (November 2014–October 2018) at Dedelow (Dd), Bad Lauchstädt (BL), and Selhausen (Se)

	T_a			R						
Location	**Annual mean**	**Fall/winter[a]**	**Spring/summer[a]**	**Annual mean**	**Fall/winter[a]**	**Spring/summer[a]**	**ET_0**	**AI**	**R_g**	**WS**
	°C			mm			mm yr^{-1}		kWh m^{-2} yr^{-1}	m s^{-1}
Dd	9.8	4.5	14.9	591	286	305	727	1.23	1,105.8	2.7
BL	10.8	5.2	16.1	514	224	290	755	1.47	1,170.8	2.6
Se	11.4	6.7	15.8	649	324	325	625	0.96	1,217.7	2.4
Δ(BL-Dd)	+1.0	+0.7	+1.2	−77	−63	−15	+28	+0.24	+65.1	−0.1
Δ(Se-Dd)	+1.6	+2.2	+0.9	+58	+38	+20	−102	−0.27	+111.9	−0.3

Note. Both Δ(BL-Dd) and Δ(Se-Dd) denote the differences between the respective sites and those at Dedelow

[a]According to the definition of astronomical season for fall/winter (23 September until 19 March) and spring/summer (20 March until 22 September).

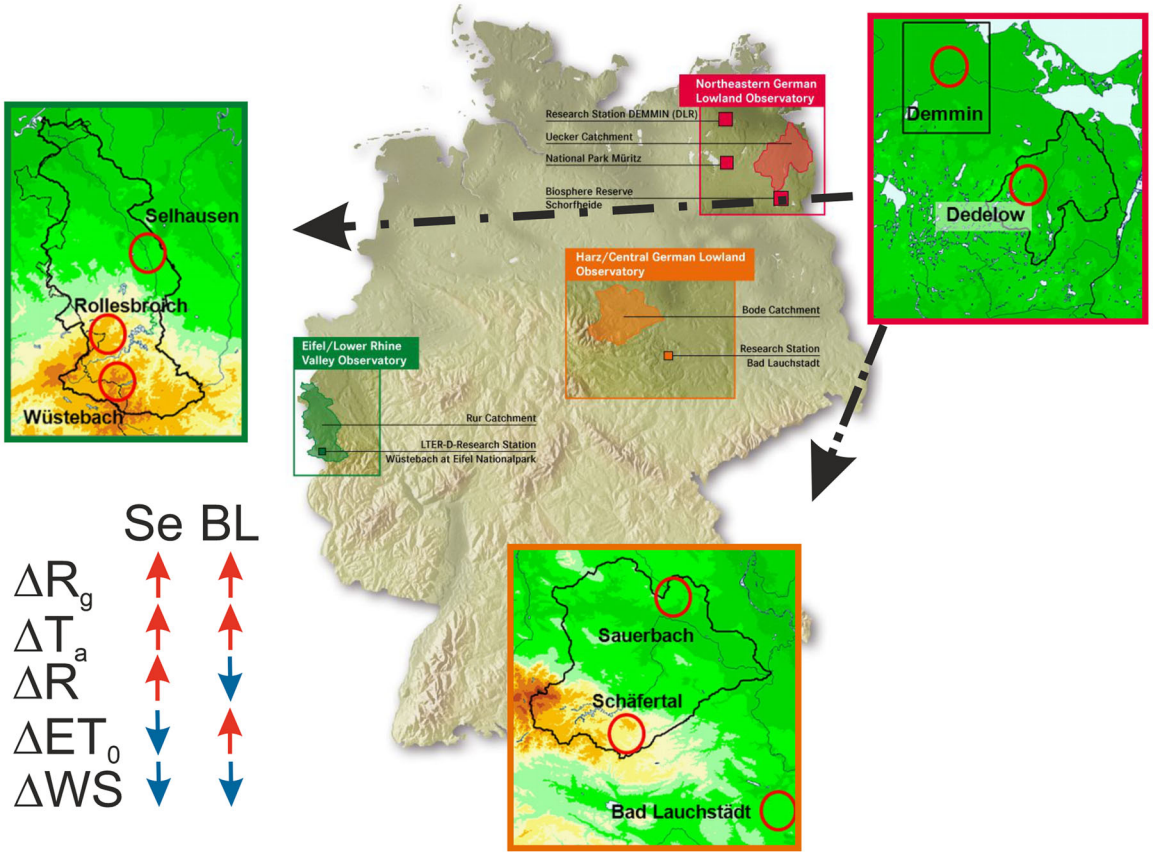

FIGURE 1 Map of selected TERENO (TERrestrial ENvironmental Observatories) SOILCan sites and associated hydrological watersheds in the northern part of Germany (adapted from Pütz et al., 2016) indicating the transfer routes of the soil monoliths from Dedelow (Dd) to Bad Lauchstädt (BL) and Selhausen (Se). The table in the left corner shows the differences (denoted as Δ) in climatic conditions for the hydrological years November 2014–October 2018 related to those at the Dedelow site as arrows (up = larger, down = smaller) for Δ annual global radiation (R_g), Δ mean annual rainfall (R), Δ mean annual grass reference evapotranspiration (ET_0), Δ mean air temperature (T_a), and Δ mean daily wind speed (WS)

the sites and the four cropping seasons. The Dd site was characterized by relatively large variability in annual R for the hydrological years 2016–2017 (858 mm yr^{-1}) and 2017–2018 (443 mm yr^{-1}) as compared with 2014/15–2015/16 (531 and 532 mm yr^{-1}). Regarding crop productivity (Supplemental Table S2), the highest grain yield (GY) for all crops was observed for the Dd lysimeters at the original Dd site; in terms of grain equivalent units, it outperformed the overall crop productivity of the Se lysimeters by almost 100%.

2.2 | Model comparison strategy

Eleven models participated in the study to predict environmental and agronomic variables; AgroC (AC) (Klosterhalfen et al., 2017), DailyDayCent (DC) (Del Grosso et al., 2001), Daisy (DY) (Hansen et al., 2012), HERMES (HE) (Kersebaum, 2007), MONICA (MO) (Nendel et al., 2011), THESEUS (TH) (Wegehenkel et al., 2019), Expert-N (Priesack et al., 2006) coupled to CERES (CE), GECROS (GE), SPASS (SP), and SUCROS (SU), and one hydrological model HydroGeoSphere (HG) (AQUANTY, 2013). A multi-model mean (MM) was calculated considering the corresponding simulation outputs of the 10 crop models, excluding HG, which does not simulate crop growth.

Daily values of actual evapotranspiration (ET_a), R, and net drainage water flux at the bottom boundary of the lysimeter at 1.5-m depth (NetQ) were determined from the high precision weighing lysimeters directly (c.f., Groh, Stumpp, et al., 2018), which have a surface area of 1 m^2. Volumetric soil water content was measured with time domain reflectometry (CS610, Campbell Scientific) and soil water pressure heads with tensiometers (TS1, METER), with pairs of both types of sensors installed at 0.1-, 0.3-, and 0.5-m depths, and a TS1 sensor at 1.4-m depth. Observations on volumetric soil water content were aggregated to a mean soil water content (SWC), which represents the soil's main root zone (from 0- to 0.6-m soil depth).

The crop management and development data for each site included sowing and harvest dates, crop type, information on

FIGURE 2 Photo of (a) the Dedelow site before, and (b) after lysimeter extraction July 2010. (c) Typical soil profile with Ap-, Bt-, and C-horizons of the Luvisol. In the back of Photos a and b, the instrument for lysimeter extraction can be seen. In Photo a, an air-suspended truck is shown, which was used to transport the excavated soil monoliths to the experimental research field station in Bad Lauchstädt and Selhausen. Photo by Wilfried Hierold, Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg

the previous crop, seed density, emergence, tillage operations, crop residual management (Supplemental Table S3), the crop development in BBCH stages (Biologische Bundesanstalt, Bundessortenamt and Chemical industry), and applications of fertilizers (Supplemental Table S4), plant growth regulators, and pesticides (Supplemental Tables S5 and S6) were provided to set up the simulation. Daily values of T_a (minimum, maximum, and mean), WS, relative humidity, barometric pressure, and R_g were also provided for the simulation. ET_0 was calculated for hourly time steps according to the FAO56 Penman–Monteith method (Allen et al., 2006) and summed up to obtain daily values. We have to note that R input provide from lysimeter observations also contained inputs from nonrainfall events (i.e., dew and hoar frost; Groh, Slawitsch, et al., 2018). Pressure heads (hPa) measured in each lysimeter by tensiometers at 1.4-m soil depth or the depth to water table (m; converted from pressure head observations) were used to define the lower boundary of the corresponding soil profile. Soil physical and chemical characteristics were obtained from laboratory data derived from intact core samples and disturbed bulk soil samples (Supplemental Tables S7 and S8). These data were used to parameterize the soil hydraulic properties and chemical characteristics (e.g., total N content, C/N ratio) of the soil profiles and were provided to set up the simulation. More details on the determination on those soil characteristics and initial conditions for the simulation can be found in Groh, Diamantopoulos, et al. (2020).

The calibration of the crop model parameters for each crop was based solely on the data from the Dd site. The provided data to calibrate the models included end-of season agronomic variables GY and aboveground biomass (AgBio), and in-season observations of leaf area index (LAI) (for measurement methods for LAI see Groh, Diamantopoulos, et al. [2020]) over four cropping seasons. Additionally, the daily main environmental variables (ET_a, NetQ, and SWC) from lysimeters at Dd, including the three Dd lysimeters and two additional lysimeters of the present study (eroded Luvisols; c.f., Groh, Diamantopoulos, et al., 2020), were provided for the calibration process.

For the predictive simulations of the Dd soil–ecosystem under a D-W and W-W climate, the measured soil physical and chemical characteristics of each individual lysimeter and the atmospheric conditions, as well as lower boundary information for each lysimeter (site), were used. During model calibration, each modeling group identified and adjusted a single set of plant related model parameters for each crop (winter wheat [*Triticum aestivum* L.], winter rye [*Secale cereale* L.], and oat [*Avena sativa* L.]) for the soils at the reference site in Dd. This approach was chosen because the crop parameters in models are plant specific but remain the same for all soil types. We followed recommendations from Wallach, Palosuo, Thorburn, Hochman, Gourdain, et al. (2021) to document the calibration process for each model (see Supplemental Tables S9–S18), including information on the choice of fitting parameters, the estimated parameter values (i.e.,

optimal and range), and the software or method used to obtain the parameter values.

The crop model parameters determined during calibration to describe crop development and soil water fluxes for the different boundary conditions at BL and Se, but similar crop management (see Supplemental Table S3) was used in a forward simulation to validate the predictive ability of the models. Also, for these sites, climate and crop management data as well as the soil hydraulic properties and chemical characteristics were provided. The "Data set Classification software" developed by Kersebaum et al. (2015) was used to quantitatively analyze and classify our dataset for calibration and validation of crop models. The estimated ranking showed a score of >200 points ("Platinum"), which classifies our calibration dataset as highly suitable for parametrizing, calibrating, and improving crop models.

The individual model and MM performance for the calibration and prediction was evaluated for the agronomic related in- and end-season variables (LAI, AgBio, and GY) and the main in-season environmental variables (ET_a, NetQ, and SWC). The normalized root mean square error (nRMSE) was used as the objective function for model calibration and as criterion for model evaluation as

$$\mathrm{nRMSE} = \frac{100}{\mathrm{SD\,(Obs)}} \sqrt{\frac{\sum_{i=1}^{N} \left(\mathrm{Sim}_i - \mathrm{Obs}_i\right)^2}{N}} \quad (1)$$

where SD is the standard deviation of GY, AgBio, LAI, ET_a, NetQ, and SWC; N is the number of observations; and Sim and Obs are the simulated and observed values. Note that we have used SD instead of the mean value to take into account variables such as NetQ, which include negative (drainage) and positive values (upward directed water flux). To evaluate the overall model performance, we calculated the combined nRMSE for each lysimeter $\mathrm{nRMSE_{Total}}$ as

$$\mathrm{nRMSE_{Total}} = \frac{\left(\begin{array}{l}\mathrm{nRMSE_{GY}} + \mathrm{nRMSE_{AgBio}} + \mathrm{nRMSE_{LAI}} \\ + \mathrm{nRMSE_{SWC}} + \mathrm{nRMSE_{ET_a}} + \mathrm{nRMSE_{NetQ}}\end{array}\right)}{M} \quad (2)$$

where M is the number of individual available measured variables (i.e., 6, if all measured variables were considered).

The water use efficiency (WUE) (kg m^{-3}) at the ecosystem level was described as

$$\mathrm{WUE} = \frac{\mathrm{GY}}{\mathrm{ET_a}} \quad (3)$$

and represents the measure of consumed water during the growing season (i.e., from sowing until harvest) of the corresponding crop (Katerji et al., 2008). As a second crucial agricultural index, the harvest index (HI) was calculated (Hay, 1995) as

$$\mathrm{HI} = \frac{\mathrm{GY}}{\mathrm{AgBio}} \quad (4)$$

The HI index was used here to compare how crop models perform under different environmental conditions (e.g. water stress) (Fereres & González-Dugo, 2009) and crop management decisions (e.g., growth regulator) (Hütsch & Schubert, 2018).

3 | RESULTS

3.1 | Model calibration

The mean deviation of model results from data in terms of nRMSE after calibration (Figure 3) indicates characteristic differences between the tested models for the boundary fluxes (ET_a, NetQ), crop development and yield (LAI, GY, AgBio), and soil state variables (SWC). Except for LAI, the model performances (Figure 3) were relatively similar for all models, including the MM. For the crop related variables, the nRMSE values ranged from 15 to 121% for GY, 13 to 72% for AgBio, and 44 to 192% for LAI. The lowest nRMSE values were obtained with the MO model for GY and LAI and with the SP model for AgBio (Figure 3).

