Crop rotation modelling—A European model intercomparison

Chris Kollas[a,c,*], Kurt Christian Kersebaum[a], Claas Nendel[a], Kiril Manevski[b], Christoph Müller[c], Taru Palosuo[d], Cecilia M. Armas-Herrera[e], Nicolas Beaudoin[e], Marco Bindi[r], Monia Charfeddine[g], Tobias Conradt[c], Julie Constantin[h], Josef Eitzinger[i], Frank Ewert[k], Roberto Ferrise[r], Thomas Gaiser[k], Iñaki Garcia de Cortazar-Atauri[l], Luisa Giglio[g], Petr Hlavinka[m,u], Holger Hoffmann[k], Munir P. Hoffmann[n], Marie Launay[l], Remy Manderscheid[o], Bruno Mary[e], Wilfried Mirschel[a], Marco Moriondo[s], Jørgen E. Olesen[b], Isik Öztürk[b], Andreas Pacholski[j,o], Dominique Ripoche-Wachter[l], Pier Paolo Roggero[f], Svenja Roncossek[b], Reimund P. Rötter[d], Françoise Ruget[p], Behzad Sharif[b], Mirek Trnka[m,u], Domenico Ventrella[g], Katharina Waha[c,t], Martin Wegehenkel[a], Hans-Joachim Weigel[o], Lianhai Wu[q]

[a] *Institute of Landscape Systems Analysis, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany*
[b] *Department of Agroecology, Aarhus University, Blichers Alle 20, 8830, Tjele, Denmark*
[c] *Potsdam Institute for Climate Impact Research, P.O. Box 601203, 14412 Potsdam, Germany*
[d] *Natural Resources Institute Finland (Luke), Viikinkari 4, 00790 Helsinki, Finland*
[e] *INRA UR1158 AgroImpact, 02000 Barenton-Bugny, France*
[f] *Nucleo di Ricerca sulla Desertificazione e Dipartimento di Agraria, University of Sassari, 07100 Sassari, Italy*
[g] *Consiglio per la Ricerca in Agricoltura e l'analisi dell'economia agraria, Unità di ricerca per i sistemi colturali degli ambienti caldo-aridi (CRA-SCA), Via Celso Ulpiani 5, 70125 Bari, Italy*
[h] *INRA, UMR1248 AGIR, F-31326 Castanet-Tolosan, France*
[i] *Institute of Meteorology, Department for Water, Atmosphere and Environment, University of Natural Resources and Applied Life Sciences, Peter Jordan Strasse 82, A-1190 Vienna, Austria*
[j] *Graduate School/Inkubator, Leuphana University Lüneburg, Scharnhorststr. 1, 21335 Lüneburg, Germany*
[k] *INRES, University of Bonn, Katzenburgweg 5, Bonn 53115, Germany*
[l] *INRA, US1116 AgroClim, F-84914 Avignon, France*
[m] *Global Change Research Centre AS CR, v.v.i., Bělidla 986/4a, 603 00 Brno, Czech Republic*
[n] *Crop Production Systems in the Tropics, Georg-August-Universität Göttingen, Grisebachstr. 6, 37077 Göttingen, Germany*
[o] *Thünen Institute of Biodiversity, Bundesallee 50, 38116 Braunschweig, Germany*
[p] *INRA, UMR1114 EMMAH, F-84914 Avignon, France*
[q] *Rothamsted Research, North Wyke, Okehampton EX20 2SB, UK*
[r] *Department of Agri-food Production and Environmental Sciences, University of Florence, P.le delle Cascine 18, 50144 Firenze, Italy*
[s] *Institute of Biometeorology of the National Research Council (CNR-IBIMET), via Caproni 8, 50145 Firenze, Italy*
[t] *CSIRO Agriculture, 306Carmody Road, 4067 St. Lucia QLD, Australia*
[u] *Institute of Agrosystems and Bioclimatology, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic*

* Corresponding author. Fax: +49 33432 82212.
E-mail address: chris.kollas@zalf.de (C. Kollas).

1. Introduction

In many European countries, field crops are traditionally grown in rotation (a varying sequence of different crops, often including intermediate crops) for a number of economic and environmental reasons.

Firstly, most crops benefit from the nutrients released by mineralising residues of the preceding crop. This substitution of mineral fertiliser by the strategic use of nutrient-rich crop residues is especially important in low-fertilised cropping systems, where intermediate crops are grown for the purpose of catching and recycling nutrients, in particular nitrogen (N), from deeper soil layers (catch crops, see Askegaard and Eriksen, 2008). Furthermore, intermediate crops prevent rainfall from percolating through the soil and leaching nutrients out of the rooting zone of the main crops, also preventing erosion (cover crops, Boardman and Favis-Mortlock, 2014; Simoes et al., 2014). In addition, intermediate crops are used as fertility-building crops or green manure to add further nutrients to the soil (Kirkegaard et al., 2008). Secondly, a well-designed crop rotation can contribute organic matter to the soil to compensate for decomposition under the prevailing environmental conditions (Kay, 1990), maintaining long-term soil fertility and habitat quality for soil organisms. Sufficient soil organic matter also reduces the risk of erosion and nutrient losses while increasing the potential water supply to the crop by increasing the soil water storage capacity. Thirdly, the risk of phyto-sanitary problems in the crop rotation can be minimised by accounting for the ability of different species to repel pests or diseases by interrupting the life cycle of host-specific pathogens via so-called break crops (Angus et al., 2011). The same applies to the control of weeds, the unwanted growth of which can be hampered effectively by selecting a suitable preceding crop, especially in cropping systems where resistance to herbicides has developed (Stevenson and Van Kessel, 1996).

In a wider context, crop rotations – and thus crop diversification – have been identified as prominent measures for increasing the resilience of the agricultural system (Reidsma et al., 2009; Lin, 2011; Smith et al., 2008), developing mitigative adaptation strategies to climate change (Olesen et al., 2011) and improving ecosystem services (Hauck et al., 2014). Consequently, the recent common agricultural policy (CAP) of the European Commission (2011) considers the diversification of crop rotations a key measure for more sustainable agriculture. While in organic farming the design of crop rotations is driven by the idea of reducing N losses by transferring it among crops and improving soil fertility, conventional crop rotations are driven by economic and political boundary conditions. Agricultural policies have a strong potential to modify cropping trends (European Commission 2010), such as the use of catch crops, which is promoted by policy incentives within agro-environmental action plans (e.g. European Water Framework directive, Uthes et al., 2010).

Crop models provide an explicit representation of fundamental bio-physical processes such as crop development and growth (photosynthesis, leaf area and canopy expansion, dry matter partitioning and root growth), and water/N cycles in a single crop season or within a crop rotation (Wallach et al., 2006). In the past, crop rotation modelling was regularly applied as a tool for investigating the soil water balance among certain crop sequences (Post et al., 2007; Salado-Navarro and Sinclair, 2009) and for estimating the amount of nitrate that leaches over long periods (Beaudoin et al., 2008; Kersebaum and Beblik, 2001; Kovács et al., 1995). In addition, growing interest in the carbon storage capacity of agricultural soils and crops has led to the application of crop rotation modelling studies focusing on long-term soil carbon dynamics (Li et al., 1994; Blombäck et al., 2003; Hlavinka et al., 2013). Last but not least, crop rotation models have also been used to study the development of above-/below-ground biomass and yields (Berntsen et al., 2006) as well as nitrogen uptake (Nendel et al., 2013). These examples show that the precise simulation of crop rotations will help us to better address a wide range of challenges facing society, e.g. soil and water conservation, carbon sequestration, mitigation of greenhouse gas emissions, sustainable intensification of cropping practice, and food security.

One drawback of such models is that they do not deal specifically with uncertainty in explaining data, measurements or conditions in agro-climatic regions; instead, they inherently evoke uncertainty in the model predictions (Asseng et al., 2013). For this reason, emphasis has recently been placed on the multi-model ensemble methodology, which was recommended as a valuable tool for assessing and reducing uncertainties in crop simulations (Rosenzweig and Wilbanks, 2010; Rötter et al., 2011; Challinor et al., 2014). In fact, previous studies demonstrated the strength of model intercomparisons (Palosuo et al., 2011; Rötter et al., 2012; Asseng et al., 2013, 2014; Bassu et al., 2014; Li et al., 2014; Martre et al., 2014). Palosuo et al. (2011) showed that, with minimal calibration, none of the models involved were robust enough and sufficiently accurate across a range of environments in a winter-wheat crop model comparison exercise. Furthermore, all of the above studies showed that the multi-model mean of simulations is a better estimator of the mean crop yield than single-model simulations.

However, most climate impact studies on crop production focus methodologically on simulating single years and single crops (Asseng et al., 2013; Bassu et al., 2014; Palosuo et al., 2011; Rötter et al., 2012), although in situ crop performance depends strongly on the crop's position within the sequence of crops (see arguments above). In these studies, the initial conditions of the soil in terms of water, organic matter and nutrients were kept constant at the onset of each of the growing seasons; carry-over effects such as N mineralising from the harvest residues of the previous year or altered soil water content due to evaporation from cover crops were ignored, which is viewed as a drawback of climate impact studies (Ewert et al., 2014). In contrast, Teixeira et al. (in press) recently demonstrated the advantage of simulating continuous crop rotations compared to single crops and years for a single location in new Zealand, particularly under limited growing conditions. Here, we hypothesise that the continuous simulation of crop growth across years will improve yield predictions at different locations across Europe and rotations compared to simulating crop growth in single years only.

All of the multi-model comparisons were undertaken for the globally most important staple crops, such as wheat, barley, rice and maize (Asseng et al., 2013; Bassu et al., 2014; Li et al., 2014; Palosuo et al., 2011; Rötter et al., 2012); less attention was paid to other key crops such as sugar beet and oat (White et al., 2011). Moreover, intermediate crops are largely disregarded in crop simulation studies. With a few exceptions (i.e. some legumes), these crops generate a limited or no direct commercial product, but have a significant impact on soil fertility and the growth of the following crop (Blombäck et al., 2003). Consequently, many current crop models exhibit a limited ability to simulate continuous farming systems.

