BURLEIGH DODDS SERIES IN AGRICULTURAL SCIENCE

Improving data flow and integration in models assessing the impact of climate change on agriculture

Claas Nendel, Roland Baatz, Michael Berg-Mohnicke and Gohar Ghazaryan, Leibniz Centre for Agricultural Landscape Research (ZALF), Germany; Sander Janssen, Wageningen Environmental Research, The Netherlands; Pierre Martre, INRAE, France; and Cheryl Porter, University of Florida, USA

burleigh dodds SCIENCE PUBLISHING

Improving data flow and integration in models assessing the impact of climate change on agriculture

Claas Nendel, Roland Baatz, Michael Berg-Mohnicke and Gohar Ghazaryan, Leibniz Centre for Agricultural Landscape Research (ZALF), Germany; Sander Janssen, Wageningen Environmental Research, The Netherlands; Pierre Martre, INRAE, France; and Cheryl Porter, University of Florida, USA

1 Introduction

A growing world population depends for more and more of its daily needs on the food and agriculture system, while planetary resources are infinite and bounded (O'Neill et al., 2018; Steffen et al., 2015). Today's food and agriculture systems are facing multiple challenges (e.g. climate change, biodiversity loss, and soil degradation), so with the same or even fewer resources in terms of fossil fuel and land, more demands have to be met. Some relationships between the abovementioned challenges and agriculture systems have been more extensively studied than other relationships. For example, there is a fairly good and quantitative understanding of the interconnection between climate

http://dx.doi.org/10.19103/AS.2022.0115.03

Chapter taken from: Nendel, C. (ed.), *Modelling climate change impacts on agricultural systems*, pp. 39–74,
Burleigh Dodds Science Publishing, Cambridge, UK, 2023, (ISBN: 978 1 80146 174 0; www.bdspublishing.com)

change and agricultural production (Asseng et al., 2015), i.e. temperature increase and productivity effects, which also implies that these have been extensively modelled (e.g. Ewert et al., 2015; Rosenzweig et al., 2013). This is different for biodiversity loss, where there is still a limited understanding of how species loss and species diversity would impact agricultural systems and their productivity. Even the challenges are themselves interconnected, so a holistic approach towards the agricultural system is required to capture relevant development pathways and not be blinded by a deep understanding of one domain (i.e. agronomy, economy, supply chains), while missing important interactions across domains.

With these challenges occurring at the same time, and the food and agriculture system as multi-scale, multi-domain and multi-sectoral, integrated modelling approaches for agriculture and food systems are crucial (Harris, 2002; Parker et al., 2002). The integrated modelling approaches and the consequent modelling results can be used as a method for deepening the understanding of the different system analysis, an ex-ante tool to evaluate (policy) options for future development, and for sustainability analysis of existing agricultural systems. In the past these integrated modelling approaches have been most actively developed by linking models presenting domains together into an integrated framework, especially in the 2000s; see, for example, van Ittersum et al. (2008) and Janssen et al. (2011). In the early 2010s the focus shifted from model linking to ensemble analysis of the system components models, like economic and crop models, through initiatives such as the Agricultural Model Intercomparison and Improvement Project (AgMIP; Rosenzweig et al., 2013) or the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Warszawski et al., 2014). At the same time there has been a lot of focus on the potential of digitalisation and data-driven technologies to release new data sources, and new analysis techniques (see e.g. Holzworth et al., 2015; Janssen et al., 2017; Lokers et al., 2016), which could be beneficial to new approaches to (integrated) modelling.

It is therefore timely to review the issues involved in integrated modelling, especially with respect to an already well-studied relationship between climate change and agricultural systems. Even so, many elements require further elaboration and discovery in this relationship between climate change and agricultural systems.

2 Model-data integration

Every model is created on data. Beyond the model-building process, the classical way of informing any kind of model with data is to provide the data as input for the model before the run. While statistical models would need input in exactly the form with which the models were created, mechanistic models often

Published by Burleigh Dodds Science Publishing Limited, 2023.

provide a bit more flexibility with respect to the amount of data that they could use for their run. For this chapter, we concentrate on mechanistic, i.e. process-based simulation models, as at present they still represent the vast majority of models in use for climate change impact assessments in agriculture. For these models, we will first look at data that are needed during model development, starting with the calibration process, and then move on to the initialisation of models. In this section, we will concentrate on data provisioning for large-area applications. Finally, we will address assimilation of data as a method to inform models while running the simulations, a technique that allows integrating more and better data as soon as they become available and improve the simulation underway.

2.1 Multi-criteria calibration of mechanistic agro-ecosystem models

Mechanistic simulation models for agro-ecosystems (AEMs) are commonly composed of a multitude of algorithms that make up subcomponents (modules), which then are merged to build the system that the model shall represent. Each of these algorithms has been derived from individual experiments that were set up to find relationships between two or sometimes more variables that interact. These experiments are mostly designed in a way that other influencing factors are excluded, creating more or less artificial conditions, which as such do not appear in the real system. At each level of aggregation, from algorithms to modules and from modules to model, distinct target variables are computed. To ensure that algorithm, module and model do what they are supposed to do, the performance of the component needs to be tested, against experimentally observed target variables, or target variables monitored under real-world conditions. Looking back into the historical literature on model calibration, often studies report on such performance checks against one variable only, while others consider multiple variables, but no systematic procedure to arrive at the best calibration. Furthermore, there are studies that calibrate for optimal performance and then test against previously unused (independent) data obtained from the same system under similar conditions. Very often, the model testing yields very satisfying results. However, agro-ecosystem models compute several potentially interesting target variables, and all of them are output of the same simulated system and therefore dependent on each other. Calibrating only to the optimal simulation of one of them often comes at the expense of low performances in simulating others, often without recognising that there is a problem ('right for the wrong reason'). It is therefore necessary to test the performance of the model in simulating several target variables simultaneously to make sure that the whole system's behaviour is well captured by the model (Archontoulis et al., 2014).

2.1.1 Methods

Simultaneous calibration for multiple target variables requires appropriate methods, which differ from the concepts used for calibration towards one single target variable. Multi-objective approaches account for compensating effects in the calibration procedure. Considering multiple variables in global optimisation decreases the risk of the search algorithm getting trapped in a local minimum, which may result in a good fit for one variable but in an otherwise inappropriate parameter set for others. In many studies with integrative aspects, different observed variables are optimised in a single-objective calibration process. The multi-objective approach (Groh et al., 2018; Wöhling et al., 2015) has so far received less attention in the application of agro-ecosystem models, although widely accepted in the hydrology community (Hernandez-Suarez et al., 2021; Houska et al., 2017; Mostafaie et al., 2018).

Multi-objective optimisation approaches use trade-offs to determine a set of non-dominated parameters that cannot be improved for one objective without compromising the other objective. For this purpose, they spread the search within the parameter space to identify feasible parameter sets (solutions) with acceptable trade-offs along the Pareto front (Kamali et al., 2022). Since different parameters belong to different processes in the model, the solutions along the Pareto front are all realistic and robust parameter estimates yielding similar model performances.

The most common algorithms for multi-objective calibration include particle swarm optimisation (Kennedy and Eberhart, 1995), genetic algorithms (Fonseca and Fleming, 1993) and complex evolution (Yapo et al., 1998). While most multi-objective methods offer superior performance compared to single-objective calibration (Kamali et al., 2013), they also suffer from their inability to provide information on the uncertainty of model predictions. Bayesian approaches can account for parameter uncertainty in optimisation, and their superior performance has recently been demonstrated for eco-hydrological models (Tang et al., 2018; Wöhling et al., 2013) where mostly two variables (leaf area index and soil moisture) have been targeted.

2.1.2 Harmonised data sets

Preparation of data for use in process-based crop models requires that the data are aggregated to the correct spatial and temporal scales; transformed into the correct dimensional units; and checked for completeness, semantic and syntactic alignment, and to ensure quality (Janssen et al., 2017). With the increased availability of data from remote and proximal sensing, the need for interoperability of these data sources increases. The primary data requirements for process-based crop models are weather, typically daily records of rainfall, temperature and solar radiation; soil physical and chemical properties for

surface and subsoil; management event details such as land preparation, sowing, application of inputs, and organic matter management; and genetic parameters (Hoogenboom et al., 2012).

Data preparation for large-scale, coordinated modelling activities, including ensemble and gridded modelling, requires some level of automation of these data preparation processes, but there are currently no interoperability standards for setting temporal and spatial scales, variable notation and units, file formats or data quality in large-scale crop modelling activities (van Evert, 2019). The choice of spatial scale is generally determined by the resolution of available data and the intent of the modelling activity. Temporal scale is limited by the time step of the model, generally daily, but in most large-scale applications, seasonal or annual values are reported due to data volume and analysis computational constraints.

Harmonisation of the semantics and syntax of agricultural datasets was addressed by the AgMIP (agmip.org) (Porter et al., 2014) to streamline ensemble modelling activities for Regional Integrated Assessments (Antle et al., 2015). The AgMIP approach describes a standardised data exchange mechanism using the ICASA vocabulary (White et al., 2013) implemented in a flexible data schema and allowing gap-filling of missing data. However, the AgMIP data interoperability protocols were developed for use with site-based field crop experimental data or farm survey data and were thus associated with individual plots or farms rather than spatially distributed data and these standards have not been adapted for gridded modelling activities.

Numerous vocabularies and ontologies have been developed for agricultural research, but no standard has emerged for use in spatial modelling activities. The robust agronomic vocabulary developed with the ICASA standards has limited provision for spatial applications. Several ontologies (Jonquet et al., 2015) have been developed for agricultural use, but none contain the full range of terms required for crop modelling activities yet.

The Global Gridded Crop modelling Intercomparisons (GGCMI phase 1) (Müller et al., 2017); GGCMI phase 2 (Franke et al., 2020); GGCMI phase 3 (Jagermeyr et al., 2021) may be the best examples of large-scale, ensemble modelling activities using harmonised data. These model intercomparisons were performed at a spatial resolution of 0.5 arc-degrees. Much of the data preparation was not automated and required extensive effort on the part of the study organisers. However, the resulting datasets have been shared under open-source licences.

For these types of gridded modelling efforts, the model input data are stored in geo-referenced files, typically one variable per file. NetCDF (Network Common Data Form) has emerged as the *de facto* standard for efficiently storing multidimensional arrays for crop modelling inputs which vary spatially but often also with time (weather data) or depth (soil data). In the case of GGCMI

simulations, model input data were provided to teams in NetCDF format but each team was responsible for converting data to their specific model formats for simulations. Model outputs from each team were converted to NetCDF format for analysis and intercomparison.

