Version définitive du manuscrit publié dans / Final version of the manuscript published in : European Journal of Agronomy, 2011, in press , DOI: 10.1016/j.eja.2011.05.

Simulation of winter wheat yield and its variability in different climates of Europe. A comparison of eight crop growth models

Authors: Taru Palosuo[1], Kurt Christian Kersebaum[2], Carlos Angulo[3], Petr Hlavinka[4], Marco Moriondo[5], Jørgen E. Olesen[6], Ravi H. Patil[6], Françoise Ruget[7], Christian Rumbaur[3,I], Jozef Takáč[8], Miroslav Trnka[4], Marco Bindi[9], Barış Çaldağ[10], Frank Ewert[3], Roberto Ferrise[5], Wilfried Mirschel[2], Levent Şaylan[10], Bernard Šiška[11], Reimund Rötter[1 *]

[1] MTT Agrifood Research Finland, Lönnrotinkatu 5, 50100 Mikkeli, Finland, taru.palosuo@mtt.fi, reimund.rotter@mtt.fi

[2] Leibniz-Centre for Agricultural Landscape Research (ZALF), Institute of Landscape Systems Analysis, Eberswalder Str. 84, 15374 Müncheberg, Germany, ckersebaum@zalf.de, wmirschel@zalf.de

[3] University of Bonn, Institute of Crop Science and Resource Conservation, Katzenburgweg 5, D-53115 Bonn, Germany, klav@uni-bonn.de, frank.ewert@uni-bonn.de, christian.rumbaur@uni-hohenheim.de

[4] Institute of Agrosystems and Bioclimatology, Mendel University in Brno, Zemedelska 1, Brno, 613 00, Czech Republic, mirek_trnka@yahoo.com, phlavinka@centrum.cz

[5] National Research Council of Italy, IBIMET-CNR, Institute of Biometeorology, via Caproni 8, 50145, Florence, Italy, marco.moriondo@unifi.it, roberto.ferrise@unifi.it

[6] Department of Agroecology and Environment, Aarhus University, DK-8830 Tjele, Denmark, Ravi.Patil@agrsci.dk, JorgenE.Olesen@agrsci.dk

[7] INRA, UMR 1114 Environnement et Agronomie, F-84000 Avignon, France, ruget@avignon.inra.fr

[I] Hohenheim University, International Research Training Group (769), Schwerzstr. 31, 70599 Stuttgart, Germany

[8] Soil Science and Conservation Research Institute, Gagarinova 10, 827 13 Bratislava, Slovak Republic, j.takac@vupop.sk

[9] University of Florence, DIPSA, Department of Plant, Soil and Environmental Science, Piazzale delle Cascine 18, 50144, Florence, Italy, marco.bindi@unifi.it

[10] Istanbul Technical University, Faculty of Aeronautics and Astronautics, Dpt. of Meteorology, 34469, Maslak, Istanbul, Turkey, saylan@itu.edu.tr, caldagb@itu.edu.tr

[11] Department of Biometeorology and Hydrology, Slovak University of Agriculture in Nitra, Hospodárska 7, 949 01 Nitra, Slovak Republic, bernard.siska@uniag.sk

* corresponding author (phone +358 40 353 4506; fax +358 15 226 578, reimund.rotter@mtt.fi)

Abstract

We compared the performance of eight widely used, easily accessible and well-documented crop growth simulation models (APES, CROPSYST, DAISY, DSSAT, FASSET, HERMES, STICS and WOFOST) for winter wheat (*Triticum aestivum* L.) during 49 growing seasons at eight sites in northwestern, Central and southeastern Europe. The aim was to examine how different process-based crop models perform at the field scale when provided with a limited set of information for model calibration and simulation, reflecting the typical use of models for large-scale applications, and to present the uncertainties related to this type of model application. Data used in the simulations consisted of daily weather statistics, information on soil properties, information on crop phenology for each cultivar, and basic crop and soil management information.

Our results showed that none of the models perfectly reproduced recorded observations at all sites and in all years, and none could unequivocally be labelled robust and accurate in terms of yield prediction across different environments and crop cultivars with only minimum calibration. The best performance regarding yield estimation was for DAISY and DSSAT, for which the RMSE values were lowest (1428 and 1603 kg ha^{-1}) and the index of agreement (0.71 and 0.74) highest. CROPSYST systematically underestimated yields (MBE -1186 kg ha^{-1}), whereas HERMES, STICS and WOFOST clearly overestimated them (MBE 1174, 1272 and 1213 kg ha^{-1}, respectively). APES, DAISY, HERMES, STICS and WOFOST furnished high total above-ground biomass estimates, whereas CROPSYST, DSSAT and FASSET provided low total above-ground estimates. Consequently, DSSAT and FASSET produced very high harvest index values, followed by HERMES and WOFOST. APES and DAISY, on the other hand, returned low harvest index values. In spite of phenological observations being provided, the calibration results for wheat phenology, i.e. estimated dates of anthesis and maturity, were surprisingly variable, with the largest RMSE for anthesis being generated by APES (20.2 days) and for maturity by HERMES (12.6).

The wide range of grain yield estimates provided by the models for all sites and years reflects substantial uncertainties in model estimates achieved with only minimum calibration. Mean predictions from the eight

models, on the other hand, were in good agreement with measured data. This applies to both results across all sites and seasons as well as to prediction of observed yield variability at single sites – a very important finding that supports the use of multi-model estimates rather than reliance on single models.

Keywords

Climatic variability, Crop growth model, Model comparison, Simulation, Winter wheat, Yield prediction

1 Introduction

Decision making and planning in agriculture increasingly makes use of various model-based decision support tools, particularly in relation to changing climate issues. The crop growth simulation models applied are mostly mechanistic, i.e. they attempt to explain not only the relationship between parameters and simulated variables, but also the mechanism of the described processes (Challinor et al., 2009, Nix, 1985, Porter and Semenov, 2005).

Even though most crop growth simulation models (hereafter referred to as crop models) have been developed and evaluated at the field scale, and were not originally meant for assessing large areas, it has become common practice to apply them in assessing agricultural impacts and adaptation to climate variability and change, from the field to a (supra-) national scale (e.g. Parry et al., 2005, Rosenberg, 2010). We hypothesize that many large-scale crop model applications that assess climate impacts and adaptation options for crops involve huge uncertainties related to the model parameters and model structure. For example, the models applied have often not been thoroughly calibrated for the conditions of the application; they have not been evaluated for their capacity to capture the effect of climatic variability on yield, either under the conditions for which the model was developed or for the conditions of the application. Moreover, most model users are not familiar with the range of model limitations and specificities for their proper application.

Comparison of different modelling approaches and models can reveal the uncertainties related to crop growth and yield predictions, including also the uncertainty related to model structure, which is the most difficult source of uncertainty to quantify (Chatfield, 1995). Comparisons can help to identify those parts in models that produce systematic errors and require improvements (see e.g. Porter et al., 1993). Since the 1980s, there have been many studies comparing different mechanistic crop models with respect to their performance in predicting yield and yield variability in response to climate and other environmental factors

(Diekkrüger et al., 1995, Eitzinger et al., 2004, Ewert et al., 2002, Jamieson et al., 1998, Kersebaum et al., 2007, Wolf et al., 1996), and many comparisons have been made for wheat models (e.g. Goudriaan et al., 1994, Landau et al., 1998, Meinke et al., 1998, Porter et al., 1993). However, for more than a decade, neither at the European nor at a global level has there been a comparison involving more than just a handful of the major accessible crop models (see Goudriaan et al., 1994), at least not for those that are most widely used for assessing impacts of climate variability and changes in field (cereal) crops.

The aim of this study was (1) to examine how different process-based crop models perform at the field scale when provided with limited information for model calibration and simulation, reflecting the typical situation in which these models are applied to large areas, and (2) to present and discuss the different sources of uncertainty involved in this kind of model application. To this end, eight crop models were run for 49 growing seasons at eight different study sites across Europe: in the Czech Republic, Denmark, Germany, Slovakia and Turkey. Winter wheat (*Triticum aestivum* L.) was used as the test crop as it is Europe's dominant cereal crop.

2 Material and methods

2.1 Models

The eight crop simulation models included in the comparison were APES, CROPSYST, DAISY, DSSAT, FASSET, HERMES, STICS and WOFOST. Details of these models can be obtained from the references gathered in Table 1. Table 2 provides an overview of the various modelling approaches applied regarding the major processes that determine crop growth and development.

All the eight models are applicable to winter wheat and they are capable of simulating crop phenology, total above-ground and root biomass, leaf area, grain yield, and field water balance components in daily time steps. However, they clearly differ with respect to their complexity and algorithms applied.

The eight crop simulation models can be grouped in terms of the detail with which they treat the following major crop growth processes (see also Table 2):

(1) *Leaf area development and light interception.* Most of the models simulate leaf area dynamics dependent on crop phenological stage, acknowledging that e.g. temperature and light affect differently the leaf expansion at different stages (Spitters, 1990). APES, CROPSYST and DSSAT are simpler in this respect. They base their leaf area calculations on a specific leaf area at emergence and biomass partitioning factors, or apply a forcing function with an exogenously defined maximum leaf area index (LAI) (Ewert, 2004). LAI in the FASSET model is primarily driven by nitrogen uptake in the vegetative period (Olesen et al., 2002a).

(2) *Light utilization.* DAISY, HERMES and WOFOST contain detailed descriptions of leaf photosynthesis, respiration, development-stage-dependent dry matter allocation patterns and scaling up of dry matter increase at canopy level (e.g. van Ittersum et al., 2003). Other models apply a simpler approach, using the radiation use efficiency (RUE) concept (Monteith and Moss, 1977).

(3) *Crop phenology.* Most of the models included have detailed phenological sub-routines that consider more than two phases in describing relationships between temperature and crop development. They include the effect of temperature, day length and vernalisation, the latter being important for winter wheat (see e.g. Mirschel et al., 2005, Slafer and Rawson, 1996). STICS is the only model in which water and nutrient stress could affect development rate, but that feature was not activated in this study. WOFOST and FASSET exclude the effect of vernalisation.