In the present study, therefore, we ask the following research questions concerning a minimal model calibration of fifteen crop growth simulation models:

(1) How accurately can a crop model ensemble simulate the crop yields of various crop rotations? Furthermore, which of the two modes of simulation – continuous or year-by-year – performs better in terms of accuracy?
(2) Which crops common to European agriculture are simulated more accurately, and which crops create major deficiencies in reproducing yields?
(3) Is a crop model ensemble capable of reproducing the effects of sites and treatments on crop yield?

2. Methods

2.1. Experimental crop rotations

Five experimental crop rotation datasets, each containing a different set of treatments (Table 1), were selected for the present study. The datasets cover the European environmental zones of the Atlantic North, Atlantic Central, Continental and Pannonia (lowlands, valleys and mountain peripheries on the Middle- and the Lower-Danube Plains and the Black Sea area), according to Metzger et al. (2005) (Fig. 1). Overall, the study provided experimental data on 301 growing seasons and ten distinct crops.

At the experimental site in Foulum, Denmark, the effects of catch crops in crop rotations and the strategies of tillage and crop residues at harvest were investigated from 2002 to 2012. For our study, six treatments encompassing two rotations were selected, namely a rotation of winter barley (*Hordeum vulgare* L.) – winter oilseed rape (*Brassica napus* L.) – winter wheat (*Triticum aestivum* L.) and a rotation of winter wheat/grass (*Lolium perenne* L.) – spring barley/grass–pea (*Pisum sativum* L.) – winter wheat–winter wheat–spring barley/oilseed radish (*Raphanus sativus* var. *oleiformis*) – spring oat (*Avena sativa* L.); two tillage regimes (ploughed and no tillage); and two residue managements (retention of straw at the site and removal of straw). The experimental site and setting are described in detail in Munkholm et al. (2008).

A second experiment in Müncheberg, Germany, was designed to study management intensities, irrigation, biomass development and inter-annual variation in crop rotations (Mirschel et al., 2007). The dataset, consisting of one crop rotation from 1992 to 1998, is composed of sugar beet (*Beta vulgaris* L.), winter wheat, winter barley, winter rye (*Secale cereale* L.) and oilseed radish (catch crop). The rotation was carried out in four parallel plots with a shift of one year to establish each crop each year. Treatments included rainfed agriculture vs. an irrigated regime.

The Braunschweig Free-Air Carbon Dioxide Enrichment (FACE) experiment was set up to study the interactive effects of CO_2 concentration and N fertilisation on crop production (Weigel and Manderscheid, 2012). The crop rotation consisted of winter barley – a mixture of three different ryegrass cultivars (*Lolium multiflorum* Lam.) as a cover crop – sugar beet–winter wheat, a sequence grown in two consecutive cycles starting in autumn 1999. Treatments included an ambient (374 ppm) and an enriched (550 ppm) concentration of atmospheric CO_2, both with a standard and a reduced (−50%) supply of nitrogen (N) fertiliser.

The fourth experiment focused on agricultural management practice concerning soil water drainage and nitrate leaching, using lysimeters in the agricultural region in Marchfeld, Austria (Eitzinger et al., 2004). Here, the crop rotation involved mustard (*Sinapis alba,* catch crop) – spring wheat–mustard–barley–winter wheat–mustard–potato (*Solanum tuberosum* L.) – winter wheat–maize (*Zea mays* L.) – winter wheat. The crops were grown on three soil types (Calcic Chernozem, shallow and sandy Calcaric Phaeozem and Gleyic Phaeozem) in order to study the water cycle, and the influence of soil type and rotation.

Finally, the Thibie experiment in France combined the effects of catch crops (catch crop vs. no autumn/winter crop cover) and nitrogen fertiliser (conventional vs. reduced N fertilisation) on yields involving a medium-term experiment from 1991 to 2003 (Constantin et al., 2010; Constantin et al., 2010). It consisted of a split-plot design of pea, winter wheat and sugar beet crops in rotation. All crops were present each year. The catch crops grown during the period under investigation included grass, oilseed radish and barley.

2.2. Crop models

Fifteen European modelling teams participated in the present study. The models varied considerably in complexity and functionality, ranging from a dynamic global vegetation model to agroecosystem models designed for field-scale application (Table 2). A list describing the physiological processes and approaches taken by the models can be found in Supplement A. The models were grouped according to their capability of simulating continuous crop growth. Three models (Theseus, SWIM and APSIM) performed continuous runs only, i.e. they simulated the multi-year datasets without reinitialising subroutines at the onset of each growing season, hereafter called ROTATION. The five models DSSAT1/DSSAT2, LPJmL, SPACSYS and WOFOST performed single-season crop growth only, i.e. they simulated each crop in the rotation separately, hereafter called SINGLE. Crops were simulated separately either because the model was unable to reproduce rotations (e.g. cover periods without crop) or because specific crops had not yet been implemented in that respective model. In the latter case, models skipped the corresponding crop or dataset. Finally, half of the models (DAISY, FASSET, HERMES, MONICA, STICS, LINTUL and CROPSYST) provided results for both modes, ROTATION and SINGLE (Table 2).

2.3. Simulation task with standardised input data

The simulation task for all modellers was designed to reproduce the field experimental treatments. Hence, the modelling teams were requested to simulate each treatment at each site, using observed information on daily weather (precipitation, minimum and maximum temperature, mean relative humidity, global radiation and mean wind speed), information on daily field management (previous crops, tillage, sowing, irrigation, fertilisation and harvest) and soil properties (bulk density, texture, organic matter and water capacity parameters) as the driving variables for the models. All of the variables were provided synchronously. In order to evaluate any differences between the SINGLE and ROTATION modes of simulation, the modelling teams were supplied with initial values of soil water content and soil mineral N (at a date close to sowing) for each treatment for the first year only. Thus, the ROTATION simu-

Table 1
Characteristics of the study sites.

Location	Position (latitude/longitude/elevation a.s.l.)	Precipitation[a] [mm $year^{-1}$]	Temperature[b] [°C]	Soil type	Period	Crop rotation[c]	Treatment (all)	Treatment (tested)	Minimal calibration[d]
Foulum (FO)	56.49/9.57/52 m	670	7.9	Mollic luvisol	2002–2012	BAR/RAP/WHB WHB/GRV/BAR/GRV/PEA/ WHB/WHB/BAR/RAD/OAT/ WHB/RAD/BAR/RAD/OAT	6 (tillage, rotation, residuals)	No till vs. plough Retention of resid. vs. removal of resid.	Phen/1 treat
Münche-berg (MU)	52.52/14.12/62 m	564	8.4	Eutric Cambisol	1992–1998	SBT/WHB/BAR/RYE/RAD	8 (irrigation, inter-year variation)	Irrigated vs rainfed	Biom/1 treat
Braun-schweig (BR)	52.3/10.45/79 m	642	10.0	Luvisol	1999–2005	BAR/GRV/SBT/WHB	4 (fertiliser, CO_2)	Elevated CO_2 vs ambient concentr. High vs. low fertilisation	Phen/4 years
Hirsch-stetten (HI)	48.2/16.57/150 m	495	11.0	Gleyic phaeozem/ Calcaric phaeozem/ Calcic chernozem	1998–2004	MUS/WHB/MUS/BAR/ WBH/MUS/POT/WHB/ MAZ/WHB	3 (soil type)	cPhaeo vs. gPhaeo Cherno vs. gPhaeo	Phen/1 treat
Thibie (TH)	48.93/4.23/110 m	657	10.9	Eutric Cambisol	1991–2003	PEA/WHB/SBT PEA+GRV/WHB/RAD/ SBT/BAR	12 (catch crops, inter-year variation, fertiliser)	High vs. low fertilisation Catch crops vs. bare soil	Harv/1 treat

[a] Average annual precipitation during period of observation.
[b] Average annual temperature during period of observation.
[c] BAR = barley, RAP = oilseed rape, WHB = wheat, GRV = grass vegetation, PEA = pea, RAD = oilseed radish, OAT = oat, RYE = rye, MUS = mustard, POT = potato, MAZ = maize, SBT = sugar beet.
[d] Phen = phenology, treat = treatment, Biom = biomass, Harv = harvest date.

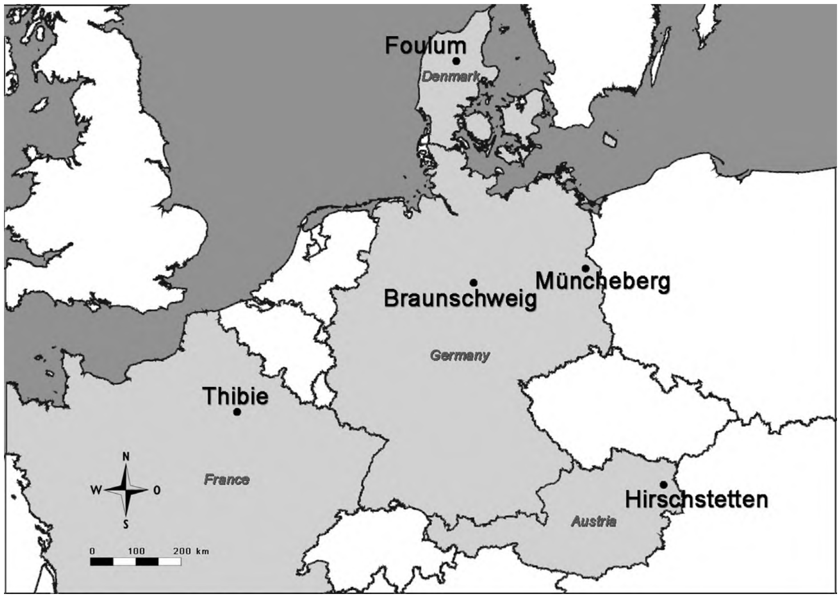

Fig. 1. Locations of the five experimental study sites in Europe.

Table 2
Models applied in this study and the web addresses of their websites.

