Environmental Modelling & Software 72 (2015) 402–417

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Analysis and classification of data sets for calibration and validation of agro-ecosystem models☆

K.C. Kersebaum [a,*], K.J. Boote [b], J.S. Jorgenson [c], C. Nendel [a], M. Bindi [d], C. Frühauf [e], T. Gaiser [f], G. Hoogenboom [g], C. Kollas [a], J.E. Olesen [h], R.P. Rötter [i], F. Ruget [j], P.J. Thorburn [k], M. Trnka [l], M. Wegehenkel [a]

[a] Leibniz Centre for Agricultural Landscape Research, Institute of Landscape Systems Analysis, D-15374 Müncheberg, Germany
[b] University of Florida, Dept. of Agronomy, Gainesville, FL 32611-0500, USA
[c] University of Reading, School of Agriculture, Policy & Development, Reading, RG6 6AR, UK
[d] University of Florence, Department of Agri-food Production and Environmental Sciences, I-50144 Florence, Italy
[e] Deutscher Wetterdienst DWD, Centre for Agrometeorological Research, D-38116 Braunschweig, Germany
[f] University of Bonn, Institute of Crop Science and Resource Conservation, D-53115 Bonn, Germany
[g] Washington State University, AgWeatherNet, Prosser, WA 99350-8694, USA
[h] Aarhus University, Dept. of Agroecology, DK-8830 Tjele, Denmark
[i] Natural Resources Institute Finland (Luke), FI-50100 Mikkeli, Finland
[j] INRA, UMR1114 Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes, F-84914 Avignon Cedex 9, France
[k] CSIRO Ecosystem Sciences, Queensland 4102, Australia
[l] Mendel University Brno, Institute of Agriculture Systems and Bioclimatology, CZ-61300 Brno, Czech Republic

A R T I C L E I N F O

Article history:
Received 5 February 2014
Received in revised form
31 March 2015
Accepted 26 May 2015
Available online 13 June 2015

Keywords:
Field experiments
Data quality
Crop modelling
Data requirement
Minimum data
Software

A B S T R A C T

Experimental field data are used at different levels of complexity to calibrate, validate and improve agro-ecosystem models to enhance their reliability for regional impact assessment. A methodological framework and software are presented to evaluate and classify data sets into four classes regarding their suitability for different modelling purposes. Weighting of inputs and variables for testing was set from the aspect of crop modelling. The software allows users to adjust weights according to their specific requirements. Background information is given for the variables with respect to their relevance for modelling and possible uncertainties. Examples are given for data sets of the different classes. The framework helps to assemble high quality data bases, to select data from data bases according to modellers requirements and gives guidelines to experimentalists for experimental design and decide on the most effective measurements to improve the usefulness of their data for modelling, statistical analysis and data assimilation.

Software availability

Name of the software: Dataset classification software
Developer: Jason Scott Jorgenson, Kurt-Christian Kersebaum, Chris Kollas
License: GPL v2
Year first available: 2014
Contact: ckersebaum@zalf.de
Hardware requirements: Intel/AMD PC, 4 GB RAM
Software requirements: Microsoft Windows operating system
Availability: ftp://tran.zalf.de/pub/out/lsa/kersebaum/DatasetRanker.zip

1. Introduction

Soil–crop–atmosphere interactions play a central role in the multiple functions of agro-ecosystems and rural landscapes such as

☆ Special Issue on Agricultural Systems Modeling & Software.

* Corresponding author. Leibniz Centre for Agricultural Landscape Research, Institute of Landscape Systems Analysis, Eberswalder Str. 84, D-15374 Müncheberg, Germany. Tel.: +49 33432 82 394; fax: +49 33432 82 334.

E-mail addresses: ckersebaum@zalf.de (K.C. Kersebaum), kjboote@ufl.edu (K.J. Boote), j.s.jorgenson@reading.ac.uk (J.S. Jorgenson), nendel@zalf.de (C. Nendel), marco.bindi@unifi.it (M. Bindi), Cathleen.Fruehauf@dwd.de (C. Frühauf), tgaiser@uni-bonn.de (T. Gaiser), gerrit.hoogenboom@wsu.edu (G. Hoogenboom), chris.kollas@zalf.de (C. Kollas), JorgenE.Olesen@agrsci.dk (J.E. Olesen), reimund.rotter@luke.fi (R.P. Rötter), francoise.ruget@avignon.inra.fr (F. Ruget), Peter.Thorburn@csiro.au (P.J. Thorburn), mirek_trnka@yahoo.com (M. Trnka), mwegehenkel@zalf.de (M. Wegehenkel).

http://dx.doi.org/10.1016/j.envsoft.2015.05.009
1364-8152/

food and energy production, carbon sequestration, soil properties, biodiversity or conservation of water resources. Emissions from agriculture are also seen as a threat to the global climate system, which makes agriculture one of the key handles for climate change mitigation. There is an increasing need to better understand these complex systems, and to develop and utilize reliable process-based models for scenario analyses as a basis for policy and management decisions. Agro-ecosystem models are increasingly applied beyond the point and field scales to support decision-making (van Ittersum et al., 2003; Jones et al., 2003; Brisson et al., 2003; Stöckle et al., 2003; Keating et al., 2003), assess the impact of climate change (Holzworth et al., 2015, position paper of thematic issue), and to derive adaptation and mitigation strategies for the sustainable use and management of land and other natural resources (Hammer et al., 2002; White et al., 2011). Integrated Assessment and Modelling as suggested by Parker et al. (2002) requires the integration of dispersed data sources in a consistent and spatially and temporarily complete data set to provide necessary model inputs for decision making (Janssen et al., 2009) and to transfer site-based knowledge to regions and continents. With increasing size of the area under investigation, input data tend to become more uncertain relative to the point data of experimental sites, which were the original basis of development for the majority of agro-ecosystem models. Hence, model uncertainty also increases with the area under investigation since data of relevant state variables for testing and evaluation are not commonly available.

Critical to the evaluation, improvement, and use of crop models is the availability of high quality data from field observations. There is a mismatch between the rising demand by users for tested models and research budgets for suitable experimental research and monitoring, which tend to be decreasing (Rötter et al., 2011).

Since field experimental data sets are usually not recorded for modelling purposes, their level of detail, quality of records, variables considered as well as their number of spatial and temporal replicates vary enormously (Nix, 1985; Groot and Verberne, 1991) Therefore, their suitability for modelling is often insufficient for different reasons. White et al. (2013) proposed a standard approach for describing and identifying variables of management, environmental conditions, soils, and crop measurements, all for the purpose of developing, testing, and applying crop simulation models. In general, datasets used for model calibration and validation consist of data describing a) the initial soil conditions, b) the crop-specific management and c) the seasonal weather conditions (Palosuo et al., 2011; Rötter et al., 2012; White et al., 2013). Additionally, data on phenology of the crop, yields and nutrient contents from intermediate harvests, intra-seasonal soil conditions and measurements of fluxes of energy, water and CO_2 may be provided.

The international community of agricultural system modellers, e.g., in the Agricultural Model Intercomparison and Improvement Project AgMIP (Rosenzweig et al., 2013) or the European MACSUR (**M**odelling European **A**griculture with **C**limate Change for Food **S**ec**UR**ity) project (Rötter et al., 2013) are currently building harmonized data bases for the purpose of model testing and improvement including the opportunity to create model-specific interfaces for various models (Porter et al., 2014). In order to find suitable experimental data for specific applications in the context of modelling out of the vast offer of available data sets, a transparent method of screening and pre-selection is demanded, which highlights specific positive and negative features of a data set with respect to the intended application. To evaluate and select data sets, Rosenzweig et al. (2013) proposed different classes of data for so-called "Sentinel Sites", which represent specific sites with experimental data suitable for different levels of model testing and improvement. However, specific for a transparent classification were not provided. A joint community effort lead to the development of a qualitative (Boote et al., 2015) and a quantitative framework (this publication) to evaluate the quality of field experimental data sets for crop modelling according to robust and accepted criteria.

The aim of this paper is to provide a quantitative classification framework by which the consistency and quality of agricultural datasets can be evaluated. Variables under consideration are weighted according to both their importance and their quality, and justified by literature describing variance and errors of the different state variables and measurement methods. The objective of such a classification framework of data evaluation and labelling is (i) to allow data base managers to pre-check the quality of data sets before integrating them into their data base, (ii) to support the creation and use of international publicly available benchmarked data sets for model evaluation, inter-comparison and improvement, (iii) to enable modellers to select appropriate data according to their requirements, (iv) to give guidance to experimentalists for designing their experiments with respect to aspects that go beyond their primary research question, allowing for a broader use of experimental data for systems analysis and modelling.

2. Definitions and terminology

Parameterisation means the estimation of fixed model parameter values (e.g., diffusion coefficient of a substance in water) for single processes under controlled conditions.

Calibration means the adjustment of values of model parameters outside the model code (e.g. thermal sums for phenological development in external parameter files) to fit their output to a set of measured state variables or fluxes (Penning de Vries and van Laar, 1982). According to Van Keulen (1976), the main purpose of calibration is to adapt weak or unknown parameters or relations. Parameter values should preferably have a real and measurable background and their values should be adjusted within a reasonable range. Calibrated parameter values are valid only for the model configuration that was used for the calibration. Introducing new processes or algorithms usually requires re-calibration of at least parts of the parameter set.

Validation (or *falsification* if model application is beyond its limits) is the examination whether a model derived from analyses of some systems is capable of describing other systems, or simply, the test of a calibrated model against an independent data set that has not been used for calibration (De Wit, 1982).

State variables represent the status of a specific dynamic system variable (e.g. soil water content) at a particular time and location or compartment. The variable can be expressed as a total amount or concentration in a pre-defined compartment (e.g. soil layer or crop organ).

Fluxes are defined as a transition of matter or energy across a defined compartment border. They can be observed cumulatively over a specific time period and compared to corresponding simulations.

3. Data requirements for model calibration and validation

Application of a model in a new geographic/climatic environment or for a new crop requires new parameterisation and eventually modifications of the model, e.g. by consideration of additional processes; otherwise parameter adjustment to fit observed data becomes a pure tuning or curve fitting exercise (De Wit, 1982). Such extension of a model requires suitable data to identify and parameterise processes and it sometimes requires a re-calibration of parameters of other processes or modules if processes interact strongly.

Calibration and validation of agro-ecosystem models require consistent data sets of multiple state variables and fluxes (Nix, 1985). For field research, the assessment of "effective" parameters representing a whole field is more suitable than measurements in the laboratory using samples from single spatial points, which might represent only a few centimetres. High spatial as well as temporal resolution of measurements is required to understand, how processes at a more detailed scale affect particular phenomena at the scale of interest (Wagenet, 1998). While spatial variability complicates model calibration if the underlying spatial structure of inputs is unknown, the knowledge of these structures also provides a chance to evaluate the sensitivity of a model's response to variable inputs in a field, which often contains most of the variability of a whole landscape (De Wit and van Keulen, 1987). Although model applications have demonstrated and contributed to a better explanation of spatial patterns, Baveye and Laba (2014) criticise the missing benefit of such analyses. Nevertheless, model evaluations (estimating the discrepancy between model results and measured values) must consider the uncertainties of inputs as well as the small-scale variability of observations.

The parameterisation and calibration, e.g. for new crops, cannot be seen isolated from the relevant boundary conditions, e.g., soil hydrological conditions, and from the purpose of application. For example, it makes no sense to adjust crop model parameters to fit only data of annual yields, if soil water dynamics are not realistically simulated, since this may lead to erroneous parameter settings in another part of the model and erroneous predictions in a new environment. Kersebaum et al. (2007) documented that different models applied to the same data set were able to satisfactorily simulate a specific target variable, but showed considerable differences in the quantification of different underlying processes. Therefore, a sound calibration of a model requires using a balanced data set which includes observed state and flux variables representing as many of the processes and states of the model as possible at temporal (and spatial) resolutions that allow process parameters in the model to be adjusted and model assumptions to be tested.