(4) Soil moisture dynamics. Apart from the fact that the eight crop models deal with the soil profile at different degrees of resolution (e.g. different number of soil layers and soil characteristics considered), they use either a simpler capacity or tipping bucket approach (seven models out of

eight), or a more detailed Richards approach for soil water movement (DAISY) (van Ittersum et al., 2003). Models also require different numbers and types of weather variables, mostly depending on the evapotranspiration formulae applied (Penman-Monteith, Priestley-Taylor, Turc, etc.). Their assumptions regarding root distribution over depth and related water uptake vary (Wu and Kersebaum, 2008).

(5) *Nitrogen balance.* Detailed sub-routines for nitrogen balance, such as those applied in DAISY, FASSET and STICS, calculate all important nitrogen processes dynamically throughout the growing season. WOFOST, on the other hand, was used here to provide only water-limited production estimates because it does not include modules that calculate full nitrogen balance of soil and plants.

2.2 Study sites

For the comparison of models we used previously unpublished experimental data from winter wheat production sites that represented diverse agro-ecological conditions in Europe. Since modelling groups were supposed to perform "blind model runs", all published data sets were discarded. Detailed data required by the models further limited the number of sites. Moreover, growing seasons during which the yields were reported to be severely affected by pests, diseases or lodging, in spite of plant protection measures, were excluded from the study.

Ultimately, the model comparison was carried out by applying data from eight experimental field sites in Denmark, Germany, the Czech Republic, Slovakia and Turkey (Fig. 1). Basic characteristics of the study sites are summarized in Table 3. The data derived from 49 growing seasons of winter wheat. Data for the longest time series, 14 years, were available for the two Czech sites. Data from two to four years were available for the remaining sites. The cultivars used differed among countries but were generally the same within each country. Soils varied widely in their moisture retention characteristics, ranging from less favourable sandy soils (Müncheberg, Germany) to favourable silt loams (Czech Republic and Slovakia).

2.3 Setup of model comparison

2.3.1 Information available for modelling groups

The current study was implemented as a "blind test", i.e. the model users were not provided with the measured information for the variables they were asked to deliver as model results (i.e. measured yields, biomass etc.) until they had submitted their results for the comparison. For the simulations, the distributed data consisted of those for daily weather, (i.e. precipitation, mean, minimum and maximum temperature, mean relative humidity, early morning vapour pressure, global radiation and mean wind speed), crop and soil management (e.g. information on previous crop, tillage, sowing, irrigation, fertilization and harvest) and basic information on soil properties (e.g. bulk density, water capacity parameters and soil data in Table 3). In addition, model users were provided with the information on various important stages, i.e. dates of sowing, emergence, flowering, ripening and harvest, for the winter wheat crops grown at the various locations and for the years to be simulated. Available phenological information varied somewhat among sites.

2.3.2 Minimum calibration

Models were calibrated for crop phenology for each cultivar. While all model users applied the given phenological and weather information by adjusting their phenology-related parameters to match the observed phenological stages in the experiments, exactly how phenological information was interpreted and converted into parameter values was not specified. A pre-condition, however, was that only one parameter set per cultivar was created and used unchanged across all sites and years for a specific winter wheat cultivar.

All other crop parameters possibly needed for the models were taken from earlier applications of the models that were assessed to be geographically and physiologically comparable. They were kept unchanged across all sites and years.

2.4 Methods used for evaluating and comparing model performance

We performed the statistical analysis for the model-estimated grain yields by applying several indices following the approach of e.g. Bellocchi et al. (2009) and Willmott (1981). Model-estimated grain yields were compared with observed grain yields. All grain yields are reported as dry matter.

The coefficient of variation (CV) of the root mean square error (*CV(RMSE)*) was taken as a measure of the relative average difference between the model estimates and measurements. Mean bias error (MBE) was taken as an indicator to inform whether the models under- or overestimated measured yields, i.e. the direction and magnitude of bias. The variance of the distribution of differences (s_d^2) was used to quantify the error variability. Overall systematic error relative to total mean square error (MSE_S/MSE) was used to identify how much or what proportion of RMSE was systematic in nature. It was calculated as a share between the systematic error and mean square error. Index of agreement (IA), developed by Willmott (1981), was used as a more general indicator of modelling efficiency. IA can have values within the range [0,1], and values close to 1 indicate high simulation quality. Additionally, for comparison, the traditional r^2 regression statistic (least-squares coefficient of determination) was calculated even though it does not take into account model bias, which is central when assessing the performance of simulation models.

For the two longest time series (Lednice and Verovany) we had measurements from replicate plots (N=3, except in years 1992, 1993, 1996 and 1997 where N=4) that allowed a rough analysis of the uncertainties in observed yields resulting from errors in yield measurements and heterogeneity of site conditions. The ranges of simulated values, as well as the means of model-estimates, were compared with observations.

3 Results

3.1 Assessment of model performance

Crop phenology

Calibration results for wheat phenology show that the mean bias of the estimates for date of start of anthesis (Zadoks 61) and date of physiological maturity (yellow ripeness - Zadoks 90) over all studied growing seasons (Fig. 2a), and the average absolute deviations (Fig. 2b), were surprisingly large. It should be noted, however, that some of the phenological data used to compare the simulations were estimates rather than accurate observations. With HERMES and CROPSYST the grain filling periods were shorter than with other models. For CROPSYST it was because of late estimates of anthesis and for HERMES because of early estimates of maturity. The most accurate estimates of the phenological stages were provided by DAISY and DSSAT.

Grain yield, above-ground biomass and harvest index

APES, DAISY and DSSAT estimates of grain yield were closest to the observed values (Fig. 3a). CROPSYST mainly underestimated the yields, whereas HERMES, STICS and WOFOST clearly overestimated the yields. Interestingly, the relatively good agreement between the simulated and observed grain yields of APES, DAISY and DSSAT are combined with very contrasting total above-ground biomass values (Fig. 3b) and thereby with harvest indices (Fig. 3c). APES, DAISY, HERMES, STICS and WOFOST generated high total above-ground biomass estimates, whereas CROPSYST, DSSAT and FASSET provided low total above-ground estimates (Fig. 3b). Consequently, DSSAT and FASSET gave very high harvest index values, followed by HERMES and WOFOST. APES and DAISY, on the other hand, produced low harvest index values. CROPSYST produced a narrow range of harvest index values, from 0.41 to 0.48, whereas the largest range of harvest indices was associated with HERMES, from 0.28 to 0.61. It is to be noted that the harvest indices shown are calculated as ratios of the simulated grain yield (0% moisture) to simulated maximum total above-ground dry matter. In that context, the very high harvest index values

generated by DSSAT and FASSET are close, and even partly exceed, the upper limit of the harvest index values reported for winter wheat.

Root biomass

Maximum root biomass estimates of the models (not available for APES and CROPSYST) divided the models into two groups (Fig. 3d). DAISY, FASSET and STICS provided root biomass estimates on average higher than 3000 kg ha^{-1}, whereas DSSAT, HERMES and WOFOST provided estimates on average lower than 1000 kg ha^{-1}. DAISY's root biomass estimates had the largest range (from 1600 to 4300 kg ha^{-1}) and DSSAT's the smallest (from 300 to 1200). The only measured root biomass available to allow evaluation of the model results in this respect was for the Foulum study site for 2008 measured on 11th June, which can be taken as the maximum root biomass estimate (see e.g. Zhang et al., 2004). The observed root biomass was then 2250 kg ha^{-1}. The closest root estimate for Foulum 2008 was given by STICS (2270 kg ha^{-1}), the highest estimate by DAISY (4090 kg ha^{-1}) and lowest estimate by DSSAT (1040 kg ha^{-1}).

Dynamics of above-ground biomass and leaf area development

There was large variation among the simulated maximum total above-ground biomass (Fig. 3b) and LAI (Fig. 3c) estimates for all sites. The 1994 Müncheberg study site, with rainfed and irrigated treatments, was used as an example to show the differences in dynamics within a growing season. For that year and for the rainfed experiment, the model-estimated maximum above-ground biomass ranged from 8200 kg ha^{-1} for DSSAT to 15400 kg ha^{-1} for DAISY (Fig. 4b) and maximum LAI from 1.33 for DSSAT to 6.76 for FASSET (Fig. 4c). The greatest difference between the total above-ground biomass production on irrigated and rainfed plots was for DAISY and HERMES, whereas APES showed only a moderate or very small response (Fig. 4a,b). The simulated soil water contents were distinctly different among models, and are higher than the seasonal patterns (Fig. 4d). Although the initial water content and the values for field capacity, wilting point and rooting depth were provided, some models produced values beyond these limits and which deviated significantly from the measured water content values (DAISY and DSSAT).

Performance statistics for grain yield

A detailed comparison of the observed and model-predicted yields showed that none of the models perfectly reproduced observations at all sites and in all years. Figure 5 and the statistical analysis (Fig. 6) show that the best performance regarding yield estimation was for DAISY and DSSAT, for which the RMSE values (1428 and 1603 kg ha^{-1}) were lowest and the IA and r^2 values highest. CROPSYST clearly systematically underestimated yields, whereas HERMES, STICS and WOFOST clearly overestimated them (Fig. 5, 6b). The error variability (Fig. 6c) and the overall or average systematic error (Fig. 6) were lowest for DAISY. IA was lowest for APES, followed by WOFOST, HERMES and STICS (Fig. 6e), indicating some high individual discrepancies between simulated and observed yields.

The observed yields were accurately reproduced by the mean estimates from all eight models for most of the growing seasons (Fig. 5). This encouraging result was further confirmed by statistical indicators (CV(RMSE) = 0.18, IA = 0.80). In addition, the observed yield variability at Lednice and Verovany was captured by the multi-model means (Fig.7). Poor results at Lednice are recorded for three years in which the models overestimated grain yields. Careful study of field logbooks revealed factors that were either not captured by the models or were particularly challenging to quantify; 1993 (spring drought), 2002 (disease severely reduced yield) and 2003 (frost). In addition, in 1992 the models mostly underestimated the yield. For Verovany in 1994, the models overestimated yields in the face of severe yield reduction due to diseases.

3.2 Uncertainties

A broad range in model estimates of grain yield (Figs. 3a, 5 and 7) indicates the magnitude of model outcome uncertainty, which represents the accumulated uncertainty from various sources (Walker et al., 2003). The same applies to estimates of LAI, total above-ground biomass and other state variables (see, e.g. Fig. 4). At Lednice and Verovany (both with total 14 years of observed yields), the range of model estimates for each year was, on average, considerably higher than the range of observed yields from different plots (N=3 or 4) (Fig. 7). However, not all uncertainties in observations resulting from errors in yield

measurement or heterogeneity in site conditions are reflected in model estimates. As an example, the variation in growing conditions within a replicate plot is not accounted for.