Model	Abbre-viation	Version	Key Reference	SINGLE/ ROTATION	Web address
DSSAT1	DS	4.6	Jones et al. (2003)	Yes/no	http://dssat.net
HERMES	HE	4.26	Kersebaum (2011); Kersebaum and Nendel (2014)	Yes/yes	http://www.zalf.de/en/forschung/institute/lsa/forschung/oekomod/hermes
MONICA	MO		Nendel et al. (2011)	Yes/yes	http://monica.agrosystem-models.com/
LPJmL	LP		Bondeau et al. (2007)	Yes/no	http://www.pik-potsdam.de/research/projects/lpjweb
Daisy	DA	5.16	Hansen et al. (2012)	Yes/yes	https://code.google.com/p/daisy-model
FASSET	FA		Olesen et al. (2002)	Yes/yes	www.fasset.dk
SPACSYS	SP	5.0	Wu, et al. (2007)	Yes/no	
Theseus	TH		Wegehenkel (2009)	No/yes	Request from mwegehenkel@zalf.de
STICS	ST	8.3.1	Brisson et al. (2003)	Yes/yes	www6.paca.inra.fr/stics_eng/
SIMPLACE, LINTUL2, SoilCN, SLIM	LI	svn3275	Addiscott and Whitmore (1991); Angulo et al. (2013); van Oijen and Lefelaar (2008)	Yes/yes	www.simplace.net
CROPSYST	CR	3.02	Stöckle et al. (2003)	Yes/yes	www.sipeaa.it/ASP/ASP2/CropSyst.asp
SWIM	SW		Krysanova et al. (2000)	No/yes	https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/models/swim
WOFOST	WO	7.1.5	Boogaard et al. (1998); Jylhä et al. (2004); Supit et al. (1994); van Ittersum et al. (2003)	Yes/no	http://www.wofost.wur.nl
DSSAT2	DT	4.5	Jones et al. (2003)	Yes/no	www.icasa.net/dssat
APSIM	AP	7.5	Keating et al. (2003)	No/yes	www.apsim.info

lation was set up once using the specified initial values, and crop growth was subsequently simulated for the period of the rotation. In contrast, the modellers set up the SINGLE mode of simulations by using the initial values to set up and run the first year of simulations and subsequently used estimates of the initial values to calculate the subsequent year's crops separately.

The modelling teams were additionally asked to provide information on how initial values were estimated at the beginning of the growing seasons in the SINGLE mode of calculation. The modellers also provided information on how "experienced" the model was in simulating each crop in terms of the number of seasons it had been applied and calibrated in the past.

2.4. Model calibration

In addition to the input data described above, the modellers were provided with limited data (separately for each site and dependent on the availability of observation data) to perform a minimal calibration of local crop varieties. The data, i.e. the variable subject to calibration for each site, is shown in Table 1. This method followed the idea of a "blind test" in order to mimic modelling practice in the event of scarce data, which is often encountered in regional climate impact studies (Palosuo et al., 2011; Rötter et al., 2012; Asseng et al., 2013; Bassu et al., 2014). The calibration data consisted of key phenological observations (dates of emergence, anthesis and maturity) for one single treatment of the datasets for Foulum and Hirschstetten (Table 1), harvest dates in Thibie, final biomass observations for Müncheberg as well as phenological observations for the first four years at Braunschweig.

After calibration, the modelling teams ran their models for all other years and treatments in the two simulation modes (ROTATION and SINGLE), and the model outputs were gathered for statistical analysis.

2.5. Evaluation of model performance

2.5.1. Crop rotations

Model performance was evaluated by calculating complementary performance indicators, as proposed by Bennett et al. (2013). The selection of indicators enables the magnitude of errors to be quantified and bias to be detected. The following model performance indicators were calculated for each model, site and mode of simulation (ROTATION and SINGLE), and then averaged for each site: mean absolute error (MAE), index of agreement (IOA), percent bias (PBIAS) and root mean square error (RMSE). MAE is calculated as the average of the absolute errors. It provides the magnitude of deviation by ignoring the direction of the deviation. IOA is a standardised measure of the degree of model prediction error, ranging from 0 to 1, with the latter indicating a perfect fit (Willmott, 1982). PBIAS (%) was calculated as:

$$\mathrm{PBIAS} = 100 \times \frac{\Sigma_{i=1}^{n}(S_i - O_i)}{\Sigma_{i=1}^{n} O_i} \quad (1)$$

where S_i is simulated crop yield and O_i observed crop yield at each harvest date. PBIAS measures the tendency of the model to overestimate or underestimate the measured values. An optimal PBIAS value is 0.0; positive values indicate an overestimation and negative values are indicative of an underestimation. Finally, RMSE represents the sample standard deviation of the differences between simulated and observed values. In contrast to MAE, the main drawback of RMSE is that it is sensitive to outliers. Nonetheless, it was calculated here to enable the results to be compared with earlier modelling studies. Further, we partitioned RMSE into its systematic part:

$$\mathrm{RMSE_s} = \sqrt{\frac{1}{n\Sigma_{i=1}^{n}\left(\bar{\bar{S}}_i - O_i\right)^2}} \quad (2)$$

which describes the linear bias, and its unsystematic part, where $\bar{\bar{S}}_i$ is derived from the linear regression between observed and simulated values. The random error was calculated according to Willmott (1982):

$$\mathrm{RMSE_u} = \sqrt{\frac{1}{n\sum_{i=1}^{n}\left(S_i - \bar{\bar{S}}_i\right)^2}} \quad (3)$$

2.5.2. Performance of crop rotation vs. single-year simulations

Student's t-tests were conducted separately on the three main analyses (per site, per crop, per treatment) of this study ($\alpha = 0.05$). We compared the mean of performance indicators in the ROTATION mode of simulation against the mean in SINGLE mode. In the site-specific analysis, we used performance indicators for each site as a pair in the paired t-test. In the crop-specific analysis, we used performance indicators for each crop as pairs in the paired t-test. Finally, in the treatment- specific analysis, we tested each treatment separately, using performance indicators for each model as pairs in the paired t- test.

2.5.3. Crop-specific yield

In order to evaluate the quality of the crop-specific simulation, we calculated performance indices on yield prediction (in tonnes dry matter per hectare; t DM ha^{-1}) for each crop across all models, treatments and sites. Final aboveground biomass was only evaluated for oilseed radish and grass vegetation because biomass was a better proxy for growth in these catch crops. Mustard was excluded from this analysis because no observations were available for this crop. In order to detect any differences between the predictability of crops, the normalised mean absolute error (nMAE) was calculated as follows:

$$\mathrm{nMAE} = \frac{\frac{1}{n}\left(\Sigma_{i=1}^{n}|S_i - O_i|\right)}{\bar{O}_i} \quad (4)$$

where $\bar{O}_i$ is the mean of observations of all datasets and treatments. Here, normalisation is required because mean observed yields vary considerably from crop to crop. In addition, we selected the indicator PBIAS because it measures the average tendency to overpredict or underpredict. We consciously decided against using the index of agreement or modelling efficiency (Nash and Sutcliffe, 1970), since both indicators determine the variance of the simulated and observed dataset. Since the yield simulation of each model for a certain crop type varies around a certain level/mean but observations vary on one level only, the ensemble of simulations will obviously invariably exhibit greater variance than observations.

2.5.4. Management treatments

The capability of models to reproduce a wide range of treatments (Table 1) was tested. Differences between yields of the standard treatment and the above-mentioned treatments were calculated for observations and for the ROTATION and SINGLE modes of simulation. Differences were expressed as percentages of the yields of the standard treatment, and averaged per model. In addition, performance indicators RMSE, MAE and IOA were calculated on the level of the simulations of each model, and averaged per treatment.

3. Results

3.1. Crop rotations

Yield simulations of both ROTATION and SINGLE runs of each model were compared to observations of all seasons, treatments and sites covered by the individual model. Fig. 2 provides an overview of all modelling results per site and crop. Overall, the models provided similar results for ROTATION and SINGLE simulations, and the model results showed higher yield variability than the observed results, especially in Foulum. Notable differences between the observed and simulated yields were detected for several crops, such as oat and pea at Foulum, sugar beet at Müncheberg and Thibie, and wheat and potato at Hirschstetten. The closest match between the observed and simulated mean results was achieved for crops at Braunschweig.

The largest deviations (highest RMSE and MAE values) between simulated and observed yields occurred at Thibie, which was the most diverse dataset, followed by Müncheberg, although the IOA values for these two sites were high (Table 4). In contrast, the models performed best at Braunschweig and Foulum, as shown by performance indicators RMSE and MAE, although the results of these two sites had the lowest IOA values. Here, the low IOA values were due to the fact that IOA evaluates the variance of the observations and simulations such that large variances are favoured. Since the variance of yields in Foulum is very low (no sugar beet), the index of agreement is influenced negatively in this case.

With regard to the crop yield datasets, those from Müncheberg and Thibie were systematically underestimated in ROTATION mode by 11% and 18%, respectively, whereas those from Hirschstetten were overestimated considerably (by 43% mean per site across all models). Systematic errors regularly exceeded unsystematic errors, with the exception of Foulum. This indicates that crop growth processes were reflected successfully by the models, but predictions of crop yields were considerably biased due to the minimal calibration.