3 Informing spatiotemporal simulations

Once a model has been successfully calibrated and declared fit for purpose, it can be applied in very different manners and at different scales. This implies that the model will be applied at sites to which it has not been explicitly calibrated. The quality of the simulations then often depends entirely on the quality of the input information, if no further calibration as applied at larger scales with aggregated observed data. The most fundamental drivers for mechanistic agro-ecosystem models are weather and soil data. For any spatial application of models, soil data needs to be available in sufficient detail, so that for each grid cell or point that shall be simulated, a realistic representation of soil information can be provided. This includes not only a soil type but also soil texture information and soil organic matter contents (and sometimes more, depending on the need of the model) for each soil horizon down the profile. The challenge here is that despite an often very sophisticated soil survey and sampling has been applied, the information product that is then produced from this data is often less detailed. In Germany, as one example, the most detailed soil survey was conducted during the 1930s for land use planning and taxation. Only in individual federal states, contemporary soil surveys produce soil maps at a similar degree of detail. For many other areas, detailed soil surveys have not been conducted at all, and available soil maps build on much coarser information, e.g. the Brazilian soil survey (RADAMBRASIL) of the 1970s (De Negreiros et al., 2009). And then, these maps are often not available in a digital format, ready to be used with models. Recently, first attempts have been presented to build finer resolved digital soil maps using methods of artificial intelligence to fill gaps (e.g. https://soilgrids.org; Hengl et al., 2017).

For weather data, a similar problem exists. Fine-resolved weather data products are available for areas in which meteorological observation networks are dense. For areas that are less densely covered, the quality of the weather data is lower *per se* and relies a lot on the skills of the meteorologists to interpolate reasonably the available data. This applies not only to spatial gap filling but also to temporal, as met stations sometimes fail to record. In addition to this, many met stations are not fully equipped and record temperatures and precipitation amounts, but not air humidity and solar radiation, two variables often required by crop and agro-ecosystem models. This aggravates the gap-filling problem. Especially for model applications to inform farmers on what happens on their fields or in their region (early-warning systems, irrigation and fertilisation

recommendation, etc.), fine-scaled weather information is in high demand, such as the AgERA5 10 × 10 km² for global applications (Boogaard et al., 2020) or the 1 × 1 km² weather data grid of the German Weather Service for Germany. While observed weather and short-term forecasts may be available in high resolution, long-term forecasts and climate projections are often not. Here, science and engineering are still underway towards better and more detailed information products. Until then, modellers have to live with lower-resolution weather products, which comes at the expense of a higher uncertainty. How this influences the quality of the model predictions is discussed elsewhere in this book.

Beyond soil and weather data, agro-ecosystem models require more information, e.g. on crop types, crop and soil management and other items. For these highly dynamic variables, no static maps can help, but frequent observations. For this purpose, different data sources are available, of which we illustrate some in the following subsections.

3.1 Remote sensing

Remote sensing (RS) is one of the main sources of information across different spatial and temporal scales that can be used to monitor agricultural systems and the impacts of climate change. Several parameters describing vegetation growth and surface characteristics can be derived and used to inform the mechanistic models (Novelli et al., 2019). The increasing availability of RS data with improved spatial, spectral and temporal properties allows mapping and extraction of RS-based metrics over large areas. Such metrics can be used to derive proxies of vegetation status and condition, which can be the basis for deriving parameters describing crop growth stage, condition and phenology (Meroni et al., 2021). Examples of such parameters are the start (greening), peak, amplitude and the end of the growing season (senescence; Younes et al., 2021). The parameters in turn can be further used to derive information on growing conditions (Guo and Gu, 2022) and management practices (e.g. sowing and harvest date; Rezaei et al., 2021) and crop type (Blickensdörfer et al., 2022; Griffiths et al., 2019).

Crop type information is one of the essential datasets that can be generated from RS. Different data, such as Landsat, Sentinel and MODIS, has been used to derive spatially explicit crop maps. Nevertheless, availability of the crop maps is not at the same level globally and datasets covering large areas often specify only agricultural practices such as arable land or rangeland (Hoefsloot et al., 2012). Based on the freely available RS time series and use of different machine learning models, large-scale crop maps are available for USA (Boryan et al., 2011), Germany (Blickensdörfer et al., 2022), as well as large-area mapping initiatives such as Sentinel-based crop mapping (http://www.esa-sen2agri.org/).

Besides agricultural land cover characterisation, further information on crop and land management can be derived from RS, such as irrigation events and nutrient application, e.g. nitrogen (N) concentration in the plant tissue. Irrigation is one of the main tools for increasing crop productivity and the information can improve models performance and assist the estimation of water resource demands. Several studies have shown the applicability of RS data for mapping irrigation extent and amount (Ren et al., 2021; Zappa et al., 2021). Although some advances have been made, the majority of these studies focus on the local scale, with the use of a large amount of ground data, and are concentrated on arid and semiarid regions. Many studies have applied RS data for N estimation, based on the correlation between N and some leaf traits. Fewer studies assessed P and K, mostly focusing on vegetation index-based assessment based on hyperspectral and multispectral data.

RS has also been used for soil tillage assessment, mainly by estimation of parameters related to residue cover and surface roughness using optical multi- or hyperspectral and Synthetic Aperture Radar (SAR) sensors (Begue et al., 2018). There is also great interest in the detection and early warning of crop pests and diseases (Schirrmann et al., 2022). However, such a task requires high-resolution images and high-frequency fly-overs (Ibrahim et al., 2023; Oerke et al., 2014), and only a few models are yet ready to include this information (Bregaglio et al., 2021). Lastly, RS has been used to monitor highly dynamic grassland systems and to identify mowing and grazing events based on intra-seasonal dynamics of optical and SAR-based data, such as backscatter and interferometric coherence (Andreatta et al., 2022; Schwieder et al., 2022).

3.2 Proximal sensing

Proximal sensing is another source of information about site-specific conditions of soil and crop, which can deliver model input parameters. Compared to RS, where sensors are deployed on aerial or satellite platforms, proximal sensors are often placed within 2 m from the target (Adamchuk et al., 2018; Deery et al., 2017). Visible, near and mid-infrared spectrometry is one of the most commonly used methods to measure soil and crop properties. For soil, important variables include organic matter, soil nitrogen, particle size distribution, pH and moisture (Adamchuk et al., 2018; Dhawale et al., 2015). Ion-selective potentiometric sensors have also been used to measure the chemical properties of the soil, such as pH and nutrient content (Lobsey et al., 2010). Among the soil variables, moisture is probably the most important variable, as it critically influences crop production (Kashyap and Kumar, 2021). Proximal sensing methods such as ground penetrating radar (Minet et al., 2013), electromagnetic induction (Hedley et al., 2010; Tabbagh et al., 2022), ground-based radiometer and radar systems as well as gamma-ray spectroscopy (Baldoncini et al., 2019) have been

developed for field-scale soil moisture determination (Babaeian et al., 2019; Kashyap and Kumar, 2021).

Optical sensors that record radiation reflected or absorbed by vegetation have been used to collect data non-destructively to assess plant properties such as nutrient content or leaf area index (LAI). While there is a large number of studies that estimate variables related to nitrogen content, such as crop N uptake and concentration (Corti et al., 2018; Morari et al., 2021; Padilla et al., 2018), other macro- and micronutrients remain comparably less investigated. The reflectance measurements can often be complemented with fluorescence data that can give additional information regarding plant characteristics, such as plant stress (Maxwell and Johnson, 2000). The advantage of proximal fluorescence sensing is that it is not sensitive to soil backgrounds, environmental light or biomass conditions. Aside from nitrogen, different fluorescence indices can be useful for estimating nutrient statuses such as potassium, magnesium and calcium (Holland et al., 2019).

Besides optical sensors, light detection and ranging (LiDAR) data have been used to assess the plant growth and canopy properties (El-Naggar et al., 2021). LiDAR has been shown to be useful for deriving information such as LAI (Hosoi et al., 2011), crop density and volume (Saeys et al., 2009) and water use (El-Naggar et al., 2021). In addition, infrared thermography, fluorescence and spectral sensors have been used for the assessment of the plant stress and disease and their spatiotemporal variation (Baker and Rosenqvist, 2004; Oerke et al., 2014). Lastly, there are several research activities towards an integrated use of proximal data with satellite based-observations for upscaling (Munnaf et al., 2021; Wolters et al., 2021).

3.3 Distributed data

Some information that is considered useful to drive AEM is difficult to sense and is not available in maps and other static information products. It often comprises information on crop impacts with limited spatial extent, e.g. scattered flooding, infections of pests and diseases, the appearance of rodents or larger animals that destroy fields while browsing, or farmer's preferences for certain management that is not immediately visible, as e.g. the choice of crop varieties, or the agents being used to battle pests and diseases. However, such information may be available through campaigns, surveys or observations that are distributed unevenly in space and time. Citizen science projects, for example, provide such information, but the nature of this data collection mode entails a sampling design that is far from ideal for data analysis. Experience with distributed data collection through laypersons revealed that people observe where they sojourn and not where their targets occur (Hampf et al., 2021; Kamp et al., 2016), or prefer easy-to-observe targets to others (Callaghan et al., 2021)

and thus produce biased data. Analysing such biased data requires knowledge on the quality issues and appropriate methods that deal with them (Hochachka et al., 2012).

4 Assimilation of data in spatiotemporal simulations

The great variety of input data that AEM consume for their simulations leads to a very large state-space of possible results. Uncertainties afflicted with individual data sources add up and propagate through the course of the simulation, and any method that contributes to reducing these errors is highly welcomed. An idea that emerged in early model applications was the frequent testing against observed data, and instantaneous correction of the simulations in case it went off the observed course of the respective variable (Dorigo et al., 2007; Wallach et al., 2018; de Wit and van Diepen, 2007). The process of regularly updating a running simulation is referred to as data assimilation. The key motivation for data assimilation is the optimal estimation of the modelled state-space at any point in time an observation is available.

Data assimilation into spatiotemporal simulations enables the prediction to consider a wider range of uncertainties than traditional forward, optimisation and calibration simulations. The employed AEM has been constructed on a large empirical basis and they contain complex multivariate relationships and numerical solutions for high-dimensional problems. Data assimilation considers the often neglected uncertainties in atmospheric forcing conditions to the agro-ecosystem models, uncertainties in initial and boundary conditions, model structure (process representation, process dynamics) and model parameterisation (vegetation, soil) and most importantly observation uncertainty. Data assimilation as understood in this chapter is the optimal merging of model predictions and observations under consideration of the uncertainties in both simulation and observation. The consideration of these uncertainties in the model prediction is the key difference of data assimilation to traditional optimisation, inverse modelling, calibration and machine learning approaches. The following section outlines the two main data assimilation methods used in agro-ecosystem modelling: Ensemble Kalman Filters (EnKFs) and Particle Filters.