4 Discussion

We hypothesized that use of crop simulation models in climate impact and adaptation studies with restricted calibration leads to a high degree of uncertainty in estimated impact indicators. Here we adopt the general definition of uncertainty as being "any departure from the unachievable ideal of complete determinism" (Walker et al., 2003, p. 8). Our simulation results that also involved model structure uncertainty showed a wide range for grain yield estimates (1800 to 12000 kg ha^{-1}) for all sites and seasons that were larger but not too different from observed ranges (2200 to 9500 kg ha^{-1}). However, there were substantial discrepancies in the estimates for individual sites and years among the models (Fig. 5 and Fig. 7), and in the agreement between simulated and observed yields, with RMSEs ranging from 1400 to 2300 kg ha^{-1} and index of agreement (IA) between 0.40 and 0.74 reflecting large model outcome uncertainties. These uncertainties may possibly be of the same order of magnitude as those in climate projections produced by different General Circulation Models (GCMs) (Murphy et al., 2004, Murphy et al., 2007).

The performance of individual models in predicting yields across sites cannot generally be regarded as satisfactory for the type of model use applied here with only restricted calibration. No model is perfect, and none of those included in this comparison performed clearly better than others. However, we can assume that each may at least be internally consistent and plausible. As there is no "error-free" or clearly "best" model, we looked at the possibility of applying multi-model means – currently a common practice when using climate models (Murphy et al., 2004). An important finding is that the mean model predictions (the average of eight) were in good agreement with observed yields. This applies to both results across all sites and seasons (Fig. 5) as well as to prediction of observed yield variability at single sites (Fig. 7). This is promising and would, for instance, imply a recommendation to use multi-model estimates rather than rely on

single models that are deemed to perform well in specific regions and for particular agro-ecological conditions. At least this holds true for winter wheat across northwestern and Central Europe.

4.1 Uncertainties in model simulations

Van Oijen and Ewert (1999) distinguished three major sources of simulation uncertainty that largely coincide with the locations of model uncertainty identified by Walker et al. (2003) as those related to (*i*) input data, (*ii*) parameterization and (*iii*) model structure. We discuss below these locations of uncertainty, their nature and extent regarding our exercise. In addition, we bring to the fore and discuss human errors (*iv*) related to the setup of this kind of model comparison study, but which are very difficult to quantify.

(i) Model input

The models used daily weather variables as input data, in addition to soil physical properties and initial conditions (soil water, soil organic matter and soil nitrogen) and basic crop and soil management information. The uncertainties related to input data were largely minimized in our case as we placed special emphasis on removing such sites and seasons from the comparison where model input had been considered incomplete, dubious or obviously erroneous. However, some inaccuracies in model input data, such as those related to location of the weather stations (representation error), or spatial heterogeneity of soil properties, could not be excluded. Furthermore, there are always measurement errors related to measured variables. For instance, measurements of precipitation using standard rain-gauges usually underestimate ground precipitation (e.g. Legates and Willmott, 1990) in the order of magnitude of 10%, depending on wind speed (Subedi and Fullen, 2009). Additionally, small-scale variability in annual precipitation can be in the order of 8% (Subedi and Fullen, 2009). Also, almost all weather data series had some missing values that required interpolated estimates to be made or data from nearby stations.

(ii) Parameterization

In this study, the model users were only allowed to calibrate crop phenology related parameters based on observations. In many cases, however, the available observations were only of moderate quality, meaning that those phenological dates needed for calibration were either not unequivocal or had to be estimated from records from other phenological stages. Fig. 2 shows that although all model users were provided with the same data, there were differences in how they were interpreted and translated in the models. The fact that model errors for anthesis for most models were larger than for maturity (Fig. 2b) indicates that the data for anthesis were less precise and unambiguous. Our results also gave no clear indication of the relationship between the accuracy of crop phenology and grain yield estimates (Fig. 2 and Fig. 6).

Other crop cultivar-specific parameters needed for the models were taken from default values in the models, from the literature or/and some earlier applications of the models, the quality of which depended on both the applicability (geographically near or far, same/similar varieties etc.) and quality of the initial empirical data source. For example, the underestimation of the drought effect by APES (Fig. 3) may be because the parameters used to represent the effect of water limitation on biomass referred to the less drought sensitive varieties that are typically grown in southern France (Therond et al., 2010). Due to the lack of suitable experimental data from our study sites, we could not assess the values for other crop-related parameters used in the models. Their importance for the simulated results is, however, potentially high.

(iii) Model structure

Crop simulation models are, by definition, simplifications of reality, and sometimes there are oversimplifications that lead to marked discrepancies between simulated and measured data. Recently, Adam et al. (2011) showed the impact on modelling detail of light interception and conversion into biomass. Though the various models applied in this comparison differ to some degree, and for some processes even considerably in the level of complexity (Table 2), in this study we could not satisfactorily determine effects introduced by such differences. This is because for most seasons we lacked sequential measurements of biomass, leaf area and soil moisture.

(iv) Human errors

In addition to the traditional model-related sources of variation above, we can identify some additional ones that are related to communication and interpretation of the data and model results. For example, some terms, like "rooting depth", were interpreted differently by model users either as maximum rooting depth or as "effective" rooting depth, from which water and nutrients can be completely depleted by the roots. The equivocality of the phenological data described above led to different interpretations by model users, which in turn created differences in the calibration results.

Compilation of consistent and complete datasets for the models, as well as the simulations with process-based models having numerous input parameters, is both laborious and involves the risk of human errors. The complexity of the exercise from the model user's point of view was reflected by several facts. Although the models were applied "blind" (by data providers holding back the experimental data until simulation results had been received and processed), several modelling groups needed, and were allowed to make, corrections and iterations after their first model runs. Ultimately, all models had iterations for some reason, e.g. APES was run again with the water-balance module developed while the study was proceeding, DSSAT was corrected for unreasonably high harvest indices (even higher than reported here), HERMES had problems with leaf senescence, WOFOST was accidentally used at first with incorrect soil input data for one site, and for STICS the misinterpretation of phenological data led to erroneous crop parameterization and was corrected. The human error remaining, even after making the various corrections described above, was certainly far from being negligible. One means to quantify its contribution in future could be to run the same model (version) with different model users.

4.2 Uncertainties related to observed data and site selection

Van Oijen and Ewert (1999) also distinguished the sources of variation in observed yields that the models cannot account for and which are related to heterogeneity in crop characteristics and environmental conditions, and error in yield measurements and other experimental observations.

We have no indication that yield variability within the sites due to genetic variation in the plant material or variation in seed quality led to uneven seedling emergence or crop growth at any of our sites. Only cultivars with high quality seeds were sown at the 8 locations over 49 seasons.

Heterogeneity of growing environments refers here to differences in availability of resources and occurrence of yield-reducing factors such as weeds, pests and diseases that are not taken into consideration in the models. We aimed at avoiding such sites and growing seasons during which the yield-reducing factors had significantly affected the yields. This was not fully successful because discrepancies between simulated and observed data (such as in 1999 and 2001 for Verovany in Fig. 7b) led to a critical scrutiny of the measured data, which revealed issues not noticed when selecting the sites. But even for factors that should be covered by the model, such as yield limitation due to soil water deficits, we need to bear in mind that model input data, such as those for soil characteristics, can only be given as point data for the whole experiment and thus cannot capture variation in soil properties within sites.

For the Czech sites we had three or four replicates, which at least partly reflected the extent of within-site yield variations (Fig. 7). Average CVs reported earlier by Taylor et al. (1999), based on 220 experimental wheat field trials and by Joernsgaard and Halmoe (2003) based on an intra-field variation study of 124 fields, were 13% and 10%, respectively.

In our case, sampling errors were site-specific as the harvesting techniques and conditions varied across sites. Hand harvested yield values are usually higher than values from combine harvesting due to grain losses from combining. High-yielding plots are particularly at risk of lodging, leading to an underestimation of yields harvested by a combine harvester. Kersebaum et al. (2005) reported differences between hand harvested and combine harvested wheat yields of about 2 Mg ha^{-1} (27 %) on fertilized plots after a heavy rain shortly before harvest, while non-fertilized plots furnished similar yields, irrespective of harvesting method.

4.3 Perspectives on model improvements

A novel and clear merit of this study is that nearly all the major and widely applied crop growth simulation models for winter wheat in Europe were put to the test. This has not happened to such an extent since the winter wheat model comparison by Goudriaan et al. (1994), i.e. with this large a number of models and with various climatic conditions. In future these kinds of model comparisons need to be performed with newly generated and more comprehensive datasets containing sequential measurements (see e.g. Rosenzweig et al., 2011). Similar comparisons are needed also for other crops at field and regional scales and for other regions. It is also important to compare crop models and modelling approaches for assessing yields under sub-optimum conditions, i.e. under nutrient-limitation, and also develop improved quantitative tools to simulate or otherwise assess crop yield reduction by pests, diseases, weeds and pollutants (Rosenberg, 2010, van Oijen and Ewert, 1999).

Comprehensive experimental datasets for comparing simulations with observations are scarce. Model evaluations increasingly use the same well-documented datasets (e.g. Groot and Verberne, 1991, Jamieson et al., 1998, Porter et al., 1993), which makes it at least questionable as to what the major reasons are that model results agree very well with observations. An obvious constraint to all model development and improvement is the availability of comprehensive, long-term datasets for calibrating and validating models for various crop cultivars. Longer, and better suited yield series for clearly defined growth and management conditions would also greatly enhance the outcome of model comparison studies for subsequent model improvement.

5 Conclusions

From the results obtained we conclude that application of crop simulation models with restricted calibration leads to a high degree of uncertainty about climate impacts on yield and yield variability. An important

finding is that the mean model predictions were in good agreement with observed yields. This supports earlier recommendations to use multi-model estimates rather than rely on single models deemed to perform well for specific regions and agro-ecological conditions.