Many processes show high spatial and temporal variability. Denitrification for example requires a coincidence of suitable temperatures, anoxic conditions and the presence of nitrate at a certain location. These conditions may vary within very small distances of millimetres to centimetres leading to a high spatial variability of measurements (Parkin, 1987). To integrate this variability to a larger space, e.g. field scale measurements of fluxes or crop biomass should cover a representative elementary volume or area to avoid too much noise in the observations due to effects at smaller scales (Wagenet, 1998; Boote, 1999). Also the resolution of vertical extent should be adequate for the scale, the problem to be addressed and the model approach used. Besides the number and accuracy of observed variables, one important point is the variability of experimental conditions occurring at a particular location and year, related to the specific interactions between genotype (crop/cultivar), environment (climate, soil) and agronomic management practices.

3.1. General framework of the data set classification

Four classes were defined to classify the datasets by quality: From low to high quality, these classes encompass the levels "copper", "silver", "gold" and "platinum" oriented to the classes proposed by Rosenzweig et al. (2013) and extended by Boote et al. (2015) for the classification of sentinel sites. Data sets of all classes are useful for modelling as even in the copper class they fulfil the minimum requirements of models. However, depending on the purpose of their use by modellers they differ in suitability. Data sets of the copper and silver class are usually sufficient to validate or to adjust key parameters for crop varieties when the model is already parameterised and calibrated. Model improvement, e.g. to identify new processes or calibration for a new crop require more specific data and usually data of higher density in terms of temporal or spatial resolution or more compartments, e.g. crop organs.

In our effort to develop a framework to evaluate data sets for modelling, we first differentiate between input data and state variables corresponding to model outputs to calibrate or validate the model and to prove consistency of model outputs against measured values. Additionally, we aim to address possible additional measurements that might be suitable to improve models for particular purposes. However, we have created the framework mainly from the aspect of crop modelling and it should be stressed that specific research questions usually require additional measurements. The framework encompasses six thematic groups of input variables, which are also denoted as blocks since they are evaluated separately (Table 1): meteorological data, agronomic management, soil data, initial values, previous crop, and topography.

In addition, the following thematic groups of state variables (Table 2) are needed to be compared with model outputs: phenology, crop growth variables, soil variables, and additional observations. Since some phenological stages are regarded as essential information to adjust crop parameters for a new environment (e.g., Asseng et al., 2013) phenology is evaluated as a separate block, while the other three groups are aggregated into one block (state variables, fluxes and observations) for the evaluation.

Within these groups, 65 potential variables were listed (see Tables 1 and 2) which describe the experimental agricultural datasets. For each variable, an estimation of their importance for crop modelling was determined from modeller interviews using a questionnaire. According to the selected importance, a relevance factor (RF) (5 = very important, 4 = important, 3 = relevant, 2 = special interest, 1 = additional value) was assigned to each variable (Tables 1 and 2). For instance, in crop modelling the weather variables Tmax and Tmin (the daily maximum and minimum temperature) are of major importance for many processes considered in the models. Thus, the variables get the highest relevance (5), whereas Tavg (daily average temperature) has a lower predictive power and was assigned the relevance of 3. Additionally, the relevance factors are modified individually to cover aspects of temporal and spatial resolution, representativeness, and accuracy of observation methods to estimate weighting points (WP), which are accumulated for the evaluation. The general scheme to calculate weighting points WP is:

$$WP = RF \cdot f(n_{ob}, no, d, rep, xyz) \tag{1}$$

where the elements which are considered (n_{ob} represents the number of observations during a year/season, no means the number of compartments/layers, d indicates the depth of the soil profile, rep the number of replications of the observations, and xyz the distance and elevation difference between experiment and meteorological station) in the functions vary depending on the variable.

For each block of variables and each class a minimum amount of information was defined by a sum of total weights (Fig. 1). Absence of a variable defined as a necessary part of the minimum data requirements (= essential data) leads to a total failure of the data set unless alternative information (e.g. global radiation can be substituted by sunshine duration) can be used. This means that the data set is not suitable for modelling or requires additional measurements. Classification of the whole data set is performed by either using the total sum of all variable weights or the average of the achieved classes (values from 1 for copper to 4 for platinum

Table 1
Input variables for agro-ecosystem modelling (# = minimum data requirement, (#) can be substituted by (#2)).

Variable	Relevance	Weight condition	Factor for total weight	Max. weight
Meteorological data				
Precipitation#	5	≤1 km dist.	MIN(f{dist.}; f{altitude}	5
		≤30 m altitude	see. Fig. 2, Eq. 2	
Air temperature average	3	≤5 km dist.; dto.	dto.	3
Air temperature minimum#	5	≤5 km dist.	dto.	5
Air temperature maximum#	5	≤5 km dist.	dto.	5
Rel. humidity/Tdew/vapor pr.#	4	≤5 km dist.	dto.	4
Wind speed	3	≤10 km dist.; dto.	dto.	3
Global radiation (#)	5	≤20 km dist.	f{distance}	5
Sunshine duration (#2)	3	≤ 20 km dist.	f{distance}	3
Agronomic management				
Variety	3	0/1	0/1	3
Sowing/planting date#	5	0/1	0/1	5
Harvest date#	5	0/1	0/1	5
Fertilization#	5	date, amount, type	0/1	5
Rainfed/irrigation#	5	date, amount	0/1	5
Sowing density	3	0/1	0/1	3
Tillage	2	date, depth, type	0/1	2
Soil data				
Texture#	5	0–150 cm	*MIN(no. layers/3; 1.2)	7.5
		>2 layers	*MIN(depth/1.5; 1.25)	
Bulk density	3	0–150 cm	dto.	4.5
Water retention curve	3	0–150 cm	dto.	4.5
Hydraulic conductivity curve	3	0–150 cm	dto.	4.5
Field capacity meas.	3	0–150 cm	dto.	4.5
Wilting point meas.	3	0–150 cm	dto.	4.5
Soil organic carbon#	3	0–30 cm	*MIN(no. layers; 1.2)	4.5
			*MIN(depth/0.3; 1.25)	
soil organic nitrogen	3	0–30 cm	dto.	4.5
pH	2	0–30 cm	dto.	3
Initial values				
Soil moisture#	4	0–100 cm	*MIN(no. layers/2; 1.2)	6
			*MIN(depth; 1.25)	
Soil mineral N	4	0–100 cm	dto.	6
Previous crop				
Crop	3	0/1	0/1	3
Sowing/planting date	2	0/1	0/1	2
Harvest date	3	0/1	0/1	3
Yield	2	0/1	0/1	2
Residue management	4	0/1	0/1	4
Fertilization	3	total amount	0/1	3
Irrigation	2	total amount	0/1	2
Topography				
Latitude	5	0/1	0/1	5
Longitude	3	0/1	0/1	3
Altitude	3	0/1	0/1	3
Slope/exposition	1	0/1	0/1	1

were attributed to calculate a numerical average) of the single blocks (Fig. 2). This should ensure that data sets are well balanced and relevant deficits in one block of input variables are not masked by very detailed information of another block. Thus, by listing all variables given in an experimental dataset and multiplying these variables with the above-mentioned weights, the dataset automatically becomes classified into one of the four quality classes.

However, the results so far just represent only the data of one season (or the average of several seasons). To consider the added value when data of multiple seasons and/or different treatments of a manipulated experiment are available, we created an additional block with this information (see Table 3). In this block the number of growing seasons per crop (e.g. if shifted crop rotations were run in parallel) and of applied treatments are considered. Ideally, the same cultivar of the crop should be used at multiple seasons, sites, or treatments, which could lead to deviations due to genetic coefficient variation. Additionally, data sets are upgraded if at least one treatment comes close to representing “no stress” conditions for crop growth which is seen as ideal for model calibration as stresses might alter crop phenology and biomass partitioning, e.g. root/shoot ratio (e.g. Blum, 1996). Also the added value of a treatment without nitrogen (N) fertilisation is positively regarded as this provides information on the basic N supply by soil organic matter or crop residue mineralisation and the response of crops to limited N availability. Elevated CO_2 treatment could be given higher rating if compared to ambient. If crop growth can be related to different site conditions, e.g., if different soil information is available from precision agriculture observations or an experiment is performed within parallel lysimeters with different soils or hydrological boundary conditions, the data exhibit a special value for model sensitivity testing. Weight of a specific treatment factor is calculated from the number of the specific variants. However, weights of different treatment factors are considered in an additive instead of multiplicative way. The sum of this block is used to calculate a factor that is used to modify the total sum as well as the average of block classification to estimate the final rating of the data set. In most cases, the data sets are upgraded for their additional focus data, seasons or treatments. However, it may also be downgraded if there are no additional seasons and treatments, but the experiment indicates that water, nutrient availability, weeds, and pests or diseases limited crop growth. Although the number of specific treatments is not limited (see Table 3), realistically the sum of treatment

Table 2
State variables, fluxes and observations for model testing ($^{\#}$ = minimum data requirement).

Variable	Relevance	Weight condition	Factor for total weight	Maximum weight
Phenology				
Emergence	3	0/1	0/1	3
Tillering/stem elong./other	2	0/1	0/1	2
Ear emergence/other	2	0/1	0/1	2
Flowering$^{\#}$/other	5	0/1	0/1	5
Maturity/other	3	0/1	0/1	3
Crop growth				
Yield	5	0/1	+MIN(no.repl.-1)*0.05); 1)	6
Above ground biomass	4	0/1	*MIN(no. observ/3; 1.25) *MIN(no. Replic/3; 1.2)	6
Weight single organs (leaves, stems, etc.)	3	0/1	*MIN(no. organs/3; 1.25) *MIN(no. observ/3; 1.2) *MIN(no. Replic/3; 1.1)	4.95
Root biomass	3	0/1	dto.	4.95
Nitrogen in above ground biomass	3	0/1	*MIN(no. observ/3; 1.25) *MIN(no. Replic/3; 1.2)	4.5
Nitrogen in single organs	3	0/1	*MIN(no. organs/2; 1.25) *MIN(no. observ; 1.2) *MIN(no. Replic/3; 1.1)	4.95
Leaf area index	4	0/1	*MIN(no. observ/3; 1.25) *MIN(no. Replic/3; 1.2)	6
Soil				
Soil water gravimetric	3	0/1	*MIN(no. depths/3; 1.25) *MIN(no. observ/5; 1.2) *MIN(no. replic/3; 1.1)	4.95
Pressure heads	2	0/1	*MIN(no. depths/3; 1.25) *MIN(no. observ/20; 1.2) *MIN(no. replic/3; 1.1)	3.3
Soil mineral nitrogen	3	0/1	*MIN(no. depths/3; 1.25) *MIN(no. observ/3; 1.2) *MIN(no. replic/3; 1.1)	4.95
Soil water calibrated sensor	3	0/1	*MIN(no. depths/3; 1.25) *MIN(no. observ/50; 1.2) *MIN(no. replic/3; 1.1)	4.95
Deep percolation flux	3	0/1	*MIN(no. obs./10; 1.25)	4
Deep nitrogen leaching	2		*MIN(no. obs./10; 1.25)	2.5
Surface fluxes				
Evapotranspiration	3	0/1	0/1	3
CO_2 flux (CO_2 exchange)	3	0/1	0/1	3
NH_3 flux (volatilisation)	2	0/1	0/1	2
N_2 flux (denitrification)	2	0/1	0/1	2
N_2O flux	2	0/1	0/1	2
NO flux	2	0/1	0/1	2
CH_4 flux	2	0/1	0/1	2
Additional observations (qualitatively)				
Lodging	3	0/1	0/1	3
Pest/disease effects	3	0/1	0/1	3
Weeds	3	0/1	0/1	3
Physical damages (frost/hail/animals/etc.)	3	0/1	0/1	3

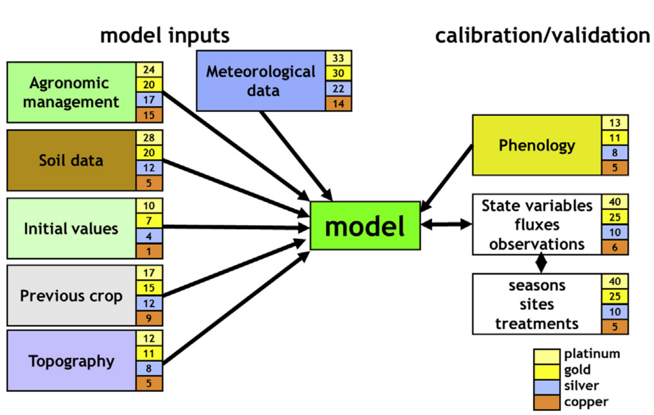

Fig. 1. Variable blocks and their minimum sums of weighting points required for each quality class.

weights will rarely lead to an upgrade over two classes. However, the factor, which is used to multiply the total sum of points and the average of block ratings, is limited to a range from 0.94 to 1.3.