The original Kalman Filter estimates the optimal position of a system state variable based on observations and observation uncertainty (Kalman, 1960). Usually a numerical model provides further constraints on the estimated system state. The EnKF is the most commonly used and most generalised variant of Kalman Filters in spatiotemporal simulations (Evensen, 1994). The EnKF estimates the system state from the spatiotemporal model forecast, the observation at forecast time, and the uncertainty of both. The update process is schematically represented in Fig. 1. Model uncertainty is calculated

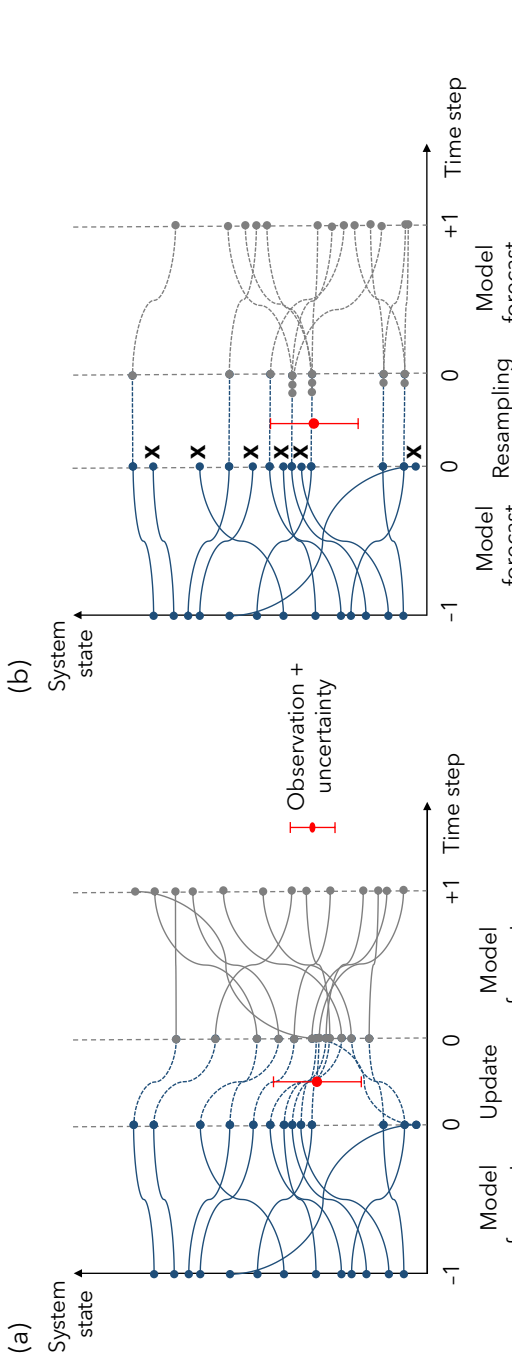

Figure 1 Updating scheme of the Ensemble Kalman Filter (a) at the time t = 0, when the observation (red) becomes available with known uncertainty (red bar). The resampling of the Particle Filter in (b) shows the removal of model realisations marked with X and recreation of new realisations based on weights. The update (a) actively alters the model state while the resampling (b) reproduces existing model states hence maintaining mass and energy balance.

from an ensemble of model realisations, which were created from uncertain and often randomly perturbed initial conditions, parameters and forcing. The observational uncertainty and model forecast uncertainty are optimally combined in the cost function to calculate the Kalman gain and hence the updated model state, from which the model is re-initialised until the next observation becomes available. Parameter updates can also be achieved by extending the state vector with the parameter vector (e.g. Baatz et al., 2017). Parameter updates usually require further assumptions and numerically robust control of parameter updates through e.g. damping factors. Further variants of the Kalman Filter introduce spatial constraints through localisation (e.g. Local Ensemble Transform). The Kalman Filter by Hunt et al. (2007) requires inflation i.e. deterioration of the forecast accuracy or simplifying the calculation of the Kalman gain via the Extended Kalman Filter (e.g. Anderson et al., 2009). This can be useful, as often the main assumption of Gaussian distribution in model and observation uncertainty is not accurate, and localisation of updates avoids non-causal geospatial correlations and inflation prevents overfitting. The EnKF is particularly well suited for large numerically complex models with small ensemble sizes due to the high cost of running numerically expensive models and calculating the Kalman gain. It comes at the expense of potentially altering model assumptions (e.g. energy and mass balance) through model state updates to continuous numerically calculated solutions.

Particle Filters optimise the model forecast with the observation under consideration of both uncertainties as is done by many Kalman Filter variants. Particle Filters consider an ensemble of model realisations as a swarm of particles where each particle is represented by one model realisation (Doucet et al., 2000). Particle Filters propagate the model forward until an observation is available. At the time of observation, the particles i.e. the ensemble of model states are resampled based on the probability density function of the model forecasts and the observation uncertainty. The resampled particle swarm is re-initialised with updated states until the next time step with an available observation. Resampling maintains, removes and reproduces model realisations based on the cost function and the weights assigned to individual realisations. The resampling process is schematically represented in Fig. 1b. Although the main principle remains the one described, a large number of Particle Filter variations were developed mainly by addressing the resampling scheme to account for various problem formulations (e.g. Penny and Miyoshi, 2016; van Leeuwen et al., 2019). Resampling was made subject to localisation, inflation was introduced to avoid overfitting, and the resampling algorithm was extended to include further time steps beyond the one at the time of observation. Particle Filters are independent of the assumption that uncertainties follow Gaussian distribution. Particle Filters typically require a larger ensemble size than EnKF approaches to quantify model uncertainty. This makes Particle Filters suitable for numerically

Published by Burleigh Dodds Science Publishing Limited, 2023.

less complex models. Particle Filters do not alter the conservation laws in the model ensemble due to the resampling of realisations.

Many examples exist where data assimilation was applied to merge forecasts of agro-ecosystem model states and remotely sensed observations in an optimal way (Ling et al., 2019). de Wit and van Diepen (2007) used the EnKF to assimilate the soil water index obtained from the ERS scatterometer at 0.25° spatial resolution into the WOFOST agro-ecosystem model, with the aim to improve maize and winter wheat yield predictions across the European continent. They found limited benefits of the data assimilation, as there were too large errors originating from sources not considered in this study that lowered the predictive performance of the model. Ines et al. (2013) investigated the combined assimilation of LAI and soil water content into DSSAT crop model forecasts with the EnKF. Interestingly, they found the LAI-only assimilation yielded favourable results under high-biomass and wet conditions, and favourable results of the combined assimilation under low-biomass and dry conditions. De Bernardis et al. (2016) applied the Particle Filter with MODIS NDVI observations to update plant phenology of a simple crop growth model, while Zare et al. (2022) applied it with Landsat 7 and 8 and Sentinel-2A and B to update LAI. While studies on vegetation indices other than LAI remain rare, consideration of the range of spectral indices is a promising avenue (e.g. Silvestro et al., 2021). Further advances can also be anticipated when considering problem-focused alternative Particle Filter developments and even merged EnKF variants with the Particle Filter (van Leeuwen et al., 2019). Given the increasing availability of high-performance computing infrastructures, preconditions for numerically complex model-data-fusion frameworks are rapidly improving.

5 Workflows for massive parallel computing

The need for executing large-scale simulations or simulations that also include data assimilation efforts in a reasonable amount of time necessitates the application of large computing resources. These are often situated in HPC facilities or can be accessed via cloud services. To utilise these resources efficiently, the user needs to understand the general structure in which these computing powers are available. In most cases the user has to deal with a potentially very large amount of single machines (tens to thousands) with each machine offering usually a certain amount of computing cores (tens – right now – to hundreds – in the near future) within one machine. In order to use all available cores within a single machine, either many programs can be executed in parallel or one single program can run many execution threads on top of the operating system's threading mechanism. As long as the need for computational resources can be satisfied with a single machine, the only restriction comes from limited access to storage or limited parallelisation potential of the program. Communication

of parallel processes can be fast via the multi-threading approach where code runs in the same address space or somewhat slower if multiple programs communicate via inter-process communication. In both cases, latency due to communication overhead is low and the potential for broken communication channels is reduced. This all changes if the need for computational resources rises and multiple machines have to be involved in the simulations. The first question to be answered is whether a simulation consists of many independent pieces. If not, users have to try to move parts of the simulation set-up closer together to speed up communication, if possible onto the same machine or even process. Independent computations can be distributed more freely. Currently, this is true for most simulations in the agricultural domain. They often use heavyweight point-based (one-dimensional) models running on a grid, with each grid cell being represented by an independent simulation (Elliott et al., 2015; Zhao et al., 2016). In this case, after some initial set-up, the model spends most of its time computing, and less for communication. The smaller the computational part becomes, the larger the ratio of communication (set-up, loading and storing of data) to computation (simulation) gets. These concepts have to be kept in mind in order to understand and evaluate the trade-offs of the general parallelisation methods described below.

Parallelisation concepts can be ordered along multiple dimensions. The first dimension distinguishes data from task parallelism. Data parallelism refers to large amounts of data being processed using a single program. Here, the user wants to apply parallel algorithms within the program to process the data efficiently. From the user's point of view, this may be seen as a single processing pipeline, but the functions being applied to the data work internally in parallel. In contrast, task parallelism is a concept applied to larger tasks (e.g. the abovementioned point model application), which run as separate programs. For a large simulation, the focus is more on how to create the necessary set-up for each task and how to run it. In the agricultural domain, this often means running complex models on millions of set-ups to reflect the changes at different locations (e.g. soil and climate data).

There is a constant gradient between data and task parallelism. Somewhere in the middle, one can find frameworks and environments for (massively) parallel data flows, e.g. Apache Spark (https://en.wikipedia.org/wiki/Apache_Spark) or Storm (https://en.wikipedia.org/wiki/Apache_Storm). Recalling the image of data parallelism as a single pipeline, each execution step may exhibit the potential for parallelisation. If datasets get larger and at some point do not fit into memory anymore, data flow pipelines can be constructed in which each component within the pipeline is specialised for particular processing steps. These components do not all live necessarily within a single machine, as they can also be distributed over many different machines. In that way, a particular data flow can be computationally scaled to the demands of the data

Published by Burleigh Dodds Science Publishing Limited, 2023.

being processed. An example of a particular short data flow is the Map-Reduce pattern (https://en.wikipedia.org/wiki/MapReduce), in which a large amount of independent data is being 'mapped' and then reduced in one or more further steps. Mapping in this case refers to the application of a function (or an equivalent to it) to every single data piece. As the data are independent, this map function can be executed on arbitrary amounts of computational units (cores, processors, machines). The results of the applied function will usually need some aggregation steps, which sequentially follow the mapping step.