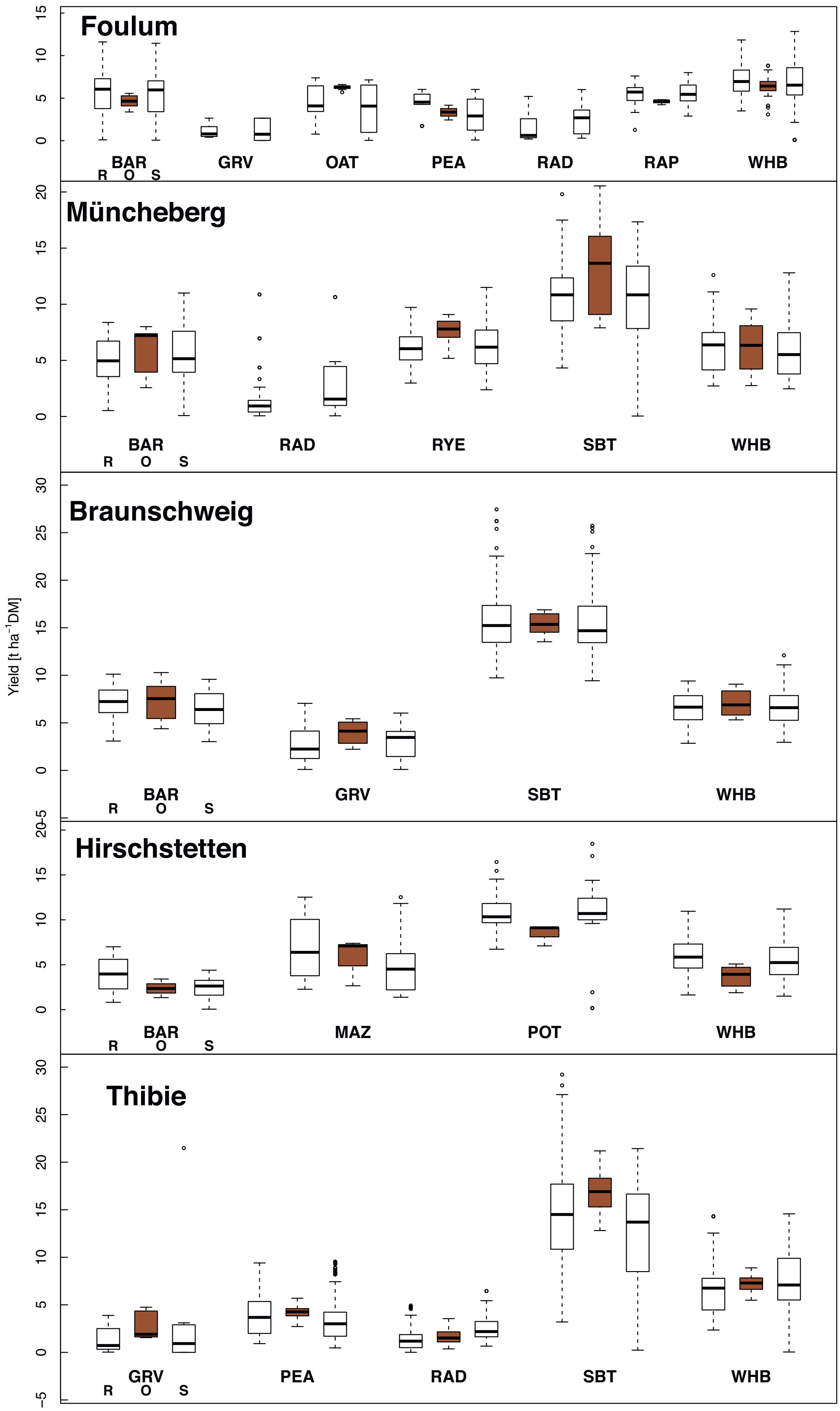

Fig. 2. Yield predictions (R-ROTATION, S-SINGLE) of 15 models and observation (O) of all treatments of the five datasets. Final biomass predictions are shown in GRV and RAD. Boxes of the box-and-whisker plots show the upper and lower quartile of the distribution and the median. The whiskers extend to the most extreme data points (minimum/maximum), which are no further away than 1.5 times the inter-quartile range whereas the circles represent outliers. For crop abbreviations, see Table 1.

Table 3
Crop-specific observations and the abilities of each model, stating the abbreviation of the crop, the number of models able to simulate the crop in multi-year (ROTATION) and single-year (SINGLE) mode, the number of models that predicted this crop for the first time, the number of datasets in which the crop appears and the number of yield observations.

Crop	Code	No. of models ROTATION/SINGLE	No. of models first time	No. of datasets	No. of observations
Maize	MAZ	7/7	2	1	3
Winter wheat	WHB	10/12	1	5	96
Winter barley	BAR	10/11	3	5	37
Rye	RYE	10/9	5	1	12
Oat	OAT	7/7	7	1	8
Sugar beet	SBT	10/9	5	3	64
Potato	POT	7/6	5	1	3
Oilseed radish	RAD	5/4	7	3	42
Pea	PEA	8/9	4	2	52
Grass vegetation	GRV	7/6	5	3	14

Table 4
Model evaluation indices describing the goodness of fit between observed and simulated yields for all sites. The results are given as the mean per site (across models), the best performing model over all sites (best model mean) and the goodness of fit of the mean of predictions by all models (multi-model mean).

Site	RMSE [t ha^{-1} DM]		MAE [t ha^{-1} DM]		IOA		PBIAS [%]	
	ROTATION**	SINGLE	ROTATION	SINGLE	ROTATION*	SINGLE*	ROTATION	SINGLE
FO	2.5 (1.7 + 1.8)	2.8 (1.9 + 1.9)	2.1	2.4	0.45	0.45	+10	−1
MU	3.3 (2.6 + 1.8)	3.2 (2.6 + 1.9)	2.5	2.5	0.71	0.66	−11	−10
BR	2.9 (2.1 + 1.7)	2.5 (1.7 + 1.7)	2.2	1.9	0.87	0.80	−5	−5
HI	3.1 (2.4 + 1.8)	3.5 (2.6 + 2.2)	2.7	2.7	0.66	0.54	+43	+22
TH	4.4 (3.5 + 2.3)	4.3 (3.6 + 1.9)	3.3	3.4	0.78	0.68	−18	−16
Best model mean	2.1(1.2 + 1.6)	2.1(1.3 + 1.4)	1.6	1.7	0.81	0.82	0	3
Multi-model mean	2.2	2.1	1.7	1.6	0.78	0.83		

* p-value < 0.05 for significance of the mean.
** Values in brackets indicate the systematic part plus random error of RMSE.

Table 5
Crop-specific performance indicators describing the accuracy of yield predictions generated by the models involved and the multi-model ensemble.

Crop	RMSE [t ha^{-1} DM]		rRMSE		nMAE		PBIAS [%]	
	ROTATION*	SINGLE*	ROTATION	SINGLE	ROTATION*	SINGLE*	ROTATION	SINGLE
MAZ	3.9	3.6	0.68	0.63	0.53	0.48	25.8	−11.8
WHB	2.6	2.9	0.40	0.44	0.31	0.36	−0.1	2.3
BAR	2.5	2.8	0.51	0.56	0.40	0.45	16.1	9.8
RYE	2.4	2.5	0.34	0.35	0.28	0.29	−18.2	−12.3
OAT	2.3	3.7	0.37	0.59	0.30	0.48	−25.5	−41.9
SBT	6.9	6.7	0.42	0.41	0.32	0.30	−21.1	−23.4
POT	3.7	7.1	0.44	0.84	0.35	0.70	16.7	40.0
RAD	1.4	1.8	0.87	1.11	0.65	0.87	−18.0	−4.0
PEA	2.1	2.3	0.51	0.57	0.41	0.47	−2.5	−16.7
GRV	2.3	3.3	0.67	0.95	0.54	0.56	−27.1	−20.6
Multi model mean	1.8	2.1	0.31	0.35	0.26	0.31		

* p-Value <0.05.

Table 6
Observed (O) and simulated (R-ROTATION, S-SINGLE) treatment effects on crop yield. DM yield change (%) is calculated from yields of a zero-treatment as a reference. Three model performance indicators are included viz. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Index of Agreement (IOA). BR-Braunschweig dataset and TH- Thibie dataset. Note that there were no significant differences in performance indicators between ROTATION and SINGLE.

Treatment	Yield change [%]			RMSE [%]		MAE [%]		IOA	
	OBSERVED	ROTATION	SINGLE	ROTATION	SINGLE	ROTATION	SINGLE	ROTATION	SINGLE
Irrigation	19.6	12.2	8.1	35	31	27	25	0.41	0.51
CO_2	11.3	10.9	12.3	13	15	10	12	0.36	0.28
N (BR)	12.7	13.3	7.5	22	19	18	15	0.37	0.41
N (TH)	2.5	2.8	0.1	16	13	10	9	0.39	0.43
Soil cPhaeo	−51.1	−11.4	−5.1	42	49	38	45	0.33	0.30
Soil Cherno	0.86	−0.1	2.1	28	23	20	15	0.28	0.43
Tillage	−3.7	0.6	−0.1	28	26	18	16	0.10	0.11
Residues	1.5	0.6	−9.7	13	22	9	19	0.15	0.19
Intermediate crop	3.5	0	−2.5	19	17	12	13	0.33	0.33
Best model mean				18	18	14	13	0.37	0.54
Multi model mean				18	19	14	15	0.36	0.40

The simulation accuracy of the single best model resembled the accuracy of the multi-model mean simulations (Tables 4 and 6). One out of three indicators (IOA) exhibited significantly more accurate results when the ROTATION simulation mode was compared to the SINGLE mode (Table 4).

3.2. Crop-specific yields

Each crop type was simulated by at least five (ROTATION) or four (SINGLE) models (Table 3). However, more models provided results for main crops such as wheat and rye, whereas fewer models were capable of simulating intermediate crops such as oilseed radish or grass vegetation. Of the crops simulated, oilseed radish and oat were simulated by seven modelling teams for the first time (Table 3), using proxies for the crop- specific parameter setting. Wheat and barley were present in each dataset and thus grown under varying environmental conditions, whereas crops such as maize, potato, rye and oat were grown at one site each only (Table 3).