The importance of specific input variables depends on their ecological significance but also on the model assumptions, the model architecture and the main interested output variable or research question. Although we provide a method here with suggested default weights for crop modelling, these weights may be subject to individual choices as demands vary depending on the specific purpose of model development and applications. Therefore, the realisation of the classification software offers the opportunity for users to change the individual weighting for variables (see example in Fig. 4). Thus, one dataset can end up with multiple rankings, and that reflects its suitability for different purposes.

The dataset ranking software was programmed as a Qt desktop application in C++ code compiled under Windows 7 and includes three components: "GUI", "Datasets" and "RankPointGenerator". The dataset classification software contains register cards, where all metadata defined in this chapter needs to be filled in manually. The

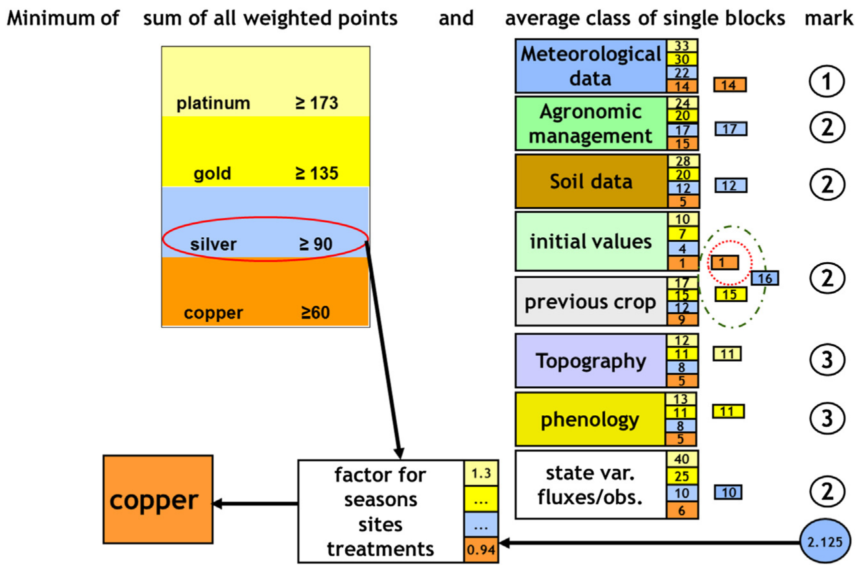

Fig. 2. Final classification scheme based on the minimum of total sum and average label of single blocks (downgraded to copper because no treatments and multiple years and not optimal growth conditions).

dataset is ranked per block as soon as data is parsed in. This metadata can be stored per evaluated dataset in an ASCII file using the JSON format (www.json.org). This data format is a language independent data interchange format. Therefore, the format can be used as an interface to automatically generate the required metadata from any data base.

3.2. Relevance of data and requirements to data

3.2.1. Input data

Quality and uncertainty of input data have a high impact on model output uncertainty and accuracy. In our evaluation framework we differentiate between different types of input data (Table 1):

3.2.1.1. Meteorological data. Meteorological input data are usually the drivers of dynamic agro-ecosystem models. Sensitivity of models to single meteorological variables differs depending on the simulated model processes. Therefore, the weight of the single variables differs in our framework (Table 1). However, weather variables have different spatio-temporal variability which has to be considered when evaluating their suitability for site-specific modelling.

3.2.1.2. Agronomic management. Data on soil and crop management are generally essential inputs for crop models. Sowing or (trans-) planting date is essential to initiate crop growth. Harvest date is required by many models to terminate the growing season. Information about water and nutrient (especially N) supply are required as both are important resources which determine crop growth. In the framework, we just evaluate if the information is available (Table 1), e.g., even to confirm rain-fed conditions. Other information such as variety, sowing density and tillage are less important.

3.2.1.3. Soil data. Soils influence crop growth in several ways, in particular as soil acts as an important storage and buffer especially for water and N. Soil texture is one of the most influencing factors for sorption properties, pore size distribution and water retention characteristics and is given a very high relevance factor (Table 1). Water availability is among the most important input information for crop models. The available water capacity is one of the most

Table 3
Calculation of multiplier for multiple seasons, treatments and optimum conditions.

Variable	Relevance	Weight condition	Factor for total weight	Max. weight
Seasons				
No. of seasons per crop	1	0/1	*MIN(no. seasons; 5)	5
Sites				
No. of site variants (e.g. soils)	1	0/1	*MIN(no. site variants.; 5)	5
Treatments/management				
Management close to optimum 1 (limited) to 5 (potential growth)	1	1–5	* value	5
zero N treatment (no = 0, yes = 1)	3	0/1	* value	3
Treatment 1 (e.g. irrigation)	1	0/1	* no. treatments	No limit
Treatment 2 (e.g. nitrogen fert.)	1	0/1	* no. treatments	No limit
Treatment 3 (e.g. CO_2)	1	0/1	* no. treatments	No limit
Treatment 4 (e.g. tillage)	1	0/1	* no. treatments	No limit
Treatment 5 (…)	1	0/1	* no. treatments	No limit
Treatment 6 (…)	1	0/1	* no. treatments	No limit
Factor estimation				
SW = MIN (Sum of weights;28)			**1+ (SW/7 -1)*0.1**	**1.3**

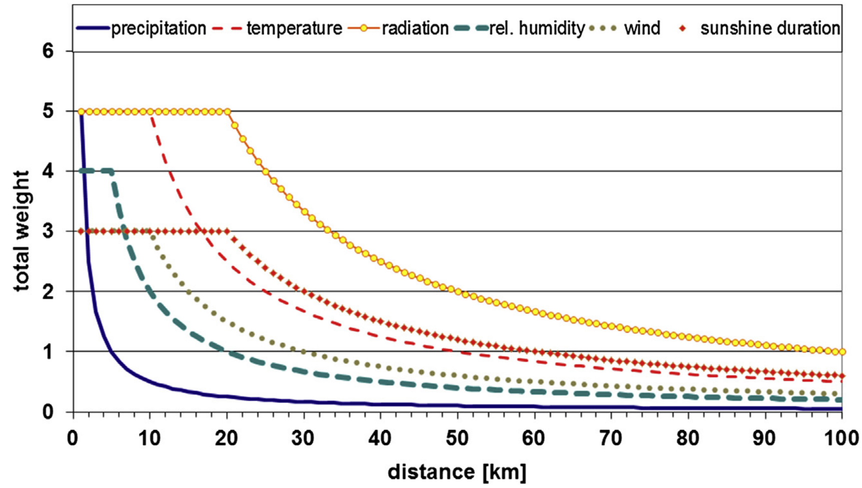

Fig. 3. Decrease of relevance for data classification of important weather variables with distance between experimental field and weather station.

sensitive yield-affecting soil properties under rain-fed conditions with significant impact on crop model output especially when water is limiting (van Keulen and Seligman, 1987; Aggarwal, 1995; Pachepsky and Acock, 1998). Beside texture, bulk density of the soil plays an important role, as compaction mainly affects the mid-size pore space which determines plant available water. Estimation of soil water retention curve and soil hydraulic conductivity functions is required for models which simulate water dynamics based on the Richards equation (Richards, 1931). However, many models are operating with a capacity approach which requires only the soil water content at field capacity and permanent wilting point.

Knowledge of soil organic carbon (C) and N are relevant inputs to assess C and N pools for C and N mineralisation (Table 1). Under limited N fertilisation typical for agriculture in low-income countries of Sub-Saharan Africa, information about total soil organic C and N is specifically important since soil organic matter decomposition is the major source of plant available N. This is also the case for many organic farming systems in developed countries (Doltra et al., 2011). A remaining problem is the fractionation of the total soil organic C and N into active versus stable soil organic C and N pools, which are used differently by various models (Gabrielle et al., 2002). Especially in sandy soils, soil organic matter contributes

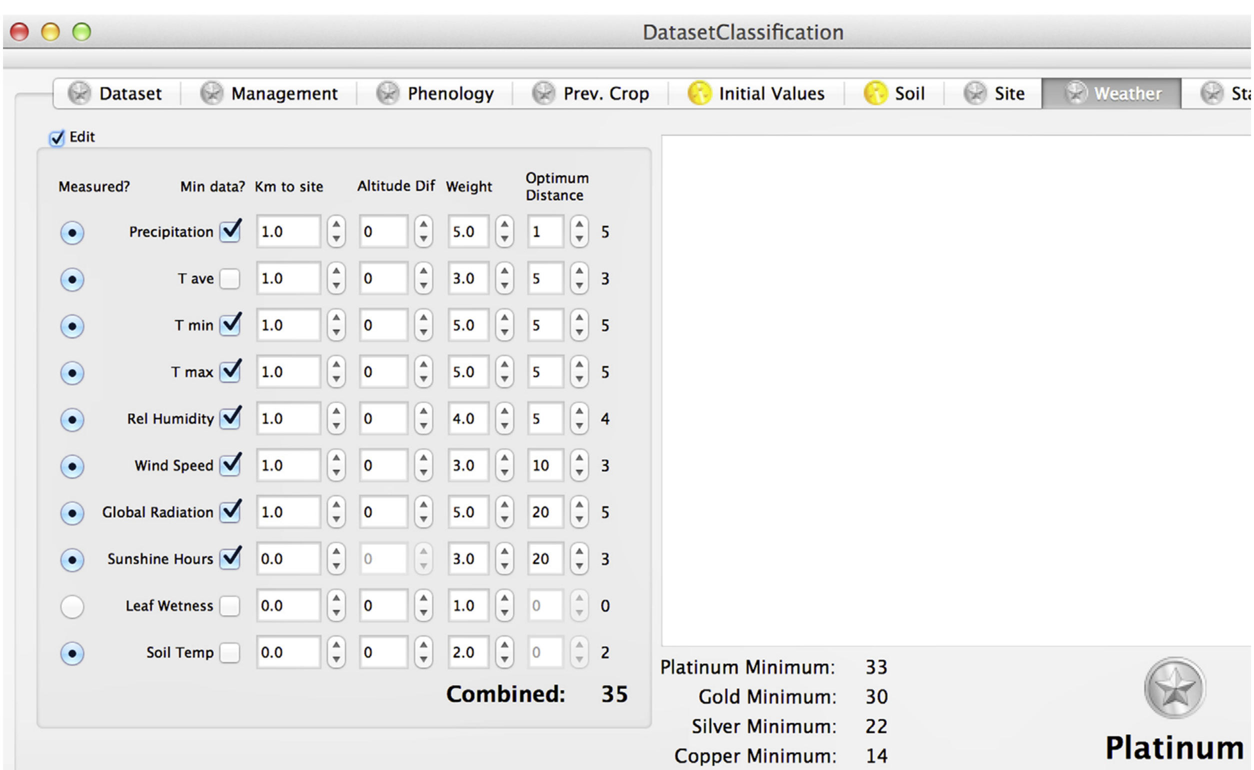

Fig. 4. Example of a sub-page for meteorological variables of the classification software representing the data situation for Müncheberg 1993–1996 and Lednice (except soil temperature).

significantly to the soil cation exchange capacity and influences its water holding capacity.