In terms of judicious use of crop models, our results confirm earlier findings that crop models need calibration for their most important parameters before they can be applied with confidence. Minimal calibration for phenological dates will not be sufficient to generate robust crop cultivar-specific yield estimates for different environments. Some models performed better than others in estimating grain yield and other crop variables, but none could unequivocally be termed robust and accurate in terms of yield prediction across different environments and for different crop cultivars. This is a strong argument for ensemble crop modelling. Good prediction of crop yield for some models came at the cost of overestimating or underestimating harvest index or total biomass. Other models showed a distinct bias towards under- or overestimating yields (Fig. 6). Unfortunately, these biases cannot solely be related to model deficiencies. Partly, we included experimental datasets that *ex post*, after careful scrutiny, turned out not to correspond to model boundary conditions (e.g. assuming the absence of pests and diseases as yield-reducing factors).

6 Acknowledgements

This study was implemented as a co-operative project under the umbrella of COST734 "Impacts of Climate Change and Variability on European Agriculture – CLIVAGRI" and the work of individual researchers was funded by various bodies as listed below:

- T. Palosuo, R. Rötter: the strategic project "Integrated Assessment Modelling and Tools (IAM-Tools) funded by MTT Agrifood Research Finland, and project "Agri-Adapt" co-funded by the Dutch Climate Change programme (BSIK), the German Academic Exchange Service (DAAD), and the Academy of Finland (decision 139270).
- K.C. Kersebaum, W. Mirschel: LandCaRe 2020 project (01LS05109) funded by the German Federal Ministry of Education and Research, co-funded by the German Federal Ministry of Consumer

Protection, Food and Agriculture, and the Ministry of Infrastructure and Agriculture of the Federal State of Brandenburg (Germany).

– M. Trnka, P. Hlavinka: Research Plan No. MSM6215648905 “Biological and technological aspects of sustainability of controlled ecosystems and their adaptability to climate change”, The Czech Science Foundation (GACR) project no. 521/09/P479 and project NAZV QI91C054.

– J. Takáč: FP7 EU under agreement No. 212535 (Climate Change – Terrestrial Adaptation and Mitigation in Europe).

– B. Šiška: FP6 EU under agreement No. 037005 (Central and Eastern European Climate Change Impact and Vulnerability Assessment).

– J.E. Olesen and R.H. Patil: The project ”Impacts and adaptation to climate change in cropping systems” funded by the Danish Ministry of Food, Agriculture and Fisheries.

– L. Şaylan and B. Çaldağ: Project “Estimation the effects of meteorological factors on crop-growth by using crop-climate model”, ITU Research-Development Foundation and Project No. 108O567 “Investigation the potential effects of climate change on crop growth by crop growth simulation models” supported by the Scientific and Technological Research Council of Turkey.

References

Abrahamsen, P., Hansen, S., 2000. Daisy: an open soil-crop-atmosphere system model. Environ. Modell. Softw. 15, 313-330.

Adam, M., van Bussel, L.G.J., Leffelaar, P.A., van Keulen, H., Ewert, F., 2011. Effects of modelling detail on simulated potential crop yields under a wide range of climatic conditions. Ecol. Model. 222, 131-143.

Bellocchi, G., Rivington, M., Donatelli, M., Matthews, K., 2009. Validation of biophysical models: issues and methodologies. A review. Agron. Sustain. Dev. 30, 109-130.

Berntsen, J., Petersen, B.M., Jacobsen, B.H., Olesen, J.E., Hutchings, N.J., 2003. Evaluating nitrogen taxation scenarios using the dynamic whole farm simulation model FASSET. Agric. Syst. 76, 817-839.

Boogaard, H.L., van Diepen, C.A., Rötter, R.P., Cabrera, J.M.C.A., van Laar, H.H., 1998. WOFOST 7.1. User's guide for the WOFOST 7.1 crop growth simulation model and WOFOST Control Center 1.5. 52. DLO Winand Staring Centre, Wageningen, 142 p.

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

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agronomie 18, 311-346.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P., Hénault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model. Eur. J. Agron. 18, 309-332.

Challinor, A.J., Ewert, F., Arnold, S., Simelton, E., Fraser, E., 2009. Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. J. Exp. Bot. 60, 2775-2789.

Chatfield, C., 1995. Model uncertainty, data mining and statistical-inference. J. Roy. Stat. Soc. A Sta. 158, 419-466.

Diekkrüger, B., Söndgerath, D., Kersebaum, K.C., McVoy, C.W., 1995. Validity of agroecosystem models a comparison of results of different models applied to the same data set. Ecol. Model. 81, 3-29.

Donatelli, M., Russell, G., Rizzoli, A.E., Acutis, M., Adam, M., Athanasiadis, I.N., et al. 2010. A component-based framework for simulating agricultural production and externalities, in Brouwer, F.M., van Ittersum, M.K. (Eds.), Environmental and Agricultural Modeling - Integrated Approaches for Policy Impact Assessment. Springer, pp. 63-108.

Eitzinger, J., Trnka, M., Hosch, J., Zalud, Z., Dubrovsky, M., 2004. Comparison of CERES, WOFOST and SWAP models in simulating soil water content during growing season under different soil conditions. Ecol. Model. 171, 223-246.

Ewert, F., Rodriguez, D., Jamieson, P., Semenov, M., Mitchell, R., Goudriaan, J., Porter, J., Kimball, B., Pinter, P., 2002. Effects of elevated CO2 and drought on wheat: testing crop simulation models for different experimental and climatic conditions. Agr. Ecosyst. Environ. 93, 249-266.

Ewert, F., 2004. Modelling Plant Responses to Elevated CO2: How Important is Leaf Area Index? Ann. Bot. 93, 619-627.

Goudriaan, J., van de Geijn, S.C., Ingram, J.S.I., 1994. GCTE Focus 3 Wheat modelling and experimental data comparison workshop report, Lunteren, The Netherlands, November 1993. GCTE Focus 3 Office, University of Oxford, Oxford, UK.

Groot, J.J.R., Verberne, E.L.J., 1991. Response of wheat to nitrogen fertilization, a data set to validate simulation models for nitrogen dynamics in crop and soil. Fert. Res. 27, 349-383.

Hansen, J., Jensen, H.E., Nielsen, N.E., Svendsen, H., 1990. DAISY – A Soil Plant System Model. Danish Simulation Model for Transformation and Transport of Energy and Matter in the Soil-Plant-Atmosphere System. National Agency for Environmental Protection, Copenhagen.

Hansen, S., 2000. Daisy, a Flexible Soil – Plant - Atmosphere System Model. Equation Section 1. The Royal Veterinary and Agricultural University, Copenhagen, 47 p.

Hoogenboom, G., Jones, J.W., Porter, C.H., Wilkens, P.W., Boote, K.J., Batchelor, W.D., Hunt, L.A., Tsuji, G.Y., 2003. Decision Support System for Agrotechnology Transfer Version 4.0. Volume 1: Overview. University of Hawaii, Honolulu, H.I.

Jamieson, P.D., Porter, J.R., Goudriaan, J., Ritchie, J.T., van Keulen, H., Stol, W., 1998. A comparison of the models AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. Field Crops Res. 55, 23-44.

Joernsgaard, B., Halmoe, S., 2003. Intra-field yield variation over crops and years. Eur. J. Agron. 19, 23-33.

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

Kersebaum, K.C., 2007. Modelling nitrogen dynamics in soil–crop systems with HERMES. Nutr. Cycl. Agroecosys. 77, 39-52.

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 Kersebaum, K.C., Hecker, J.M., Mirschel, W., Wegehenkel, M. (Eds.), Modelling Water and Nutrient Dynamics in Soil Crop Systems. Springer, Dordrecht, pp. 1-17.

Kersebaum, K.C., Beblik, A.J., 2001. Performance of a nitrogen dynamics model applied to evaluate agricultural management practices, in Shaffer, M.J., Ma, L., Hansen, S. (Eds.), Modeling Carbon and Nitrogen Dynamics for Soil Management. Lewis Publishers, Boca Raton, pp. 549-569.

Kersebaum, K.C., Lorenz, K., Reuter, H., 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, 59-67.

Kersebaum, K.C., 1995. Application of a simple management model to simulate water and nitrogen dynamics. Ecol. Model. 81, 145-156.

Landau, S., Mitchell, R.A.C., Barnett, V., Colls, J.J., Craigon, J., Moore, K.L., Payne, R.W., 1998. Testing winter wheat simulation models' predictions against observed UK grain yields. Agr. Forest Meteorol. 89, 85-99.

Legates, D.R., Willmott, C.J., 1990. Mean seasonal and spatial variability in gauge-corrected, global precipitation. Int. J. Climatol. 10, 111-127.

Meinke, H., Rabbinge, R., Hammer, G., van Keulen, H., Jamieson, P., 1998. Improving wheat simulation capabilities in Australia from a cropping systems perspective II. Testing simulation capabilities of wheat growth. Eur. J. Agron. 8, 83-99.

Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Mucher, C.A., Watkins, J.W., 2005. A climatic stratification of the environment of Europe. Global Ecol. Biogeogr. 14, 549-563.

Mirschel, W., Wenkel, K., Schultz, A., Pommerening, J., Verch, G., 2005. Dynamic phenological model for winter rye and winter barley. Eur. J. Agron. 23, 123-135.

Monteith, J.L., Moss, C.J., 1977. Climate and the efficiency of crop production in Britain. Philos. T. Roy. Soc. B 281, 277-294.

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768-772.

Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H., Webb, M.J., 2007. A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philos. T. Roy. Soc. A 365, 1993-2028.

Nix, H.A., 1985. Chapter 5. Agriculture, in Kates, R.W., Ausubel, J.H., Berberian, M. (Eds.), Climate Impact Assessment. Studies of the Interaction of Climate and Society (Scope 27). John Wiley & Sons, Chichester, U.S., pp. 105-130.

Olesen, J.E., Berntsen, J., Hansen, E.M., Petersen, B.M., Petersen, J., 2002a. Crop nitrogen demand and canopy area expansion in winter wheat during vegetative growth. Eur. J. Agron. 16, 279-294.

Olesen, J.E., Petersen, B.M., Berntsen, J., Hansen, S., Jamieson, P.D., Thomsen, A.G., 2002b. Comparison of methods for simulating effects of nitrogen on green area index and dry matter growth in winter wheat. Field Crops Res. 74, 131-149.