RMSE and nMAE showed that, across all crops, the ROTATION mode of simulation resulted in slightly more accurate results than those generated in SINGLE mode (Fig. 3, Table 5). In addition, the multi-model mean again produced good accuracy compared to the results generated by single crops (rRMSE of 0.31 and nMAE of 0.26 in ROTATION mode). With the exception of maize, yields of the main cereals (wheat, barley, rye and oat) were reproduced reasonably well with nMAE < 0.5, meaning errors less than 50% of the mean observed crop yield. Notably, sugar beet (occurring in three datasets and modelled by nine groups in ROTATION mode and ten teams in SINGLE mode) was simulated with a high degree of accuracy, as shown by the two normalised measures of accuracy (nMAE = 0.32 and rRMSE = 0.42), whereas in absolute terms it was the crop with the highest simulation error (RMSE: 6.9 t ha^{-1} DM). Potato exhibited a low deviation in ROTATION and a high deviation in SINGLE simulation mode. In contrast, intermediate crops (oilseed radish and grass vegetation) were generally simulated with a rather low degree of accuracy (nMAR > 0.5). Overall, both ROTATION and SINGLE agreed on bias across all crops with the exception of maize, where ROTATION overestimated (PBIAS > 0) yield and SINGLE underestimated it (PBIAS < 0; Table 5). A notable systematic underestimation of yields was found for oat, sugar beet (in Müncheberg and Thibie) and grass vegetation and, to a lesser extent, for rye, pea and oilseed radish; in contrast, yields were overestimated for potato and barley.

3.3. Management treatments

The observed and simulated treatment effects on crop yields are shown in Table 6, separated by modes of simulation (ROTATION and SINGLE). The DM yield change (in %) is calculated from the yields of a zero-treatment as a reference. The three indicators of simulation quality (RMSE, MAE and IOA) exhibited no significant difference between ROTATION and SINGLE simulation modes for treatment effects.

3.3.1. Irrigation

The observations at Müncheberg indicated a 19.6% mean yield increase due to irrigation (Table 6). However, the mean irrigation effect of the simulation of all models was 12.2% (ROTATION) and 8.1% (SINGLE). Hence, although the positive effect of irrigation on crop yield was simulated by the models, it was underestimated for all crops, particularly for wheat and sugar beet. The rainfed treatments were simulated more accurately than the irrigated treatments.

3.3.2. CO_2

At Braunschweig, an 11.3% mean yield increase was observed due to a 176 ppm increase in atmospheric CO_2 concentration. This effect was captured well by the models (Table 6, Fig. 4). The simulated mean effect was 10.9% (ROTATION) and 12.3% (SINGLE).

3.3.3. Nitrogen

The observed effect of the increased N application rate on yields was 12.7% in Braunschweig and 2.5% in Thibie (Table 6). This effect was captured especially well by the results generated using the ROTATION mode. In contrast, the SINGLE simulation mode underestimated this effect at Braunschweig, generating an effect of about 7%, and approximately 2% at Thibie. The accuracy of simulations was the same for both N application treatments.

3.3.4. Tillage and residues

At Foulum, the observed effect of conventional tillage on the crop yield compared to no tillage was −3.7%. The mean results generated by the models exhibited virtually no effect on yields, although there was a 0.6% simulated yield increase when using the ROTATION mode. The observed effect of residue handling (retention of straw on the field compared to removal of straw) on crop yield was negligible during the investigated period (1.5%). This was confirmed by the ROTATION simulation mode (0.6%), but not by the SINGLE mode of yearly calculations (−9.7%).

3.3.5. Soil

At Hirschstetten, the significantly lower crop yields on the Calcaric Phaeozem soil (−51.1%) compared to yields on the Gleyic Phaeozem soil were captured poorly by the simulations (−11.4% and −5.1% for ROTATION and SINGLE mode, respectively). The observed strong negative effect of the sandy soil was underestimated by the models for all crops, with the largest errors occurring for wheat and barley. In addition, the majority of errors emerged from the simulation of yields at the Calcaric Phaeozem (sandy texture) rather than from the simulation of yields at the Gleyic Phaeozem. Four models showed no effect or even an opposite soil effect on yields; five other models exhibited only very minor effects on yield reduction. Observed yields at the Calcic Chernozem differed insignificantly from the Gleyic Phaeozem (0.9% change), which was reproduced by both modelling modes (ROTATION and SINGLE).

3.3.6. Intermediate crops

The model ensemble failed to capture the slightly positive effect of introducing intermediate crops (oilseed radish and grass vegetation) on yields of the following main crops (+3.5%). The model ensemble yielded effects of 0% and −2.5% for ROTATION and SINGLE, respectively. Only two models in the ROTATION mode were able to reproduce the small positive effect.

4. Discussion

This study compared for the first time the accuracy of fifteen minimally calibrated crop models in simulating yields of various crop rotations. The overall results showed that the simulation of an entire rotation is better at estimating crop yields than single-year simulations. The study also revealed that not all crops, i.e. their yields, are simulated equally well, highlighting focal areas for future field studies and potential improvements to the models. Likewise, treatments, i.e. field management practices, were identified that require attention when represented in the models. Since the study includes a large number of models that simulate a variety of field experiments, it must be noted that individual model errors may cancel each other out (see, e.g. Fig. 2). Hence the conclusion drawn from our analysis refers to the ensemble, and cannot be related to

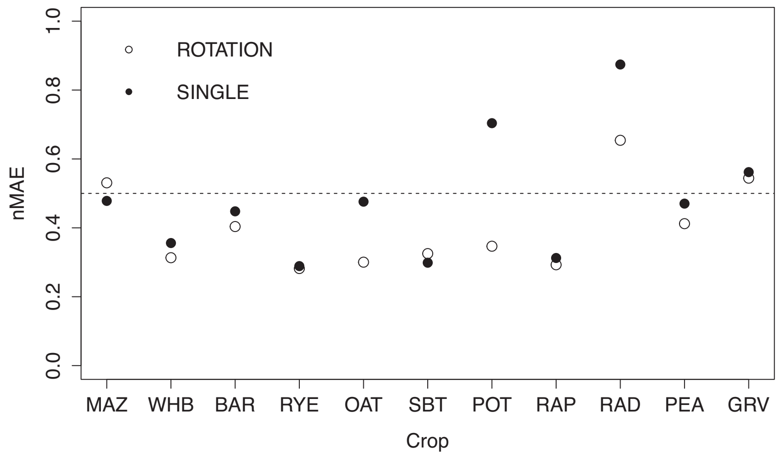

Fig. 3. Crop-specific errors (normalised mean absolute errors) of yield prediction across all models, sites and treatments. Final biomass was predicted in RAD and GRV. The area above the dashed line indicates errors exceeding 50% of mean observed yields.

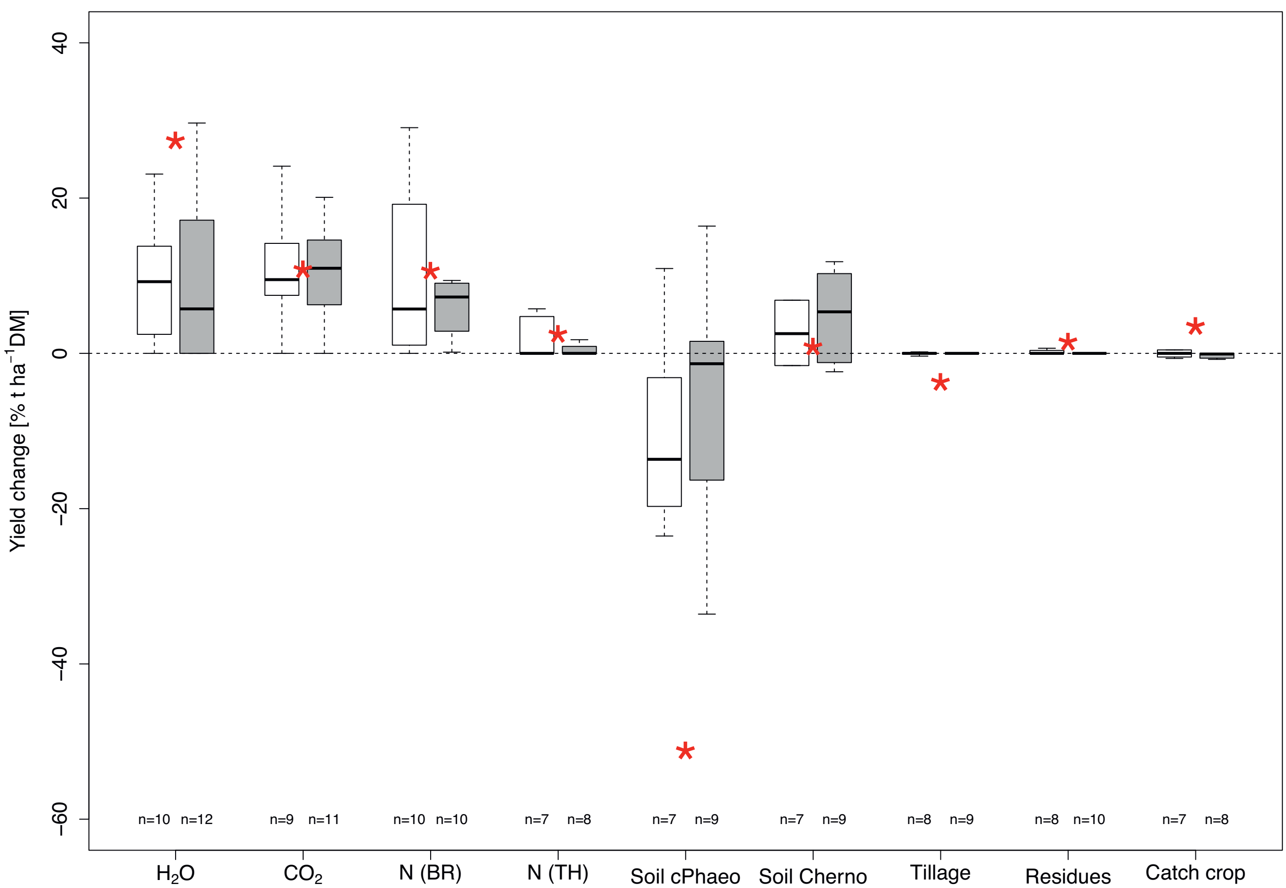

Fig. 4. Treatment effects reproduced by all models and observation (on the basis of percent yield change). H_2O = irrigation, CO_2 = increased CO_2 concentration, N = fertilisation, soil cPhaeo = calcaric phaeozem compared to gleyic phaeozem, soil cherno = calcic chernozem compared to gleyic phaeozem, tillage = no tillage compared to common plough, residues = retention of residues compared to removal of residues, catch crop = catch crop use compared to bare soil. Red stars indicate the median of observations, boxplots show the median response of each model. Boxes of the box-and-whisker plots show the upper and lower quartile of the distribution and the median. The whiskers extend to the minimal/maximal data points. Left bars: ROTATION, right bars: SINGLE.

a single model. The field experiments used in this study were originally designed to address specific questions that were of interest to stakeholders. In practice, crop rotations are often simpler and, under certain circumstances, cause environmental problems.