Soil pH has an impact mainly on biological turnover processes and some nutrient dynamics, e.g. phosphorus (Deveau et al., 2009). The weight of pH in the framework is relatively low (Table 1) as soil pH in well managed arable soils is usually not at a critical level. However, very low pH values typical of tropical soils can cause severe limitations to root growth and might require raising the weight for these soils (Marschner, 1991).

3.2.1.4. Initial values. As dynamic simulation models predict state variables for a specific time step from their status of the previous time step, the models require initialisation for all variables with reasonable start values prior to the simulation. Most crucial are soil water and soil mineral N contents because of their high spatio-temporal variability. There is a high weight given to knowing initial nitrate and ammonium in soil layers at pre-season if the crop is grown under limiting N fertilisation (Table 1). Considering the importance of water and N availability, the initial conditions of soil water status and soil mineral N have a strong influence, especially on early crop development and growth. If soils are dry, germination can be delayed significantly (Ashraf and Abu-Shakra, 1970). Also low N contents, especially in combination with dry topsoil can cause distinct N and water stress in the juvenile phase when the root system is poorly developed. However, the impact of initial conditions varies depending on the site properties (Kersebaum et al., 2002). For example high initial soil mineral N before sowing of a winter crop might have a minor contribution to crop growth and yield formation if N is leached during autumn and winter in soils of low water holding capacity. However, if N leaching is a major aspect of the simulation study, the pre-sowing mineral N content is of high relevance. In our framework, these data can be substituted by good information on previous crop and its cultivation. Other state variables like organic C and N pools can be derived from more conservative soil characteristics such as soil organic matter content.

3.2.1.5. Previous crop. The status of water and N for the simulation could be simulated by starting the model one growing season earlier with assumed initial data if observed initial values are not available. The error of this assumption will mainly disappear during this pre-simulation if the model is capable of properly simulating the dynamics. However, this requires data regarding the growing period of the previous crop and the method of residue management, which makes this data relevant for modelling (Table 1). For the simulation of long-term C and N dynamics in soil and its feedback on the growth dynamics of a crop rotation system, a pre-run of several decades to centuries may be necessary to achieve steady state conditions in the virtual pools (Bruun and Jensen, 2002). In this case, assumptions on the history of crop management including the return of crop residues to the soil also have to be made.

3.2.1.6. Topography. Information regarding the location of the experimental site such as latitude, longitude and altitude are particularly important if models calculate process-relevant status variables such as day length from latitude and day of year or correct other input data due to deviations of their locations, e.g. if weather data come from another location or altitude. Information on slope and exposure provides information if topographic shading or lateral water flow might affect the site (Reuter and Kersebaum, 2009).

3.2.2. State variables for model testing

Model performance is usually tested by comparing simulated against measured state variables or fluxes, preferably measured with an appropriate temporal frequency and spatial resolution (Table 2).

3.2.2.1. Phenology. Since phenological data provide an essential pre-information to adjust crop variety specific parameter values when applying models for new varieties or new sites, we assign these data within a separate block for evaluation. Phenology is crucial information for modelling of a crop. It controls a multitude of physiological processes in the plant, and, in turn is almost entirely controlled by external climatic and soil variables such as temperature and photoperiod, drought or N availability (van Keulen and Wolf, 1986; van Keulen and Stol, 1991; Olesen et al., 2012). This usually makes phenology the first process that is considered during the model calibration procedure. The ability to reproduce the plant's ontogenesis with the model largely determines the success of modelling other plant traits (Jeong, 2012). Phenology plays a central role in the partitioning of assimilates. Additionally, sensitivity to stresses can be different during specific growth stages. Therefore, it is important to have information at least on some key phenological stages. Flowering is very important (Table 2) for many grain crops as it initiates the period of grain filling. It also marks a stage which is very sensitive to stress, e.g. drought or heat (Savin and Nicolas, 1999; Semenov and Shewry, 2011; Eitzinger et al., 2013). Also maturity date provides relevant information on the termination of plant growth. We provide more detailed considerations on phenological observations in section 4.5.

3.2.2.2. Crop growth variables. Crop yield is one of the most relevant variables for crop model simulations (Table 2). Often reported as dry matter yield (e.g. grains for cereals), but for some crops (e.g. field vegetables, sugar beet or sugarcane) fresh matter yield is a more useful unit to work with and the conversion from one unit into the other (i.e., dry matter content) requires special attention (Nendel et al., 2009). Total crop biomass and its partitioning to different crop organs over time is of special interest to calibrate assimilate partitioning among organs in growth models (Kersebaum, 2011; Boote et al., 2015). Within many crop growth models a specific leaf area (area per unit mass) is used to transfer leaf weight into leaf area index (LAI), which forms the interface to adsorb solar radiation for photosynthesis. Therefore, measurements of LAI play an important role to control crop model behaviour. The vertical distribution of root biomass (and related root length density) is relevant information, which can be directly compared to corresponding model variables. Additionally, it indicates the soil depth exploited by the crop for water and nutrient uptake.

Crop nutrient status of above-ground biomass and of crop organs (including roots) over time provides information on the integrated nutrient fluxes in and from the crop at a specific growth stage of the plant and the distribution among crop organs. In conjunction with a stepwise increase of nutrient supply, subsequent measurements of the nutrient concentration in plants can provide useful information to derive critical nutrient concentration curves for assessing nutrient stress at specific growth stages (e.g., Greenwood et al., 1990; Plénet and Lemaire, 1999).

3.2.2.3. Soil variables. A central part of agro-ecosystem models is the simulation of soil water balance including temporal dynamics of soil water contents, and upward and downward fluxes. Crop growth and soil water and N dynamics are closely interlinked and can hardly been simulated independently from each other. To prove the validity of a crop growth model it has to be insured that the simulation of biomass production is based on the right boundary conditions in the soil. Depending on the model approach, soil water content measurements are relevant to test models in that aspect, whereas tensiometer measurements of matric potential are of special interest for potential-driven models (Table 2).

Flux measurements at the bottom of a soil profile – usually deeper than the rooting depth – are mainly restricted to the use of lysimeters or accumulated measurements by ion exchange resins (Lehmann et al., 2001). These measurements are required to evaluate simulated seepage and nutrient losses by leaching, which are relevant components of the whole water and nutrient balance and important environmental indicators.

3.2.2.4. Surface fluxes. Evapotranspiration is a key process in agroecosystem modelling. Therefore, model testing with measured water fluxes at soil or canopy level is quite important (Table 2). Measurements of gas exchange at the soil surface are especially important for environmental impact analysis, e.g. the emission of trace gases (CH_4, N_2O, CO_2) or to quantify processes related to N dynamics such as ammonia volatilisation (NH_3) and denitrification (N_2, N_2O, NO) for model calibration. CO_2 flux measurements provide vital information for testing the models' ability to describe short-term C dynamics. In the context of climate change impact assessment and model-based development of mitigation strategies, the testing of appropriate model approaches becomes increasingly important, and therefore there is an increasing need to test agroecosystem models against flux measurements.

3.2.2.5. Documentation of damages and yield losses by external factors. Most crop models do not consider crop damages by external factors (e.g. pest and diseases, lodging, hail, frost, radiation damage, damage from animals, etc.). Large deviations between observed and simulated yield can occur especially if the risk increases with increasing yield level as for cereal lodging. Therefore, such events should be recorded to explain deviations between observed and simulated biomass/yield and documentation is rated as relevant (Table 2).

4. Quality and uncertainty of observations

Field data are never free of error (random or bias), which challenges model calibration and validation (van Keulen and Seligman, 1987). This constrains desired stringency in model evaluation: If random error were the only source, it would, as a rule make sense to consider model results satisfactory, as long as they are within the standard error (SE) of the data, because the SE is the statistical measure of accuracy. However, SE provides no information on bias, so there will always be uncertainty in the judgement (van Keulen and Seligman, 1987). The number of replicates for measurements is therefore considered positively in the calculation of the weighting and conservation of replicate data is emphasized. Three replicates are assumed to be the standard for most of the measurements. Therefore, less replicates lead to a reduction of weighting according to the factor for total weight given in Table 2.

The quality of measurement can be related to the type of instruments used, thus measurement methodology and instruments used should be reported. However, the quality of instruments is not taken into account here since there are multiple opportunities to measure a specific state variable or flux and technical suitability may vary depending on site conditions. In addition, there is less value for measurements that have high variability (uncertainty), which is possible to document if replicate values are given, or if measurement error is reported.

4.1. Meteorological data

The representativeness of meteorological measurements for an experimental site must be considered in the evaluation of data as spatial variability of specific variables reduces their usefulness if not directly measured at the site. To qualify for "platinum" or "gold" meteorological measurements should be within the optimal distance of each variable (Fig. 3).

The network of synoptic stations in Europe, used for weather forecast models, has a typical inter-station distance of 50–90 km (van Diepen and van der Voet, 1998). Spatial variation in weather and climate may be larger than between two stations. For field experiments the nearest weather station must capture the same weather condition. Phenomena like thunderstorms and cumulus convection have a horizontal dimension of up to 10 or 20 km. The deviance in the meteorological elements from the local weather at the field site increases with the distance to the weather station and with the difference in altitude (Zhao et al., 2015).

Precipitation often exhibits significant spatial variations within a region (Baigorria et al., 2007; Tetzlaff and Uhlenbrock, 2005). Small-scale variability in annual precipitation can be in the order of 8% (Subedi and Fullen, 2009). Zhao et al. (2015) estimated very high spatial variability of precipitation within a few kilometres even in moderately heterogeneous regions. To achieve a deviation (root mean square error) below 200 kg ha^{-1} of water limited crop biomass by crop modelling, resolutions of 1–10 km were necessary in a predominant part of Germany. For recording the amount of water which can be used by the crops, rainfall measurements should be made as close as possible to the experimental site. We suggest a distance within one kilometre as realistic for a good representative of an experimental station. In general, wind speed measurements must be close enough that surface wind is driven by the same meso-scale driving airflow. Wind measurements made according to WMO-rule at 10 m over open terrain are synoptically comparable wind speeds and can describe wind in a region of at least 5 km diameter (Wieringa, 1998b, Table 1).

A comparison of ground-based global radiation measurements with satellite-derived hourly global irradiance was made for Switzerland, the Northeast US (Zelenka et al., 1999) and Germany (Schroedter-Homscheidt et al., 2006). The analysis shows that global radiation measurements up to 20 km around a station are more accurate and representative than satellite data, provided that surrounding obstacles remain below a few degrees above the horizon (Wieringa, 1998a). Measured ground-based global radiation is, in many cases, unavailable, and this data is frequently replaced by sunshine duration (Campbell-Stokes), cloud cover or diurnal temperature range (sometimes combined with precipitation). Although the reliability of sunshine duration (hours of bright sunshine) as a global radiation predictor is rather high (Trnka et al., 2005), under specific conditions it can significantly alter agroecosystem model performance (e.g. Trnka et al., 2007). On this basis, measured global radiation gets a higher weight compared to sunshine duration.