Parry, M., Rosenzweig, C., Livermore, M., 2005. Climate change, global food supply and risk of hunger. Philos. T. Roy. Soc. B 360, 2125-2138.

Porter, J.R., Semenov, M.A., 2005. Crop responses to climatic variation. Philos. T. Roy. Soc. B 360, 2021-2035.

Porter, J.R., Jamieson, P.D., Wilson, D.R., 1993. Comparison of the wheat simulation models AFRWHEAT2, CERES-Wheat and SWHEAT for non-limiting conditions of crop growth. Field Crops Res. 33, 131-157.

Ritchie, J.T., Otter, S., 1985. Description and performance of CERES-Wheat: a user-oriented wheat yield model. ARS Wheat Yield Project ARS-38. Natl Tech Info Serv, Springfield, Missouri, pp. 159-175.

Rosenberg, N.J., 2010. Climate change, agriculture, water resources: what do we tell those that need to know? Climatic Change 100, 113-117.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, C., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2011. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Submitted manuscript .

Slafer, G., Rawson, H.M., 1996. Responses to photoperiod change with phenophase and temperature during wheat development. Field Crops Res. 46, 1-13.

Spitters, C.J.T., 1990. Crop growth models; their usefulness and limitations. Acta Hort. 267, 349-368.

Stockle, C.O., Donatelli, M., Nelson, R., 2003. CropSyst, a cropping systems simulation model. Eur. J. Agron. 18, 289-307.

Subedi, M., Fullen, M.A., 2009. Spatial variability in precipitation within the Hilton Experimental Site, Shropshire, UK (1982–2006). Hydrol. Process. 23, 236-244.

Supit, I., Hooijer, A.A., van Diepen, C.A., 1994. System description of the WOFOST 6.0 crop simulation model implemented in CGMS. CGMS Publication 15956. EUR 15956 EN of the Office for Official Publications of the E.U., Luxembourg.

631 Taylor, S.L., Payton, M.E., Raun, W.R., 1999. Relationship between mean yield, coefficient of variation,
632 mean square error, and plot size in wheat field experiments. Commun. Soil Sci. Plant Anal. 30, 1439-
633 1447.

634 Therond, O., Hengsdijk, H., Casellas, E., Wallach, D., Adam, M., Belhouchette, H., Oomen, R., Russell, G.,
635 Ewert, F., Bergez, J., Janssen, S., Wery, J., Van Ittersum, M.K., 2010. Using a cropping system
636 model at regional scale: Low-data approaches for crop management information and model
637 calibration. Agr. Ecosyst. Environ. In Press.

638 van Diepen, C.A., Wolf, J., van Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop
639 production. Soil Use Manage. 5, 16-24.

640 van Ittersum, M.K., Leffelaar, P.A., van Keulen, H., Kropff, M.J., Bastiaans, L., Goudriaan, J., 2003. On
641 approaches and applications of the Wageningen crop models. Eur. J. Agron. 18, 201-234.

642 van Oijen, M., Ewert, F., 1999. The effects of climatic variation in Europe on the yield response of spring
643 wheat cv. Minaret to elevated CO2 and O3: an analysis of open-top chamber experiments by means
644 of two crop growth simulation models. Eur. J. Agron. 10, 249-264.

645 Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Von
646 Krauss, M.P.K., 2003. Defining uncertainty: a conceptual basis for uncertainty management in
647 model-based decision support. Integr. Assess. 4, 5-17.

648 Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184-194.

649 Wolf, J., Evans, L.G., Semenov, M.A., Eckersten, H., Iglesias, A., 1996. Comparison of wheat simulation
650 models under climate change .1. Model calibration and sensitivity analyses. Climate Res. 7, 253-270.

651 Wu, L., Kersebaum, K.C., 2008. Modeling Water and Nitrogen Interaction Responses and Their
652 Consequences in Crop Models, in Ahuja, L.R., Reddy, V.R., Saseendran, S.A., Yu, Q. (Eds.),

653 Response of Crops to Limited Water: Understanding and Modeling Water Stress Effects on Plant
654 Growth Processes. ASA-CSSA-SSSA, pp. 215-250.

655 Zhang, X., Pei, D., Chen, S., 2004. Root growth and soil water utilization of winter wheat in the North
656 China Plain. Hydrol. Process. 18, 2275-2287.

657

658

Figure captions

Fig. 1. Locations of the study sites.

Fig. 2. Model performance for phenology. (a) Mean model estimates for date of start of anthesis (Zadoks 61) (grey circles) and physiological maturity (yellow ripeness, Zadoks 90) (white circles). Dashed lines present the observed means. (b) RMSE of the model-calculated anthesis (grey bars) and maturity (white bars) date estimates. Note that the anthesis estimates for DAISY and STICS are for full flowering (Zadoks 65).

Fig. 3. Box-and-whisker plots of grain yield estimates of models and observations (a) maximum above-ground biomass estimates (b), harvest indices (c) and root biomass estimates of the models (d) among the simulated sites and years (N=49). Boxes delimit the inter-quartile range (25-75 percentiles) and whiskers show the high and low extreme values. Root biomass estimates for APES and CROPSYST were not available.

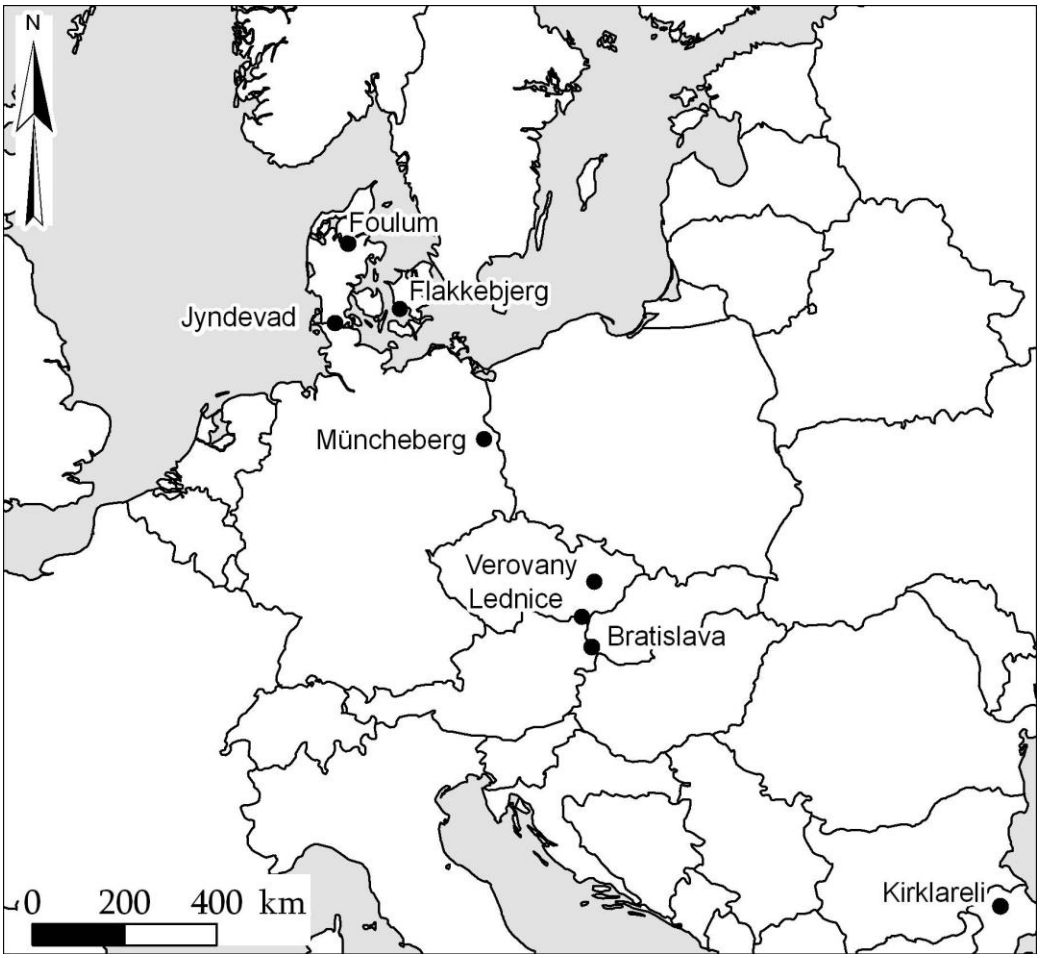
Fig. 4. Simulated and observed time course of total above-ground biomass of irrigated (a) and rainfed (b) treatments, and leaf area index (LAI) (no observations available) (c) and soil moisture content averaged over 0 - 90 cm layer (d) (not available for STICS) of rainfed treatment for winter wheat at Müncheberg study site in year 1994. Dotted horizontal lines in (d) show the field capacity (FC) and the wilting point (WP).