4.1. Crop rotations

In the present study, the site-specific index of agreement (IOA) across all models varied between 0.45 and 0.87 (per site), and is thus comparable to reported IOA values of 0.4–0.7 per model found in a single-crop study with very limited calibration (Palosuo et al., 2011). However, the IOA values determined are lower than the IOA values of 0.9–0.99 per model reported in a calibrated single-site study (Kersebaum et al., 2007).

Large simulation errors in sugar beet yields were the main driver of the high discrepancies between modelled and observed yields for the Thibie dataset. Since five modelling teams parameterised sugar beet for the first time using proxies such as potato crop parameters, these models were responsible for the wide range of simulated yields (between 0.2 and 29.2 $t\,ha^{-1}$) differing greatly from observed yields (12.8–21.2 $t\,ha^{-1}$). Thus, we identified the need to accurately parameterise sugar beet in several models in order to improve the simulation results and to meet study demands, since this crop is agro-economically and environmentally important in large parts of Europe. The best modelling results were achieved for the Braunschweig dataset, which contained only one of the less accurately simulated crops (grass vegetation), homogenous soil conditions and a relatively small number of treatments. Uncertainties between sites were therefore driven mainly by the ability of the models to capture minor crops and soil conditions.

4.2. Crop rotation vs. single-year simulation

Simulating continuous crop rotations has the potential to increase the precision of yield estimates compared to simulations of crop growth on a single-year basis. This is due to the fact that soil water, soil organic matter and nutrient conditions can be predicted more accurately in rotation, as this mode continuously updates soil conditions daily, and does not assume any soil conditions at the beginning of the growing season, as is the case in single-year simulations.

In the five selected field experiments, crops were grown under nutrient-rich conditions and generally adequate water supply, especially during the onset of the growing period, when soils were usually at field capacity. Thus, most crops began to grow under close-to-optimal environmental conditions. Under these conditions, the presumed carry-over effects, such as water and nutrient savings and transfer from the previous crop (reproduced in ROTATION mode only), could only affect the growth of the next crop to a low degree. While rotation effects under nutrient-limited conditions turned out to be significant (Smith et al., 2008), studies for wheat and maize, for instance, showed that the positive rotation effect decreased with increasing fertilisation levels (Angus et al., 2001; Berzsenyi et al., 2000; Sieling et al., 2005) and better water supply (Nevens and Reheul, 2001). Similar results were obtained by Teixeira et al. (in press) for New Zealand where consideration of continuous rotation was more important under water- and nitrogen-limited conditions. From the aspect of crop yields, at least, there was therefore limited potential for ROTATION simulation, which explicitly takes into account carry-over effects, to achieve more superior predictions than SINGLE simulations.

Nevertheless, we found that continuous simulations performed slightly better when the two methods of simulating yield (ROTATION vs. SINGLE) were compared site by site, crop by crop, and treatment by treatment (Tables 4–6). This means that the conditions of water and nutrients in the soil, and thus the emerging yield, were simulated better across all models by the continuous mode of simulation than the single-year mode. In the latter case, most modellers manually or automatically estimated soil water content and nutrients less accurately at the beginning of each growing season. We expect to see greater contrast between continuous and single-year simulations (i) when other model outputs such as water and matter fluxes are investigated; (ii) when long-term effects (>20 years) are studied; (iii) when all crops are well parameterised; and (iv) when low-fertilised/water-limited crop datasets are investigated. Although the beneficial effect of continuous simulation on crop yield results was not pronounced, the general need for continuous simulations to assess ecosystem services from cropping systems cannot be ignored.

4.3. Crop-specific yields

Considering that the calibration data provided for this modelling study was minimal, the yield simulations of the widely grown and simulated cereals wheat, barley and rye were good (RMSE between 2.4 and 2.6 $t\,ha^{-1}$ DM in the ROTATION mode). Errors are comparable with reported errors under similar conditions (Beaudoin et al., 2008). This high level of accuracy is probably due to the great deal of experience that modellers have with these crops since the development of the first crop growth simulation models in the late 1960s (e.g. Wit, 1965). In contrast, the low simulation quality in maize yields is due to the fact that modellers have little experience in simulating the wide range of maize varieties (FAO, 2015). Maize parameterisation in European models is often poor, possibly due to the calibration of crop models with observations (a) from nearby field sites only, (b) under ample water and nutrient supply, and (c) generated decades ago with lower yield potentials (Reidsma et al., 2009; Manevski et al., 2014).

Nonetheless, the RMSE of 3.9 $t\,ha^{-1}$ DM in maize is consistent with findings by Bassu et al. (2014), who compared maize crop models. Furthermore, the results of yield simulations in maize are probably biased because this crop was grown in Hirschstetten only, where yield was measured only once on three different soils. As shown below, most of these models failed to reproduce crop growth on these soils.

Yields of intermediate crops (oilseed radish and grass vegetation) were generally reproduced with a lower degree of accuracy than main crop yields. According to information provided by the modelling teams, this is mainly due to their lack of experience in dealing with these crops. This is explained by their low economic value and the fact that they are not considered as highly influential on the growth of main crops. Only about half of the modelling teams were experienced in specifically simulating oilseed radish and grass vegetation. Those who simulated intermediate crops reported difficulties in reproducing the emergence date of these crops under summer conditions.

4.4. Management treatments

4.4.1. CO_2/nitrogen/tillage/residues

Thanks to the provision of calibration data concerning both ambient and increased CO_2 concentration, the model results were in good agreement with the measured effect of increased CO_2 concentration on yield in the FACE experiment at Braunschweig. Thus, by applying a 15-model ensemble (where most models used the approach of radiation use efficiency), we were able to reproduce the results generated in other modelling studies that simulated the effect similarly well (Kartschall et al., 1995; Nendel et al., 2009; Tubiello et al., 1999).

The fertilisation experiments at Braunschweig and Thibie reduced the "normal" dose of N fertilisation to 50% and 69%, respectively. In both datasets, the effect of fertiliser was generally well reproduced, albeit underestimated slightly by 2–3%. This indicates that the models successfully reflect the effects of varying crop N nutrition. Notably, in the case of Braunschweig, the growth response to the complex interaction of CO_2, available soil water and N (see review in Wu and Kersebaum, 2008) was simulated well.

According to field data from Foulum, reducing tillage and remaining harvest residuals on the field had no short-term effect on yield. This was reproduced by the models. It is worth noting that several models did not specifically simulate ploughing, whereas those able to simulate ploughing (e.g. DAISY) do not include any effects of ploughing on the soil organic matter turnover, which has direct implications for the simulation of soil processes in ROTATION over time. Similar observations were found under comparable climatic conditions at two sites in Denmark and Germany (Deike et al., 2008), whereas short-term tillage effects have been reported under warmer and drier climatic conditions (Fischer et al., 2002; López-Bellido and López-Bellido, 2001).

Thus, we conclude that simulated responses to fertiliser and residues are described well in the models.

4.4.2. Irrigation/soil/intermediate crop

In general, dynamic crop simulation models are able to successfully reproduce irrigation effects because they were developed in order to respond to key environmental drivers such as precipitation, temperature and radiation. However, the model ensemble we evaluated strongly underestimated the effect of irrigation for all crops in the Müncheberg dataset. In particular, simulations of the irrigated treatment exhibited a high discrepancy to observations because half of the models demonstrated no or virtually no effect of irrigation. We found that the erroneous reproduction of the soil water dynamics was responsible for the mismatch. However, when comparing multi-model simulations of wheat in the same dataset, Palosuo et al. (2011) reported similarly underestimated yields due to a poor representation of the soil water dynamics. These mismatches cannot be explained by the various water limitation approaches taken by the models. Instead, it seems that water dynamics driven by light spring and/or summer drought were not well reproduced by the models for the sandy soil of Müncheberg. Four models used implausible values of field capacity; for two models, the underestimation of soil water content was related to the underestimation of yields; and one model simulated implausibly high soil water contents.

Similarly, in Hirschstetten most models failed to reproduce the very low yields of all crops on the Calcaric Phaeozem. For this dataset, minimal calibration data was provided for a rotation on Gleyic Phaeozem. When applying the same rotation to the shallow and sandy Calcaric Phaeozem, which contains >50% gravel from below 95 cm soil depth, the limited availability of measured data for a proper soil parameter calibration turns out to be responsible for the reduced response. In particular, the reduced water holding capacity provoked by the high stone content was not considered by most models. This is confirmed by the models' inaccurate soil water simulations in Hirschstetten (Supplementary B). Of the four datasets where soil water measurements were available, the simulation accuracy of this variable was lowest at Hirschstetten. In addition, the interpretation of models differed regarding rooting depth. While some models used an exponential distribution function over depth with decreasing uptake from the subsoil, others interpreted depth as an "effective rooting depth", allowing full extraction of water down to 2 m, which was the maximum rooting depth. In view of the sandy soil and the high stone content in the profile, this seems unrealistic, and led to a lower response to dry periods. Examples from precision agriculture show that models usually respond well to differences in soil conditions if rooting depth is properly considered (Kersebaum et al., 2005).