Fig. 3 shows the decrease of relevance of different weather variables with distance between experimental field and weather station. An altitude effect on weight points WP_z of the variable is calculated if the altitude difference (ΔZ) between experimental plot and weather station is $\geq$30 m:

$$WP_z = \mathrm{Max}\,(RF - 1^*(\Delta Z/30);\ 0) \quad (2)$$

The minimum of both weights from distance and altitude difference is used to calculate the final weighting points WP (Table 1).

4.2. Basic soil data

Crop response to climatic drivers strongly depends on site conditions, in particular soil properties (Kersebaum and Nendel, 2014). Spatial variability of soils is often high even within fields leading to spatial variability of crop yields (Nielsen and Bouma, 1985; de Wit and van Keulen, 1987; Kersebaum et al., 2002, 2005).

Soil texture can vary within very short distances depending on the sedimentation conditions of the parent material, relief or erosion processes. Soil organic matter is often related to clay content, but is especially prone to re-distribution due to erosion processes leading to reduced soil organic matter at shoulders and backslope and accumulation at footslope (Moore et al., 1993). Consequently, related soil properties, especially water characteristics, relevant for crop growth show a similar pattern (Kersebaum et al., 2002). Vertical distribution of texture and bulk density can affect crop growth mainly if root penetration is limited to shallow soil layers. Stones in soils have to be considered as they reduce the pore space of a given soil volume.

Estimation of soil water retention and conductivity functions are labour-intensive and require a large number of samples due to their large variability. Water content at field capacity and wilting point can be estimated from laboratory measurements or can be determined in the field if soil moisture observations cover the whole range of wet and dry conditions. However, soil hydraulic properties are often derived from primary data like texture, bulk density and organic matter using pedo-transfer functions (e.g. van Genuchten, 1980; Wösten et al., 2001) with an inherent uncertainty (Heuvelink, 1998).

4.3. Soil water status

Methods for measuring soil water content range from gravimetry, and neutron moisture probes to new sensors mainly based on the relationship of soil dielectric constant vs. volumetric soil water content such as Time Domain Reflectometry (TDR) or Frequency Domain Reflectometry (FDR; e.g. Topp et al., 2003). Sensor-based automated measurements of soil water status, enable researchers to determine water contents in soil profiles rapidly and continuously over longer time periods with a high time resolution. Such measurement techniques can well determine the wet and dry extremes of soil water contents. Accurate knowledge is required of both the wet extremes because of their important contribution to water flow and, the dry extremes in order to validate models of root water uptake. Reviews on methods for measuring soil water status are given by Robinson et al. (2008) and Vereecken et al. (2008). A site-specific calibration of TDR and FDR-probes is often needed, improving the precision to a range from ± 0.01 to ± 0.05 cm^3 cm^{-3} (Abbas et al., 2011; Paige and Keefer, 2008; Wegehenkel, 2005; Mittelbach et al., 2012). Therefore, sensor data are not considered by the evaluation framework as “calibrated” if less than two gravimetric measurements were performed.

4.4. Soil nutrient status

Soil nutrient status can vary significantly spatially as well as temporarily. Especially soil mineral N has shown tremendous spatial variability within small distances (Kersebaum et al., 2002). Giebel et al. (2006) found that 35–49 % of soil mineral N variability within a 20 ha field derived from short distance variability (<5 m) and that spatial patterns across the field are temporarily unstable. Analytic error was estimated in the range of 5–10 kg N ha^{-1}; however, measurement uncertainties of 10–26 kg N ha^{-1} for the soil depth of 60 cm can be expected. Dahiya et al. (1984) estimated a required number of sampling points of 21–33 depending on the depth to attain a sample mean within 10% of the real mean value (85% probability). The number can be reduced to 5 to 8 if the precision limit is set to 20%. Therefore, three replicate mixed samples were set as a standard requirement in the evaluation method (Table 1). Compared to auger sampling, measurements of nutrient concentration in the soil solution using suction cups have the problem that the sampling captures a very small pore space which might fail to represent nutrient status at field level. Additionally, preferential flow pathways can lead to large deviations within small distances (Grossmann and Udluft, 1991), which may not be well represented by suction cup samplings.

In our method we have not considered explicitly other nutrients like potassium or phosphorus. However, tropical regions routinely have phosphorus deficient soils. The problem here is that most models lack response to P or K fertility, and even those that do, require soil tests for P and K as model inputs and the soil testing methodology varies considerably across regions. Nevertheless, the rating whether an experiment is close to optimum in terms of nutrient supply can be used to consider such deficiencies (Table 3).

4.5. Phenology

Observing plant phenology has two major challenges: first, site heterogeneity causes some plants within the stand to develop more rapidly than others (Elmore et al., 2012), resulting in a significant time lag between an individual plant's ontogeneses, especially at later developmental stages due to propagation of heterogeneous development (Zhang and Tao, 2013; Migliavacca et al., 2012). This makes phenological monitoring for a larger population of plants difficult and highly dependent on observer skills. Second, some crucial developmental stages cannot be detected visually and are therefore rarely recorded. The double ridge stage of wheat can only be observed with an apex specimen using a microscope. However, it marks a crucial stage at which generative development begins and photoperiodism starts to significantly influence further development of wheat. Some simulation models use this stage to control other physiological processes in the virtual plant (Cao and Moss, 1997; Nendel et al., 2011). Visually well detectable phenological stages are more often recorded (Meier, 2001). However, since extensive monitoring requires frequent presence at the growing site, most data sets include only a limited set of observed phenological stages and often based on less frequent observations.

The use of different phenometrics, such as BBCH (Lorenz et al., 1994), Feekes (1941) or Zadoks et al. (1974), may cause trouble when harmonising data sets, even more if different characteristics of a developmental stage are recorded. For grapevine phenology, bud break is commonly recorded in Germany at BBCH 09 (Lorenz et al., 1994), while in France, Czechia or Slovakia, on the contrary, it is commonly recorded at BBCH 07.

4.6. Plant biomass and nutrient uptake

Plant biomass measurements represent probably the most important variables both in calibrating and in validating crop models (i.e., yields, above and below ground biomass, leaf area, leaf mass, stem mass, grain etc.). In particular, yields represent the most important model output variables for model users, stakeholders, policy makes, etc. (Table 2). Unfortunately, most of the data collected on crop biomass are not obtained for modelling applications. Thus, very often only data at harvest time are measured. This allows their use, mainly for validation purposes, but limits their application for calibrating simulations of crop growth. Therefore, the weight in the evaluation process is reduced if less than three and increased by up to 25% if more measurements were performed during the season. Measuring biomass distribution to different organs over time during the crop life cycle is required to parameterise partitioning of assimilates in models (Boote et al., 2015). Doing this for new crops where stress response functions are not already known, requires data which are obtained under optimal crop growth conditions (Table 3), because biomass partitioning can be altered if crops are suffering from water or nutrient deficit (Blum, 1996).

Crop biomass and yield may vary considerably within small distances and across fields (Kersebaum et al., 2005). Therefore, the number of plants on the area of harvesting biomass samples has to be large enough to avoid variation due to missing plants or microscale variation (Boote, 1999). We therefore presume in the evaluation framework that three replicates are required to achieve full marks for yield and biomass measurements (Table 2). This holds for the determination of nitrogen uptake since it is based on biomass measurement. Further variation is added on N uptake due to the high spatial variability of soil mineral N especially in N-limited systems, e.g. in organic farming systems or smallholder agriculture in developing countries.

5. Application of the evaluation framework on datasets

To examine the outcome of the evaluation framework we applied it to three datasets used for various modelling studies with two periods at Müncheberg/Germany (1993–1995 and 1996–1998), and one at Lednice/Czech Republic (1992–2006).

The "Müncheberg" data set, described in detail by Mirschel et al. (2007), has been used (at least partially) for model inter-comparisons (Kersebaum et al., 2007; Palosuo et al., 2011). Tables 4 and 5 provide information for the two sites on data availability and the resulting rating using the default weights. Within the table we differentiate the data set into two periods, 1993–1995 and 1996–1998, because measurements of biomass, soil water and soil mineral N were reduced towards the end due to budget cuts, a typical situation in many countries. During the whole period the field trial was managed under both rainfed and irrigated conditions. As frequency of observations varied due to different crops, we estimated an average number for both periods. During the first period five intermediate harvests divided into stem/leaf, storage organ and root biomass were performed. Additionally, soil mineral N and gravimetric water contents were sampled 4 to 8 times per year. During the later three years biomass sampling was performed just once beside the final harvest (except sugar beet where 4 samples were taken) and soil sampling occurred two times per year. Record of soil moisture by TDR probes in different depth provided about 250–300 data points per depth for each year for both periods. However, measured TDR values had to be re-calibrated using the gravimetric values (see section 4.3).

Table 4
Input data availability for data sets from Müncheberg, Germany (2 periods) and Lednice, Czech Republic and resulting classification of input data blocks (# = minimum data requirement).

Variable	Relevance	Müncheberg 1993–1995				Müncheberg 1996–1998				Lednice			
			av	WP			av	WP			av	WP	
Agronomic management													
Variety	3		1	3			1	3			1	3	
Sowing date#	5		1	5			1	5			1	5	
Harvest date#	5		1	5			1	5			1	5	
Fertilization#	5		1	5			1	5			1	5	
Rainfed/irrig.#	5		1	5			1	5			1	5	
Seed density	3		1	3			1	3			1	3	
Tillage	2		1	2			1	2			1	2	
sum/class				**28**	**plat**			**28**	**plat**			**28**	**plat**
Soil data		**dep**	**lay**			**dep**	**lay**			**dep**	**lay**		
Texture#	5	2.0	6	7.5		2.0	6	7.5		1.5	4	6	
Bulk density	3	0.9	3	1.8		0.9	3	1.8		1.5	4	3.6	
Water ret. curve	3	0	0	0		0	0	0		0	0	0	
Hyd. cond. fct.	3	0	0	0		0	0	0		0	0	0	
Field capacity	3	0.9	3	1.8		0.9	3	1.8		0	0	0	
Wilting point	3	0.9	3	1.8		0.9	3	1.8		0	0	0	
Soil org. C#	3	0.9	3	4.5		0.9	3	4.5		1.5	4	4.5	
Soil org. N	3	0.9	3	4.5		0.9	3	4.5		1.5	4	4.5	
pH	2	0.9	3	3		0.9	3	3		1.5	4	3	
sum/class				**24.9**	**gold**			**24.9**	**gold**			**21.6**	**gold**
Initial values		**dep**	**lay**			**dep**	**lay**			**dep**	**lay**		
Soil moisture#	4	0.9	3	4.3		0.9	3	4.3		0	0	0	
Soil min. N	4	0.9	3	4.3		0.9	3	4.3		0	0	0	
sum/class				**8.6**	**gold**			**8.6**	**gold**			**0**	fail
previous crop													
Crop	3		1	3			1	3			1	3	
Sowing date	2		1	2			1	2			0	0	
Harvest date	3		1	3			1	3			0	0	
Yield	2		1	2			1	2			1	2	
Residue man.	4		1	4			1	4			1	4	
Fertilization	3		1	3			1	3			0	0	
Irrigation	2		1	2			1	2			1	2	
sum/class				**19**	**plat**			**19**	**plat**			**11**	**cop**
Topography													
Latitude	5		1	5			1	5			1	5	
Longitude	3		1	3			1	3			1	3	
Altitude	3		1	3			1	3			1	3	
Slope/expos.	1		1	1			1	1			1	1	
sum/class				**12**	**plat**			**12**	**plat**			**12**	**plat**

av = available (1)/not available (0); WP = weighting points; dep = depth (m); lay = no. of layers, classes: plat (platinum), gold (gold), silv (silver), cop (copper).

Table 5
Availability of state variables, flux measurements and additional observations for agro-ecosystem model testing ($^{\#}$ = minimum data requirement).