The distribution of tasks to the available computational resources can be described using established mechanisms. The most popular is the manual distribution of the tasks and data to the computational units (mostly machines or nodes in a HPC or cloud). On these machines, a program or script will be executed to run the actual program on the set of data assigned to the machine. At the machine level, this resembles batch processing. To distribute the work automatically, cluster management and job scheduling tools like SLURM (https://slurm.schedmd.com/documentation.html) have been designed, which will acquire the necessary resources and run shell scripts on the involved compute nodes. The same tools can also be used to distribute programs, which use a message-passing interface (MPI, https://en.wikipedia.org/wiki/Message_Passing_Interface). Instead of manually taking care of distributing data and code and organising workflows at the operating system level, the user writes a program, which uses standardised interfaces to do this task. Using MPI, the same program will be started on all computational nodes, but one of these nodes is assigned to coordinate the initial distribution of work and the final aggregation of results. The author of the program does not have to be concerned with the peculiarities of the infrastructure anymore and rather writes a single program that uses the underlying MPI implementation of the infrastructure to distribute data and do the necessary communication. Depending on the kind of task, programs using MPI can scale very easily to large amounts of computing units, but with the drawback of being more difficult to debug due to its indirect execution.

Somewhere between manual batch execution and MPI-style program distribution, one can think in terms of clouds of programs. If the time to transfer the input and output data is sufficiently low compared to running the code itself, it can be advantageous to start many instances of a program as a service on remote machines, provided by either a cloud provider or an HPC cluster. Subsequently, a short pipeline is created, consisting of at least one program producing jobs (Producer) for one of the models and another program receiving the results of the model run for further processing (Consumer). This is similar to the Map-Reduce pattern but offers more flexibility in what the Producer can do, as the act of creating a job can be arbitrarily complex. This approach introduces bottlenecks in the sense that in the naive case, one task

is producing jobs for many processes and their results have to be funnelled back into a single process for aggregation. However, the advantages emerge from the fact that the Producer and Consumer processes can be developed incrementally and can be run and even debugged locally as long as remote access to the cloud of models is available. Optimising the job creation side will lead towards the MPI-approach or more complex parallel data flows. However, optimising this concept has a few trade-offs to consider, where computation speed stands against process responsibilities within the pipeline.

6 Model–model integration

Agricultural systems include a range of very different subsystems, including biophysical, socioeconomic and even mixed systems. Simulation models have been developed for almost all of these subsystems, but each based on different methods and modelling philosophies. To analyse high-level feedback loops and system behaviour, it is sometimes desired to couple some of these subsystems, represented by individual models. A simple example is the handshake between a mechanistic agro-ecosystem model providing crop yields and an agent-based economic model that uses these yields to simulate market dynamics or farm-level decision-making. Here, a so-called 'hard coupling' of the two models physically combined in computer code is often not required, as the passing-on of the yield values already suffices for the purpose ('soft coupling'). However, there are more applications imaginable, and some of them become complex. In the following subchapters, we will address some of these challenges and provide some solutions.

6.1 Obstacles for model–model integration

6.1.1 Resolution in time and space

One common issue with coupling models is the individual resolution of the processes represented in the model. As for the temporal resolution, the processes describing the dynamics in agro-ecosystems are typically resolved in days, as a compromise between the need to reflect short-term, event-based management and weather shocks (e.g. heat or frost events) and the availability of weather data to drive the models and computation time. Some of the process descriptions may benefit from a finer-temporal resolution, which is why some models have internal modules that run in an hourly or even minutely resolution. Such models are able to describe day and night processes, or the response of crop and soil to heavy rain and wind gusts. Coupling a daily resolved agro-ecosystem model with an erosion model, in order to assess the impact of a rainfall event on erosion and crop growth simultaneously, requires that both models match in their temporal resolution.

Published by Burleigh Dodds Science Publishing Limited, 2023.

If not, the two remaining options are (i) simplifying the finer resolved model so that it matches the coarser or (ii) running the two models at different paces. As an example, EROSION 3D (Schmidt, 1991) is a mechanistic model that simulates erosion events on agricultural fields in very high temporal resolution, also on marginal slopes. For this purpose, it resolves rainfall data in minutes and requires this data as input. While coupling to an agro-ecosystem model with daily resolution would be feasible at two different paces at the field scale, this concept would hit computational limitations when scaling to larger areas, as for each AEM simulation time step, 1440 simulations of EROSION 3D need to be executed. On the other hand, if reducing the resolution of EROSION 3D to daily time steps, the model would miss important information on rainfall intensity to drive the simulation of erosion events. What remains is a more general erosion approach, such as the Universal Soil Loss Equation (Wischmeier and Smith, 1965), which, however, often underestimates event-based erosion.

At the other end, feedback between soil and plant processes to explain the genesis of soils in agricultural landscapes, elaborate sinks and sources for carbon and subsequently assess carbon sequestration potentials requires the consideration of hundreds if not thousands of years. Models designed for this purpose often use only yearly time steps and thus look at soil and plant management only through frosted glass rather than a binocular. Using an AEM with a soil formation model therefore requires AEM runs for a long time, with the need to make assumptions on crop and soil management (Lugato et al., 2014).

Spatial resolution has played a more prominent role in considerations (Ewert et al., 2011) and a considerable body of literature has grown on the effect of using weather (Hoffmann et al., 2015; Nendel et al., 2013; Zhao et al., 2015), soil (Grosz et al., 2017; Hoffmann et al., 2016; Zhao et al., 2016) or management data (Constantin et al., 2019) at different scales to drive AEM. Even though the general impression from these studies is that scaling between 1 km² to 100 km² does not result in drastic effects on the simulation of crop yields or other variables, the scales below 1 km² have not yet been intensively studied. Suitable data products, especially for soils, are still rare and may not even contain the information that would be relevant, e.g. for investigating macro pore fluxes of infiltrating water after rainfall events, and the respective fast transport of water deep into the soil along cracks, earthworm passages or root channels (Christiansen et al., 2004; Jarvis et al., 1991), which has consequences for the water budget in the topsoil. In general, AEMs are sensitive to soil heterogeneity also at the subfield scale (Wallor et al., 2018). However, their application on grid cells larger than 100 km² comes always with massive simplifications in the assumptions that have to be made for soil distribution and crop management and that – despite a low scaling effect – contribute to the limitations in the

conclusions that can be drawn from such model applications (e.g. Heinicke et al., 2022).

6.1.2 Model philosophies

Different models often clash because of different philosophies being used as the foundation of their development. Using the erosion and crop model coupling as an example again, the one-dimensional nature of many crop or agro-ecosystem models makes it difficult to employ them for science cases in which proximity relationships play a role in the processes considered. While on crop yields, the condition of the neighbouring plants and soils have often little influence, an erosion event means explicit matter exchange with the neighbouring spatial unit. Coupling an AEM with an erosion model therefore requires a mass balance across space, and the ability to add or subtract soil mass to or from the soil profile, with consequences for some soil properties. Burying organic matter under accumulating substrate, or gradually exposing horizons low in organic matter to frequent organic matter inputs through erosion can then be simulated with such a coupled model.

Another example of clashing philosophies is the coupling of an AEM with a farm economy model that requires annual production from the farmer's fields as input. As such, yields can be simulated using an AEM, this coupling seems an easy undertaking to begin with, but the devil lies in the detail. Farm economy models, e.g. MPMAS (Berger, 2001), often do not care about the actual locality of the fields but only about the statistical coherence. Any decision the virtual farmer takes in such a model yields a new statistical distribution of crops being produced on that farm in the new season. However, whether crop rotation rules on individual fields are followed is often beyond focus. In contrast, AEMs do need to make sure that one crop follows another explicitly and plausibly on one field to track carry-over effects of water and nutrients. This spatial coherence is key to any simulation of irrigation water requirements, N fertiliser applications and other cost-afflicted management, which would be of interest to the farm economy model besides yields.

7 Granularity and modular design for model improvement, reuse, exchange and interoperability

The first computational models of crop and soil processes were developed in the 1960s (Jones et al., 2017). Their long history has produced a large number of crop models, which have evolved from a few landmark models such as CERES and EPIC in the USA or SUCROS and ARCWHEAT in Europe (Muller and Martre, 2019). Crop models differ according to their choice of relationships and hypotheses regarding process functioning and feedback loops and their combinations of mechanistic components. In the last decade, the AgMIP

(Rosenzweig et al., 2013) and MACSUR (Rötter et al., 2013) networks of crop modellers have carried out several benchmarking crop simulation model intercomparison studies. These studies highlighted both the capacities and the limits of current models (e.g. Asseng et al., 2013; Bassu et al., 2014; Kollas et al., 2015; Li et al., 2015; Martre et al., 2015) and have led to important model improvements (e.g. Maiorano et al., 2017; Wang et al., 2017). This process highlighted also the importance of equally considering soil processes, which is done in AEM, but often not sufficiently in crop models. However, it became also obvious that model improvement and extension (e.g. to consider crop disease and extreme weather event impacts) is complicated by the different software architectures and programming languages in which they are implemented.

When developing a model or choosing a model for a specific analysis, a key decision to be made is on the level of granularity that is required. Individual choices have to be made regarding how to represent real-world entities in the model, and the boundaries and scope of the system must be defined. All these choices may have to be balanced, including a number of non-functional requirements such as availability and performance. All these contribute to define the granularity of a model and will have important implications on its reusability, transparency, and reproducibility, and on the interpretability of the simulation results (Neveu et al., 2020).

The internal granularity of a component must be distinguished from the granularity of the whole model. The internal granularity of a component is easier to define in most cases and should be at the scale of unitary processes (often defined by a single controlling equation). The choice of the granularity of a modelling solution, which defines the number of components, should be primarily guided by reusability aspects. A model component that simulates the hourly or daily energy balance of canopies is more likely to be reused than a component that simulates the canopies' latent heat flux.

Most crop modelling platforms have a modular architecture (e.g. Brown et al., 2014; Donatelli et al., 2014; Enders et al., 2010; Jones et al., 2001; Wang et al., 2002), which facilitates the development, reuse and extension of models within each of these frameworks. However, their specificities and the different programming languages they use make the reuse of components outside the platform in which they were developed, or the extension of an existing model with components from a different platform, a difficult task. Recently, the Agricultural Model Exchange Initiative has coordinated the development of a centralised framework (Crop2ML, Crop Modelling Meta Language) to support the development of modular model and facilitate the exchange and reuse of soil, plant, and crop process components between modelling platforms (Midingoyi et al., 2020; Midingoyi et al., 2021). Crop2ML reduces the emphasis on the software part of the model, bridging the gap between computer programmers and model developers, who can better capture the biophysical processes

represented in the model. Models in the open-source format Crop2ML can be stored, retrieved, and shared on an easily accessible online open model repository (CropMRespository; http://crop2ml.org). Crop2ML does not aim at replacing existing modelling platforms or at simulating components within large modelling solutions; rather Crop2ML provides a solution to create transparent and well-documented components, which can be easily extended, reused, and shared between crop modelling platforms.