In the experiment, the effect of growing intermediate crops (oilseed radish and grass vegetation) instead of leaving the soil bare during the non-growing season resulted in higher quantities of biomass generated by the following sugar beet and winter wheat (Constantin et al., 2010). This was not reproduced by the model results (ROTATION and SINGLE mode). There may be two main reasons for the lack of model response: (a) the generally low accuracy of biomass simulation for intermediate crops (Fig. 3) due to limited experience and minimal calibration data, and (b) uncertainty in the simulated N release from mineralisation of soil organic matter and from the decomposition of crop residues, both of which contribute to the N supply to the main crop under N-limited conditions. Under the conditions of minimal calibration, a poor model performance does not coercively imply model deficiencies, as modellers were forced to assume certain conditions and processes. Thus, the question as to which model process requires improvement may better be answered after full calibration of each individual model.

5. Conclusion

Past model inter-comparisons focused on single crops only and usually on single years when drawing conclusions on the uncertainty of applying models at sites for which they had not been calibrated. In this study, we accommodate the fact that crop production is generally driven by crop rotations, where preceding crops influence the growth of the following crop due to a number of processes. As such, we raise the complexity of crop modelling to a higher system level.

The results suggest that it is a matter of urgency to model crop rotations in order to evaluate the resilience of cropping systems and their contribution to ecosystem services under changing climate conditions. However, none of the models involved was capable of reliably simulating yields of all crop species in all datasets when data for calibration was sparse. Hence, the multi-model ensemble approach minimised the error arising from simulations of single models. As hypothesised, the continuous simulation of multi-year crop rotations slightly outperformed simulations of single years with regard to crop yield. Finally, a better functional understanding and parameterisation of intermediate crops is required, especially in order to reproduce their effects on main crops.

Based on the results we obtained and the authors' expert knowledge, we propose taking the following steps in priority order when addressing future challenges in crop rotation modelling:

(1) Knowledge of carry-over effects in crop rotations remains sparse. We suggest a literature review of the various effects of the preceding crop on yields in the following crop, as well as of soil water and nutrient balances from an agro-ecosystem modeller's perspective.
(2) High-quality experimental data (potential growth and water-/nutrient-limited growth) is required as the "backbone" for developing and improving models. We identified a high demand for crop rotation datasets including measurements of soil conditions at high temporal resolution and for datasets of crops that have so far been investigated inadequately (e.g. rape seed, radish, sugar beet, oat, potato). Kersebaum et al. (2014) recently stressed the need for consistent datasets in order to improve models.
(3) We propose the careful review and improvement of existing models towards the continuous simulation of multi-year crop growth (rotations) because technical limitations continue to exist for some models.
(4) More specifically, the following processes were identified as seeming to be represented inadequately in many models and therefore in need of improvement/implementation for a sound representation of crop rotations and their treatments: N release from mineralisation of residues, effects of tillage, dynamics of soil organic matter, parameterisation of under-studied crops, low temperature and frost effects on intermediate crops.

Acknowledgements

The present study was carried out in the context of CropM within the FACCE-MACSUR knowledge hub. We thank the experimenters involved in the Thibie experimental dataset: ARVALIS Institut du Végétal (G. Briffaux, G. Aubrion) and INRA (J. Duval). KCK, CK, CN, MW and WM acknowledge financial support from the FACCE MACSUR project (2812ERA147), funded by the German Federal Ministry of Food and Agriculture (BMEL). CN received additional support by BMBF via the CARBIOCIAL research project (01LL0902 M). CM, KW and TC acknowledge financial support from the FACCE MACSUR project (031A103B), funded by the German Federal Ministry of Education and Research (BMBF). CM acknowledges financial support from the KULUNDA project (01LL0905L), funded by the BMBF. TP and RPR were supported by the NORFASYS project, funded by the Academy of Finland (decision no. 268,277) and MTT strategic projects MACSUR and MODAGS, funded by the Finnish Ministry of Agriculture and Forestry. HH, TG and FE acknowledge financial support from the FACCE MACSUR project (2812ERA115), funded by the German Federal Ministry of Food and Agriculture (BMEL). RM and HJW acknowledge financial support by BMEL. JEO, KM, IÖ, SR and BS were funded from the FACCE MACSUR project by Innovation Fund Denmark. CMAH, NB, JC, IGCA, ML, BM, DRW, FR were funded from the FACCE MACSUR project by INRA ACCAF Metaprogramme.

References

Addiscott, T., Whitmore, A., 1991. Simulation of solute leaching in soils of differing permeabilities. Soil Use Manage. 7 (2), 94–102.
Angulo, C., et al., 2013. Characteristic ‘fingerprints’ of crop model responses to weather input data at different spatial resolutions. Eur. J. Agron. 49, 104–114.
Angus, J., Kirkegaard, J., Peoples, M., 2001. Rotation, sequence and phase: research on crop and pasture systems, science and technology: delivering results for agriculture? Proceedings of the 10th Australian Agronomy Conference.
Angus, J. et al. 2011. A review of break- crop benefits of brassicas. 17th Australian Research Assembly on Brassicas: 15–17.
Askegaard, M., Eriksen, J., 2008. Residual effect and leaching of N and K in cropping systems with clover and ryegrass catch crops on a coarse sand. Agric. Ecosyst. Environ. 123, 99–108, 1).
Asseng, S., et al., 2014. Rising temperatures reduce global wheat production. Nat. Clim. Change 5 (2), 143–147.
Asseng, S., et al., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Change 3 (9), 827–832.
Bassu, S., et al., 2014. How do various maize crop models vary in their responses to climate change factors. Global Change Biol. 20 (7), 2301–2320.
Beaudoin, N., Launay, M., Sauboua, E., Ponsardin, G., Mary, B., 2008. Evaluation of the soil crop model STICS over 8 years against the on farm database of Bruyères catchment. Eur. J. Agron. 29 (1), 46–57.
Bennett, N.D., et al., 2013. Characterising performance of environmental models. Environ. Model. Software 40, 1–20.
Berntsen, J., Petersen, B.M., Olesen, J.E., 2006. Simulating trends in crop yield and soil carbon in a long-term experiment—effects of rising CO2, N deposition and improved cultivation. Plant Soil 287 (1–2), 235–245.
Berzsenyi, Z., Győrffy, B., Lap, D., 2000. Effect of crop rotation and fertilisation on maize and wheat yields and yield stability in a long-term experiment. Eur. J. Agron. 13 (2), 225–244.
Blombäck, K., Eckersten, H., Lewan, E., Aronsson, H., 2003. Simulations of soil carbon and nitrogen dynamics during seven years in a catch crop experiment. Agric. Syst. 76 (1), 95–114.
Boardman, J., Favis- Mortlock, D.T., 2014. The significance of drilling date and crop cover with reference to soil erosion by water, with implications for mitigating erosion on agricultural land in South East England. Soil Use Manage. 30 (1), 40–47.
Bondeau, A., et al., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Global Change Biol. 13 (3), 679–706.
Boogaard, H., Van Diepen, C., Rötter, R., Cabrera, J. and Van Laar, H. 1998. WOFOST 7.1: user's guide for the WOFOST 7.1 crop growth simulation model and WOFOST Control Center 1.5. DLO Winand Staring Centre Wageningen.
Brisson, N., et al., 2003. An overview of the crop model STICS. Eur. J. Agron. 18 (3), 309–332.
Challinor, A., et al., 2014. A meta-analysis of crop yield under climate change and adaptation. Nat. Clim. Change 4 (4), 287–291.
Constantin, J., et al., 2010. Effects of catch crops, no till and reduced nitrogen fertilization on nitrogen leaching and balance in three long-term experiments. Agric. Ecosyst. Environ. 135 (4), 268–278.
Deike, S., Pallutt, B., Melander, B., Strassemeyer, J., Christen, O., 2008. Long-term productivity and environmental effects of arable farming as affected by crop rotation, soil tillage intensity and strategy of pesticide use: a case-study of two long-term field experiments in Germany and Denmark. Eur. J. Agron. 29 (4), 191–199.
Eitzinger, J., Trnka, M., Hösch, J., Žalud, Z., Dubrovský, M., 2004. Comparison of CERES, WOFOST and SWAP models in simulating soil water content during growing season under different soil conditions. Ecol. Model. 171 (3), 223–246.
European Commission, 2010. Environmental impacts of different crop rotations within the European Union. Final report BioIntelligence service. (http://ec.europa.eu/environment/agriculture/pdf/BIO_crop_rotations%20final%20report_rev%20executive%20summary_.pdf).
European Commission, 2011. Greening – Results of partial analysis on impact on farm income using FADN, Annex 2D, Impact assessment – Common Agricultural Policy towards 2020. Staff Working Paper, Brussels, 2011.
Ewert, F., et al., 2014. 2014. Crop modelling for integrated assessment of risk to food production from climate change. Environ. Model. Software, http://dx.doi.org/10.1016/j.envsoft.2014.12.003, in press.
FAO, 2015. FAO – Water Development and Management Unit – Crop Water Information: Maize.
Fischer, R., Santiveri, F., Vidal, I., 2002. Crop rotation, tillage and crop residue management for wheat and maize in the sub-humid tropical highlands: II. Maize and system performance. Field Crops Res. 79 (2), 123–137.
Hansen, Søren et al., 2012. Daisy: model use, calibration, and validation. Transactions of the ASABE 55.4: 1315–1333.
Hlavinka, P., et al., 2013. Modelling of yields and soil nitrogen dynamics for crop rotations by HERMES under different climate and soil conditions in the Czech Republic. J. Agric. Sci. 152 (1), 1–17.
Hauck, J., Schleyer, C., Winkler, K.J., Maes, J., 2014. Shades of greening: reviewing the impact of the new EU agricultural policy on ecosystem services. Change Adaptation Socioecol. Syst. 1, 51–62.
Jones, J.W., et al., 2003. The DSSAT cropping system model. Eur. J. Agron. 18 (3), 235–265.
Jylhä, K., Tuomenvirta, H., Ruosteenoja, K., 2004. Climate change projections for Finland during the 21st century. Boreal Environ. Res. 9 (2), 127–152.
Kartschall, T., et al., 1995. A simulation of phenology, growth, carbon dioxide exchange and yields under ambient atmosphere and free-air carbon dioxide enrichment (FACE) Maricopa, Arizona, for wheat. J. Biogeography 22, 611–622.
Kay, B., 1990. Rates of change of soil structure under different cropping systems. Advances in Soil Science, 12. Springer, pp. 1–52.
Keating, B.A., et al., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18 (3), 267–288.
Kersebaum, K., 2011. Special features of the HERMES model and additional procedures for parameterization, calibration, validation, and applications. Methods of Introducing System Models into Agricultural Research: 65–94.
Kersebaum, K., Beblik, A.J. 2001. Performance of a nitrogen dynamics model applied to evaluate agricultural management practices.
Kersebaum, K.C., Boote, K.J., Jorgenson, J.S., Nendel, C., Bindi, M., Frühauf, C., Gaiser, T., Hoogenboom, G., Kollas, C., Olesen, J.E., Rötter, R.P., Ruget, F., Thorburn, P., Trnka, M., Wegehenkel, M., 2015. Analysis and classification of data sets for calibration and validation of agro-ecosystem models. Env. Model. Software, http://dx.doi.org/10.1016/j.envsoft.2015.05.009, in press.
Kersebaum, K., Nendel, C., 2014. Site-specific impacts of climate change on wheat production across regions of Germany using different CO_2 response functions. Eur. J. Agron. 52, 22–32.
Kersebaum, K.C., Hecker, J.-M., Mirschel, W., Wegehenkel, M., 2007. Modelling water and nutrient dynamics in soil?crop systems: a comparison of simulation models applied on common data sets. In: Modelling Water and Nutrient Dynamics in Soil–Crop Systems. Springer, pp. 1–17.
Kersebaum, K.C., Lorenz, K., Reuter, H.I., Schwarz, J., Wegehenkel, M., Wendroth, O., 2005. Operational use of agro-meteorological data and GIS to derive site specific nitrogen fertilizer recommendations based on the simulation of soil and crop growth processes. Phys. Chem. Earth 30 (1–3), 59–67.
Kirkegaard, J., Christen, O., Krupinsky, J., Layzell, D., 2008. Break crop benefits in temperate wheat production. Field Crops Res. 107 (3), 185–195.
Kovács, G., Németh, T., Ritchie, J., 1995. Testing simulation models for the assessment of crop production and nitrate leaching in Hungary. Agric. Syst. 49 (4), 385–397.
Krysanova, F., Wechsung, J., Arnold, R., Srinivasan, J. and Williams, 2000. SWIM (Soil and Water Integrated Model), User Manual.
Li, C., Frolking, S., Harriss, R., 1994. Modeling carbon biogeochemistry in agricultural soils. Global Biogeochem. Cycles 8 (3), 237–254.
Li, T., et al., 2015. Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Global Change Biol. 21 (3), 1328–1341.
Lin, B.B., 2011. Resilience in agriculture through crop diversification: adaptive management for environmental change. Bioscience 61, 183–193.
López-Bellido, R., López-Bellido, L., 2001. Efficiency of nitrogen in wheat under Mediterranean conditions: effect of tillage, crop rotation and N fertilization. Field Crops Res. 71 (1), 31–46.