Variable	Relevance	Müncheberg 1993–1995					Müncheberg 1996–1998					Lednice				
		no	n_{ob}	rep	WP		no	n_{ob}	rep	WP		no	n_{ob}	rep	WP	
Phenology																
Emergence	3		1		3			1		3			1		3	
Tillering/stem elong./other	2		1		2			1		2			1		2	
Ear emer./oth.	2		1		2			0		0			0		0	
Flowering$^{\#}$	5		1		5			1		5			1		5	
Maturity/oth.	3		1		3			0		0			1		3	
sum/class					**15**	**plat**				**10**	**silv**				**13**	**plat**
Crop growth																
Yield	5	1$^{¥}$	1	3	6		1$^{¥}$	1	3	6		1$^{¥}$	1	3	6	
Ab. gr. biomass	4	1$^{¥}$	5	3	5		1$^{¥}$	1	3	1.33			0		0	
Mass single organs	3	2$^{¥}$	5	3	2.4		2$^{¥}$	1	3	0.67			0		0	
Root biomass	3	1$^{¥}$	4	3	1.2		0	0	3	0			0		0	
N ab. gr. biomass	3	1$^{¥}$	5	3	3.6		1$^{¥}$	1	3	1			0		0	
N in single organs	3	2$^{¥}$	5	3	3.6		2$^{¥}$	1	3	3			0		0	
Leaf area index	4		0		0			0		0			0		0	
Soil																
Soil water gravimet.	3	3$^{\equiv}$	6	3	3.6		3$^{\equiv}$	2	3	1.2			0		0	
Pressure heads	2	5$^{\equiv}$	100	3	3		5$^{\equiv}$	0	3	0			0		0	
Soil mineral N	3	3$^{\equiv}$	6	3	3.6		3$^{\equiv}$	2	3	2			0		0	
Soil wat. cal. sensor	3	3$^{\equiv}$	300	3	3.6		3$^{\equiv}$	280	3	3.6			0		0	
Deep percol. flux	2		0		0			0	3	0			0		0	
Deep N leaching	2		0		0			0	2	0			0		0	
Surface fluxes																
Evapotranspiration	3		0		0			0		0			0		0	
CO_2 flux	3		0		0			0	3	0			0		0	
NH_3 flux	2		0		0			0	2	0			0		0	
N_2 flux	2		0		0			0	0	0			0		0	
N_2O flux	2		0		0			0	5	0			0		0	
NO flux	2		0		0			0	0	0			0		0	
CH_4 flux	2		0		0			0		0			0		0	
Add. observations (qualitatively)																
Lodging	3		1		3			1	3	3			1		3	
Pest/disease effects	3		1		3			1	2	3			1		0	
Weeds	3		1		3			1	3	3			1		0	
Physical damages	3		1		3			1	2	3			1		2	
sum/class					**47.6**	**plat**				**30.8**	**gold**				**18**	**silv**

no = no. of $^{¥}$ organs or $^{\equiv}$ layers; ob = no. of observations per year; rep = no. of replications; WP = weighting points; classes: plat (platinum), gold (gold), silv (silver), cop (copper).

Meteorological data are within the optimal distance of 1 km from the investigation sites for Müncheberg and Lednice leading to a "platinum" rating for both (Fig. 4).

Rating of the single blocks results in "platinum" classes for nearly all blocks except for initial values and soil data, which achieved a "gold" mark mainly because samples are limited to 90 cm depth and soil water characteristic (water retention and conductivity curves) was not measured (Table 4). As the crop rotation was replicated with one year shift, at least two more seasons for each crop are available. In combination with the second treatment, the total sum of weights and the average of block classes were multiplied by 1.07 (Table 6) which did not change the overall mark "platinum". As only frequency of state variable measurements was reduced, the second period received mainly the same marks for the classes except for the state variable block (Table 5), which was downgraded to gold and the phenology block (Table 5), which was downgraded to silver as observations were reduced to three per season (no observation for sugar beets). As the second period still benefits from the basic investigations of the first period it achieved still a "gold" mark.

The data set from Lednice, Czech Republic was used in the model-inter-comparison of Palosuo et al. (2011). It provides crop yield data of one cultivar, basic soil data, management and key phenological stages of winter wheat for 15 years (1992–2006). "Platinum" was achieved for the agronomical management, phenology, and weather information (33 WP), while site and soil information achieved "gold" due to missing measured soil hydraulic properties (Tables 4 and 5), which had to be derived from texture. The block of initial soil conditions failed since measurements were missing. However, as data from previous crop are sufficient to assess at least initial soil water content, the combined mark reached "copper". Only yield was measured in three replicates in the state variable block (Table 5), which achieved "silver" due to well documented ratings of damages. The large number of 15 seasons with a constant variety is a strong aspect to increase the total weight sum by the seasons/treatments factor. There were no additional treatments and growth conditions were not always at optimum since the trial was not irrigated and dry spells occur during some of the seasons. Therefore, the rating for optimum conditions was set to 3 out of 5. In summary the seasons/treatments factor modifies the total sum and the block average by only +3% (Table 6). Overall, the data set closely failed to achieve "gold" and is marked as "silver". This would be not suitable to parameterise, calibrate and improve models, but offers very good opportunities to test and validate models especially regarding their sensitivity to annual weather variability, which was the main intention of the study of Palosuo et al., 2011).

Many field trials which are mainly focussed on yield estimation related to variations of varieties (variety trials) or fertiliser or water supply may fall into the "copper" class or even fail to be suitable for modelling purposes since they are lacking relevant soil information, having missing or distant weather data, or lacking information

Table 6
Calculation of multiplier (according to Table 3) for multiple seasons, treatments and optimum conditions for three data set examples.

Variable ×	Relevance	Müncheberg 1993–1995			Müncheberg 1996–1998			Lednice		
		Value of ×	WP	Multiplier	No of#	WP	Multiplier	No of#	WP	Multiplier
Crop seasons										
No. of seasons per crop	1	6	5		6	5		15	5	
Sites										
No. of site variants (e.g. soils)	1	1	1		1	1		1	1	
Treatments										
Management close to optimum × = 1 (limited) to 5 (potential growth)	1	4	4		4	4		3	3	
zero N treatment (× = 0 for no, 1 for yes)	3	0	0		0	0		0	0	
No. of treatment 1 (e.g. irrigation)	1	2	2		2	2		0	0	
No. of treatment 2 (e.g. nitrogen fert.)	1	0	0		0	0		0	0	
No. of treatment 3 (e.g. CO_2)	1	0	0		0	0		0	0	
No. of treatment 4 (e.g. tillage)	1	0	0		0	0		0	0	
No. of treatment 5 (…)	1	0	0		0	0		0	0	
No. of treatment 6 (…)	1	0	0		0	0		0	0	
Sum			**12**	**1.07**		**12**	**1.07**		**9**	**1.03**

WP = weighting points.

on previous crop or initial soil conditions although they investigate a very useful range of treatments. Some of these data could be measured even after the experiment, e.g. conservative soil properties like texture or SOC, to upgrade the data and make them useful for modelling.

6. Discussion

Several attempts have been made to describe data requirements for agro-ecosystem modelling. Minimum data requirements for crop modelling were defined by Nix (1984) and were manifested in the IBSNAT data base (Tsuji et al., 2002). Some data quality requirements were inherently included in the IBSNAT and ICASA protocol and best methods for growth sampling and specified minimal land-area for growth samples and minimum land-areas for final yields were specified, to insure data quality (Boote, 1999; Hunt et al., 2001). Recently, White et al. (2013) published an approach for describing and identifying variables of management, environmental conditions, soils, and crop measurements for improvement and testing of crop models. Although Rosenzweig et al. (2013) presented a general classification hierarchy for data for the AgMIP project, no quantitative weighting criteria were assigned to the classes. Boote et al. (2015) recently presented a qualitative method to evaluate data sets for their suitability for crop modelling. The comprehensive quantitative evaluation framework and software presented here provides a tool to evaluate data regarding their suitability for modelling including aspects of data quality. This helps data base managers to evaluate data regarding their suitability for modelling, and modellers to select data appropriate to their requirements to either parameterise, or calibrate or validate their models. Additionally, experimentalists have the opportunity to screen their experimental design and decide on the most valuable and effective measurements to improve the usefulness of their data for modelling.

Although, the weight of the variables was set mainly under the aspect of crop modelling, users can adjust weighting factors according to their specific objectives and needs to cover different aspects of agro-ecosystem analysis. The examples show that the method was capable to differentiate data for the different purposes of modelling. While in all three examples most of the standard input data for crop modelling were available, the difference came from information provided on initial values or previous crop management and from the range and extent of observed state variables.

The framework presented here should not be seen as a final product and is open for extension. Several topics are currently under discussion for model improvement, which require additional information to fill existing knowledge gaps. Quality of marketable yields is an issue, which has large potential for improvement in many models. Wheat protein composition for example is relevant to describe baking quality. Further development of models in this direction requires investigations on protein fractionation under controlled conditions. Moreover, yield quality parameters (i.e. protein content and composition, sugar and acid concentrations, oil content, digestibility, etc.) are often more important than quantity ones. For such purposes, failure to collect these data makes datasets less appropriate for modelling applications and requires adequate attention for future experimental design and measurement protocols.

The possibility to apply agro-climatic models in climate change impact assessments is strongly related to the availability of data on the impact of CO_2 concentrations on crop growth and yield (Ewert et al., 2015; Webber et al., 2014). Such data have been collected from experiments performed in a wide variety of settings (e.g. greenhouses, chambers, free air enrichments), and results have not always been consistent across the different experimental conditions (Ainsworth et al., 2008; Kimball et al., 2002). Thus, their use in simulated impact assessments has been criticized for overestimating the effect of elevated CO_2 concentration on wheat growth and yield (Long et al., 2006), while others have shown that the simulations are consistent with FACE experiments (Tubiello et al., 2007). Bunce (2013) emphasized the need to monitor high frequency fluctuations of CO_2 during the free air experiments. Combined effects of increasing temperature and CO_2 concentration are increasingly investigated. This should also include the above mentioned yield quality aspects.

The impact of various adverse weather events e.g., late frost, severe frost with absent snow cover, heat stress during anthesis, heat stress during the reproductive phase affecting leaf senescence, signs of drought stress and wilting, effect of lodging or overly wet soil leading to root anoxia should be monitored. They should be clearly described and ranked using standardized tables (e.g. 1–9 scale). While these events have high impact (both in experiments but also on "real" farms) they tend to be absent in most field data sets; however, the ability of crop models to capture such events is highly demanded (e.g., Asseng et al., 2011; Lobell et al., 2013). The usual practice is to dismiss experiments damaged by adverse weather event from the database as non-representative. The usefulness of such data, however, should not be under-estimated; for example, Trnka et al. (2010), demonstrated how crop model performance benefits from the introduction of a simple snow cover model. It can be of high value to make use of such experiments, which are significantly damaged, to learn how to deal with such events in model approaches. However, without proper data of field conditions as described above, the further development of models capable of encapsulating these situations will remain problematic at best (see Rötter et al., 2011; Semenov and Shewry, 2011).

Measured data can be erroneous. Therefore, the presented classification framework has to be accomplished by a test on plausibility of data. Application of a limited number of selected agro-ecosystem models can help to identify errors in observations, e.g. of water balance, and to eliminate erroneous observations before making the data set publicly available.