For the water balance related variables, the nRMSE values range from 48 to 87% for ET_a, 84 to 234% for NetQ, and 33 to 102% for SWC, with lowest nRMSE values achieved always by the MM (Figure 3). The standard deviation in nRMSE values between all models after calibration was relatively large for NetQ (15%), mainly caused by few models with low predictive performance (HE, MO, DC, and GE), smallest for ET_a (2%), and relatively similar for GY (10%), AgBio (7%), LAI (7%), and SWC (8%). Performance for models describing the soil water balance through a capacitance approach (i.e., in the non-Richards-based models HE, MO, and DC) showed large variability between single soil profiles reflected by higher standard deviations of nRMSE values for NetQ, SWC, and GY (Figure 3). Most models showed a good fit for the variables GY, AgBio, and ET_a (i.e., nRMSE values ≤50%). On the other hand, nRMSE for NetQ and LAI were relatively large (≥100%). The overall agreement between observations and simulation after calibration for the Dd soil at the Dd site was best for the MM (see blue asterisks in Figure 3), based on the multiobjective criterion $\mathrm{nRMSE_{Total}}$.

3.2 | Model prediction ability

For the evaluation of the predictive model performance (Figure 4), the deviations between simulated and measured

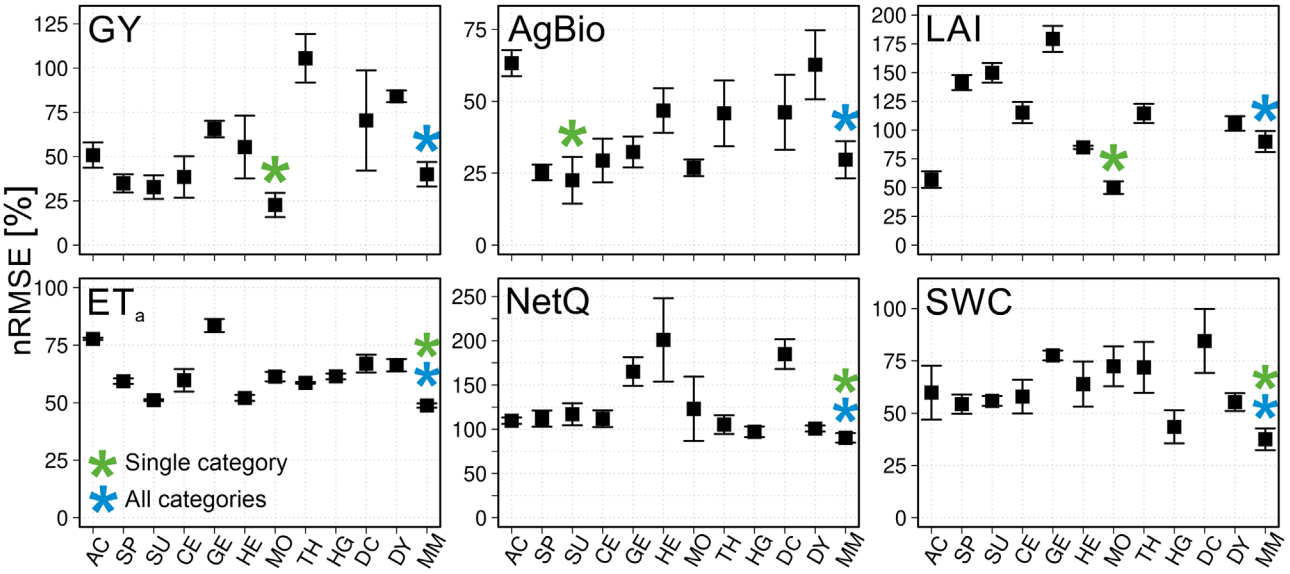

FIGURE 3 Individual model (*x* axis) performance in terms of the mean (symbols) and standard deviation (error bars) of the normalized RMSE (nRMSE, *y* axis) for grain yield (GY), aboveground biomass (AgBio), leaf area index (LAI), actual evapotranspiration (ET_a), net water flux at 1.5-m soil depth (NetQ), and mean soil water content down to 0.6-m depth (SWC) resulting from the calibration step using the data of the five soil lysimeters at Dedelow. For each of the soils, one nRMSE was calculated over all observations of the respective variable. Models include AgroC (AC), Expert-N SPASS (SP), Expert-N SUCROS (SU), Expert-N CERES (CE), Expert-N GECROS (GE), HERMES (HE), MONICA (MO), THESEUS (TH), HydroGeoSphere (HG), DailyDayCent (DC), Daisy (DY), and multi-model mean (MM); the asterisks denote the best model describing the respective crop or soil variable (green) and the model with the best performance over all variables (blue)

data were used to calculate the MnRMSE, which is the variable specific total mean nRMSE averaged over the entire crop rotation and averaged over the three replicate lysimeters at each site. The MnRMSE values from the calibration at the Dd site were compared with MnRMSE values from predictions for BL and Se site (Figure 3). Similar MnRMSE values in calibration and prediction models can also predict crop growth and environmental fluxes under a different climate. In case of stronger deviations, this means that calibrated model parameters are not able to describe the observed processes as well. As expected, the MnRMSE values averaged for each model over the three soil profiles per site (Figure 4) are mostly larger for the validation at BL and Se sites than those obtained at the calibration site Dd.

The MnRMSE values for the prediction of GY, ET_a, and NetQ at BL (Dd soil at W-D climate) were relatively similar for most models except for GE, HE, and DC (Figure 4). For the latter models, the performance even improved for NetQ. The MnRMSE values for the prediction of AgBio and SWC, however, deviated largely from the 1:1 line for most models at BL site (Figure 4). For Se, the MnRMSE values for most predictions of GY were relatively large, whereas AgBio was captured well (as compared with BL) by most crop models except for DC, DY, and SP (Figure 4). Note that results for LAI were not included in the analysis here, because of a lack in LAI observations at Se and BL. For ET_a, the MnRMSE values were relatively similar among all three sites, with slight shifts from the 1:1 line indicating overestimation mainly at Se (Figure 4). For NetQ, most models achieved relatively similar or even better (i.e., DC, GE, HE) quality measures at the validation site. For SWC, the MnRMSE values were relatively large at the validation sites except for models AC, HE, TH, HG, and DY, where the prediction quality was similar to that obtained at the Dd calibration site.

3.3 | Crop growth and yield

For GY (Figure 5) most models described the measured yield relatively well at the calibration site Dd (i.e., the reference: Dd soil at Dd site). Despite water stress during high temperature periods in June and July 2018 (Dr. G. Verch, ZALF Research Station Dedelow, personal communication, 2019), the predictions of oat GY matched the observation with an early ripening of the oat well, except for model CE. The simulated GY at the validation site BL was also mostly in agreement with observations except for the winter wheat predicted by the GE and MO models and for the winter rye GY predicted with AC, HE, TH, DC, and GE models. For oat, predictions largely matched the observations at BL (Figure 5). In contrast, the predicted GY for the Dd soil lysimeters at the validation site Se (Figure 5) was overestimated, in particular for winter wheat in the growing season 2014–2015, by all crop models except for the DC and TH model. Furthermore, models GE and SU

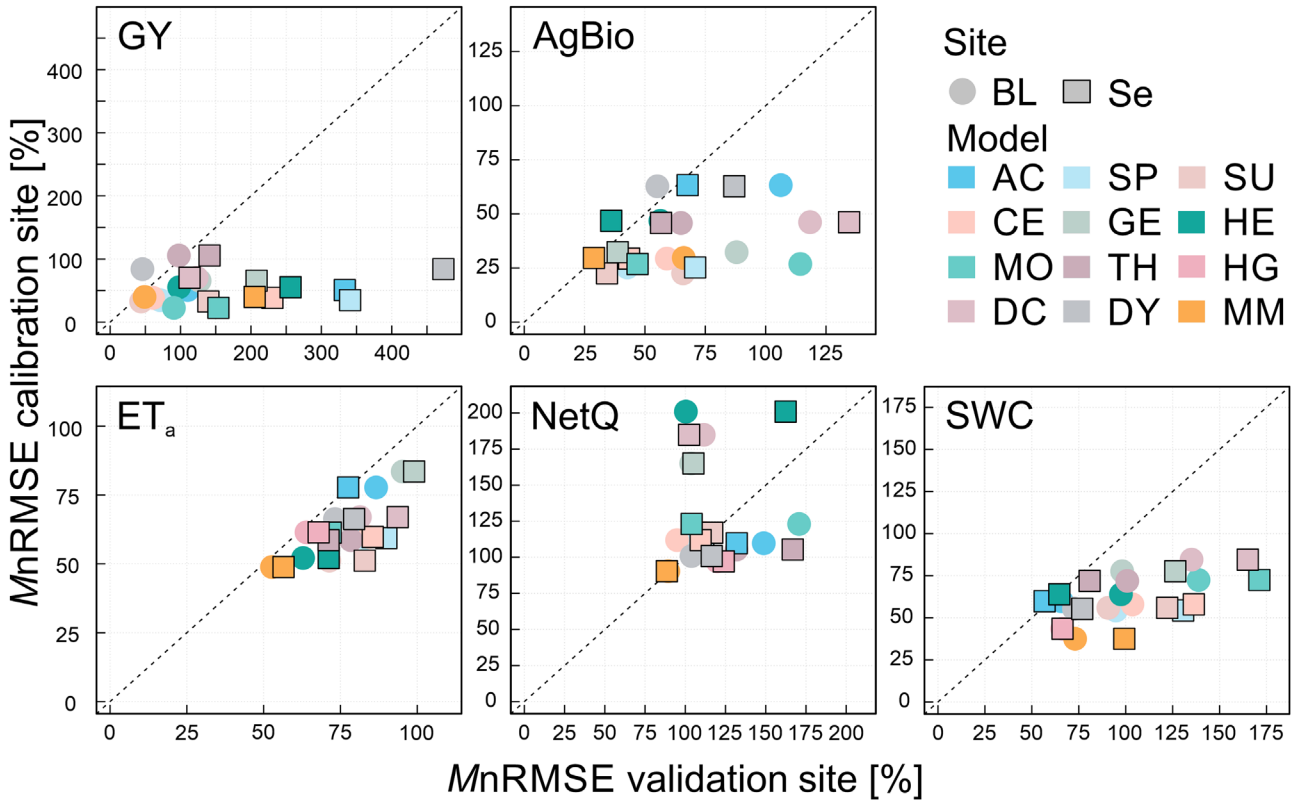

FIGURE 4 Comparison of the model performance in terms of the normalized RMSE (nRMSE) for simulations of lysimeter data at the reference site Dedelow (i.e., model calibration) with those transferred to the locations Bad Lauchstädt (BL) and Selhausen (Se) for grain yield (GY), aboveground biomass (AgBio), leaf area index (LAI; no data at BL and Se), actual evapotranspiration (ET_a), net drainage (NetQ) at 1.5-m soil depth, and average soil water content at 0-to-60-cm depth (SWC). The nRMSE for each site was averaged over the nRMSE values of the three profiles per model and site (MnRMSE). Models include AgroC (AC), Expert-N SPASS (SP), Expert-N SUCROS (SU), Expert-N CERES (CE), Expert-N GECROS (GE), HERMES (HE), MONICA (MO), THESEUS (TH), Theseus-HydroGeoSphere (HG), DailyDayCent (DC), Daisy (DY), and the multi-model mean (MM)

predicted the GY for oat and winter wheat relatively well, but not the GY for winter rye. The mismatch for the MM for GY was relatively small with average MnRMSE values over all crops of 40% at Dd and 49% at BL. However, at Se predictions of GY were clearly overestimated by the MM and achieved a MnRMSE value of 205%.

The AgBio could be described well by most models at the calibration site Dd (Supplemental Figure S1). One exception is perhaps the winter wheat in the first growing season 2014–2015, which was underestimated by AC, DC, and DY models and overestimated by TH. Predictions of AgBio at BL were less accurate than those at Se. Here, AgBio was underestimated for winter rye by all models except for SP and for the winter wheat crop mainly by AC, DC, and MO models. In contrast to GY, the prediction of AgBio at Se were mostly in agreement with the observations (Supplemental Figure S1), with only DC underestimating and AC and DY overestimating AgBio. AgBio predictions deviated remarkably for winter rye by SP model and for winter wheat by TH model (Supplemental Figure S1). The MM achieved lowest MnRMSE values at Se site and the SP model at Dd site and the SU model at BL site (both belonging to the Expert-N model family) performed best in simulating AgBio with regard to other models. However, MnRMSE values for AgBio for MM were higher at the validation sites BL (66%) as compared with the reference site Dd (30%) and validation site Se (29%).