Manevski, K., et al., 2014. Reduced nitrogen leaching by intercropping maize with red fescue on sandy soils in North Europe: a combined field and modeling study. Plant Soil 388 (1), 67–85.
Martre, P., et al., 2015. Multimodel ensembles of wheat growth: many models are better than one. Global Change Biol. 21 (2), 911–925.
Metzger, M., Bunce, R., Jongman, R., Mücher, C., Watkins, J., 2005. A climatic stratification of the environment of Europe. Global Ecol.Biogeography 14 (6), 549–563.
Mirschel, W., et al., 2007. Müncheberg field trial data set for agro-ecosystem model validation. In: Modelling Water and Nutrient Dynamics in Soil–Crop Systems. Springer, pp. 219–243.
Munkholm, L.J., Hansen, E.M., Olesen, J.E., 2008. The effect of tillage intensity on soil structure and winter wheat root/shoot growth. Soil Use Manage. 24 (4), 392–400.
Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I—a discussion of principles. J. Hydrol. 10 (3), 282–290.
Nendel, C., et al., 2011. The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. Ecol. Model. 222 (9), 1614–1625.
Nendel, C., et al., 2009. Testing different CO_2 response algorithms against a FACE crop rotation experiment. NJAS-Wagen. J. Life Sci. 57 (1), 17–25.
Nendel, C., et al., 2013. The performance of the EU-Rotate_N model in predicting the growth and nitrogen uptake of rotations of field vegetable crops in a Mediterranean environment. J. Agric. Sci. 151 (04), 538–555.
Nevens, F., Reheul, D., 2001. Crop rotation versus monoculture; yield, N yield and ear fraction of silage maize at different levels of mineral N fertilization. NJAS-Wagen. J. Life Sci. 49 (4), 405–425.
Olesen, J.E., et al., 2002. Comparison of methods for simulating effects of nitrogen on green area index and dry matter growth in winter wheat. Field Crops Res. 74 (2), 131–149.
Olesen, J.E., et al., 2011. Impacts and adaptation of European crop production systems to climate change. Eur. J. Agron. 34 (2), 96–112.
Palosuo, T., et al., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. Eur. J. Agron. 35 (3), 103–114.
Post, J., et al., 2007. Evaluation of water and nutrient dynamics in soil?crop systems using the eco-hydrological catchment model SWIM. In: Modelling Water and Nutrient Dynamics in Soil–Crop Systems. Springer, pp. 129–146.
Reidsma, P., Ewert, F., Boogaard, H., van Diepen, K., 2009. Regional crop modelling in Europe: the impact of climatic conditions and farm characteristics on maize yields. Agric. Syst. 100 (1), 51–60.
Rosenzweig, C., Wilbanks, T.J., 2010. The state of climate change vulnerability, impacts, and adaptation research: strengthening knowledge base and community. Clim. Change 100 (1), 103–106.
Rötter, R.P., Carter, T.R., Olesen, J.E., 2011. Crop–climate models need an overhaul. Nat. Clim. Change 1, 175–177.
Rötter, R.P., et al., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: a comparison of nine crop models. Field Crops Res. 133, 23–36.
Salado-Navarro, L.R., Sinclair, T.R., 2009. Crop rotations in Argentina: analysis of water balance and yield using crop models. Agric. Syst. 102 (1), 11–16.
Sieling, K., Stahl, C., Winkelmann, C., Christen, O., 2005. Growth and yield of winter wheat in the first 3 years of a monoculture under varying N fertilization in NW Germany. Eur. J. Agron. 22 (1), 71–84.
Simoes, M.P., Belo, A.F., Pinto-Cruz, C., Pinheiro, A.C., 2014. Natural vegetation management to conserve biodiversity and soil water in olive orchards. Span. J. Agric. Res. 12 (3), 633–643.
Smith, R.G., Gross, K.L., Robertson, G.P., 2008. Effects of crop diversity on agroecosystem function: crop yield response. Ecosystems 11, 355–366.
Stevenson, F., Van Kessel, C., 1996. A landscape-scale assessment of the nitrogen and non-nitrogen rotation benefits of pea. Soil Sci. Soc. Am. J. 60 (6), 1797–1805.
Stöckle, C.O., Donatelli, M., Nelson, R., 2003. CropSyst, a cropping systems simulation model. Eur. J. Agron. 18 (3), 289–307.
Teixeira, E.I., Brown, H.E., Sharp, J.M., Meenken, E.D., Ewert, F. In press. Evaluating methods to simulate crop rotations for climate impact assessments—a case study on the Canterbury plains of New Zealand. Environmental Modelling & Software. In press. doi:10.1016/j.envsoft.2015.05.012.
Uthes, S., Matzdorf, B., Müller, K., Kächele, H., 2010. Spatial targeting of agri-environmental measures—cost-effectiveness and distributional consequences. Environ. Manage. 46, 494–509.
Supit, I., Hooijer, A. and Van Diepen, C. 1994. System description of the Wofost 6.0 crop simulation model implemented in CGMS. Joint research centre; European Commission.
Tubiello, F.N., et al., 1999. Testing CERES-wheat with free-air carbon dioxide enrichment (FACE) experiment data: CO_2 and water interactions. Agron. J. 91 (2), 247–255.
van Ittersum, M.K., et al., 2003. On approaches and applications of the Wageningen crop models. Eur. J. Agron. 18 (3), 201–234.
van Oijen, M., Lefelaar, P., 2008. Lintul-2: Water Limited Crop Growth: A Simple General Crop Growth Model for Water-limited Growing Conditions. Waageningen University, the Netherlands.
Wallach, D., Makowski, D., Jones, J.W., 2006. Working with Dynamic Crop Models. Elsevier.
Wegehenkel, M., 2009. Test of a modelling system for simulating water balances and plant growth using various different complex approaches. Ecol. Model. 129, 39–64.
Weigel, H.-J., Manderscheid, R., 2012. Crop growth responses to free air CO_2 enrichment and nitrogen fertilization: rotating barley, ryegrass, sugar beet and wheat. Eur. J. Agron. 43, 97–107.
White, J.W., Hoogenboom, G., Kimball, B.A., Wall, G.W., 2011. Methodologies for simulating impacts of climate change on crop production. Field Crops Res. 124 (3), 357–368.
Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteorol. Soc. 63 (11), 1309–1313.
Wit, C.T., 1965. Photosynthesis of Leaf Canopies. Centre for Agricultural Publications and Documentation.
Wu, L. and Kersebaum, K.C. 2008. Modeling water and nitrogen interaction responses and their consequences in crop models. Response of Crops to Limited Water: Understanding and Modeling Water Stress Effects on Plant Growth Processes: 215–249.
Wu, L., McGechan, M.B., McRoberts, N., Baddeley, J.A., Watson, C.A., 2007. SPACSYS: integration of a 3D root architecture component to carbon, nitrogen and water cycling—model description. Ecol. Model. 200, 343–359.