Fig. 5. Simulated and observed grain yield estimates [kg ha^{-1}, dry matter] for 49 studied growing seasons. Simulation results are shown for the eight individual models and as multi-model means. Different study sites are depicted with different symbols; filled blue tetragons = Lednice, open blue tetragons = Verovany, light blue squares = Bratislava, filled black squares = Müncheberg rainfed, open black squares = Müncheberg

685 irrigated, red triangles = Flakkebjerg, red windows = Jyndevad, red circles = Foulum, Grey circles =
686 Kirklareli. The 1:1 line is shown, representing perfect agreement.
687
688 **Fig. 6.** Graphical representation of statistics describing the performance of models in simulating all study
689 sites and growing seasons (*N*=49); (a) normalized root mean square error CV(RMSE) [0,1], (b) mean bias
690 error (MBE), (c) variance of model residuals, (d) systematic error (MSE_S/MSE) [0,1], (e) index of
691 agreement (IA) [0,1], (f) least-squares coefficient of determination (r^2) [0,1].
692
693 **Fig. 7.** Means and ranges of model-estimated (black tetragons and lines, eight models) and observed (open
694 circles and grey rectangles, three or four measured plots) yields for the studied growing seasons (N=14) in
695 Lednice (a) and Verovany (b) study sites.
696