Crop model predictions can be used to calculate agronomic measures such as the HI (Figure 6). For the reference site Dd, HI values were obtained with reasonable accuracy by most models except for DC, DY, GE, and TH (MnRMSE > 150%). Note that the DC model used a fixed GY/biomass ratio for all crops. The predicted HI values were overestimated by most models for the W-D site BL compared with the reference site Dd, except for DC and SU models. Largest deviations of HI at BL were observed (MnRMSE = 472%) by the MO model (Figure 6), overestimating the HI for winter rye and underestimating for winter wheat for one of the soils. At the Se validation site (Figure 6), most models failed to predict HI values. The model performance was generally low and MnRMSE values ranged from 144 to 544% because of the overestimation of GY. The predictions by all models failed to describe

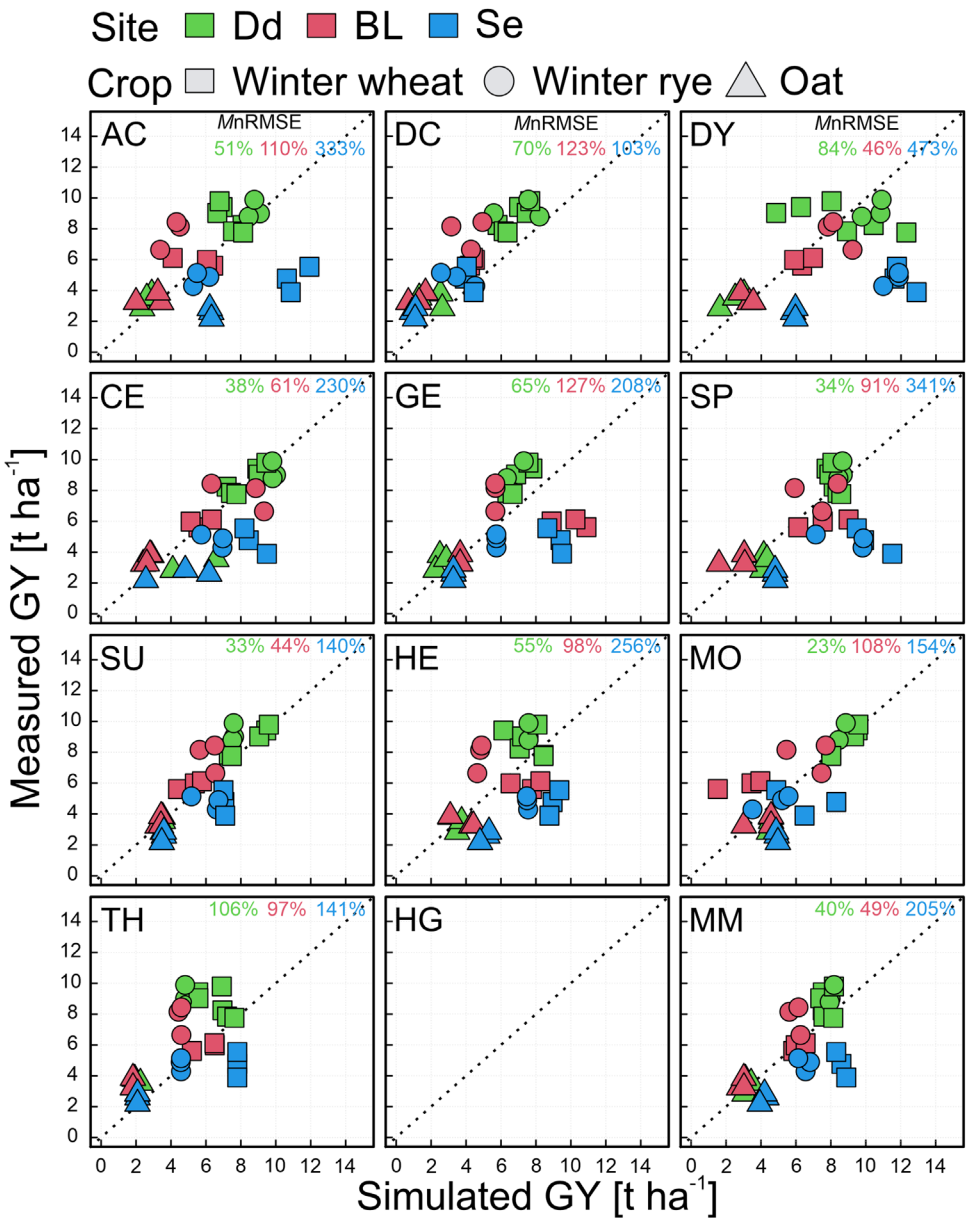

FIGURE 5 Comparison of simulated vs. measured grain yield (GY) for lysimeters at the reference site Dedelow (Dd, green symbols; i.e., model calibration) with those transferred to the locations Bad Lauchstädt (BL, red symbols) and Selhausen (Se, blue symbols) for winter wheat (squares), winter rye (circles), and oat (triangles). Models include AgroC (AC), DailyDayCent (DC), Daisy (DY), Expert-N SPASS (SP), Expert-N SUCROS (SU), Expert-N CERES (CE), Expert-N GECROS (GE), HERMES (HE), MONICA (MO), THESEUS (TH), THESEUS-HydroGeoSphere (HG; no data), and the multi-model mean (MM). MnRMSE denotes the average normalized RMSE over three lysimeters at each site

the decrease in HI values in the order from the Dd reference site (0.5–0.6) to the validation sites BL (0.4–0.5) and Se (<0.4).

3.4 | Crop water use and soil ecosystem fluxes

The majority of the models underestimated ET_a (crop wise) for the growing seasons (Figure 7) at the three sites and achieved relatively similar MnRMSE values (e.g., for MM, 49% at Dd, 53% at BL, and 56% at Se). The ET_a for oat in 2018 was only captured by DY. The models AC, MO, GE, and HG more or less systematically underestimated ET_a during the whole crop rotation, whereas other models (DC, CE, SP, SU, TH, and the MM) for certain crops under and over-predicted measured ET_a (Figure 7).

In contrast with ET_a, the actual bare soil evaporation (E_a) during periods without vegetation (Supplemental Figure S2) is controlled by weather and the soil hydraulic properties,

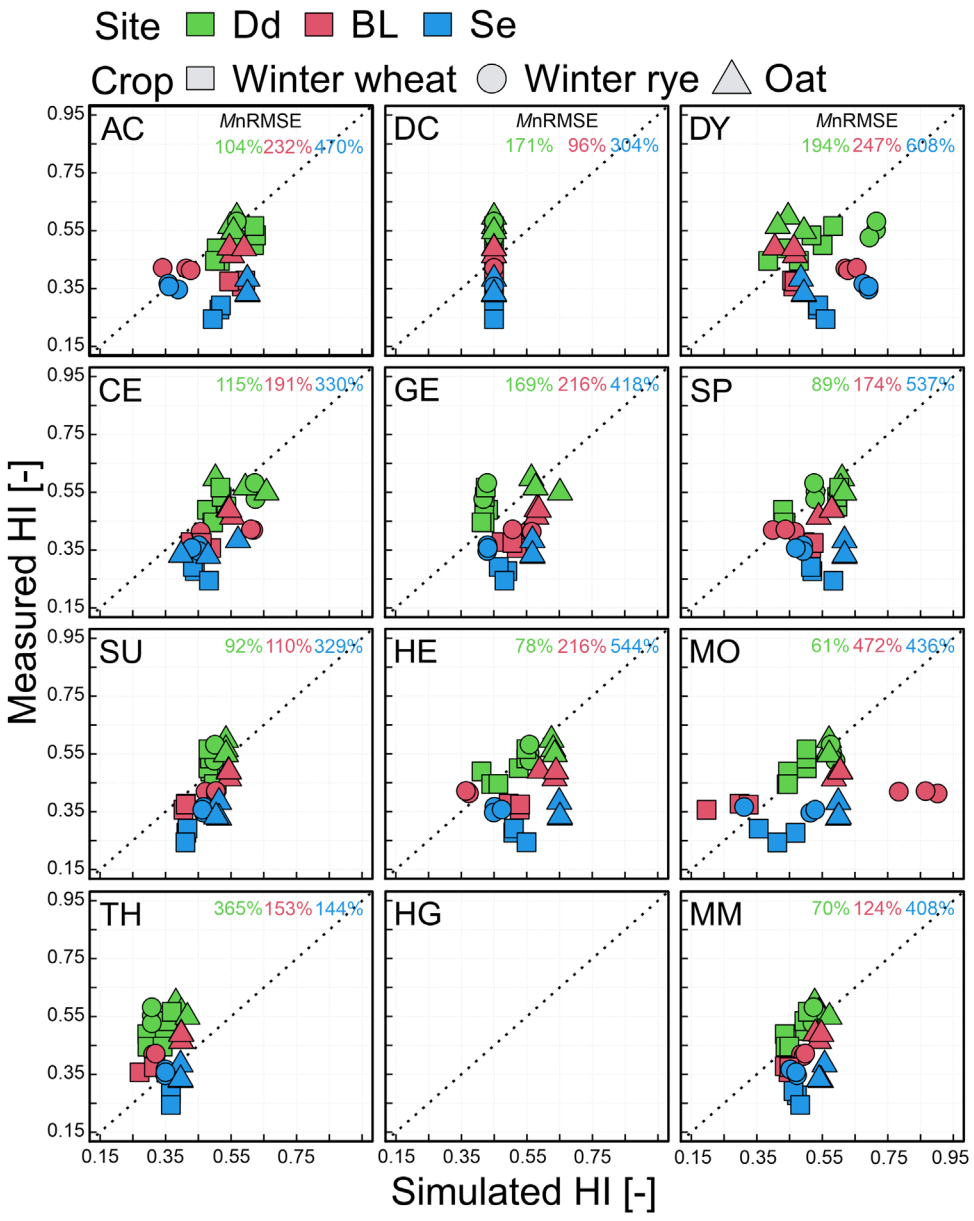

FIGURE 6 Comparison of simulated vs. measured harvest index (HI; i.e., ratio between grain yield and total biomass) for lysimeters at the reference site Dedelow (Dd, green symbols; i.e., model calibration) with those transferred to the locations Bad Lauchstädt (BL, red symbols) and Selhausen (Se, blue symbols) for winter wheat (squares), winter rye (circles), and oat (triangles). Models include AgroC (AC), DailyDayCent (DC), Daisy (DY), Expert-N SPASS (SP), Expert-N SUCROS (SU), Expert-N CERES (CE), Expert-N GECROS (GE), HERMES (HE), MONICA (MO), THESEUS (TH), THESEUS-HydroGeoSphere (HG; no data), and the multi-model mean (MM). MnRMSE denotes the average normalized RMSE over three lysimeters at each site

but not by plant properties. The E_a was here assigned to a specific cropping period by cumulating daily E_a values of the calendric periods before planting and after harvest of the crop. Simulated E_a values agreed with observations fully for the HG model and MM approach, partly for DC, DY, HE, MO, TH, and AC (Supplemental Figure S2). The AC, HE, and TH models underestimated E_a, whereas the Expert-N models (CE, GE, SP, and SU) overestimated E_a, especially before and after oat cropping at both validation sites. These deviations of predicted E_a from observation were largest for relatively long bare soil periods before and after the oat crop (August 2017 until March 2018).

The predicted WUE did not match well for most crop models at the reference site Dd, and model performance values even increase considerably for the validation sites (Figure 8). For all crops, WUE was overestimated by AC and MO and underestimated by the DC model. For oat, CE and SP overestimated WUE, and for winter wheat, the DY model showed relatively large scatter (Figure 8). Predictions for the validation sites underestimated WUE for winter rye (AC,

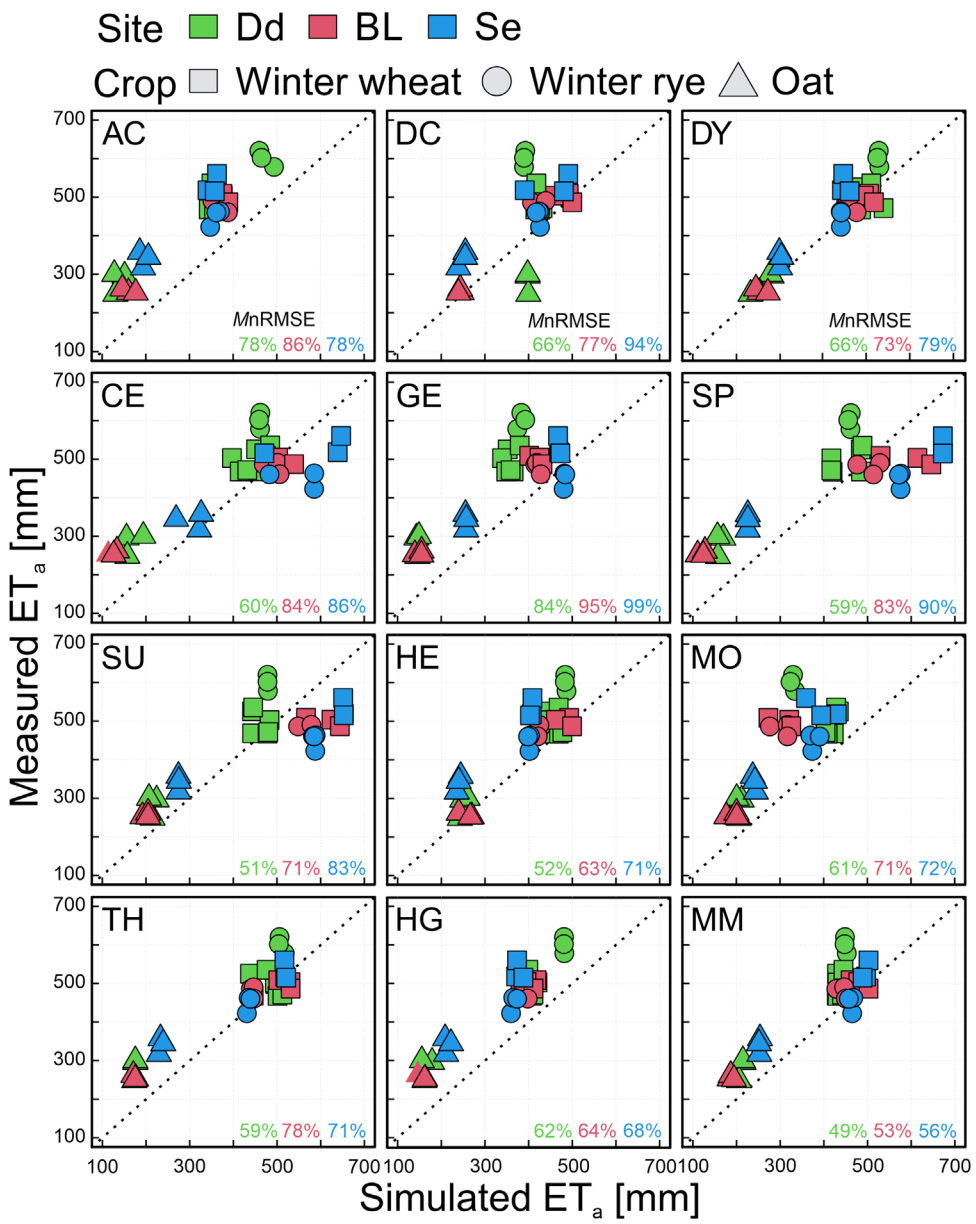

FIGURE 7 Comparison of simulated vs. measured cumulative actual evapotranspiration (i.e., ET_a) for lysimeters at the reference site Dedelow (Dd; green symbols) and lysimeters transferred to Bad Lauchstädt (BL; red symbols) and Selhausen (Se; blue symbols) for winter wheat (squares), winter rye (circles), and oat (triangles). Models include AgroC (AC), DailyDayCent (DC), Daisy (DY), Expert-N SPASS (SP), Expert-N SUCROS (SU), Expert-N CERES (CE), Expert-N GECROS (GE), HERMES (HE), MONICA (MO), THESEUS (TH), THESEUS-HydroGeoSphere (HG), and the multi-model mean (MM). MnRMSE denotes the average normalized RMSE over three lysimeters at each site, calculated based on daily values

DC, CE, SU, and HE), and overestimated WUE for winter wheat (AC, GE), winter rye (MO), and oat (AC CE, GE, and SP). The WUE was overestimated similarly to DC's GY for almost all crops. A similar overestimation of WUE for oat at Dd by CE and SP models was predicted at BL. Except for DC, the crop models mostly overestimated the WUE at Se (Figure 8), because WUE depends on GY, which was overpredicted by most models at this site. The MM was in better agreement with observations, in particular for values at Dd and BL, somehow reflecting a compensation of over- and underestimations of WUE by the different crop models.