7. Conclusions

Modeller groups have so far mainly used their specific data sets and formats to develop and evaluate their models (Holzworth et al., 2015). Some groups have access to large data bases to evaluate their models over a wide range of conditions and cropping systems (e.g. Coucheney et al., 2015). The different formats and requirements aggravate the joint usage of data by different models and hampers collaboration between modeller groups (Holzworth et al., 2015). The common use of data by multi-model ensembles enables to evaluate robustness and behaviour of different model approaches under a wide range of conditions and allows the exchange of knowledge and algorithms to improve individual models and to reduce the uncertainty expressed by the variability of model ensemble results (Asseng et al., 2013). Holzworth et al. (2015) emphasised the necessity of standardized publicly available benchmark data sets to evaluate models and algorithms at different scales. Therefore, one of the goals of international initiatives such as AgMIP and MACSUR is to adopt a common data storage format and to provide tools to convert the data to the specific input requirements of the various models (Porter et al., 2014). These community efforts ensure that the data sets are made publicly available and stored with a trustworthy and long-lasting provider. Availability for all users is one of the prime criteria for defining standard data sets, which the community could later develop out of the platinum-ranked sites into mandatory data sets for model development and compulsory benchmarks for model testing. The appointment of such mandatory data sets could stimulate joint efforts in the data providing communities across the globe to produce a new generation of experimental data sets or monitoring series, which are then well concerted with the model developers. A rising awareness of a secondary use of experimental data by modelling studies is already tangible among the experimentalists and observers and the presented data evaluation framework will aid them to design multiple use field experiments and monitoring programs. However, up to now, the number of platinum-ranked data sets is scarce and the model improvement efforts carried forward in the community networks continuously identifies knowledge gaps, which – once being filled – could boost the quality of model results.

Acknowledgements

The study was financially supported by the related national European ministries contributing to the MACSUR project (2812ERA_147; _154; _115; _42; _17; _189; _92; _99; _196) under FACCE-JPI (031A103B), and it was further an integrated activity under AgMIP (Agricultural Model Intercomparison and Improvement Project) supported by the US Department of Agriculture and the United Kingdom Department for International Development.

References

Abbas, F., Fares, A., Fares, S., 2011. Field calibrations of soil moisture sensors in a forested watershed. Sensors 11, 6354–6369.

Aggarwal, P.K., 1995. Uncertainties in crop, soil and weather inputs used in growth models: implications for simulated outputs and their applications. Agric. Syst. 48, 361–384.

Ainsworth, E.A., Leakey, A.D.B., Ort, D.R., Long, S.P., 2008. FACE-ing the facts: inconsistencies and interdependence among field, chamber and modeling studies of elevated [CO2] impacts on crop yield and food supply. New. Phytol. 179, 5–9.

Ashraf, C.M., Abu-Shakra, S., 1970. Wheat seed germination under low temperature and moisture stress. Agron. J. 70, 135–139.

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

Asseng, S., Foster, I., Turner, N.C., 2011. The impact of temperature variability on wheat yields. Glob. Change Biol. 17, 997–1012.

Baigorria, G.A., Jones, J.W., O'Brien, J.J., 2007. Understanding rainfall spatial variability in southeast USA at different timescales. Int. J. Climatol. 27, 749–760.

Baveye, P.C., Laba, M., 2014. Moving away from the geostatistical lamppost: why, where, and how does the spatial heterogeneity of soils matter? Ecol. Model. http://dx.doi.org/10.1016/j.ecolmodel.2014.03.018 (online first).

Blum, A., 1996. Crop responses to drought and the interpretation of adaptation. Plant Growth Regul. 20, 135–148.

Boote, K.J., 1999. Data requirements for model evaluation and techniques for sampling crop growth and development. In: Hoogenboom, G., Wilkens, P.W., Tsuji, G.Y. (Eds.), DSSAT Version 3. A Decision Support System for Agrotechnology Transfer, vol. 4. G. University of Hawaii, Honolulu, HI, pp. 201–216.

Boote, K.J., Porter, C., Jones, J.W., Thorburn, P.J., Kersebaum, K.C., Hoogenboom, G., White, J.W., Hatfield, J.L., 2015. Sentinel site data for crop model improvement – definition and characterization. In: Hatfield, J.L., Fleisher, D. (Eds.), Improving Modeling Tools to Assess Climate Change Effects on Crop Response, Advances in Agricultural Systems Modeling. ASA, CSSA, and SSSA, Madison, WI, USA (in press).

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, F.B., Sinoquet, H., 2003. An overview of the crop model STICS. Eur. J. Agron. 18, 309–332.

Bruun, S., Jensen, L.S., 2002. Initialisation of the soil organic matter pools of the Daisy model. Ecol. Model. 153, 291–295.

Bunce, J.A., 2013. Effects of pulses of elevated carbon dioxide concentration on stomatal conductance and photosynthesis in wheat and rice. Physiol. Plant. 149, 214–221.

Cao, W., Moss, D.N., 1997. Modelling phasic development in wheat: a conceptual integration of physiological components. J. Agric. Sci. 129, 163–172.

Coucheney, E., Buis, S., Launay, M., Constantin, J., Mary, B., García de Cortázar-Atauri, I., Ripoche, D., Beaudoin, N., Ruget, F., Andrianarisoa, K.S., Le Bas, C., Justes, E., Léonard, J., 2015. Accuracy, robustness and behavior of the STICS soil–crop model for plant, water and nitrogen outputs: evaluation over a wide range of agro-environmental conditions in France. Environ. Model. Softw. 64, 177–190.

Dahiya, I.S., Kersebaum, K.C., Richter, J., 1984. Spatial variability of some nutrient constituents of an Alfisol from loess. I. Classical statistical analysis. Z. Pflanzenernn. Bodenkd. 147, 695–703.

Deveau, N., Le Cadre, E., Hinsinger, P., Jaillard, B., Gérard, F., 2009. Soil pH controls the environmental availability of phosphorus: experimental and mechanistic modelling approaches. Appl. Geochem. 24, 2163–2174.

De Wit, C.T., 1982. Simulation of living systems. In: Penning de Vries, F.W.T., van Laar, H.H. (Eds.), Simulation of Plant Growth and Crop Production. Simulation Monographs, Wageningen, The Netherlands, pp. 3–8.

De Wit, C.T., Van Keulen, H., 1987. Modelling production of field crops and its requirements. Geoderma 40, 253–265.

Doltra, J., Lægdsmand, M., Olesen, J.E., 2011. Cereal yield and quality as affected by N availability in organic and conventional crop rotations in Denmark: a combined modeling and experimental approach. Eur. J. Agron. 34, 83–95.

Eitzinger, J., Thaler, S., Schmid, E., Strauss, F., Ferrise, R., Moriondo, M., Bindi, M., Palosuo, T., Rötter, R., Kersebaum, K.C., Olesen, J., Patil, R., Saylan, L., Caldag, B., Caylak, O., 2013. Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. J. Agric. Sci. 151 (6), 813–835.
Elmore, A.J., Guinn, S.M., Minsley, B.J., Richardson, A.D., 2012. Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests. Glob. Change Biol. 18, 656–674.
Ewert, F., Rötter, R.P., Bindi, M., Webber, H., Trnka, M., Kersebaum, K.C., Olesen, J.E., van Ittersum, M.K., Janssen, S., Rivington, M., Semenov, M.A., Wallach, D., Porter, J.R., Stewart, D., Verhagen, J., Gaiser, T., Palosuo, T., Tao, F., Nendel, C., Roggero, P.P., Bartosova, L., Asseng, S., 2015. Crop modelling for integrated assessment of climate change risk to food production. Environ. Model. Softw. 72, 287–303.
Feekes, W., 1941. De tarwe en haar milieu. In: Verslagen van de Technische Tarwe Commissie, pp. 523–888. Groningen 17.
Gabrielle, B., Mary, B., Roche, R., Smith, P., Gosse, G., 2002. Simulation of carbon and nitrogen dynamics in arable soils: a comparison of approaches. Eur. J. Agron. 18, 107–120.
Giebel, A., Wendroth, O., Reuter, H.I., Kersebaum, K.C., Schwarz, J., 2006. How representatively can we sample soil mineral nitrogen? J. Plant Nutr. Soil Sci. 169, 52–59.
Greenwood, D.J., Lemaire, G., Gosse, G., Cruz, P., Draycott, A., Neeteson, J.J., 1990. Decline of percentage N of C_3 and C_4 crops with increasing plant mass. Ann. Bot. 66, 425–436.
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, 349–381.
Grossmann, J., Udluft, P., 1991. The extraction of soil water by the suction-cup method: a review. J. Soil Sci. 42, 83–93.
Hammer, G.L., Kropff, M.J., Sinclair, T.R., Porter, J.R., 2002. Future contributions of crop modelling—from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. Eur. J. Agron. 18, 15–31.
Heuvelink, G.B.M., 1998. Error Propagation in Environmental Modelling with GIS. Research Monographs in GIS Series, London, 127.
Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: current status and future prospects. Environ. Model. Softw. 72, 276–286.
Hunt, L.A., White, J.W., Hoogenboom, G., 2001. Agronomic data: advances in documentation and protocols for exchange and use. Agric. Syst. 70, 477–492.
Janssen, S., Andersen, E., Athanasiadis, I.N., van Ittersum, M.K., 2009. A database for integrated assessment of European agricultural systems. Environ. Sci. Pollut. 12, 573–587.
Jeong, S.J., Medvigy, D., Shevliakova, E., Malyshev, S., 2012. Uncertainties in terrestrial carbon budgets related to spring phenology. J. Geophys. Res. Biogeosci. 117, G01030. http://dx.doi.org/10.1029/2011JG001868.
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.
Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18, 267–288.
Kersebaum, K.C., 2011. Special features of the HERMES model and additional procedures for parameterization, calibration, validation, and applications. In: Ahuja, L.R., Ma, L. (Eds.), Advances in Agricultural Systems Modeling Series 2Amer. Soc. Agron., Madison, WI, USA, pp. 65–94.
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., 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, 59–67.
Kersebaum, K.C., Reuter, H.I., Lorenz, K., Wendroth, O., 2002. Modelling crop growth and nitrogen dynamics for advisory purposes regarding spatial variability. In: Ahuja, L., Ma, L., Howell, T. (Eds.), Agricultural System Models in Field Research and Technology Transfer. Lewis Publishers, Boca Raton, pp. 229–252.
Kersebaum, K.C., Nendel, C., 2014. Site-specific impacts of climate change on wheat production across regions in Germany using different CO_2 response functions. Eur. J. Agron. 52, 22–32.
Kimball, B.A., Kobayashi, K., Bindi, M., 2002. Responses of agricultural crops to free-air CO_2 enrichment. Adv. Agron. 77, 293–368.
Lehmann, J., Kaiser, K., Peter, I., 2001. Exchange resin cores for the estimation of nutrient fluxes in highly permeable tropical soils. J. Plant Nutr. Soil Sci. 164, 57–64.
Lobell, D.B., Hammer, G.L., McLean, G., Messina, C., Roberts, M.J., Schlenker, W., 2013. The critical role of extreme heat for maize production in the United States. Nat. Clim. Change 497–501.
Long, S.P., Ainsworth, E.A., Leakey, A.D.B., Nösberger, J., Ort, D.R., 2006. Food for thought: lower-than-expected crop yield stimulation with rising CO_2 concentrations. Science 312, 1918–1921.
Lorenz, D.H., Eichhorn, K.W., Bleiholder, H., Klose, R., Meier, U., Weber, E., 1994. Phänologische Entwicklungsstadien der Weinrebe (*Vitis vinifera* L. ssp. vinifera) – Codierung und Beschreibung nach der erweiterten BBCH Skala. Vitic. Enol. Sci. 49, 66–70.
Marschner, H., 1991. Mechanisms of adaptation of plants to acid soils. Plant Soil 134, 1–20.
Meier, U., 2001. Growth Stages of Mono-and Dicotyledonous Plants, second ed. In: BBCH Monograph Federal Biological Research Centre for Agriculture and Forestry, Germany, p. 158.
Migliavacca, M., Sonnentag, O., Keenan, T., Cescatti, A., O'Keefe, J., Richardson, A., 2012. On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model. Biogeosciences 9, 2063–2083.
Mirschel, W., Wenkel, K.-O., Wegehenkel, M., Kersebaum, K.C., Schindler, U., Hecker, J.-M., 2007. Müncheberg field trial data set for agro-ecosystem model validation. In: Kersebaum, K.-C., Hecker, J.-M., Mirschel, W., Wegehenkel, M. (Eds.), Modelling Water and Nutrient Dynamics in Soil Crop Systems. Springer, Dordrecht, pp. 219–243.
Mittelbach, H., Lehner, I., Seneviratne, S.I., 2012. Comparison of four soil moisture sensor types under field conditions in Switzerland. J. Hydrol. 430–431, 39–49.
Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57, 443–452.
Nendel, C., Berg, M., Kersebaum, K.C., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K.O., Wieland, R., 2011. The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. Ecol. Model. 222, 1614–1625.
Nendel, C., Schmutz, U., Venezia, A., Piro, F., Rahn, C.R., 2009. Converting simulated total dry matter to fresh marketable yield for field vegetables at a range of nitrogen supply levels. Plant Soil 325, 319–334.
Nielsen, D.R., Bouma, J., 1985. Soil Spatial Variability. Proc. Workshop of the ISSS and the SSSA, Las Vegas, 1984. Pudoc, Wageningen, p. 243.
Nix, H., 1984. Minimum datasets for agrotechnology transfer. In: ICRISAT, Proc. Int. Symposium on Minimum Data Sets for Agrotechnology Transfer. ICRISAT Center, Patencheru, A. P., India, pp. 181–188.
Nix, H., 1985. 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, pp. 105–130.
Olesen, J.E., Børgesen, C.D., Elsgaard, L., Palosuo, T., Rötter, R.P., Skjelvåg, A.O., Peltonen-Sainio, P., Börjesson, T., Trnka, M., Ewert, F., Siebert, S., Brisson, N., Eitzinger, J., van Asselt, E.D., Oberforster, M., van der Fels-Klerx, H.J., 2012. Changes in time of sowing, flowering and maturity of cereals in Europe under climate change. Food Add. Contamin. 29, 1527–1542.
Pachepsky, Y., Acock, B., 1998. Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates. Geoderma 85, 213–229.
Paige, G.B., Keefer, T.O., 2008. Comparison of field performance of multiple soil moisture sensors in a semi-arid rangeland. J. Am. Water Res. Assoc. 44, 121–135.
Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., Trnka, M., Bindi, M., Çaldağ, B., Ewert, F., Ferrise, R., Mirschel, W., Şaylan, L., Šiška, B., Rötter, R.P., 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, 103–114.
Parker, P., Letcher, R., Jakeman, A., Beck, M.B., Harris, G., Argent, R.M., Hare, M., Pahl-Wostl, C., Voinov, A., Janssen, M., Sullivan, P., Scoccimarro, M., Friend, A., Sonnenshein, M., Barker, D., Matejicek, L., Odulaja, D., Deadman, P., Lim, K., Larocque, G., Tarikhi, P., Fletcher, C., Put, A., Maxwell, T., Charles, A., Breeze, H., Nakatani, N., Mudgal, S., Naito, W., Osidele, O., Eriksson, I., Kautsky, U., Kautsky, E., Naeslund, B., Kumblad, L., Park, R., Maltagliati, S., Girardin, P., Rizzoli, A., Mauriello, D., Hoch, R., Pelletier, D., Reilly, J., Olafsdottir, R., Bin, S., 2002. Progress in integrated assessment and modelling. Environ. Model. Softw. 17, 209–217.
Parkin, T.B., 1987. Soil microsites as a source of denitrification variability. Soil Sci. Soc. Am. J. 51, 1194–1199.
Penning de Vries, F.W.T., Van Laar, H.H., 1982. Simulation of Plant Growth and Crop Production. Simulation Monographs, Wageningen, The Netherlands.
Plénet, D., Lemaire, G., 1999. Relationships between dynamics of nitrogen uptake and dry matter accumulation in maize crops. Plant Soil 216, 65–82.
Porter, C.H., Villalobos, C., Holzworth, D., Nelson, R., White, J.W., Athanasiadis, I.N., Janssen, S., Ripoche, D., Cufi, J., Raes, D., Zhang, M., Knapen, R., Sahajpal, R., Boote, K.J., Jones, J.W., 2014. Harmonization and translation of crop modeling data to ensure interoperability. Environ. Model. Softw. 62, 495–508.
Reuter, H.I., Kersebaum, K.C., 2009. Applications in precision agriculture. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry, Developments in Soil Science, vol. 33, pp. 623–636.
Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. Phys. J. Gen. Appl. Phys. 1, 318–333.
Robinson, D.A., Campbell, C.S., Hopmans, J.W., Hornbuckle, B.K., Jones, S.B., Knight, R., Ogden, F., Selker, J., Wendroth, O., 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: a review. Vadose Zone J. 7, 358–389.
Rötter, R., Carter, T.R., Olesen, J.E., Porter, J.R., 2011. Crop-climate models need an overhaul. Nat. Clim. Change 1, 175–177.
Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Takáč, J.,