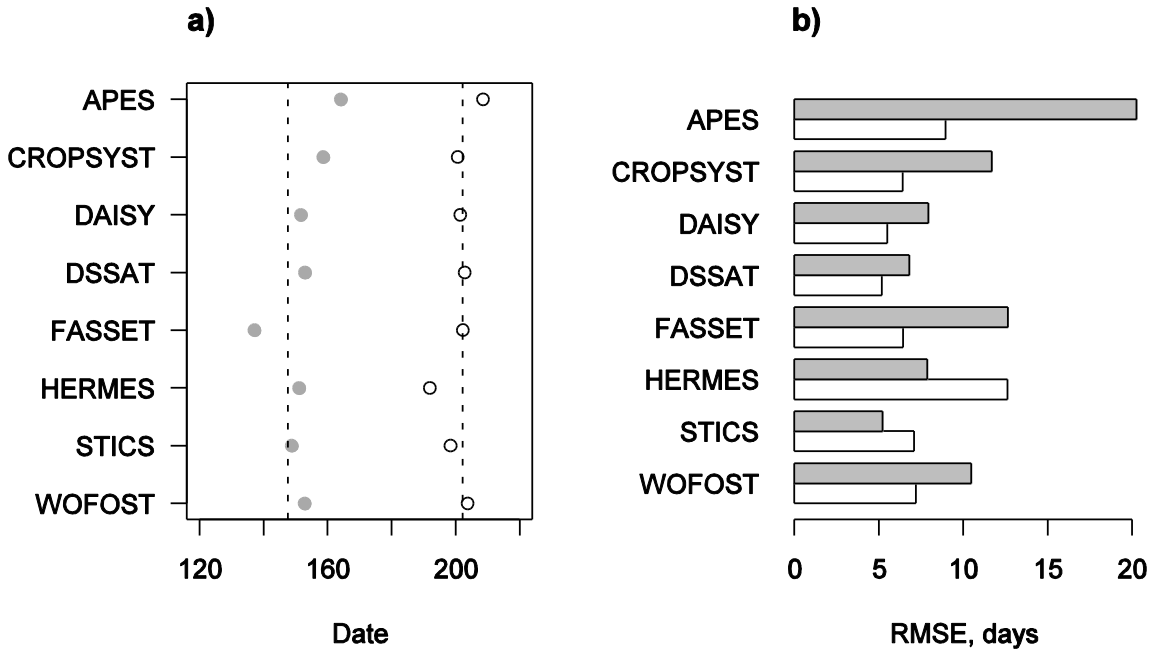
Figures

Figure 1



703 ***Figure 2***

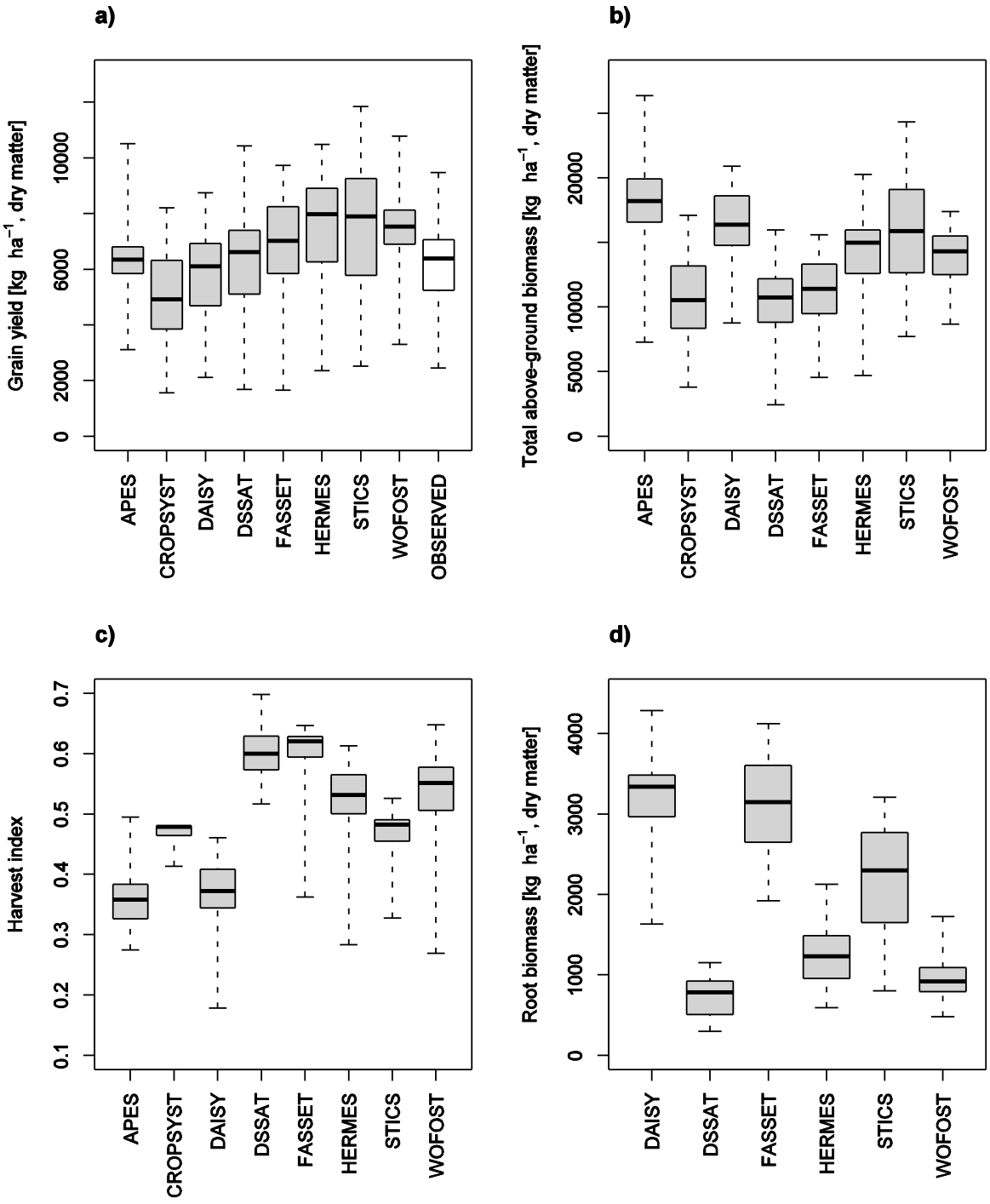
704



705

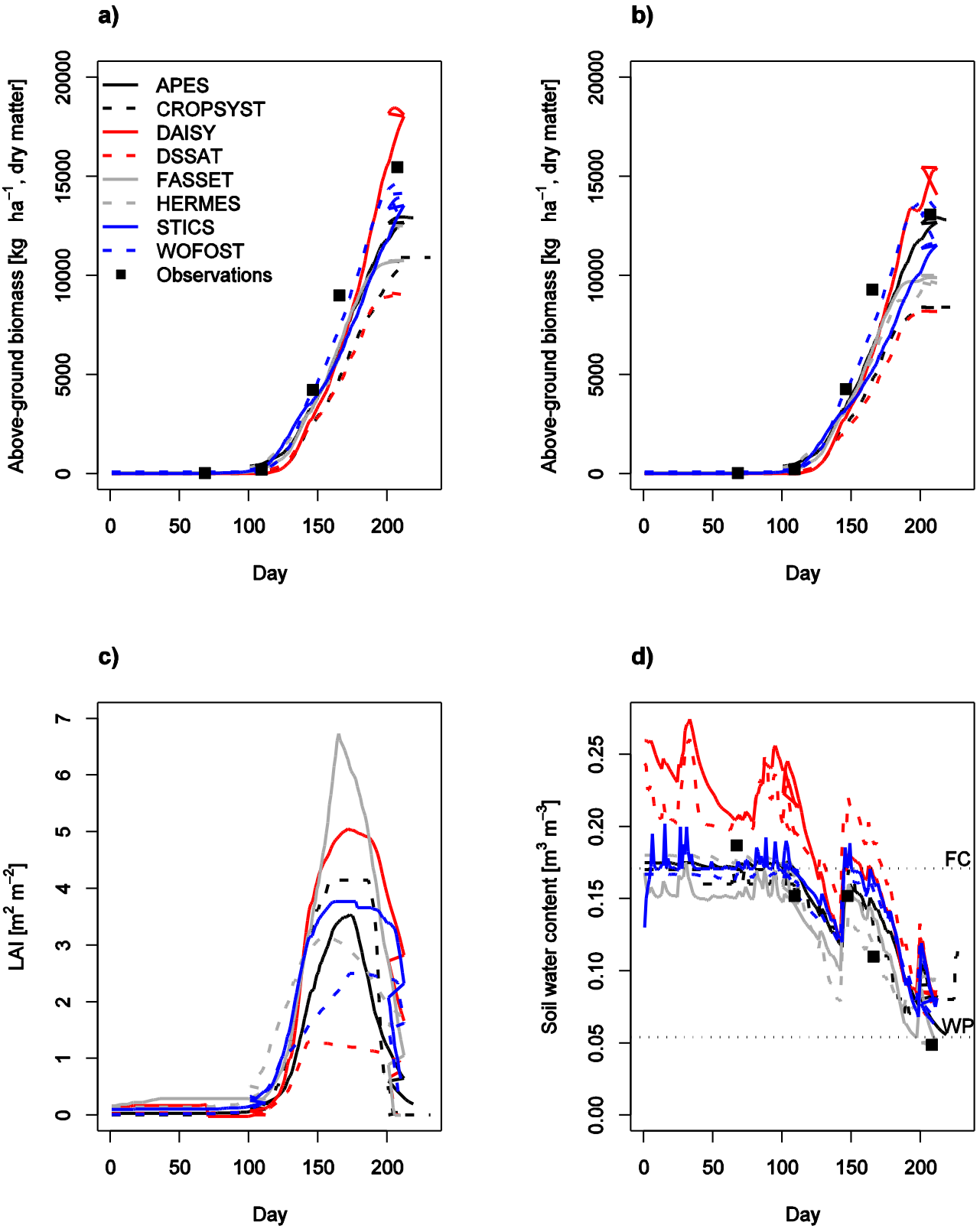
706

Figure 3



711 ***Figure 4***

712

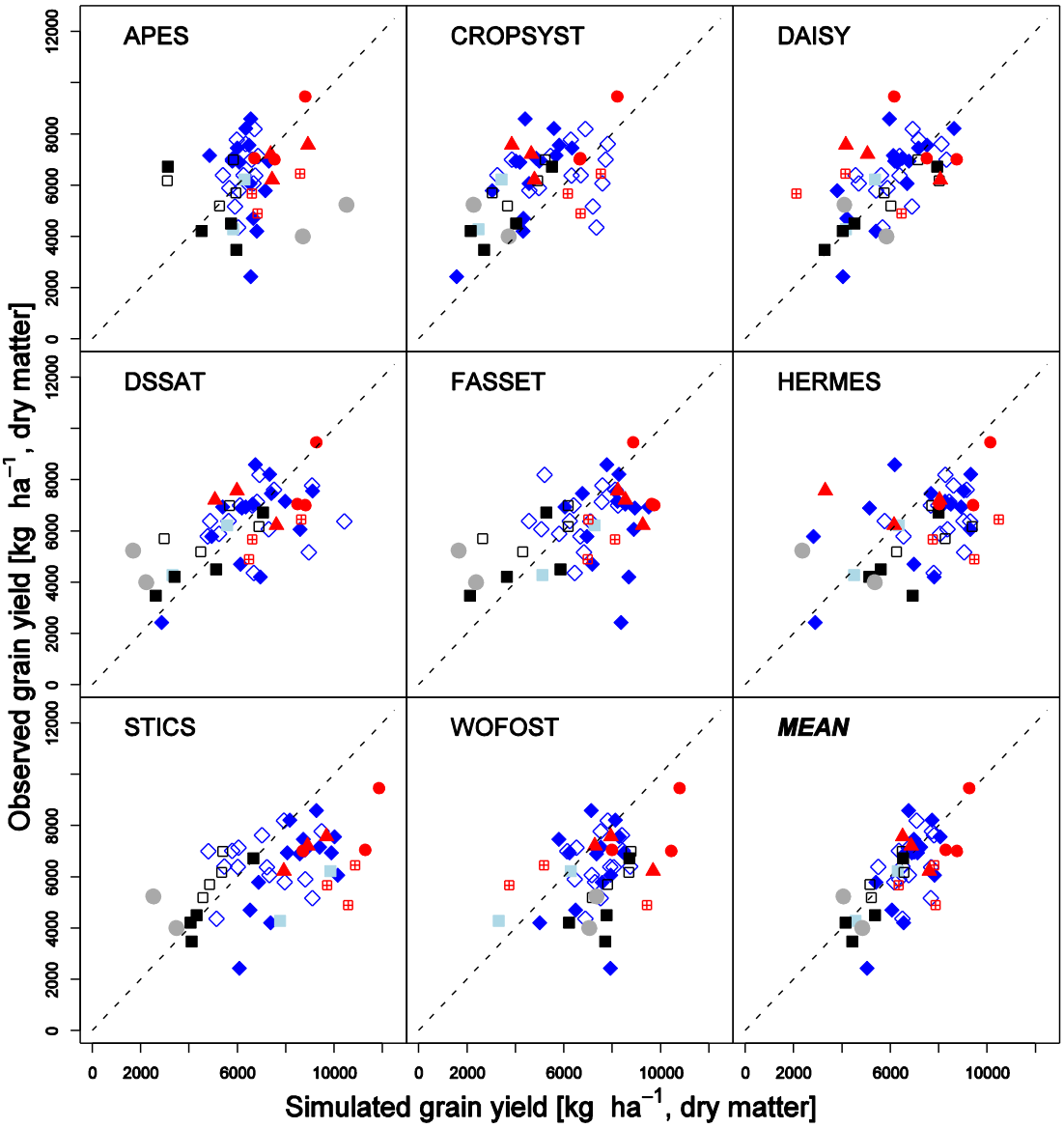


713
714
715

716

717 ***Figure 5***

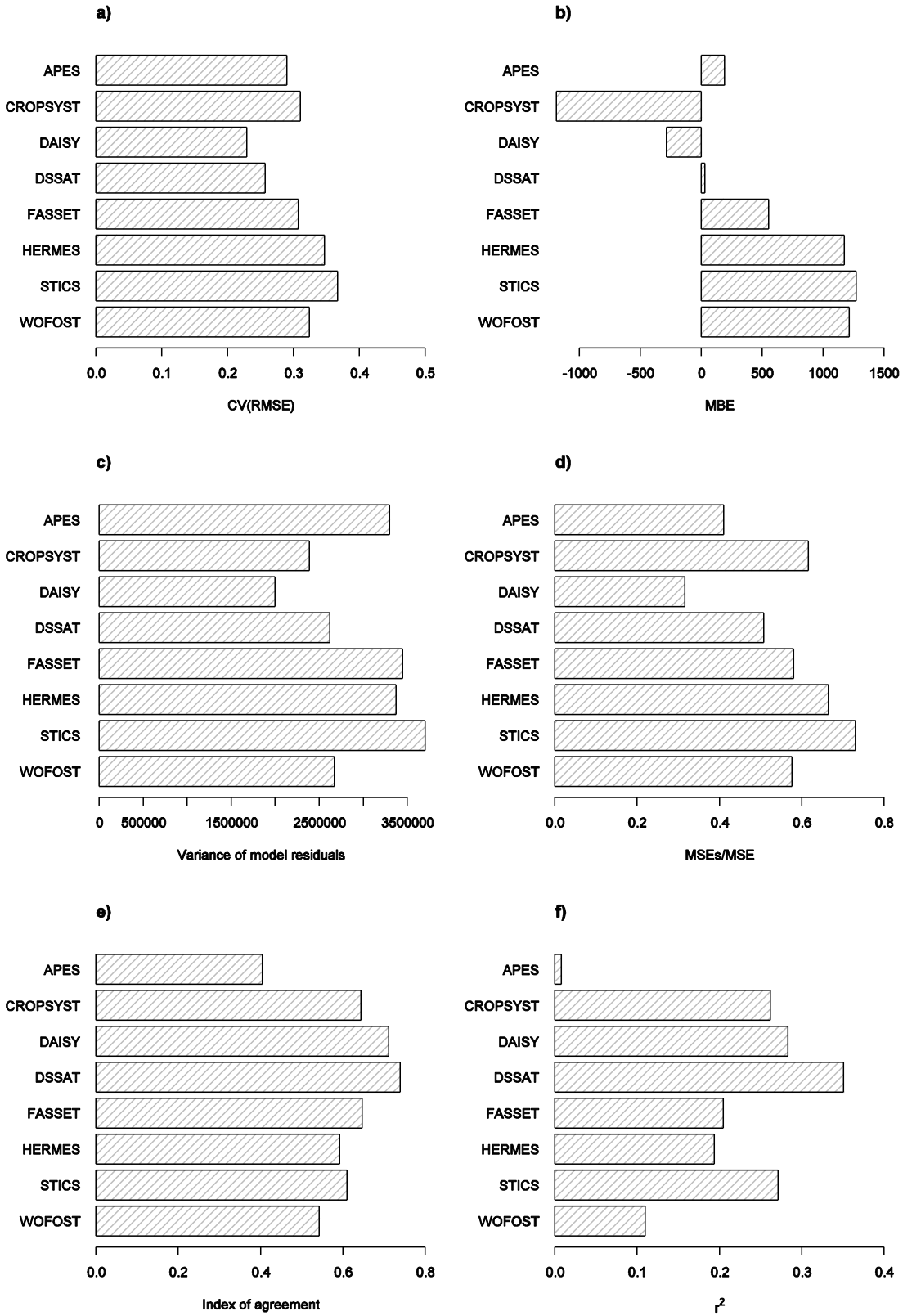
718



719

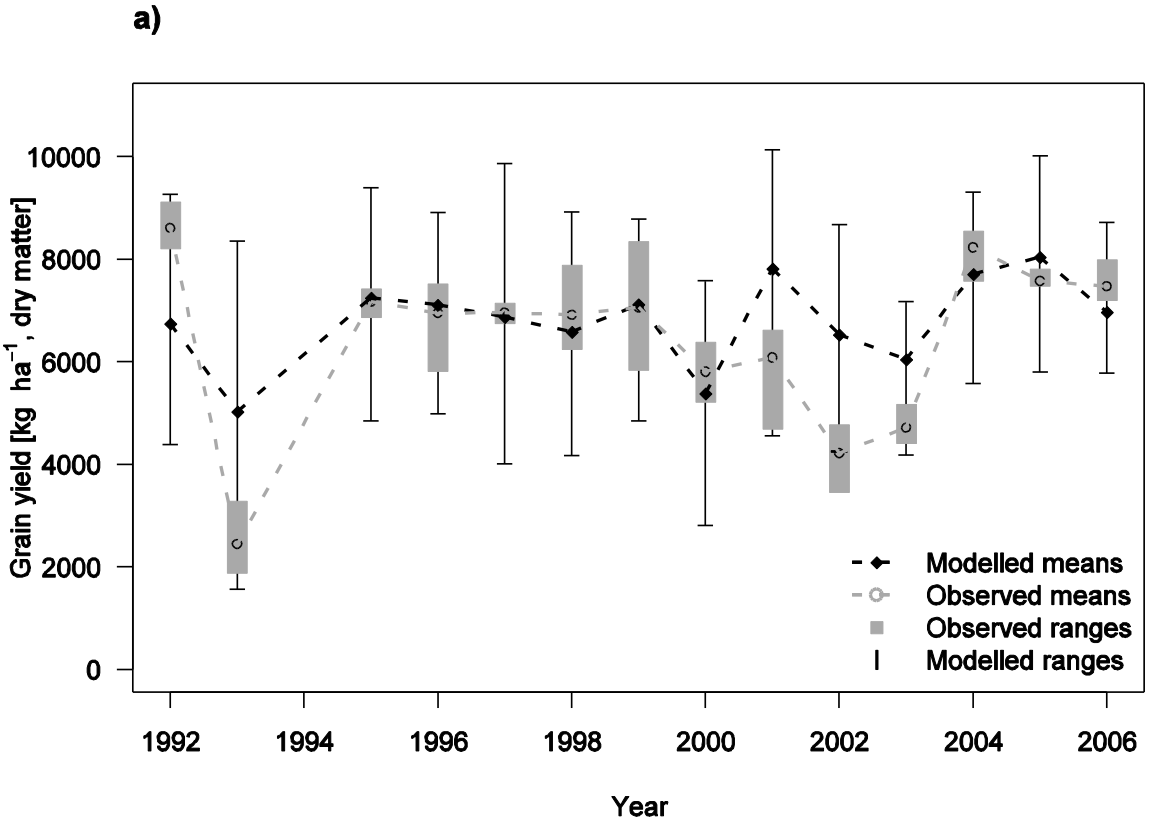
720

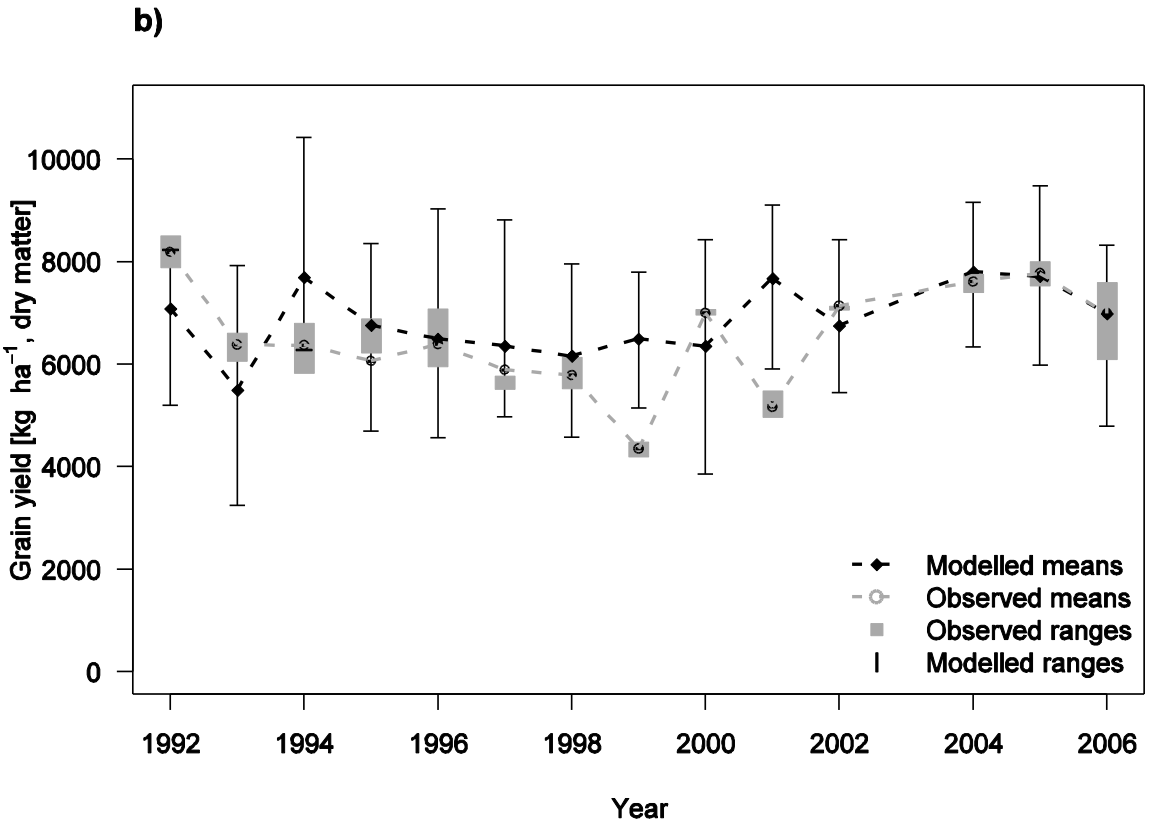
721 ***Figure 6***



722
723

724 ***Figure 7***





725
726

Tables

Table 1

Table 1. Model version applied in this study, references to papers with model descriptions and model web address.

Model	Version	Description	Web address
APES	V. 0.9.0.0	(Donatelli et al., 2010)	http://www.apesimulator.it/default.aspx
CROPSYST	V. 3.04.08	(Stockle et al., 2003)	http://www.bsyse.wsu.edu/CS_Suite/CropSyst/index.html
DAISY	V. 4.01	(Abrahamsen and Hansen, 2000, Hansen et al., 1990, Hansen, 2000)	http://code.google.com/p/daisy-model/
DSSAT	V. 4.0.1.0	(Hoogenboom et al., 2003, Jones et al., 2003, Ritchie and Otter, 1985)	http://www.icasa.net/dssat/
FASSET	V. 2.0	(Berntsen et al., 2003, Olesen et al., 2002a, Olesen et al., 2002b)	http://www.fasset.dk
HERMES	V. 4.26	(Kersebaum and Beblik, 2001, Kersebaum, 2007, Kersebaum, 1995)	Request from ckersebaum@zalf.de
STICS	V. 6.9	(Brisson et al., 1998, Brisson et al., 2009, Brisson et al., 2003)	http://www.avignon.inra.fr/agroclim_stics_eng/
WOFOST	V. 7.1	(Boogaard et al., 1998, Supit et al., 1994, van Diepen et al., 1989)	http://www.wofost.wur.nl

Table 2

Table 2. Modelling approaches applied in this study regarding the major processes determining crop growth and development.

	APES	CROPSYST	DAISY	DSSAT	FASSET	HERMES	STICS	WOFOST
Leaf area development and light interception[a]	D	S	D	S	D	D	D	D
Light utilization[b]	RUE	RUE	P-R	RUE	RUE	P-R	RUE	P-R
Yield formation[c]	Y(Prt)	Y(HI,B)	Y(Prt)	Yield(HI(Gn),B)	Y(HI,B)	Y(Prt)	Y(HI(Gn),B)	Y(Prt,B)
Crop phenology[d]	f(T, DL, V)	f(T, DL, V)	f(T, DL, V)	f(T, DL, V)	f(T, DL)	f(T, DL, V)	f(T, DL, V)	f(T, DL)
Root distribution over depth[e]	EXP	LIN	EXP	EXP	EXP	EXP	SIG	LIN
Stresses involved[f]	W, N	W, N	W, N	W, N	W, N	W, N, A	W, N	W, N[k]
Water dynamics[g]	C	C	R	C	C	C	C	C[l]
Evapo-transpiration[h]	P	PT	PM	PT	Makk	PM, TW[j]	P, PT or SW (here)	P
Soil CN-model[i]	CN, P(3)	N, P(1)	CN, P(6), B	CN, P(4), B	CN, P(6), B	N, P(2)	C, P(3); B	-

[a] Leaf area development and light interception; Simple (=S) or Detailed (=D) approach;

[b] Light utilization or biomass growth: RUE = Simple (descriptive) Radiation use efficiency approach, P-R = Detailed (explanatory) Gross photosynthesis – respiration; (for more details, see e.g. Adam et al. (2011))

[c] Y(x) yield formation depending on: HI = fixed harvest index, B = total (above-ground) biomass, Gn = number of grains, Prt = partitioning during reproductive stages