The predictions of the cumulative net drainage flux out of the lysimeters (NetQ) agreed with observations at the reference site Dd for AC, DY, CE, SP, SU, MO, TH, and HG (Supplemental Figure S3). For the validation site BL, simulated NetQ values fully agreed with observations for DC, DY, CE, SP, and HE model (Supplemental Figure S3). Note that NetQ simulated by the HE model was always zero because capillary rise was restricted for cases of deeper groundwater levels. For the second validation site Se, with a W-W climate, the same crop models either overestimated (AC, HG, TH, MO) or underestimated (CE, GE, SP, SU) NetQ.

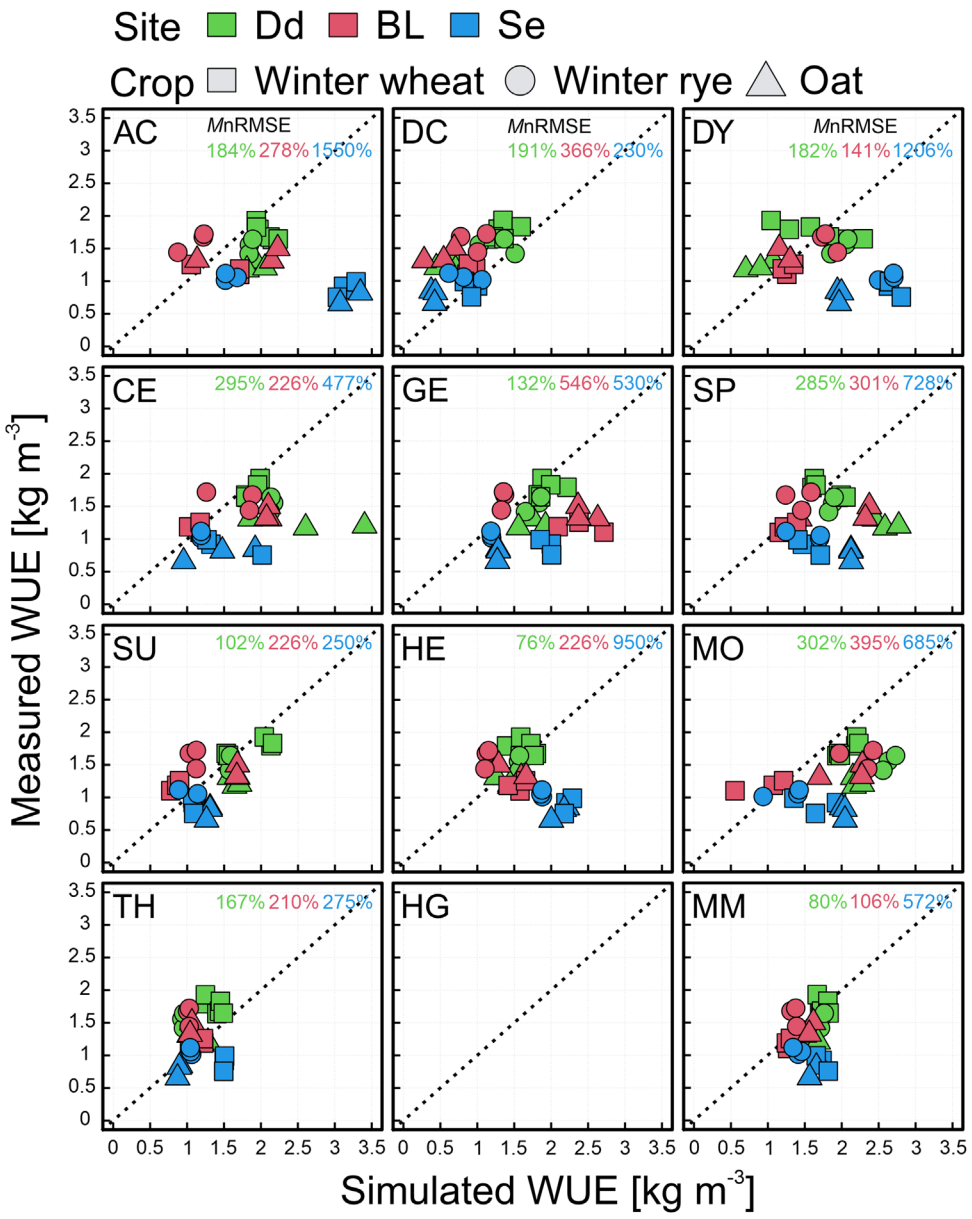

FIGURE 8 Comparison of simulated vs. measured water use efficiency (WUE, defined as grain yield [kg m^{-2}] per volume of crop water use [m^3 m^{-2}] in terms of actual evapotranspiration) for lysimeters at the reference site Dedelow (Dd; green symbols) and lysimeters transferred to Bad Lauchstädt (BL; red symbols) and Selhausen (Se; blue symbols) for the growing seasons of winter wheat (squares), winter rye (circles), and oat (triangles). Models include AgroC (AC), DailyDayCent (DC), Daisy (DY), Expert-N SPASS (SP), Expert-N SUCROS (SU), Expert-N CERES (CE), Expert-N GECROS (GE), HERMES (HE), MONICA (MO), THESEUS (TH), THESEUS-HydroGeoSphere (HG; no data), and the multi-model mean (MM). MnRMSE denotes the average normalized RMSE over three lysimeters at each site

Predictions of NetQ by DY and partly also by TH and HE were close to the 1:1 line. The NetQ predicted by DY model and the MM agreed best for all sites. The MnRMSE values of MM were similar among the reference site Dd (90%) and validation sites Se (89%) and BL (89%).

The mean seasonal SWC in the upper 0.6 m agreed with observations when looking at the values for the reference site Dd (Supplemental Figure S4), with a MnRMSE value across all models of 63%. The match for predicted SWC among the models at the validation site BL was less (MnRMSE of 97%) because DC, CE, SP, SU, and the MO models tended to underestimate the SWCs. Note, that the scatter between the SWC values of the three soil profiles (e.g., during the oat cropping season) at BL was related to a limited availability of SWC data. The tendency of the models to underestimate SWC was even more pronounced at the second validation site Se (MnRMSE value of 109%), indicating that these values were either not well represented in the measurements used for model calibration and that relevant processes describing the SWC dynamics outside the calibration range could not be

parameterized during calibration. Both, the Richards- (AC, DY, TH) and the non-Richards-based (HE) soil water models seem to comparably describe the SWC dynamics at the Se site.

3.5 | Combined assessment of ecosystem productivity and fluxes

The overall simulation performance of the models was separated in a crop-related part that consisted of the arithmetic mean nRMSE values describing productivity (GY, AgBio, LAI) and in an environmental part describing soil water fluxes and soil water content (ET_a, NetQ, SWC). The model performance for the agronomic variables at reference site Dd (nRMSE range: 13–121%) and the site BL (nRMSE range: 12–140%) was for most models in a relatively similar range of nRMSE values (Figure 9). For the validation site Se, nRMSE values were larger (range nRMSE: 16–530%), especially for individual crop models and the MM (Figure 9). Here, individual models such as MO for reference site Dd, DY and SU for the validation sites BL and Se, outperformed the MM for the agronomic variables. Again, the agronomic nRMSE was only calculated for GY and AgBio for the validation sites, as no LAI observations were available. The variables ET_a, NetQ, and SWC had a similar model performance in terms of the mean environmental nRMSE at the reference site as for the validation sites. For the environmental nRMSE, the MM outperformed the results of the individual models at both the reference and the validation sites. However, considering both agronomic and environmental nRMSE values, the MM performed better than any single model, except for the validation site Se. Here, the model SU showed the overall best performance and outperformed the MM, as the simulation results of other models for the agronomic part were partially very poor (nRMSE > 150%).

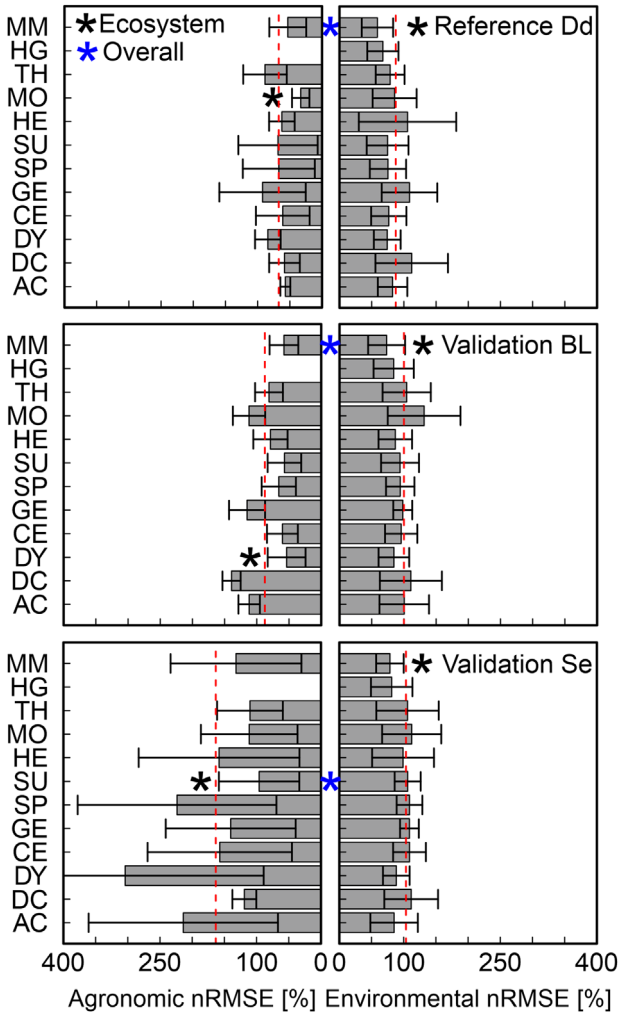

FIGURE 9 Comparison of model performance in terms of normalized RMSE (nRMSE) for crop-related variables (bars oriented to the left) and for the ecosystem-related variables (bars oriented to the right) resulting from the 4-yr simulations (Nov. 2014–Oct. 2018) for the reference site Dedelow (Dd; top) and soils transferred to Bad Lauchstädt (BL; mid) and Selhausen (Se, bottom) for models AgroC (AC), DailyDayCent (DC), Daisy (DY), Expert-N SPASS (SP), Expert-N SUCROS (SU), Expert-N CERES (CE), Expert-N GECROS (GE), HERMES (HE), MONICA (MO), THESEUS (TH), THESEUS-HydroGeoSphere (HG), and the multi-model mean (MM). Error bars indicate the variability between individual lysimeters and the red dashed line the mean nRMSE over all models, excluding HG and MM

4 | DISCUSSION

4.1 | Model calibration

The use of a larger and more encompassing dataset containing in- and end-season agronomic variables as well as environmental data for model calibration improved the performance of crop models in general for the same sites and years (Figure 9). It seems that the calibration of phenology, which controls important variables like GY, AgBio, and LAI, does not contain enough information to predict all agronomic and environmental states and fluxes well, when comparing calibration to phenology only, as discussed by Groh, Diamantopoulos, et al. (2020). This was already shown for GY of wheat (Asseng et al., 2013), but not for a more comprehensive set of environmental variables. Here, it has to be noted that it was not intended in our study to compare the different embedded functions describing the effect of the environment on plant production and environmental fluxes, rather to compare amongst different models. A comprehensive comparison and evaluation of the different model embedded functions and ways to calibrate them harmonically among different crop models is still ongoing research and would be out of scope here. The improvement by calibration

based on all available data was particularly visible for the end-season-related agronomic variables GY (Figure 4) and AgBio (Supplemental Figure S1). This can be explained by the fact that mainly crop-related parameters were calibrated within this study. The GY of the winter wheat between both cropping periods differed substantially (Supplemental Table S2), which was captured only by some of the crop models (i.e., CE, SU, and MO). A recent study using two long-term experiments (1969–2018) suggested that the choice of cultivar can, apart from other management decisions (e.g., crop growth regulator) (Hütsch & Schubert, 2018) or weather conditions, substantially influence GY of the crop winter wheat (Aula et al., 2020). As only a single set of plant parameters was estimated for the two different wheat cultivars grown in the two cropping periods, our study confirmed that the choice of cultivar as well as the use of crop growth regulators is important to reproduce GY of both winter wheat cultivars. The predicted GY of the oat for the reference site Dd in 2018 agreed well with observations (Figure 4), which showed that the model parameterization obtained by calibration was able to cope with the effect of water stress on GY for these soils. This is also confirmed by the improvement of the average nRMSE for oat by 182% as compared with simulations where solely phenological stages were used for calibration (Groh, Diamantopoulos, et al., 2020). However, other important parameters affecting the seasonal LAI development (Figure 3, especially maximum LAI, date of flowering) were not captured well by most of the Expert-N family models (e.g., SP, SU, and GE), resulting in low model performance values for LAI.