Several other initiatives in other modelling communities are also developing solutions to facilitate the exchange and the coupling of models. For example, in the earth's surface modelling, the Community Surface Dynamics Modelling System (CSDMS; Tucker et al., 2022) provides cyber-infrastructure to promote the quantitative modelling of earth surface processes and distributes software tools and models. The CSDMS developed the Basic Model Interface (BMI), a set of standard query and control functions that make models both easier to learn and easier to couple with other software (Hutton et al., 2020; Peckham et al., 2013). As each community of modellers is developing its own system for coupling and exchanging modelling solutions or model components within its own community, a challenge for more integrated modelling is to ease the coupling of models developed by different communities. The Open Modelling Foundation (https://openmodelingfoundation.github.io) is working in this direction by coordinating and administrating a common, community-developed body of standards and best practices among diverse communities of modelling scientists to support the exchange, reuse, and interoperability of human and natural system models across communities.

8 Concepts for distributed modelling

In the last decades, the world of computing has moved from single machines, which execute program code sequentially, to a world of interconnected devices of all kinds and sizes, in which the mode of operation is increasingly concurrent, even within the devices themselves. There are already thousands of operational systems available, in which continuous data streams (e.g. from sensors) are fed into cloud-computing infrastructures which run large-scale computations in a highly parallel manner. Those systems can be accessed using mobile devices, which as well are able to do multiple things in parallel, for instance accessing different data sources or interacting with remote environments and services.

In today's networked world all components of such a workflow can potentially be distributed, i.e. made living on separate processes, machines or networks. In the extreme case, this distribution is even applied at the level of model components, even though there is a limit to a meaningful granularity of the distributed components. A distributed system needs a way to address and access its remote components. To do this uniformly, a defined protocol has to

be used. In the internet of today, such a widely used protocol is the HyperText Transfer Protocol (HTTP), which allows remote resources to be addressed via Uniform Resource Locators (URL) or more generally as Uniform Resource Identifiers (URI). Taking the HTTP protocol to its logical conclusion leads to the well-established REST-based (REpresentational State Transfer) web-services. This allows the creation of very flexible services built upon the widely supported HTTP mechanism. A disadvantage though lies in its text-based nature, which is less space-efficient and requires many latency-increasing network round trips for even simple service interactions. Diverging from the clean and pure REST mechanism, optimising for some aspects, e.g. transfer efficiency, leads to the RPC (Remote Procedure Call) paradigm. Here, a remote component offers some kind of interface that defines a set of procedures, equivalent to a function in a programming language, to be called. RPC implementations do not have to comply with HTTP rules or text-based data transfer, which is why they often use binary data encoding. Depending on the design goals, these data representations can be more compact for reducing the required transfer bandwidth or optimised for efficient handling, e.g. for minimising copy operations. In many cases, schema languages are used to describe the remote interfaces (set of procedures) as well as the data structures.

The divergence from REST allows for more specialised and optimised interfaces, reducing latency. Nevertheless, complex remote Application Programming Interfaces (API) still suffer from many necessary network round trips. However, the CapTP protocol (http://www.erights.org/elib/distrib/captp/index.html) and its implementations offer a way out, regaining some of the efficiency local procedure calls offer and even getting back some of the flexibility and cleanliness the REST-based approach exhibited, but without the drawbacks. One such implementation is Cap'n Proto, a fast data interchange format and capability-based RPC system (https://capnproto.org). A Cap'n Proto schema file will describe the messages that a remote object will understand and the structure data in these messages will have. To make a model or component remotely accessible it has to implement the according interface. The possibility to have interfaces implemented by different models and components allows them to interoperate. Contrary to most RPC systems, Cap'n Proto allows interfaces themselves to be sent or received in messages. In this context, the interfaces are called capabilities and are the powerful foundation for the Capability Security paradigm (Miller et al., 2003).

One challenge neglected so far is how to manage access and authority to remote objects (models, components and services). In a traditional RPC or REST-based setting access to an URL/URI has to be authenticated. This can happen via usernames and passwords, systems like OAuth (https://en.wikipedia.org/wiki/OAuth) or access tokens (a kind of capability). Except for the last one, these methods make use of ambient authority, i.e. the authority and the means to

exercise this authority are separated. A consequence of this is that a system hosting remote objects and employing ambient authority has to rely on a central authority (an administrative user) to manage access rights. This makes it difficult and sometimes impossible to apply the Principle of Least Authority (https://en.wikipedia.org/wiki/Principle_of_least_privilege), which states that to execute a certain task a user should be given the required authority, but nothing more than that. In our context, that task could refer to running a remote computation on some input data. The actual authority needed is just the ability to send a message to the remote object, which represents this computation. This logic is well supported by Object Capability systems (e.g. Cap'n Proto), in which a capability is the reference to a remote object. Possessing this reference means being authorised to communicate with the remote object. Figure 2 abstractly illustrates this concept: the object *Alice* holds the right to communicate with remote object *Carol*. To allow also remote object *Bob* to talk to remote object *Carol*, *Alice* just has to send *Bob* a message referencing *Carol* (Berg-Mohnicke and Nendel, 2022).

With these foundations in place, ubiquitous connected systems, fast data interchange and capability-secured remote procedure calls it is conceivable to allow for distributed modelling, model application, and more generally distributed workflows. Models or data flows can be composed of remote components without the need to know where they are executed as long as they conform to a common interface and can understand each other. All that is needed is a capability to a remote component in order to use it. Of course, this comes at the cost of efficiency and can be prohibitive in some contexts, for example in applications due to the demand for large amounts of data (potentially without

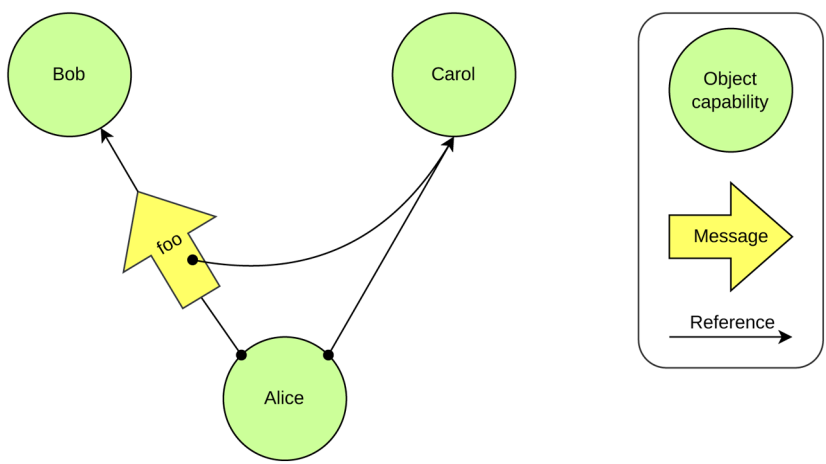

Figure 2 Three owners of an object capability (*Alice*, *Bob* and *Carol*) communicating via the message *foo*. Source: Miller et al. (2003).

data locality), tight coupling with a lot of interactions per time unit (latency being the limiting factor) or reliance on remote availability and less control over the execution environment (e.g. slow hardware, no parallelisation or optimisation possible). Some issues can be mitigated for instance by moving code and data close together. However, while speed is often an issue with distributed modelling, distributing models and components offer new possibilities, such as enabling new ways of collaboration, local testing and debugging, access to large remote datasets or the participation in remote workflows.

9 Future trends and conclusion

Simulation models for different components of the agricultural system have been available already for a while, and often model developers hesitate about touching again a model that has once been satisfactorily coded, parameterised and calibrated. However, increasingly available high-performing computing resources and big data gradually offer more and more application opportunities for models, and this increases the need to adapt structure, code and framework of the models to benefit from this development. Here, standards for data interoperability and cross-platform communication at the model-data and model–model interface are needed and currently under development. Big data especially opens opportunities for large-area applications of models in sufficiently high resolution to allow not only simulations that are closer to the real world that the models represent but also better investigations of the model's representativeness in time and space and the error that the models and their drivers still carry along. Continuous data streams in principle allow data assimilation to improve simulations under limited system knowledge, but the multivariate and highly intertwined nature of the models that we deal with in the agricultural domain often poses challenges with regard to multi-objective optimisation problems. Some driving variables may remain limited in the future, but in the bid to reproduce spatiotemporal patterns of target variables, the use of artificial intelligence may soon help to replace some of the unknown input data. Machine learning has already now demonstrated its power in this realm (Guilpart et al., 2022; Shahhosseini et al., 2021; van Klompenburg et al., 2020; Webber et al., 2020) but remains difficult to interpret (Lischeid et al., 2022).

10 Where to look for further information

Current developments in NetCDF:

- https://www.unidata.ucar.edu/software/netcdf/.

Cap'n Proto overview

- https://capnproto.org.

11 References

Adamchuk, V. I., Ji, W., Viscarra Rossel, R., Gebbers, R. and Trembley, N. (2018). Proximal soil and plant sensing. In: Sudduth, K. and Clay, D. (Eds.), *Precision Agriculture Basics* (pp. 119–140). American Society of Agronomy, Madison, WI.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R. and Avellano, A. (2009). The data assimilation research testbed - A community facility. *Bulletin of the American Meteorological Society* 90(9), 1283–1296. https://doi.org/10.1175/2009BAMS2618.1.

Andreatta, D., Gianelle, D., Scotton, M., Vescovo, L. and Dalponte, M. (2022). Detection of grassland mowing frequency using time series of vegetation indices from Sentinel-2 imagery. *GIScience and Remote Sensing* 59(1), 481–500. https://doi.org/10.1080/15481603.2022.2036055.

Antle, J. M., Valdivia, R. O., Boote, K. J., Janssen, S., Jones, J. W., Porter, C. H., Rosenzweig, C., Ruane, A. C. and Thorburn, P. J. (2015). AgMIP's transdisciplinary agricultural systems approach to regional integrated assessment of climate impacts, vulnerability, and adaptation. In: Rosenzweig, C. and Hillel, D. (Eds.), *Handbook of Climate Change and Agroecosystems* (pp. 27–44). Imperial College Press & American Society of Agronomy. The Agricultural Model Intercomparison and Improvement Project. https://doi.org/10.1142/p970.

Archontoulis, S. V., Miguez, F. E. and Moore, K. J. (2014). A methodology and an optimization tool to calibrate phenology of short-day species included in the APSIM PLANT model: Application to soybean. *Environmental Modelling and Software* 62, 465–477. https://doi.org/10.1016/j.envsoft.2014.04.009.

Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Prasad, P. V. V., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P. J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z. and Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change* 5(2), 143–147. https://doi.org/10.1038/nclimate2470.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P. J., Rotter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P. K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Muller, C., Kumar, S. N., Nendel, C., O'Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J. W., Williams, J. R. and Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change* 3(9), 827–832. https://doi.org/10.1038/nclimate1916.

Baatz, R., Franssen, H. J. H., Han, X. J., Hoar, T., Bogena, H. R. and Vereecken, H. (2017). Evaluation of a cosmic-ray neutron sensor network for improved land surface model prediction. *Hydrology and Earth System Sciences* 21(5), 2509–2530. https://doi.org/10.5194/hess-21-2509-2017.

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H. and Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics* 57(2), 530–616. https://doi.org/10.1029/2018RG000618.

Baker, N. R. and Rosenqvist, E. (2004). Applications of chlorophyll fluorescence can improve crop production strategies: An examination of future possibilities. *Journal of Experimental Botany* 55(403), 1607–1621. https://doi.org/10.1093/jxb/erh196.

Baldoncini, M., Alberi, M., Bottardi, C., Chiarelli, E., Raptis, K. G. C., Strati, V. and Mantovani, F. (2019). Biomass water content effect on soil moisture assessment via proximal gamma-ray spectroscopy. *Geoderma* 335, 69–77. https://doi.org/10.1016/j.geoderma.2018.08.012.

Bassu, S., Brisson, N., Durand, J. L., Boote, K., Lizaso, J., Jones, J. W., Rosenzweig, C., Ruane, A. C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A. R., Kersebaum, K. C., Kim, S. H., Kumar, N. S., Makowski, D., Muller, C., Nendel, C., Priesack, E., Pravia, M. V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D. and Waha, K. (2014). How do various maize crop models vary in their responses to climate change factors? *Global Change Biology* 20(7), 2301–2320. https://doi.org/10.1111/gcb.12520.

Begue, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., Ferraz, R. P. D., Lebourgeois, V., Lelong, C., Simoes, M. and Veron, S. R. (2018). Remote sensing and cropping practices: A review. *Remote Sensing* 10(2), Article 99. https://doi.org/10.3390/rs10010099.

Berg-Mohnicke, M. and Nendel, C. (2022). A case for object capabilities as the foundation of an environmental model and simulation infrastructure. *Environmental Modelling and Software* 156, 105471. https://doi.org/10.1016/j.envsoft.

Berger, T. (2001). Agent-based spatial models applied to agriculture: A simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics* 25(2–3), 245–260. https://doi.org/10.1111/j.1574-0862.2001.tb00205.x.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S. and Hostert, P. (2022). Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sensing of Environment* 269, Article 112831. https://doi.org/10.1016/j.rse.2021.112831.

Boogaard, H., Schubert, J., De Wit, A., Lazebnik, J., Hutjes, R. and Van der Grijn, G. (2020). *Agrometeorological Indicators From 1979 to Present Derived From Reanalysis Version 1.0*. https://doi.org/10.24381/cds.6c68c9bb.

Boryan, C., Yang, Z. W., Mueller, R. and Craig, M. (2011). Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International* 26(5), 341–358. https://doi.org/10.1080/10106049.2011.562309.

Bregaglio, S., Willocquet, L., Kersebaum, K. C., Ferrise, R., Stella, T., Ferreira, T. B., Pavan, W., Asseng, S. and Savary, S. (2021). Comparing process-based wheat growth models in their simulation of yield losses caused by plant diseases. *Field Crops Research* 265, Article 108108. https://doi.org/10.1016/j.fcr.2021.108108.

Brown, H. E., Huth, N. I., Holzworth, D. P., Teixeira, E. I., Zyskowski, R. F., Hargreaves, J. N. G. and Moot, D. J. (2014). Plant modelling framework: Software for building and running crop models on the APSIM platform. *Environmental Modelling and Software* 62, 385–398. https://doi.org/10.1016/j.envsoft.2014.09.005.

Callaghan, C. T., Poore, A. G. B., Hofmann, M., Roberts, C. J. and Pereira, H. M. (2021). Large-bodied birds are over-represented in unstructured citizen science data. *Scientific Reports* 11(1), Article 19073. https://doi.org/10.1038/s41598-021-98584-7.

Constantin, J., Raynal, H., Casellas, E., Hoffman, H., Bindi, M., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Haas, E., Kersebaum, K. C., Klatt, S., Kuhnert, M., Lewan, E., Maharjan, G. R., Moriondo, M., Nendel, C., Roggero, P. P., Specka, X., Trombi, G., Villa, A., Wang, E. L., Weihermuller, L., Yeluripati, J., Zhao, Z. G., Ewert, F. and Bergez, J. E. (2019). Management and spatial resolution effects on yield and water balance at regional scale in crop models. *Agricultural and Forest Meteorology* 275, 184–195. https://doi.org/10.1016/j.agrformet.2019.05.013.

Corti, M., Cavalli, D., Cabassi, G., Gallina, P. M. and Bechini, L. (2018). Does remote and proximal optical sensing successfully estimate maize variables? A review. *European Journal of Agronomy* 99, 37–50. https://doi.org/10.1016/j.eja.2018.06.008.

De Bernardis, C., Vicente-Guijalba, F., Martinez-Marin, T. and Lopez-Sanchez, J. M. (2016). Particle filter approach for real-time estimation of crop phenological states using time series of NDVI images. *Remote Sensing* 8(7), Article 610. https://doi.org/10.3390/rs8070610.

De Negreiros, G. H., Schlesinger, P., Nepstad, D. C. and Lefebvre, P. A. (2009). Pre-LBA RADAMBRASIL project data. https://doi.org/10.3334/ORNLDAAC/941.

de Wit, A. J. W. and van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalmann filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology* 146(1–2), 38–56. https://doi.org/10.1016/j.agrformet.2007.05.004.

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X. and Furbank, R. (2017). Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy* 4(3), 349–379. https://doi.org/10.3390/agronomy4030349.

Dhawale, N. M., Adamchuk, V. I., Prasher, S. O., Rossel, R. A. V., Ismail, A. A. and Kaur, J. (2015). Proximal soil sensing of soil texture and organic matter with a prototype portable mid-infrared spectrometer. *European Journal of Soil Science* 66(4), 661–669. https://doi.org/10.1111/ejss.12265.

Donatelli, M., Bregaglio, S., Confalonieri, R., De Mascellis, R. and Acutis, M. (2014). A generic framework for evaluating hybrid models by reuse and composition - A case study on soil temperature simulation. *Environmental Modelling and Software* 62, 478–486. https://doi.org/10.1016/j.envsoft.2014.04.011.

Dorigo, W. A., Zurita-Milla, R., de Wit, A. J. W., Brazile, J., Singh, R. and Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation* 9(2), 165–193. https://doi.org/10.1016/j.jag.2006.05.003.

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 10(3), 197–208. https://doi.org/10.1023/A:1008935410038.

El-Naggar, A. G., Jolly, B., Hedley, C. B., Horne, D., Roudier, P. and Clothier, B. E. (2021). The use of terrestrial LiDAR to monitor crop growth and account for within-field variability of crop coefficients and water use. *Computers and Electronics in Agriculture* 190, 106416.

Elliott, J., Muller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Buchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K.,

Rosenzweig, C., Ruane, A. C. and Sheffield, J. (2015). The global gridded crop model intercomparison: Data and modeling protocols for phase 1 (v1.0). *Geoscientific Model Development* 8(2), 261–277. https://doi.org/10.5194/gmd-8-261-2015.

Enders, A., Diekkrüger, B., Laudien, R., Gaiser, T. and Bareth, G. (2010). The IMPETUS spatial decision support systems. In: Speth, P., Christoph, M. and Diekkrüger, B. (Eds.), *Impacts of Global Change on the Hydrological Cycle in West and Northwest Africa* (pp. 360–393). Springer.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99(C5), 10143–10162. https://doi.org/10.1029/94JC00572.

Ewert, F., Rotter, 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. and Asseng, S. (2015). Crop modelling for integrated assessment of risk to food production from climate change. *Environmental Modelling and Software* 72, 287–303. https://doi.org/10.1016/j.envsoft.2014.12.003.

Ewert, F., van Ittersum, M. K., Heckelei, T., Therond, O., Bezlepkina, I. and Andersen, E. (2011). Scale changes and model linking methods for integrated assessment of agri-environmental systems. *Agriculture, Ecosystems and Environment* 142(1–2), 6–17. https://doi.org/10.1016/j.agee.2011.05.016.

Fonseca, C. M. and Fleming, P. J. (1993). Genetic algorithms for multiobjective optimization: Formulation, discussion, and generalization. In: *Proceedings of the 5th International Conference on Genetic Algorithms*.

Franke, J. A., Muller, C., Elliott, J., Ruane, A. C., Jagermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P. D., Folberth, C., Francois, L., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Jones, C., Khabarov, N., Koch, M., Li, M., Liu, W. F., Olin, S., Phillips, M., Pugh, T. A. M., Reddy, A., Wang, X. H., Williams, K., Zabel, F. and Moyer, E. J. (2020). The GGCMI Phase 2 experiment: Global gridded crop model simulations under uniform changes in CO2, temperature, water, and nitrogen levels (protocol version 1.0). *Geoscientific Model Development* 13(5), 2315–2336. https://doi.org/10.5194/gmd-13-2315-2020.

Griffiths, P., Nendel, C. and Hostert, P. (2019). Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment* 220, 135–151. https://doi.org/10.1016/j.rse.2018.10.031.

Groh, J., Stumpp, C., Lucke, A., Putz, T., Vanderborght, J. and Vereecken, H. (2018). Inverse estimation of soil hydraulic and transport parameters of layered soils from water stable isotope and Lysimeter data. *Vadose Zone Journal* 17(1), Article 170168. https://doi.org/10.2136/vzj2017.09.0168.

Grosz, B., Dechow, R., Gebbert, S., Hoffmann, H., Zhao, G., Constantin, J., Raynal, H., Wallach, D., Coucheney, E., Lewan, E., Eckersten, H., Specka, X., Kersebaum, K. C., Nendel, C., Kuhnert, M., Yeluripati, J., Haas, E., Teixeira, E., Bindi, M., Trombi, G., Moriondo, M., Doro, L., Roggero, P. P., Zhao, Z. G., Wang, E. L., Tao, F. L., Rotter, R., Kassie, B., Cammarano, D., Asseng, S., Weihermuller, L., Siebert, S., Gaiser, T. and Ewert, F. (2017). The implication of input data aggregation on up-scaling soil organic carbon changes. *Environmental Modelling and Software* 96, 361–377. https://doi.org/10.1016/j.envsoft.2017.06.046.

Guilpart, N., Iizumi, T. and Makowski, D. (2022). Data-driven projections suggest large opportunities to improve Europe's soybean self-sufficiency under climate change. *Nature Food* 3(4), 255–265. https://doi.org/10.1038/s43016-022-00481-3.