[d] Crop phenology is a function (f) of: T = temperature, DL = photoperiod (day length), V = vernalisation; O = other water/nutrient stress effects considered

[e] Root distribution over depth: linear (LIN), exponential (EXP), sigmoidal (SIG)

[f] Stresses involved: W = water stress, N = nitrogen stress, A = oxygen stress

[g] Water dynamics approach: C = capacity approach, R = Richards approach

[h] Method to calculate evapo-transpiration: P = Penman; PM = Penman-Monteith, PT = Priestley -Taylor, TW = Turc-Wendling, Makk = Makkink, HAR = Hargreaves, SW= Shuttleworth and Wallace (resistive model)

[i] Soil CN model, N = N model, P(x)= x number of organic matter pools, B = microbial biomass pool

[j] applied for Müncheberg site

[k] nitrogen-limited yields can be calculated for given soil Nitrogen supply and N fertilizer applied

[l] only two soil layers (top- and subsoil) are distinguished

Table 3

Table 3. Characteristics of the study sites.

Location	**Position**	**Precipi-**	**Tempera-**	**Period**	**Crop**	**Sand**$^{+}$	**Silt**	**Clay**	**C_{org}**	**Root**	**fc**	**wp**	**Soil name**
Environmental	**Latitude/longitude/**	**tation ***	**ture** $^{\#}$		**variety**					**depth**	**root zone°**		
zone$^{\$}$	**altitude a.s.l**	**[mm yr^{-1}]**	**[°C]**			**[%]**	**[%]**	**[%]**	**[%]**	**[cm]**	**[mm]**		
Lednice (CZ)	48°48´/16°48´/176m	539	10.0	1992-1993	Samanta	T: 17	61	22	1.41	150	474	216	Chernozem
CON 2				1995-2006		S: 18	62	20					
Verovany (CZ)	49°28´/17°17´/214m	576	9.0	1992–2002	Samanta	T: 17	66	17	1.17	150	480	215	Chernozem
CON 2				2003-2006		S: 15	64	21					
Bratislava (SK)	48°10´/17°/131m	523	10.0	1993	Hana	T: 15	63	22	1.49	120	320	109	Chernozem
PAN 2				1997	Astella	S: 16	64	20					
Müncheberg (D)	52°51´/14°07´/62m	603	8.9	1994	Bussard	T: 83	9	8	0.58	70	120	38	Eutric
CON 5				1996-1998		S: 93	6	1					Cambisol
Foulum (DK)	56°30´/9°35´/54m	694	8.8	2006-2008	Tommi	T: 78	13	9	2.15	130	329	90	Mollic
ATN3						S: 75	12	13					Luvisol
Flakkebjerg (DK)	55°11´/11°14´/32m	607	9.6	2006-2008	Tommi	T: 73	12	15	0.98	160	406	163	Glossic
CON 5						S: 69	12	19					Phaeozem
Jyndevad (DK)	54°54´/9°08´/14m	864	9.6	2006-2008	Tommi	T: 92	4	4	1.13	60	78	19	Humic
ATN3						S: 93	3	4	0.6				Podzol
Kirklareli (TR)	41°41´/27°13´/174m	734	12.7	1998	Atilla	T: 55	27	18	1.0	100	272	125	Fluvisols

MDN 3	1999	Pehlivan	S: 55	26	19

$ Environmental zone according to Metzger et al. (2005)

* average annual precipitation for period of simulations

average annual temperature for period of simulations

+ texture is given as average for T= topsoil (ploughing zone) and S = subsoil (until given root depth)

° mm water at field capacity (fc) and wilting point (wp) in specific root zone