The model performance for the environmental states and fluxes improved only slightly using the full dataset for calibration as compared with calibrating on phenology only (c.f., Groh, Diamantopoulos, et al., 2020). Only a significant improvement was achieved for NetQ by the DY and DC model. The variability of the prediction from each individual model, expressed as scatter of the nRMSE, highlights that the non-Richards-based models tend to vary much more for the variables NetQ, SWC, and GY compared with the Richards-based models (Table 2). Note that one can expect that the agronomic variables would be less dispersed, as essential plant parameters were adjusted by calibration. All of the environmental variables were less affected because the hydraulic parameters were fixed and only parameters for the root water uptake (e.g., AC) or the actual crop coefficient factor (K_c) for ET_0 (e.g., AC) were estimated. The larger variability for NetQ and SWC might be related to the different implementation of the given soil hydraulic parameters, as the Richards equation-based models used the soil hydraulic parameters (Mualem–van Genuchten parameters), whereas the non-Richards equation-based models used the water contents at saturation, field capacity and wilting point instead. Additionally, the type of lower boundary conditions affects the prediction of NetQ. This agrees well with previous investigations, which demonstrate that describing variably saturated flow with the Richards equation improves predictions of the SWC dynamics compared with a bucket-type approach (Longo et al., 2021; MacBean et al., 2020; Soldevilla-Martinez et al., 2014).

Most of the models greatly underestimated ET_a, probably due to the way root water uptake, partitioning of ET_0 into transpiration and evaporation, or soil hydraulic properties were parameterized, or due to the way the K_c factors, which scale ET_0 to a crop-type-specific ET_a, were set or constrained. We can contest that the spread among the crop models in K_c values was large. Some models used a constant value around one over the entire period for all crops (e.g., DC, TH, and all Expert-N related models), whereby in other models individual K_c values for each crop was set (AC, DY, MO, and HE,). As mentioned, the K_c factor was calibrated in for the AC model to obtain a better match with ET_a, whereby the fitted mean cropping period K_c factor was 1.09, 1.53, and 1.48 for oat, rye, and wheat, respectively. Hereby, the fitted K_c factors for rye and wheat are at the maximum end or even exceed physically consistent or literature values (c.f., Allen et al., 1998; Pereira et al., 2021).

The mean nRMSE values of the MM were lower than for any other single crop models. The specific averaged nRMSE for the soils at Dd (Figure 9) confirms that the MM better predicts the overall combined assessment of agronomic productivity, as already reported by Martre et al. (2015) and Wallach, Martre, et al. (2018), and for environmental water fluxes and states as reported by Groh, Diamantopoulos, et al. (2020). This was also shown for MM and the variable NetQ, which was previously not included in crop model intercomparison studies due to the lack of data. Many recent studies, which compared the model performance of the MM with single crop models, have found that MM predicts single (Wallach, Palosuo, Thorburn, Gourdain, et al., 2021) or multiple target variables (Bassu et al., 2014; Martre, Wallach, et al., 2015) successfully well. The reason for the superiority of MM lies in the fact that different models have different errors, which are consequently averaged over different models when forming an average (Wallach, Martre, et al., 2018).

One may also consider that the observations used in this study differed in terms of the scale from other crop model intercomparison studies that used frequently data from larger scale field experiments (Wallor et al., 2018). The discrepancy between the lysimeter scale (1 m^2) and the field scale (>100 m^2) may have affected results. However, any such effect would influence all the results of models in the same way and would thus not affect the model comparison. It should be noted here that the recent study by Groh, Stumpp, et al. (2018) confirmed that lysimeters can be representative of a larger area with respect to environmental variables (e.g., NetQ, SWC, and tracers), but similar studies for agronomic

TABLE 2 Average standard deviation of normalized RMSE (nRMSE) values for each variable of the model calibration for all, non-Richards-based (number of models = 3), and Richards-based models (number of models = 8)

Variable	Avg. standard deviation of nRMSE		
	All models	Non-Richards-based models	Richards-based models
	%		
GY	10	18	7
AgBio	7	8	7
LAI	7	3	8
ET_a	2	2	2
NetQ	15	33	9
SWC	8	12	7

Note. GY, grain yield; AgBio, aboveground biomass; LAI, leaf area index; ET_a, actual evapotranspiration; NetQ, net drainage water flux at the bottom boundary of the lysimeter at 1.5-m depth; SWC, soil water content.

variables obtained from high precision weighable lysimeter are still lacking.

4.2 | Modeling agroecosystems

Holistic agroecological modeling requires consideration of both, agronomic aspects (e.g., GY, LAI, AgBio) and environmental water fluxes and states (e.g., ET_a, NetQ, SWC) simultaneously. This investigation is one of the rare studies that include both parts for calibration and validation. A previous study has shown that soil properties, such as rooting zone thickness, varying texture, and soil hydraulic properties need to be adequately parametrized in crop models (Groh, Diamantopoulos, et al., 2020). This highlights that the soil part including vertical soil heterogeneity (horizonation) should be considered when simulating crop ecosystems. As data for the detailed calibration are often missing, the results from this study provide some estimates for the expected prediction error. However, calibrating only plant-related parameters in crop models does not necessarily improve the simulation of soil water flux-related key variables. The existence of compensating effect during parameter calibration in crop models (Sima et al., 2020) leading to nonuniqueness, as defined already for hydrological models in the early 1990s (Beven, 1993) is still a challenging aspect in the calibration approach. The problem that models are often right for the wrong reason (Kirchner, 2006), when only a single criterion is considered in model calibration (Houska et al., 2021), is added here as a challenge to calibration.

When the focus is on increasing food production (Hamidov et al., 2018), soil processes are often compromised although soil functions have important feedbacks on ecosystem productivity, nutrient cycling, energy, and water (Deckmyn et al., 2020; Zwetsloot et al., 2020). This emphasizes the need to better calibrate the soil part in crop models beyond a limited consideration of classical agronomic variables such as only GY, AgBio, LAI, HI, or plant nutrient uptake and instead including soil related observations with high temporal and if feasible spatial (depth) resolution (Kersebaum et al., 2015).

4.3 | Model prediction ability

A comparison of the predictive performance of agronomic variables of Dd soils between the original climate in Dd and a D-W climate in BL showed that under D-W conditions at the BL site (nRMSE = 88%), similar predictive performance was achieved as for calibration site Dd (nRMSE = 67%, see red lines in Figure 9). In contrast, predictions diverged strongly from the observations for the wet and warm region (W-W) at the Se site (nRMSE = 163%). The observed GY decreased substantially under W-W conditions compared with the calibration site at Dd or validation site at BL (D-W climate). Unfortunately, the crop models used in this study, including the MM, were not able to capture the effect of climate and management (e.g., different application of N fertilizer and crop growth regulator) on GY.

The observations showed a higher water use efficiency (WUE) and reproductive output (i.e., HI) of plants growing under conditions with lower plant-available soil water, confirming previous results for several soils, including those from Dd translocated to BL and Se (Groh, Vanderborght, et al., 2020); in this study, the lower WUE at the ecosystem level (see Equation 3) was hereby mainly driven by higher evaporation under W-W climatic conditions, as differences in ET_a were large during times where evaporation rather than transpiration is dominating ET_a. This might explain the partially large underestimation of ET_a or E_a by some of the crop models, because most crop models were not able to simulate E_a well. The E_a occurs in two different stages. In Stage 1, E_a is limited by the atmospheric evaporative demand. With further drying of the soil, when the surface soil water content is depleted, a change in the E_a rate occurs and this rate

drops below the potential evaporation rate (Stage 2 evaporation). Here, specific model parameters, weather conditions, intercept evaporation, and evaporation during wet soil conditions (Stage 1) and soil hydraulic parameters during dry soil conditions (Stage 2) are crucial. Thus, a better agreement of evaporation during bare soil conditions and vegetation period can be expected if in situ data are used to determine the soil hydraulic properties (e.g., Schneider et al., 2021). On the other hand, the ET_a underestimation might be in addition partially related to oxygen stress under wetter climate conditions (i.e., W-W with high SWC) by water logging, as for the HE model transpiration is reduced when the air filled pores are temporarily below 8% according to Supit, Hoojer, and Diepen (1994). Other factors such as differences between stomata conductivity, resistance in stem, and other rhizosphere processes might explain the underestimation of ET_a.

In any case, the question remains why crop models were not able to capture the reduction in GY under W-W climate. The simulation error for GY is likely caused by the soil status (SWC and nutrients) and altered thresholds under the new climatic regime. The climatic conditions at the site where the lysimeter have been transferred to were outside the range of historical observation and thus not accounted for in the calibration. This discrepancy between environmental conditions for calibration and validation can affect the functional relationships (e.g., *R* and GY) of the ecosystem as shown by Knapp, Carroll, et al. (2018) for grassland ecosystems. One further possible improvement would be a more comparable calibration of the crop models (e.g., choice of parameters, software to optimize parameter, etc.), but this was beyond the scope of this study. Several crop modeling studies have already used different ranges of environmental conditions for model calibration and evaluation (Asseng et al., 2015, 2019; Biernath et al., 2011; Hussain et al., 2018) to test models' ability to predict target variables outside the range of historical observations. Wallach, Palosuo, Thorburn, Gourdain, et al. (2021) referred to these types of studies as "extrapolation" studies, which is a common procedure in modeling crop growth under, for example, future climate conditions (Seidel et al., 2018). Thus, changing environmental conditions and differences in stress situations such as increasing temperature and heat stress (Asseng et al., 2015; Pszczółkowska et al., 2020; Rezaei et al., 2018; Webber, White, et al., 2018), temporal water limitations (Sadras, Villalobos, et al., 2016), or excess (e.g. water-logging Mäkinen, Kaseva, et al., 2018) during a specific crop development stages might have affected the crop development and consequently reduce GY of the specific crop under W-W climatic conditions. This explains why the calibrated models in Dd work well under similar conditions in BL, where the thresholds and their effects on the results were calibrated indirectly. In addition to the transpiration effect caused by higher air temperature, heat stress can serve as a possible explanation for lower GY, when a critical temperature is exceeded (Eyshi Rezaei et al., 2015). Such critical temperatures can be in particular harmful during anthesis, because the anthesis period will be shortened, which will lead to earlier flowering (L.-X. Tao et al., 2008). Additionally, higher temperatures near anthesis will reduce the grain number, because it affects pollen fertility, increases grain abortion, and reduces grains per spike due to accelerated crop development (Eyshi Rezaei et al., 2015).

Observations on GY and air temperature confirm for soils, which were transferred from Dd to BL and Se that the effect of heat stress days during flowering increased (Figure 10a). The average number of days with heat stress increase from 5 d at Dd to over 10 d at BL and Se. The average duration of heat stress during this period was in general longer in Se (2.4 d) than in BL (1.8 d). Hence, additional model parameters, or even model processes accounting for heat stress are needed. Secondly, plant parameters might not have been correctly estimated during calibration in the environment without or with less heat stress to be used to predict GY under W-W, because limited data range might lead to a miscalibration of the response curve or parameter sensitivity changes under different climatic conditions or climate change (Melsen & Guse, 2021). The GY simulation results support this assumption, as for the site BL, which has a similar climate to Dd, the influence of heat stress (days with max air temperature > 27°C [Semenov & Shewry, 2011] for a period of 32 d after flowering) on the yield could be reproduced well but was significantly underestimated under a W-W climate at the Se site (see Figure 10a). Gilardelli et al. (2018) showed for the WOFOST (WOrld FOod STudies) model, that under unfavorable climate conditions for crop growth, predictions of GY reacts sensitively to certain crop parameters, such as those related to respiration ratio or specific leaf area. Observed GY and number of heat stress days at flowering confirms a significant decrease in GY across all sites (Figure 10a). In addition, the reduction of GY showed a high correlation ($r = .87$; $p < .001$; Figure 10b) with decreasing number of ears per m^2. The observed 1,000-kernel weight remained similar at all sites, implying that the decrease in GY can be associated with the number of grains. This suggest that the higher temperatures and heat stress around anthesis negatively affected GY especially at Se, due to decreasing pollen fertility or abortion and sterile grains (Chaturvedi et al., 2021; Farooq et al., 2012; Shenoda et al., 2021).

The challenges discussed above indicate that the parameters fitted outside the prediction environment or processes not accounted for in the calibration environment might play a crucial role for agricultural projections, especially in the context of climate change and on the broader feedback that soils have on climate. F. Tao et al. (2018) already summarized that model structure and model parameters are an important source of uncertainty when predicting effects of climate change on future ecosystem services, which implies long-term

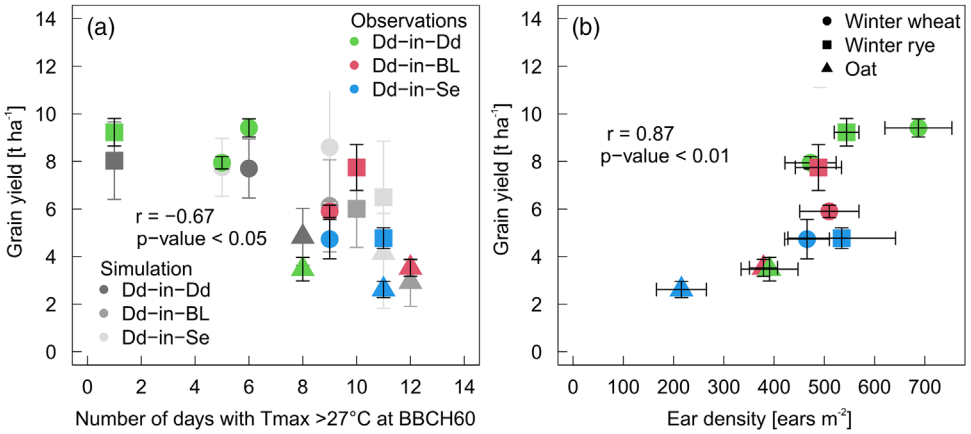

FIGURE 10 Relationship between (a) mean grain yield (observed and simulated) and the number of days with heat stress at each site for the period of 32 d after flowering. A threshold of 27°C maximum daily air temperature was used to define days with heat stress (Semenov & Shewry, 2011). (b) Relationship between the observed grain yield and ear density. Mean values are indicated by symbols and standard deviation by error bars. Sites include Dedelow (Dd), Bad Lauchstädt (BL), and Selhausen (Se). Tmax is maximum temperature

projections with altered climatic conditions outside the historical range where the models have been calibrated on. In consequence and consistent to our hypothesis, crop models are able to predict the site-specific temporal and spatial variables when calibrated on longer historical time series only when those "new" climatic conditions are within the range of the historically observed climate.