Trnka, M., 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.
Rötter, R.P., Ewert, F., Palosuo, T., Bindi, M., Kersebaum, K.C., Olesen, J.E., Trnka, M., van Ittersum, M.K., Janssen, S., Rivington, M., Semenov, M., Wallach, D., Porter, J.R., Stewart, D., Verhagen, J., Angulo, C., Gaiser, T., Nendel, C., Martre, P., de Wit, A., 2013. Challenges for agro-ecosystem modelling in climate change risk assessment for major European crops and farming systems. In: Impacts World 2013 Conference Proceedings. Potsdam Institute for Climate Impact Research, Potsdam, pp. 555–564. http://dx.doi.org/10.2312/pik.2013.001.
Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric. For. Met. 170, 166–182.
Savin, R., Nicolas, M.E., 1999. Effects of timing of heat stress and drought on growth and quality of barley grains. Aust. J. Agric. Res. 50, 357–364.
Schroedter-Homscheidt, M., Bofinger, S., Breitkreuz, H., Heilscher, G., Stettler, S., 2006. Usage of Earth Observation for Solar Energy Market Development – Lessons Learned. Proc. ISEC 2006, Denver.
Semenov, M.A., Shewry, P.R., 2011. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. Sci. Rep. 1. http://dx.doi.org/10.1038/srep00066.
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.
Tetzlaff, D., Uhlenbrock, S., 2005. Significance of spatial variability in precipitation for process-oriented modelling: results from two nested catchments using radar and ground station data. Hydrol. Earth Syst. Sci. 9, 29–41.
Topp, G.C., Davis, J.L., Annan, A.P., 2003. The early development of TDR for soil measurements. Vadose Zone J. 2, 492–499.
Trnka, M., Eitzinger, J., Kapler, P., Dubrovský, M., Semerádová, D., Žalud, Z., Formayer, H., 2007. Effect of estimated daily global solar radiation data on the results of crop growth models. Sensors 7, 2350–2383.
Trnka, M., Kocmánková, E., Balek, J., Eitzinger, J., Ruget, F., Formayer, H., Hlavinka, P., Schaumberger, A., Horáková, V., Možný, M., Žalud, Z., 2010. Simple snow cover model for agrometeorological applications. Agric. For. Met. 150, 1115–1127.
Trnka, M., Žalud, Z., Eitzinger, J., Dubrovský, M., 2005. Quantification of uncertainties introduced by selected methods for daily global solar radiation estimation. Agric. For. Met. 131, 54–76.
Tsuji, G.Y., du Toit, A., Jintrawet, A., Jones, J.W., Bowen, W.T., Ogoshi, R.M., Uehara, G., 2002. Benefits of models in research and decision support: the IBSNAT experience. In: Ahuja, L.R., Ma, L., Howell, T.A. (Eds.), Agricultural System Models in Field Research and Technology Transfer. Lewis Publishers, New York, pp. 71–89.
Tubiello, F.N., Amthor, J.S., Boote, K.J., Donatelli, M., Easterling, W., Fischer, G., Gifford, R.M., Howden, M., Reilly, J., Rosenzweig, C., 2007. Crop response to elevated CO_2 and world food supply: a comment on “Food for Thought…” by Long et al., Science 312:1918–1921, 2006. Eur. J. Agron. 26, 215–223.
Van Diepen, C.A., van der Voet, P., 1998. Application of simple interpolation methods in agrometeorology. In: Gozzini, B., Hims, M. (Eds.), EUR 18473–COST 79–Workshop on Dealing on Spatialisation. European Commission, pp. 3–17.
Van Genuchten, M.T., 1980. Aclosed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44, 892–898.
Van Ittersum, M.K., Leffelaar, P.A., Van Keulen, H., Kropff, M.J., Bastiaans, L., Goudriaan, J., 2003. On approaches and applications of the Wageningen crop models. Eur. J. Agron. 18, 289–307.
Van Keulen, H., 1976. Evaluation of models. In: Arnold, G.W., de Wit, C.T. (Eds.), Critical Evaluation of System Analysis in Ecosystem Management. Simulation Monographs, Wageningen, The Netherlands, pp. 22–29.
Van Keulen, H., Seligman, 1987. Simulation of Water Use, Nitrogen Nutrition and Growth of a Spring Wheat Crop. Simulation Monographs, Wageningen, Netherlands.
Van Keulen, H., Stol, W., 1991. Quantitative aspects of nitrogen nutrition in crops. Fert. Res. 27, 151–160.
Van Keulen, H., Wolf, J., 1986. Modelling of Agricultural Production: Weather, Soils and Crops. Simulation Monographs, Wageningen, The Netherlands.
Vereecken, H., Huisman, J.A., Bogena, H., Vanderborght, J., Vrugt, J.A., Hopmans, J.W., 2008. On the value of soil moisture measurements in vadose zone hydrology: a review. Water Resour. Res. 44, W00D06. http://dx.doi.org/10.1029/2008WR006829.
Wagenet, R.J., 1998. Scale issues in agroecological research chains. Nutr. Cycl. Agroecosyst. 50, 23–34.
Webber, H., Gaiser, T., Ewert, F., 2014. What role can crop models play in supporting climate change adaptation decisions to enhance food security in Sub-Saharan Africa? Agric. Syst. 127, 161–177.
Wegehenkel, M., 2005. Validation of a soil water balance model using soil water content and pressure head data. Hydrol. Process. 19, 1139–1164.
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, 357–368.
White, J.W., Hunt, L.A., Boote, K.J., Jones, J.W., Koo, J., Kim, S., Porter, C.H., Wilkens, P.W., Hoogenboom, G., 2013. Integrated description of agricultural field experiments and production: the ICASA version 2.0 data standards. Comput. Electron. Agric. 96, 1–12.
Wieringa, J., 1998a. Climatological use of station network data. In: Bindi, M., Gozzini, B. (Eds.), EUR 18472 – COST 79 – Workshop on Data Spatial Distribution in Meteorology and Climatology. European Commission, pp. 17–27.
Wieringa, J., 1998b. Representativity problems of wind stations. In: Bindi, M., Gozzini, B. (Eds.), EUR 18472 – COST 79 – Workshop on Data Spatial Distribution in Meteorology and Climatology. European Commission, pp. 29–43.
Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251, 123–150.
Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421.
Zelenka, A., Perez, R., Seals, R., Renné, D., 1999. Effective accuracy of satellite-derived hourly irradiances. Theor. Appl. Climatol. 62, 199–207.
Zhang, S., Tao, F.L., 2013. Modeling the response of rice phenology to climate change and variability in different climatic zones: comparisons of five models. Eur. J. Agron. 45, 165–176.
Zhao, G., Siebert, S., Enders, A., Eyshi Rezaei, E., Yan, C., Ewert, F., 2015. Demand for multi-scale weather data for regional crop modelling. Agric. For. Met. 200, 156–171.