Guo, W. X. and Gu, H. B. (2022). Advances in remote/aerial sensing of crop water status. In: Lobsey, C. and Biswas, A. (Eds.), *Advances in Sensor Technology for Sustainable Crop Production*. Burleigh Dodds Science Publishing, Cambridge, UK.

Hampf, A. C., Nendel, C., Strey, S. and Strey, R. (2021). Biotic yield losses in the Southern Amazon, Brazil: Making use of smartphone-assisted plant disease diagnosis data. *Frontiers in Plant Science* 12, Article 621168. https://doi.org/10.3389/fpls.2021.621168.

Harris, G. (2002). Integrated assessment and modelling: An essential way of doing science. *Environmental Modelling and Software* 8152(1), 201–207. https://doi.org/10.1016/s1364-8152(01)00058-5.

Hedley, C. B., Yule, I. J., Zuohy, M. P. and Kusomo, B. H. (2010). *Proximal Sensing Methods for Mapping Soil Water Status in an Irrigated Maize Field*. Springer. https://doi.org/10.1007/978-90-481-8859-8_32.

Heinicke, S., Frieler, K. A., Jagermeyr, J. and Mengel, M. (2022). Global gridded crop models underestimate yield responses to droughts and heatwaves. *Environmental Research Letters* 17(4), Article 044026. https://doi.org/10.1088/1748-9326/ac592e.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G., Ribeiro, E., Wheeler, I., Mantel, S. and Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* 12(2), Article e0169748. https://doi.org/10.1371/journal.pone.0169748.

Hernandez-Suarez, J. S., Nejadhashemi, A. P. and Deb, K. (2021). A novel multi-objective model calibration method for ecohydrological applications. *Environmental Modelling and Software* 144, Article 105161. https://doi.org/10.1016/j.envsoft.2021.105161.

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K. and Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology and Evolution* 27(2), 130–137. https://doi.org/10.1016/j.tree.2011.11.006.

Hoefsloot, P., Ines, A. V. M., van Dam, J., Duveiller, G., Kayitakire, F. and Hansen, J. W. (2012). *Combining Crop Models and Remote Sensing for Yield Prediction: Concepts, Applications and Challenges for Heterogeneous Smallholder Environments*.

Hoffmann, H., Zhao, G., Asseng, S., Bindi, M., Biernath, C., Constantin, J., Coucheney, E., Dechow, R., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Heinlein, F., Kassie, B. T., Kersebaum, K. C., Klein, C., Kuhnert, M., Lewan, E., Moriondo, M., Nendel, C., Priesack, E., Raynal, H., Roggero, P. P., Rotter, R. P., Siebert, S., Specka, X., Tao, F. L., Teixeira, E., Trombi, G., Wallach, D., Weihermuller, L., Yeluripati, J. and Ewert, F. (2016). Impact of spatial soil and climate input data aggregation on regional yield simulations. *PLoS ONE* 11(4), Article e0151782. https://doi.org/10.1371/journal.pone.0151782.

Hoffmann, H., Zhao, G., van Bussel, L. G. J., Enders, A., Specka, X., Sosa, C., Yeluripati, J., Tao, F., Constantin, J., Raynal, H., Teixeira, E., Grosz, B., Doro, L., Zhao, Z., Wang, E., Nendel, C., Kersebaum, K. C., Haas, E., Kiese, R., Klatt, S., Eckersten, H., Vanuytrecht, E., Kuhnert, M., Lewan, E., Rotter, R., Roggero, P. P., Wallach, D., Cammarano, D., Asseng, S., Krauss, G., Siebert, S., Gaiser, T. and Ewert, F. (2015). Variability of effects

of spatial climate data aggregation on regional yield simulation by crop models. *Climate Research* 65, 53–69. https://doi.org/10.3354/cr01326.

Holland, J. E., Cammarano, D., Fitzgerald, G. J., Perry, E. M., Poile, G. and Conyers, M. K. (2019). Proximal fluorescence sensing of potassium responsive crops to develop improved predictions of biomass, yield and grain quality of wheat and barley. *Precision Agriculture* 20(2), 379–397. https://doi.org/10.1007/s11119-018-09629-3.

Holzworth, D. P., Snow, V., Janssen, S., Athanasiadis, I. N., Donatelli, M., Hoogenboom, G., White, J. W. and Thorburn, P. (2015). Agricultural production systems modelling and software: Current status and future prospects. *Environmental Modelling and Software* 72, 276–286. https://doi.org/10.1016/j.envsoft.2014.12.013.

Hoogenboom, G., Jones, J. W., Traore, P. C. S. and Boote, K. J. (2012). Experiments and data for model evaluation and application. In: Kihara, J., Fatondji, D., Jones, J. W., Hoogenboom, G., Tabo, R. and Bationo, A. (Eds.), *Improving Soil Fertility Recommendations in Africa Using the Decision Support System for Agrotechnology Transfer (DSSAT)*. Springer. https://doi.org/10.1007/978-94-007-2960-5_2.

Hosoi, F., Nakabayashi, K. and Omasa, K. (2011). 3-D modeling of tomato canopies using a high-resolution portable scanning Lidar for extracting structural information. *Sensors* 11(2), 2166–2174.

Houska, T., Kraft, P., Liebermann, R., Klatt, S., Kraus, D., Haas, E., Santabarbara, I., Kiese, R., Butterbach-Bahl, K., Müller, C. and Breuer, L. (2017). Rejecting hydro-biogeochemical model structures by multi-criteria evaluation. *Environmental Modelling and Software* 93, 1–12.

Hunt, B. R., Kostelich, E. J. and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena* 230(1–2), 112–126. https://doi.org/10.1016/j.physd.2006.11.008.

Hutton, E. W. H., Piper, M. D. and Tucker, G. E. (2020). The basic model interface 2.0: A standard interface for coupling numerical models in the geosciences. *Journal of Open Source Software* 5(51), 2317.

Ibrahim, E. S., Nendel, C., Kamali, B., Gajere, E. N. and Hostert, P. (2023). Predicting potato diseases for smallholder agriculture in Nigeria using remote sensing and climate data. *PhytoFrontiers*.

Ines, A. V. M., Das, N. N., Hansen, J. W. and Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment* 138, 149–164. https://doi.org/10.1016/j.rse.2013.07.018.

Jagermeyr, J., Muller, C., Ruane, A. C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J. A., Fuchs, K., Guarin, J. R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A. K., Kelly, D., Khabarov, N., Lange, S., Lin, T. S., Liu, W. F., Mialyk, O., Minoli, S., Moyer, E. J., Okada, M., Phillips, M., Porter, C., Rabin, S. S., Scheer, C., Schneider, J. M., Schyns, J. F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F. and Rosenzweig, C. (2021). Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nature Food* 2(11), 875–. https://doi.org/10.1038/s43016-021-00400-y.

Janssen, S., Athanasiadis, I. N., Bezlepkina, I., Knapen, R., Li, H. T., Dominguez, I. P., Rizzoli, A. E. and van Ittersum, M. K. (2011). Linking models for assessing agricultural land use change. *Computers and Electronics in Agriculture* 76(2), 148–160. https://doi.org/10.1016/j.compag.2010.10.011.

Janssen, S. J. C., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W. and Antle, J. M. (2017). Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. *Agricultural Systems* 155, 200–212. https://doi.org/10.1016/j.agsy.2016.09.017.

Jarvis, N. J., Jansson, P. E., Dik, P. E. and Messing, I. (1991). Modeling water and solute transport in macroporous soil. 1. Model description and sensitivity analysis. *Journal of Soil Science* 42(1), 59–70. https://doi.org/10.1111/j.1365-2389.1991.tb00091.x.

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C. and Wheeler, T. R. (2017). Brief history of agricultural systems modeling. *Agricultural Systems* 155, 240–254. https://doi.org/10.1016/j.agsy.2016.05.014.

Jones, J. W., Keating, B. A. and Porter, C. H. (2001). Approaches to modular model development. *Agricultural Systems* 70(2–3), 421–443. https://doi.org/10.1016/S0308-521X(01)00054-3.

Jonquet, C., Dzalé-Yeumo, E., Arnaud, E. and Larmande, P. (2015). *AgroPortal: A proposition for ontology-based services in the agronomic domain. IN-OVIVE'15: atelier INtégration de sources/masses de données hétérogènes et ontologies*. Montpellier: Dans: le Domaine des Sciences du VIVant et de l'Environnement.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82(1), 35–45. https://doi.org/10.1115/1.3662552.

Kamali, B., Mousavi, S. J. and Abbaspour, K. C. (2013). Automatic calibration of HEC-HMS using single-objective and multi-objective PSO algorithms. *Hydrological Processes* 27(26), 4028–4042. https://doi.org/10.1002/hyp.9510.

Kamali, B., Stella, T., Berg-Mohnicke, M., Pickert, J., Groh, J. and Nendel, C. (2022). Improving simulations of permanent grasslands across Germany by multi-objective uncertainty-based calibration of plant-water dynamics. *European Journal of Agronomy* 134, 126464. https://doi.org/10.1016/j.eja.2022.126464.

Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T. and Donald, P. F. (2016). Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions* 22(10), 1024–1035. https://doi.org/10.1111/ddi.12463.

Kashyap, B. and Kumar, R. (2021). Sensing methodologies in agriculture for soil moisture and nutrient monitoring. *IEEE Access* 9, 14095–14121. https://doi.org/10.1109/ACCESS.2021.3052478.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In: *Proceedings of the ICNN'95 - International Conference on Neural Networks* (Vol. 4, pp. 1942–1948).

Kollas, C., Kersebaum, K. C., Nendel, C., Manevski, K., Muller, C., Palosuo, T., Armas-Herrera, C. M., Beaudoin, N., Bindi, M., Charfeddine, M., Conradt, T., Constantin, J., Eitzinger, J., Ewert, F., Ferrise, R., Gaiser, T., de Cortazar-Atauri, I. G., Giglio, L., Hlavinka, P., Hoffmann, H., Hoffmann, M. P., Launay, M., Manderscheid, R., Mary, B., Mirschel, W., Moriondo, M. and Olesen, J. E. (2015). Crop rotation modelling: A European model intercomparison. *European Journal of Agronomy* 70, 98–111. https://doi.org/10.1016/j.eja.2015.06.007.

Li, T., Hasegawa, T., Yin, X. Y., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida, M., Nakagawa, H., Oriol, P., Ruane, A. C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, F. L., Wilkens, P., Yoshida, H., Zhang, Z. and Bouman, B. (2015). Uncertainties in predicting rice yield by current crop

models under a wide range of climatic conditions. *Global Change Biology* 21(3), 1328–1341. https://doi.org/10.1111/gcb.12758.