4.4 | Modeling crop–soil–ecosystem response to changing boundary conditions

The study challenged the problem of improving predictions of agroecosystem productivity and environmental fluxes and states under changing climatic conditions for the most likely condition that changes of environmental conditions are outside the historical range of observations. This includes not only the absolute values but also the temporal distribution of rainfall events or temperature dynamics during the season. Such information seems essential for both the calibration and validation steps of models, but also for predicting ecosystem functioning under changing global climatic conditions. In addition to the space-for-time substitution approach, one possibility to deal with a limited site-specific range of observations for model calibration could be to use data from manipulative experiments to broaden the training dataset for model calibration. For example, Forstner et al. (2021) showed that only the combined use of observational and manipulative climate change experiments was able to identify how altering individual and multiple drivers (i.e., reduced R, elevated CO_2, and surface temperatures, increase in ET_0) affect ecosystem functioning of grassland ecosystems on changing climate conditions. Still, many challenges remain in addition to those discussed above, such as consideration of changes in soil properties due to climatic changes and the effects of these changes on model parameters or hints on missing model structures, depending on changing climatic conditions (e.g., Robinson et al., 2016, 2019).

5 | CONCLUSIONS

The present study compared the performance of commonly used crop models for both agronomic crop-related and soil ecosystem-related datasets from weighing lysimeters with intact soil monoliths in situ with those translocated to a warmer and drier and a warmer and wetter region. The aim was to predict the crop development and ecosystem fluxes for lysimeters with the same soil that were transferred in regions with different climatic conditions after calibration at the original site. As expected, the degree of model complexity (i.e., non-Richard- vs. Richards-based models) in describing vadose zone processes largely explained variations in the model performance between individual crop models. The relatively similar crop model predictions at the validation site BL indicated that the climatic conditions of the calibration site Dd were in a similar range. The larger deviations of the model predictions compared with lysimeter data at the Se site suggested that either the effects of climatic conditions outside of the range of Dd site-specific variability could not be described by the calibrated crop models or that the underlying process representations have not been included in the models. The observed decrease in grain yield, not captured by most models, could possibly be related to an increase in heat stress.

Thus, the presently tested crop models were not able to predict the effect of such effects of changes in climatic conditions represented in the space-for-time substitution approach. The results confirm the hypothesis that crop models are unable to

predict site-specific temporal and spatial variability of agronomic crop and soil ecosystem variables and fluxes when climatic conditions are outside the range of the data time series used for calibration. The implications of this finding are important for future applications of crop models with respect to predictions of agronomic perspectives and interrelations with ecosystem response to changing climatic conditions. The study clearly demonstrated that a key for improving the predictive capability of crop models in the soil–vegetation–atmosphere continuum is the quantification of soil related data for calibration and model testing. This would improve predictions of above- and belowground ecosystem processes in crop models and enhance future risk assessment related to ecosystem functions. The results suggest the need to develop improved concepts for model calibration procedures that incorporate experimental data sets from scenarios with assumed climatic conditions.

ACKNOWLEDGMENTS

We acknowledge the support of TERENO and SoilCan, which were funded by the Helmholtz Association (HGF) and the Federal Ministry of Education and Research (BMBF). Kurt-Christian Kersebaum acknowledges funding by the Ministry of Education, Youth and Sports of Czech Republic through SustES—Adaption strategies for sustainable ecosystem services and food security under adverse environmental conditions (CZ.02.1.01/0.0/0.0/16_019/000797). Evelyn Wallor was funded by the German Federal Ministry of Education and Research (BMBF) through the BonaRes project "I4S" (031B0513I). Jannis Groh acknowledges the German Research Foundation (DFG; Project no. 460817082) for funding. We thank Jörg Haase, Dr. Gernot Verch, Ingrid Onasch, Gudrun Buddrus, Ralf Gründling, Sylvia Schmögner, Werner Küpper, Ferdinand Engels, Rainer Harms, Philipp Meulendick, and Leander Fürst for data collection and maintenance of the experimental setup at the corresponding research station.

AUTHOR CONTRIBUTIONS

Jannis Groh: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Visualization; Writing-original draft; Writing-review & editing. Efstathios Diamantopoulos: Data curation; Formal analysis; Resources; Writing-review & editing. Xiaohong Duan: Data curation; Formal analysis; Resources; Writing-review & editing. Frank Ewert: Writing-review & editing. Florian Andreas Heinlein: Data curation; Formal analysis; Resources; Writing-review & editing. Michael Herbst: Data curation; Formal analysis; Resources; Writing-review & editing. Maja Holbak: Data curation; Formal analysis; Resources; Writing-review & editing. Bahareh Kamali: Data curation; Formal analysis; Resources; Writing-review & editing. Kurt-Christian Kersebaum: Data curation; Formal analysis; Resources; Writing-review & editing. Matthias Kuhnert: Data curation; Formal analysis; Resources; Writing-review & editing. Claas Nendel: Data curation; Formal analysis; Resources; Writing-review & editing. Eckart Priesack: Data curation; Formal analysis; Resources; Writing-review & editing. Jörg Steidl: Data curation; Formal analysis; Resources; Writing-review & editing. Michael Sommer: Writing-review & editing. Thomas Pütz: Writing-review & editing. Jan Vanderborght: Writing-review & editing. Harry Vereecken: Writing-review & editing. Evelyn Wallor: Data curation; Formal analysis; Resources; Writing-review & editing. Tobias Karl David Weber: Data curation; Formal analysis; Resources; Writing-review & editing. Martin Wegehenkel: Data curation; Resources; Writing-review & editing. Lutz Weihermüller: Data curation; Formal analysis; Resources; Writing-review & editing. Gerke Horst: Conceptualization; Data curation; Formal analysis; Funding acquisition; Writing-review & editing.

REFERENCES

Abd-Elmabod, S. K., Muñoz-Rojas, M., Jordán, A., Anaya-Romero, M., Phillips, J. D., Jones, L., Zhang, Z., Pereira, P., Fleskens, L., Van Der Ploeg, M., & De La Rosa, D. (2020). Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. *Geoderma*, *374*, 114453. https://doi.org/10.1016/j.geoderma.2020.114453

Adam, M., Boote, K. J., Falconnier, G. N., Porter, C., Rezaei, E. E., & Webber, H. (2020). Modeling the effects of climate change on agriculture: A focus on cropping systems. In D. Deryng (Ed.), *Climate change and agriculture* (1st ed.). Burleigh Dodds Science Publishing.

Ainsworth, E. A., & Rogers, A. (2007). The response of photosynthesis and stomatal conductance to rising [CO2]: Mechanisms and environmental interactions. *Plant, Cell & Environment*, *30*, 258–270. https://doi.org/10.1111/j.1365-3040.2007.01641.x

Alfieri, J. G., Kustas, W. P., Prueger, J. H., Hipps, L. E., Evett, S. R., Basara, J. B., Neale, C. M. U., French, A. N., Colaizzi, P., Agam, N., Cosh, M. H., Chavez, J. L., & Howell, T. A. (2012). On the discrepancy between eddy covariance and lysimetry-based surface flux measurements under strongly advective conditions. *Advances in Water Resources*, *50*, 62–78. https://doi.org/10.1016/j.advwatres.2012.07.008

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration: Guidelines for computing crop water requirements*. FAO.

Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J. B., Smith, M., Pereira, L. S., Raes, D., Perrier, A., Alves, I., Walter, I., & Elliott, R. (2006). A recommendation on standardized surface resistance for hourly calculation of reference ET_o by the FAO56 Penman-Monteith method. *Agricultural Water Management*, *81*, 1–22. https://doi.org/10.1016/j.agwat.2005.03.007

AQUANTY. (2013). *HGS 2013: HydroGeoShere user manual*. Aquanty.

Archontoulis, S. V., Castellano, M. J., Licht, M. A., Nichols, V., Baum, M., Huber, I., Martinez-Feria, R., Puntel, L., Ordóñez, R. A., Iqbal, J., Wright, E. E., Dietzel, R. N., Helmers, M., Vanloocke, A., Liebman, M., Hatfield, J. L., Herzmann, D., Córdova, S. C., Edmonds, P., … Lamkey, K. R. (2020). Predicting crop yields and soil-plant nitrogen

dynamics in the US Corn Belt. *Crop Science*, *60*, 721–738. https://doi.org/10.1002/csc2.20039

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., … Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, *5*, 143–147. https://doi.org/10.1038/nclimate2470

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P. J., Rötter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P. K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A. J., Doltra, J., … Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, *3*, 827. https://doi.org/10.1038/nclimate1916

Asseng, S., Martre, P., Maiorano, A., Rötter, R. P., O'leary, G. J., Fitzgerald, G. J., Girousse, C., Motzo, R., Giunta, F., Babar, M. A., Reynolds, M. P., Kheir, A. M. S., Thorburn, P. J., Waha, K., Ruane, A. C., Aggarwal, P. K., Ahmed, M., Balkovič, J., Basso, B., … Ewert, F. (2019). Climate change impact and adaptation for wheat protein. *Global Change Biology*, *25*, 155–173. https://doi.org/10.1111/gcb.14481

Aula, L., Omara, P., Eickhoff, E., Oyebiyi, F., Dhillon, J. S., & Raun, W. R. (2020). Effect of winter wheat cultivar on grain yield trend under different nitrogen management. *Agrosystems, Geosciences & Environment*, *3*, e20017. https://doi.org/10.1002/agg2.20017

Bassu, S., Brisson, N., Durand, J.-L., Boote, K., Lizaso, J., Jones, J. W., Rosenzweig, C., Ruane, A. C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., … Waha, K. (2014). How do various maize crop models vary in their responses to climate change factors? *Global Change Biology*, *20*, 2301–2320. https://doi.org/10.1111/gcb.12520

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*, 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Biernath, C., Gayler, S., Bittner, S., Klein, C., Högy, P., Fangmeier, A., & Priesack, E. (2011). Evaluating the ability of four crop models to predict different environmental impacts on spring wheat grown in open-top chambers. *European Journal of Agronomy*, *35*, 71–82. https://doi.org/10.1016/j.eja.2011.04.001

Bogena, H. R., Montzka, C., Huisman, J. A., Graf, A., Schmidt, M., Stockinger, M., Von Hebel, C., Hendricks-Franssen, H. J., Van Der Kruk, J., Tappe, W., Lücke, A., Baatz, R., Bol, R., Groh, J., Pütz, T., Jakobi, J., Kunkel, R., Sorg, J., & Vereecken, H. (2018). The TERENO-Rur hydrological observatory: A multiscale multi-compartment research platform for the advancement of hydrological science. *Vadose Zone Journal*, *17*, 1–22. https://doi.org/10.2136/vzj2018.03.0055

Boote, K. J., Allen, L. H. J., Prasad, P. V. V., & Jones, J. W. (2010). Testing effects of climate change in crop models. In D. Hillel & C. Rosenzweig, (Eds.), *Handbook of climate change and agroecosystems: Impacts, adaptation, and mitigation* (pp. 109–129). Imperial College Press.

Challinor, A. J., Müller, C., Asseng, S., Deva, C., Nicklin, K. J., Wallach, D., Vanuytrecht, E., Whitfield, S., Ramirez-Villegas, J., & Koehler, A.-K. (2018). Improving the use of crop models for risk assessment and climate change adaptation. *Agricultural Systems*, *159*, 296–306. https://doi.org/10.1016/j.agsy.2017.07.010

Chaturvedi, P., Wiese, A. J., Ghatak, A., Záveská Drábková, L., Weckwerth, W., & Honys, D. (2021). Heat stress response mechanisms in pollen development. *New Phytologist*, *231*, 571–585. https://doi.org/10.1111/nph.17380

Chisanga, C. B., Phiri, E., Chinene, V. R. N., & Chabala, L. M. (2020). Projecting maize yield under local-scale climate change scenarios using crop models: Sensitivity to sowing dates, cultivar, and nitrogen fertilizer rates. *Food and Energy Security*, *9*, e231. https://doi.org/10.1002/fes3.231

Deckmyn, G., Flores, O., Mayer, M., Domene, X., Schnepf, A., Kuka, K., Van Looy, K., Rasse, D. P., Briones, M. J. I., Barot, S., Berg, M., Vanguelova, E., Ostonen, I., Vereecken, H., Suz, L. M., Frey, B., Frossard, A., Tiunov, A., Frouz, J., … Curiel Yuste, J. (2020). KEYLINK: Towards a more integrative soil representation for inclusion in ecosystem scale models. I. review and model concept. *PeerJ*, *8*, e9750. https://doi.org/10.7717/peerj.9750

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Hartman, M. D., Brenner, J., Ojima, D. S., et al. (2001). Simulated interaction of carbon dynamics and nitrogen trace gas fluxes using the DAYCENT model. In M. J. Shaffer, L. Ma, & S. Hansen (Eds.), *Modeling carbon and nitrogen dynamics for soil management* (pp. 303–332). CRC Press.

Deltedesco, E., Keiblinger, K. M., Naynar, M., Piepho, H.-P., Gorfer, M., Herndl, M., Bahn, M., Pötsch, E. M., & Zechmeister-Boltenstern, S. (2019). Trace gas fluxes from managed grassland soil subject to multifactorial climate change manipulation. *Applied Soil Ecology*, *137*, 1–11. https://doi.org/10.1016/j.apsoil.2018.12.023

Eyshi Rezaei, E., Webber, H., Gaiser, T., Naab, J., & Ewert, F. (2015). Heat stress in cereals: Mechanisms and modelling. *European Journal of Agronomy*, *64*, 98–113. https://doi.org/10.1016/j.eja.2014.10.003

Falconnier, G. N., Corbeels, M., Boote, K. J., Affholder, F., Adam, M., Maccarthy, D. S., Ruane, A. C., Nendel, C., Whitbread, A. M., Justes, É., Ahuja, L. R., Akinseye, F. M., Alou, I. N., Amouzou, K. A., Anapalli, S. S., Baron, C., Basso, B., Baudron, F., Bertuzzi, P., … Webber, H. (2020). Modelling climate change impacts on maize yields under low nitrogen input conditions in sub-Saharan Africa. *Global Change Biology*, *26*, 5942–5964. https://doi.org/10.1111/gcb.15261

Farooq, M., Hussain, M., Wahid, A., & Siddique, K. H. M. (2012). Drought stress in plants: An overview. In R. Aroca (Ed.) *Plant responses to drought stress*(pp. 1–33). Springer.

Fereres, E., & González-Dugo, V. (2009). Improving productivity to face water scarcity in irrigated agriculture. In V. Sadras & D. Calderini (Eds.), *Crop physiology* (pp. 122–143). Academic Press.

Forstner, V., Groh, J., Vremec, M., Herndl, M., Vereecken, H., Gerke, H. H., Birk, S., & Pütz, T. (2021). Response of water fluxes and biomass production to climate change in permanent grassland soil ecosystems. *Hydrology and Earth System Sciences*, *25*, 6087–6106. https://doi.org/10.5194/hess-25-6087-2021

Franke, J. A., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Snyder, A., Dury, M., Falloon, P. D., Folberth, C., François, L., Hank, T., Izaurralde, R. C., Jacquemin, I., Jones, C., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T. A. M., … Moyer, E. J. (2020). The GGCMI Phase 2 emulators: Global gridded crop model responses to changes in CO_2, temperature, water, and nitrogen (version 1.0). *Geoscientific Model Development*, *13*, 3995–4018. https://doi.org/10.5194/gmd-13-3995-2020

Gebler, S., Hendricks Franssen, H. - J., Pütz, T., Post, H., Schmidt, M., & Vereecken, H. (2015). Actual evapotranspiration and precipitation measured by lysimeters: A comparison with eddy covariance and

tipping bucket. *Hydrology and Earth System Sciences*, *19*, 2145–2161. https://doi.org/10.5194/hess-19-2145-2015

Gilardelli, C., Confalonieri, R., Cappelli, G. A., & Bellocchi, G. (2018). Sensitivity of WOFOST-based modelling solutions to crop parameters under climate change. *Ecological Modelling*, *368*, 1–14. https://doi.org/10.1016/j.ecolmodel.2017.11.003

Groh, J., Diamantopoulos, E., Duan, X., Ewert, F., Herbst, M., Holbak, M., Kamali, B., Kersebaum, K.-C., Kuhnert, M., Lischeid, G., Nendel, C., Priesack, E., Steidl, J., Sommer, M., Pütz, T., Vereecken, H., Wallor, E., Weber, T. K. D., Wegehenkel, M., … Gerke, H. H. (2020). Crop growth and soil water fluxes at erosion-affected arable sites: Using weighing lysimeter data for model intercomparison. *Vadose Zone Journal*, *19*, e20058. https://doi.org/10.1002/vzj2.20058

Groh, J., Pütz, T., Gerke, H. H., Vanderborght, J., & Vereecken, H. (2019). Quantification and prediction of nighttime evapotranspiration for two distinct grassland ecosystems. *Water Resources Research*, *55*, 2961–2975. https://doi.org/10.1029/2018wr024072

Groh, J., Slawitsch, V., Herndl, M., Graf, A., Vereecken, H., & Pütz, T. (2018). Determining dew and hoar frost formation for a low mountain range and alpine grassland site by weighable lysimeter. *Journal of Hydrology*, *563*, 372–381. https://doi.org/10.1016/j.jhydrol.2018.06.009

Groh, J., Stumpp, C., Lücke, A., Pütz, T., Vanderborght, J., & Vereecken, H. (2018). Inverse estimation of soil hydraulic and transport parameters of layered soils from water stable isotope and lysimeter data. *Vadose Zone Journal*, *17*, 1–19. https://doi.org/10.2136/vzj2017.09.0168

Groh, J., Vanderborght, J., Pütz, T., & Vereecken, H. (2016). How to control the lysimeter bottom boundary to investigate the effect of climate change on soil processes? *Vadose Zone Journal*, *15*, 1–25. https://doi.org/10.2136/vzj2015.08.0113

Groh, J., Vanderborght, J., Pütz, T., Vogel, H.-J., Gründling, R., Rupp, H., Rahmati, M., Sommer, M., Vereecken, H., & Gerke, H. H. (2020). Responses of soil water storage and crop water use efficiency to changing climatic conditions: A lysimeter-based space-for-time approach. *Hydrology and Earth System Sciences*, *24*, 1211–1225. https://doi.org/10.5194/hess-24-1211-2020

Hamidov, A., Helming, K., Bellocchi, G., Bojar, W., Dalgaard, T., Ghaley, B. B., Hoffman, C., Holman, I., Holzkämper, A., Krzeminska, D., Kværnø, S. H., Lehtonen, H., Niedgrist, G., Øygarden, L., Reidsma, P., & Schönhart, M. (2018). Impacts of climate change adaptation options on soil functions: A review of European case-studies. *Land Degradation & Development*, *29*, 2378–2389. https://doi.org/10.1002/ldr.3006

Hansen, S., Abrahamsen, P., Petersen, C. T., & Styczen, M. (2012). Daisy: Model use, calibration, and validation. *Transactions of the ASABE*, *55*, 1315–1334. https://doi.org/10.13031/2013.42244

Haselow, L., Meissner, R., Rupp, H., & Miegel, K. (2019). Evaluation of precipitation measurements methods under field conditions during a summer season: A comparison of the standard rain gauge with a weighable lysimeter and a piezoelectric precipitation sensor. *Journal of Hydrology*, *575*, 537–543. https://doi.org/10.1016/j.jhydrol.2019.05.065

Hay, R. K. M. (1995). Harvest index: A review of its use in plant breeding and crop physiology. *Annals of Applied Biology*, *126*, 197–216. https://doi.org/10.1111/j.1744-7348.1995.tb05015.x

Heinrich, I., Balanzategui, D., Bens, O., Blasch, G., Blume, T., Böttcher, F., Borg, E., Brademann, B., Brauer, A., Conrad, C., Dietze, E., Dräger, N., Fiener, P., Gerke, H. H., Güntner, A., Heine, I., Helle, G., Herbrich, M., Harfenmeister, K., … Wilken, F. (2018). Interdisciplinary geo-ecological research across time scales in the Northeast German Lowland Observatory (TERENO-NE). *Vadose Zone Journal*, *17*, 1–25. https://doi.org/10.2136/vzj2018.06.0116

Hilton, T. W., Loik, M. E., & Campbell, J. E (2019). Simulating international drought experiment field observations using the community land model. *Agricultural and Forest Meteorology*, *266–267*, 173–183. https://doi.org/10.1016/j.agrformet.2018.12.016

Houska, T., Kraft, P., Jehn, F. U., Bestian, K., Kraus, D., & Breuer, L. (2021). Detection of hidden model errors by combining single and multi-criteria calibration. *Science of the Total Environment*, *777*, 146218. https://doi.org/10.1016/j.scitotenv.2021.146218

Hussain, J., Khaliq, T., Ahmad, A., & Akhtar, J. (2018). Performance of four crop model for simulations of wheat phenology, leaf growth, biomass and yield across planting dates. *PLOS ONE*, *13*, e0197546. https://doi.org/10.1371/journal.pone.0197546

Hütsch, B. W., & Schubert, S. (2018). Maize harvest index and water use efficiency can be improved by inhibition of gibberellin biosynthesis. *Journal of Agronomy and Crop Science*, *204*, 209–218. https://doi.org/10.1111/jac.12250

Katerji, N., Mastrorilli, M., & Rana, G. (2008). Water use efficiency of crops cultivated in the Mediterranean region: Review and analysis. *European Journal of Agronomy*, *28*, 493–507. https://doi.org/10.1016/j.eja.2007.12.003

Kersebaum, K. C. (2007). Modelling nitrogen dynamics in soil–crop systems with HERMES. *Nutrient Cycling In Agroecosystems*, *77*, 39–52. https://doi.org/10.1007/s10705-006-9044-8

Kersebaum, K. C., Boote, K. J., Jorgenson, J. S., Nendel, C., Bindi, M., Frühauf, C., Gaiser, T., Hoogenboom, G., Kollas, C., Oleson, J. E., Rötter, R. P., Ruget, F., Thorburn, M., Trnka, M., & Wegehenkel, M. (2015). Analysis and classification of data sets for calibration and validation of agro-ecosystem models. *Environmental Modelling & Software*, *72*, 402–417. https://doi.org/10.1016/j.envsoft.2015.05.009

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*, W03S04. https://doi.org/10.1029/2005WR004362

Klosterhalfen, A., Herbst, M., Weihermüller, L., Graf, A., Schmidt, M., Stadler, A., Schneider, K., Subke, J.-A., Huisman, J. A., & Vereecken, H. (2017). Multi-site calibration and validation of a net ecosystem carbon exchange model for croplands. *Ecological Modelling*, *363*, 137–156. https://doi.org/10.1016/j.ecolmodel.2017.07.028

Knapp, A. K., Carroll, C. J. W., Griffin-Nolan, R. J., Slette, I. J., Chaves, F. A., Baur, L. E., Felton, A. J., Gray, J. E., Hoffman, A. M., Lemoine, N. P., Mao, W., Post, A. K., & Smith, M. D. (2018). A reality check for climate change experiments: Do they reflect the real world? *Ecology*, *99*, 2145–2151. https://doi.org/10.1002/ecy.2474

Köhler, I. H., Huber, S. C., Bernacchi, C. J., & Baxter, I. R. (2019). Increased temperatures may safeguard the nutritional quality of crops under future elevated CO_2 concentrations. *The Plant Journal*, *97*, 872–886. https://doi.org/10.1111/tpj.14166

Kollas, C., Kersebaum, K. C., Nendel, C., Manevski, K., Müller, C., Palosuo, T., Armas-Herrera, C. M., Beaudoin, N., Bindi, M., Charfeddine, M., Conradt, T., Constantin, J., Eitzinger, J., Ewert, F., Ferrise, R., Gaiser, T., Cortazar-Atauri, I. G. D. e, Giglio, L., Hlavinka, P., … Wu, L. (2015). Crop rotation modelling: A European model intercomparison. *European Journal of Agronomy*, *70*, 98–111. https://doi.org/10.1016/j.eja.2015.06.007

Lenka, N. K., Lenka, S., Thakur, J. K., Yashona, D. S., Shukla, A. K., Elanchezhian, R., Singh, K. K., Biswas, A. K., & Patra, A. K. (2020). Carbon dioxide and temperature elevation effects on crop evapotranspiration and water use efficiency in soybean as affected by different nitrogen levels. *Agricultural Water Management*, *230*, 105936. https://doi.org/10.1016/j.agwat.2019.105936

Loik, M. E., Lesage, J. C., Brown, T. M., & Hastings, D. O. (2019). Drought-Net rainfall shelters did not cause nondrought effects on photosynthesis for California central coast plants. *Ecohydrology*, *12*, e2138. https://doi.org/10.1002/eco.2138

Longo, M., Jones, C. D., Izaurralde, R. C., Cabrera, M. L., Dal Ferro, N., & Morari, F. (2021). Testing the EPIC Richards submodel for simulating soil water dynamics under different bottom boundary conditions. *Vadose Zone Journal*, 20, e20142. https://doi.org/10.1002/vzj2.20142

MacBean, N., Scott, R. L., Biederman, J. A., Ottlé, C., Vuichard, N., Ducharne, A., Kolb, T., Dore, S., Litvak, M., & Moore, D. J. P. (2020). Testing water fluxes and storage from two hydrology configurations within the ORCHIDEE land surface model across US semi-arid sites. *Hydrology and Earth System Sciences*, *24*, 5203–5230. https://doi.org/10.5194/hess-24-5203-2020

Mäkinen, H., Kaseva, J., Trnka, M., Balek, J., Kersebaum, K. C., Nendel, C., Gobin, A., Olesen, J. E., Bindi, M., Ferrise, R., Moriondo, M., Rodríguez, A., Ruiz-Ramos, M., Takáč, J., Bezák, P., Ventrella, D., Ruget, F., Capellades, G., & Kahiluoto, H. (2018). Sensitivity of European wheat to extreme weather. *Field Crops Research*, *222*, 209–217. https://doi.org/10.1016/j.fcr.2017.11.008

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., … Wolf, J. (2015). Multimodel ensembles of wheat growth: Many models are better than one. *Global Change Biology*, *21*, 911–925. https://doi.org/10.1111/gcb.12768

Maurer, T., & Gerke, H. H. (2016). Processes and modeling of initial soil and landscape development: A review. *Vadose Zone Journal*, *15*, 1–20. https://doi.org/10.2136/vzj2016.05.0048

Melsen, L. A., & Guse, B. (2021). Climate change impacts model parameter sensitivity: Implications for calibration strategy and model diagnostic evaluation. *Hydrology and Earth System Sciences*, *25*, 1307–1332. https://doi.org/10.5194/hess-25-1307-2021

Mulla, S., Singh, S. K., Singh, K. K., & Praveen, B. (2020). Climate change and agriculture: A review of crop models. In V. Venkatramanan, S. Shah, & R. Prasad (Eds.), *Global climate change and environmental policy: Agriculture perspectives* (pp. 423–435). Springer.

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. *Ecological Modelling*, *222*, 1614–1625. https://doi.org/10.1016/j.ecolmodel.2011.02.018

Peng, B., Guan, K., Tang, J., Ainsworth, E. A., Asseng, S., Bernacchi, C. J., Cooper, M., Delucia, E. H., Elliott, J. W., Ewert, F., Grant, R. F., Gustafson, D. I., Hammer, G. L., Jin, Z., Jones, J. W., Kimm, H., Lawrence, D. M., Li, Y., Lombardozzi, D. L., … Zhou, W. (2020). Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nature Plants*, *6*, 338–348. https://doi.org/10.1038/s41477-020-0625-3

Pereira, L. S., Paredes, P., López-Urrea, R., Hunsaker, D. J., Mota, M., & Mohammadi Shad, Z. (2021). Standard single and basal crop coefficients for vegetable crops, an update of FAO56 crop water requirements approach. *Agricultural Water Management*, *243*, 106196. https://doi.org/10.1016/j.agwat.2020.106196

Priesack, E., Gayler, S., & Hartmann, H. P. (2006). The impact of crop growth sub-model choice on simulated water and nitrogen balances. *Nutrient Cycling in Agroecosystems*, *75*, 1–13. https://doi.org/10.1007/s10705-006-9006-1

Pszczółkowska, A., Okorski, A., Fordoński, G., Kotecki, A., Kozak, M., & Dzienis, G. (2020). Effect of weather conditions on yield and health status of faba bean seeds in Poland. *Agronomy*, *10*, 48. https://doi.org/10.3390/agronomy10010048

Pütz, T. h., Kiese, R., Wollschläger, U., Groh, J., Rupp, H., Zacharias, S., Priesack, E., Gerke, H. H., Gasche, R., Bens, O., Borg, E., Baessler, C., Kaiser, K., Herbrich, M., Munch, J.-C., Sommer, M., Vogel, H.-J., Vanderborght, J., & Vereecken, H. (2016). TERENO-SOILCan: A lysimeter-network in Germany observing soil processes and plant diversity influenced by climate change. *Environmental Earth Sciences*, *75*, 1242. https://doi.org/10.1007/s12665-016-6031-5

Rezaei, E. E., Siebert, S., Manderscheid, R., Müller, J., Mahrookashani, A., Ehrenpfordt, B., Haensch, J., Weigel, H.-J., & Ewert, F. (2018). Quantifying the response of wheat yields to heat stress: The role of the experimental setup. *Field Crops Research*, *217*, 93–103. https://doi.org/10.1016/j.fcr.2017.12.015

Robinson, D. A., Hopmans, J. W., Filipovic, V., van Der Ploeg, M., Lebron, I., Jones, S. B., Reinsch, S., Jarvis, N., & Tuller, M. (2019). Global Environmental changes impact soil hydraulic functions through biophysical feedbacks. *Global Change Biology*, 0. https://doi.org/10.1111/gcb.14626

Robinson, D. A., Jones, S. B., Lebron, I., Reinsch, S., Domínguez, M. T., Smith, A. R., Jones, D. L., Marshall, M. R., & Emmett, B. A. (2016). Experimental evidence for drought induced alternative stable states of soil moisture. *Scientific Reports*, *6*, 20018. https://doi.org/10.1038/srep20018

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 3268–3273. https://doi.org/10.1073/pnas.1222463110

Roy, J., Rineau, F., De Boeck, H. J., Nijs, I., Pütz, T., Abiven, S., Arnone, J. A., Barton, C. V. M., Beenaerts, N., Brüggemann, N., Dainese, M., Domisch, T., Eisenhauer, N., Garré, S., Gebler, A., Ghirardo, A., Jasoni, R. L., Kowalchuk, G., Landais, D., … Milcu, A. (2021). Ecotrons: Powerful and versatile ecosystem analysers for ecology, agronomy and environmental science. *Global Change Biology*, *27*(7), 1387–1407. https://doi.org/10.1111/gcb.15471

Ruane, A. C., Rosenzweig, C., Asseng, S., Boote, K. J., Elliott, J., Ewert, F., Jones, J. W., Martre, P., Mcdermid, S. P., Müller, C., Snyder, A., & Thorburn, P. J. (2017). An AgMIP framework for improved agricultural representation in integrated assessment models. *Environmental Research Letters*, *12*, 125003. https://doi.org/10.1088/1748-9326/aa8da6

Sadras, V. O., Villalobos, F. J., Orgaz, F., & Fereres, E. (2016). Effects of water stress on crop production. In F. J. Villalobos & E. Fereres (Eds.), *Principles of agronomy for sustainable agriculture* (pp. 189–204). Springer International Publishing.

Schleussner, C.-F., Deryng, D., Müller, C., Elliott, J., Saeed, F., Folberth, C., Liu, W., Wang, X., Pugh, T. A. M., Thiery, W., Seneviratne, S.

I., & Rogelj, J. (2018). Crop productivity changes in 1.5 ˚C and 2 ˚C worlds under climate sensitivity uncertainty. *Environmental Research Letters*, *13*, 064007. https://doi.org/10.1088/1748-9326/aab63b

Schneider, J., Groh, J., Pütz, T., Helmig, R., Rothfuss, Y., Vereecken, H., & Vanderborght, J. (2021). Prediction of soil evaporation measured with weighable lysimeters using the FAO Penman–Monteith method in combination with Richards' equation. *Vadose Zone Journal*, *20*, e20102. https://doi.org/10.1002/vzj2.20102

Seidel, S. J., Palosuo, T., Thorburn, P., & Wallach, D. (2018). Towards improved calibration of crop models – Where are we now and where should we go? *European Journal of Agronomy*, *94*, 25–35. https://doi.org/10.1016/j.eja.2018.01.006

Semenov, M. A., & Shewry, P. R. (2011). Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Scientific Reports*, *1*, 66. https://doi.org/10.1038/srep00066

Shenoda, J. E., Sanad, M. N. M. E., Rizkalla, A. A., El-Assal, S., Ali, R. T., & Hussein, M. H. (2021). Effect of long-term heat stress on grain yield, pollen grain viability and germinability in bread wheat (*Triticum aestivum* L.) under field conditions. *Heliyon*, *7*, e07096. https://doi.org/10.1016/j.heliyon.2021.e07096

Sima, M. W., Fang, Q. X., Qi, Z., & Yu, Q. (2020). Direct assimilation of measured soil water content in Root Zone Water Quality Model calibration for deficit-irrigated maize. *Agronomy Journal*, *112*, 844–860. https://doi.org/10.1002/agj2.20088

Soldevilla-Martinez, M., Quemada, M., López-Urrea, R., Muñoz-Carpena, R., & Lizaso, J. I. (2014). Soil water balance: Comparing two simulation models of different levels of complexity with lysimeter observations. *Agricultural Water Management*, *139*, 53–63. https://doi.org/10.1016/j.agwat.2014.03.011

Supit, I., Hoojer, A. A., & Diepen, C. A. (1994). System description of the Wofost 6.0 crop simulation model implemented in CGMS. Volume 1: Theory and AlgorithmsEC *(Publication EUR 15956)*. Office for Official Publications of the European Communities.

Tao, F., Palosuo, T., Rötter, R. P., Díaz-Ambrona, C. G. H., Inés Mínguez, M., Semenov, M. A., Kersebaum, K. C., Cammarano, D., Specka, X., Nendel, C., Srivastava, A. K., Ewert, F., Padovan, G., Ferrise, R., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J. G., … Schulman, A. H. (2020). Why do crop models diverge substantially in climate impact projections? A comprehensive analysis based on eight barley crop models. *Agricultural and Forest Meteorology*, *281*, 107851. https://doi.org/10.1016/j.agrformet.2019.107851

Tao, F., Rötter, R. P., Palosuo, T., Gregorio Hernández Díaz-Ambrona, C., Mínguez, M. I., Semenov, M. A., Kersebaum, K. C., Nendel, C., Specka, X., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J. G., Salo, T., Ferrise, R., … Schulman, A. H. (2018). Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. *Global Change Biology*, *24*, 1291–1307. https://doi.org/10.1111/gcb.14019

Tao, L.-X., Tan, H.-J., Wang, X., Cao, L.-Y., Song, J., & Cheng, S.-H. (2008). Effects of high-temperature stress on flowering and grain-setting characteristics of Guodao 6. *Acta Agronomica Sinica*, *34*, 609–674. https://doi.org/10.1016/S1875-2780(08)60027-9

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P. J., Ittersum, M., Aggarwal, P. K., Ahmed, M., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Dumont, B., Eyshi Rezaei, E., Fereres, E., Fitzgerald, G. J., Gao, Y., … Zhang, Z. (2018). Multimodel ensembles improve predictions of crop–environment–management interactions. *Global Change Biology*, *24*, 5072–5083. https://doi.org/10.1111/gcb.14411

Wallach, D., Palosuo, T., Thorburn, P., Gourdain, E., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hochman, Z., Hoek, S., Hoogenboom, G., Horan, H., … Seidel, S. J. (2021). How well do crop modeling groups predict wheat phenology, given calibration data from the target population? *European Journal of Agronomy*, *124*, 126195. https://doi.org/10.1016/j.eja.2020.126195

Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dumont, B., Ferrise, R., Gaiser, T., Gayler, S., Hiremath, S., Hoek, S., Horan, H., Hoogenboom, G., Huang, M., Jabloun, M., … Seidel, S. J. (2021). Multi-model evaluation of phenology prediction for wheat in Australia. *Agricultural and Forest Meteorology*, *298–299*, 108289. https://doi.org/10.1016/j.agrformet.2020.108289

Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., …, Seidel, S. J. (2021). The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environmental Modelling & Software*, *145*, 105206. https://doi.org/10.1016/j.envsoft.2021.105206

Wallor, E., Kersebaum, K.-C., Ventrella, D., Bindi, M., Cammarano, D., Coucheney, E., Gaiser, T., Garofalo, P., Giglio, L., Giola, P., Hoffmann, M. P., Iocola, I., Lana, M., Lewan, E., Maharjan, G. R., Moriondo, M., Mula, L., Nendel, C., Pohankova, E., … Trombi, G. (2018). The response of process-based agro-ecosystem models to within-field variability in site conditions. *Field Crops Research*, *228*, 1–19. https://doi.org/10.1016/j.fcr.2018.08.021

Webber, H., Hoffmann, M., & Rezaei, E. E. (2018). Crop models as tools for agroclimatology. In J. L. Hatfield, M. V. K. Sivakumar, & J. H. Prueger (Eds.), *Agroclimatology: Linking agriculture to climate* (Vol. 60, pp. 519–546). ASA and CSSA. https://doi.org/10.2134/agronmonogr60.2016.0025

Webber, H., White, J. W., Kimball, B. A., Ewert, F., Asseng, S., Eyshi Rezaei, E., Pinter, P. J., Hatfield, J. L., Reynolds, M. P., Ababaei, B., Bindi, M., Doltra, J., Ferrise, R., Kage, H., Kassie, B. T., Kersebaum, K.-C., Luig, A., Olesen, J. E., Semenov, M. A., … Martre, P. (2018). Physical robustness of canopy temperature models for crop heat stress simulation across environments and production conditions. *Field Crops Research*, *216*, 75–88. https://doi.org/10.1016/j.fcr.2017.11.005

Wegehenkel, M., Luzi, K., Sowa, D., Barkusky, D., & Mirschel, W. (2019). Simulation of long-term soil hydrological conditions at three agricultural experimental field plots compared with measurements. *Water*, *11*, 989. https://doi.org/10.3390/w11050989

Wollschläger, U., Attinger, S., Borchardt, D., Brauns, M., Cuntz, M., Dietrich, P., Fleckenstein, J. H., Friese, K., Friesen, J., Harpke, A., Hildebrandt, A., Jäckel, G., Kamjunke, N., Knöller, K., Kögler, S., Kolditz, O., Krieg, R., Kumar, R., Lausch, A., … Zacharias, S. (2016). The Bode hydrological observatory: A platform for integrated, interdisciplinary hydro-ecological research within the TERENO Harz/Central German Lowland Observatory. *Environmental Earth Sciences*, *76*, 29. https://doi.org/10.1007/s12665-016-6327-5

Yin, X., Kersebaum, K. C., Kollas, C., Baby, S., Beaudoin, N., Manevski, K., Palosuo, T., Nendel, C., Wu, L., Hoffmann, M., Hoffmann, H., Sharif, B., Armas-Herrera, C. M., Bindi, M., Charfeddine, M.,

Conradt, T., Constantin, J., Ewert, F., Ferrise, R., … Olesen, J. E. (2017). Multi-model uncertainty analysis in predicting grain N for crop rotations in Europe. *European Journal of Agronomy*, *84*, 152–165. https://doi.org/10.1016/j.eja.2016.12.009

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., … Vereecken, H. (2011). A network of terrestrial environmental observatories in Germany. *Vadose Zone Journal*, *10*, 955–973. https://doi.org/10.2136/vzj2010.0139

Zeng, X., Shen, S. S. P., Zeng, X., & Dickinson, R. E. (2004). Multiple equilibrium states and the abrupt transitions in a dynamical system of soil water interacting with vegetation. *Geophysical Research Letters*, *31*, L05501. https://doi.org/10.1029/2003GL018910

Zwetsloot, M. J., van Leeuwen, J., Hemerik, L., Martens, H., Simó Josa, I., Broek, M., Debeljak, M., Rutgers, M., Sandén, T., Wall, D. P., Jones, A., & Creamer, R. E. (2020). Soil multifunctionality: Synergies and trade-offs across European climatic zones and land uses. *European Journal of Soil Science*, *72*, 1640–1654. https://doi.org/10.1111/ejss.13051

Žydelis, R., Weihermüller, L., & Herbst, M. (2021). Future climate change will accelerate maize phenological development and increase yield in the Nemoral climate. *Science of the Total Environment*, *784*, 147175. https://doi.org/10.1016/j.scitotenv.2021.147175

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Groh, J., Diamantopoulos, E., Duan, X., Ewert, F., Heinlein, F., Herbst, M., Holbak, M., Kamali, B., Kersebaum, K.-C., Kuhnert, M., Priesack, E., Steidl, J., Sommer, M., Pütz, T., Vanderborght, J., Vereecken, H., Wallor, E., Weber, T. K. D., … Gerke, H. H. (2022). Same soil, different climate: Crop model intercomparison on translocated lysimeters. *Vadose Zone Journal*, e20202. https://doi.org/10.1002/vzj2.20202