Ling, X. L., Fu, C. B., Yang, Z. L. and Guo, W. D. (2019). Comparison of different sequential assimilation algorithms for satellite-derived leaf area index using the data assimilation testbed. *Geological Science Development* 12, 3119–3133.

Lischeid, G., Webber, H., Sommer, M., Nendel, C. and Ewert, F. (2022). Machine learning in crop yield modelling: A powerful tool, but no surrogate for science. *Agricultural and Forest Meteorology* 312, Article 108698. https://doi.org/10.1016/j.agrformet.2021.108698.

Lobsey, C. R., Viscara Rossel, R. A. and McBratney, A. B. (2010). Proximal soil nutrient sensing using electrochemical sensors. In: Viscarra Rossel, R. A., McBratney, A. B. and Minasny, B. (Eds.), *Proximal Soil Sensing: Progress in Soil Science* (pp. 77–88). Springer. https://doi.org/10.1007/978-90-481-8859-8_6.

Lokers, R., Knapen, R., Janssen, S., van Randen, Y. and Jansen, J. (2016). Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling and Software* 84, 494–504. https://doi.org/10.1016/j.envsoft.2016.07.017.

Lugato, E., Panagos, P., Bampa, F., Jones, A. and Montanarella, L. (2014). A new baseline of organic carbon stock in European agricultural soils using a modelling approach. *Global Change Biology* 20(1), 313–326. https://doi.org/10.1111/gcb.12292.

Maiorano, A., Martre, P., Asseng, S., Ewert, F., Muller, C., Rotter, R. P., Ruane, A. C., Semenov, M. A., Wallach, D., Wang, E. L., Alderman, P. D., Kassie, B. T., Biernath, C., Basso, B., Cammaran, D., Challinor, A. J., Doltra, J., Dumont, B., Rezaei, E. E., Gayler, S., Kersebaum, K. C., Kimball, B. A., Koehler, A. K., Liu, B., O'Leary, G. J., Olesen, J. E., Ottman, M. J., Priesack, E., Reynolds, M., Stratonovitch, P., Streck, T., Thorburn, P. J., Waha, K., Wall, G. W., White, J. W., Zhao, Z. G. and Zhu, Y. (2017). Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles. *Field Crops Research* 202, 5–20. https://doi.org/10.1016/j.fcr.2016.05.001.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J. W., Rotter, R. P., Boote, K. J., Ruane, A. C., Thorburn, P. J., Cammarano, D., Hatfield, J. L., Rosenzweig, C., Aggarwal, P. K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R. F., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Muller, C., Kumar, S. N., Nendel, C., O'Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stockle, C. O., Stratonovitch, P., Streck, T., Supit, I., Tao, F. L., Travasso, M., Waha, K., White, J. W. and Wolf, J. (2015). Multimodel ensembles of wheat growth: Many models are better than one. *Global Change Biology* 21(2), 911–925. https://doi.org/10.1111/gcb.12768.

Maxwell, K. and Johnson, G. N. (2000). Chlorophyll fluorescence - A practical guide. *Journal of Experimental Botany* 51(345), 659–668. https://doi.org/10.1093/jxb/51.345.659.

Meroni, M., d'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., Seguini, L. and Verhegghen, A. (2021). Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and-2. *Remote Sensing of Environment* 253, Article 112232. https://doi.org/10.1016/j.rse.2020.112232.

Midingoyi, C. A., Pradal, C., Athanasiadis, I. N., Donatelli, M., Enders, A., Fumagalli, D., Garcia, F., Holzworth, D., Hoogenboom, G., Porter, C., Raynal, H., Thorburn, P. J. and

Martre, P. (2020). Reuse of process-based models: Automatic transformation into many programming languages and simulation platforms. *In Silico Plants* 2(1).

Midingoyi, C. A., Pradal, C., Enders, A., Fumagalli, D., Raynal, H., Donatelli, M., Athanasiadis, I. N., Porter, C., Hoogenboom, G., Holzworth, D., Garcia, F., Thorburn, P. and Martre, P. (2021). Crop2ML: An open-source multi-language modeling framework for the exchange and reuse of crop model components. *Environmental Modelling and Software* 142, Article 105055. https://doi.org/10.1016/j.envsoft.2021.105055.

Miller, M. S., Yee, K. P. and Shapiro, J. (2003). *Capability Myths Demolished*.

Minet, J., Verhoest, N. E. C., Lambot, S. and Vanclooster, M. (2013). Temporal stability of soil moisture patterns measured by proximal ground-penetrating radar. *Hydrology and Earth System Sciences Discussions* 10(4), 4063–4097.

Morari, F., Zanella, V., Gobbo, S., Bindi, M., Sartori, L., Pasqui, M., Mosca, G. and Ferrise, R. (2021). Coupling proximal sensing, seasonal forecasts and crop modelling to optimize nitrogen variable rate application in durum wheat. *Precision Agriculture* 22(1), 75–98. https://doi.org/10.1007/s11119-020-09730-6.

Mostafaie, A., Forootan, E., Safari, A. and Schumacher, M. (2018). Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily GRACE data. *Computational Geosciences* 22(3), 789–814. https://doi.org/10.1007/s10596-018-9726-8.

Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W. F., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X. H., de Wit, A. and Yang, H. (2017). Global gridded crop model evaluation: Benchmarking, skills, deficiencies and implications. *Geoscientific Model Development* 10(4), 1403–1422. https://doi.org/10.5194/gmd-10-1403-2017.

Muller, B. and Martre, P. (2019). Plant and crop simulation models: Powerful tools to link physiology, genetics, and phenomics. *Journal of Experimental Botany* 70(9), 2339–2344. https://doi.org/10.1093/jxb/erz175.

Munnaf, M. A., Haesaert, G. and Mouazen, A. M. (2021). Map-based site-specific seeding of seed potato production by fusion of proximal and remote sensing data. *Soil and Tillage Research* 206, Article 104801. https://doi.org/10.1016/j.still.2020.104801.

Nendel, C., Wieland, R., Mirschel, W., Specka, X., Guddat, C. and Kersebaum, K. C. (2013). Simulating regional winter wheat yields using input data of different spatial resolution. *Field Crops Research* 145, 67–77. https://doi.org/10.1016/j.fcr.2013.02.014.

Neveu, P., David, R. and Jonquet, C. (2020). Improving data identification and tagging for more effective decision making in agriculture. In: Armstrong, L. (Ed.), *Improving Data Management and Decision Support Systems in Agriculture*. Burleigh Dodds Science Publishing, Cambridge, UK.

Novelli, F., Spiegel, H., Sandén, T. and Vuolo, F. (2019). Assimilation of Sentinel-2 leaf area index data into a physically-based crop growth model for yield estimation. *Agronomy* 9(5), Article 255. https://doi.org/10.3390/agronomy9050255.

O'Neill, D. W., Fanning, A. L., Lamb, W. F. and Steinberger, J. K. (2018). A good life for all within planetary boundaries. *Nature Sustainability* 1(2), 88–95. https://doi.org/10.1038/s41893-018-0021-4.

Oerke, E.-C., Mahlein, A.-K. and Steiner, U. (2014). Proximal sensing of plant diseases. In: Gullino, M. L. and Bonants, P. J. M. (Eds.), *Detection and Diagnostics of Plant Pathogens* (pp. 55–68). Springer Netherlands.

Padilla, F. M., Gallardo, M., Pena-Fleitas, M. T., de Souza, R. and Thompson, R. B. (2018). Proximal optical sensors for nitrogen management of vegetable crops: A review. *Sensors* 18(7), Article 2083. https://doi.org/10.3390/s18072083.

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., Baker, 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. and Bin, S. (2002). Progress in integrated assessment and modelling. *Environmental Modelling and Software* 8152(1), 209–217. https://doi.org/10.1016/s1364-8152(01)00059-7.

Peckham, S. D., Hutton, E. W. H. and Norris, B. (2013). A component-based approach to integrated modeling in the geosciences: The design of CSDMS. *Computers and Geosciences* 53, 3–12. https://doi.org/10.1016/j.cageo.2012.04.002.

Penny, S. G. and Miyoshi, T. (2016). A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes in Geophysics* 23(6), 391–405. https://doi.org/10.5194/npg-23-391-2016.

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. and Jones, J. W. (2014). Harmonization and translation of crop modeling data to ensure interoperability. *Environmental Modelling and Software* 62, 495–508. https://doi.org/10.1016/j.envsoft.2014.09.004.

Ren, J., Shao, Y., Wan, H., Xie, Y. H. and Campos, A. (2021). A two-step mapping of irrigated corn with multi-temporal MODIS and Landsat analysis ready data. *ISPRS Journal of Photogrammetry and Remote Sensing* 176, 69–82. https://doi.org/10.1016/j.isprsjprs.2021.04.007.

Rezaei, E. E., Ghazaryan, G., Gonzalez, J., Cornish, N., Dubovyk, O. and Siebert, S. (2021). The use of remote sensing to derive maize sowing dates for large-scale crop yield simulations. *International Journal of Biometeorology* 65(4), 565–576. https://doi.org/10.1007/s00484-020-02050-4.

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. and Winter, J. M. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology* 170, 166–182. https://doi.org/10.1016/j.agrformet.2012.09.011.

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. and de Wit, A. (2013, 27–30 May). Challenges for agro-ecosystem modelling in climate change risk assessment for major European crops and farming systems. In: *Impacts World Conference Potsdam*, Germany.

Saeys, W., Lenaerts, B., Craessaerts, G. and De Baerdemaeker, J. (2009). Estimation of the crop density of small grains using LiDAR sensors. *Biosystems Engineering* 102(1), 22–30.

Schirrmann, M. (2023). Advances in remote sensing technologies for assessing crop health. In: Lobsey, C. and Biswas, A. (Eds.), *Advances in Sensor Technology for Sustainable Crop Production* (pp. 43–64). Burleigh Dodds Science Publishing, Cambridge, UK.

Schmidt, J. (1991). A mathematical model to simulate rainfall erosion. In: Bork, H.-R., De Ploey, J. and Schick, A. P. (Eds.), *Erosion, Transport and Deposition Processes—Theories and Models CATENA* 19.

Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C. and Hostert, P. (2022). Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sensing of Environment* 269, Article 112795. https://doi.org/10.1016/j.rse.2021.112795.

Shahhosseini, M., Hu, G. P., Huber, I. and Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports* 11(1), Article 1606. https://doi.org/10.1038/s41598-020-80820-1.

Silvestro, P. C., Casa, R., Hanus, J., Koetz, B., Rascher, U., Schuettemeyer, D., Siegmann, B., Skokovic, D., Sobrino, J. and Tudoroiu, M. (2021). Synergistic use of multispectral data and crop growth modelling for spatial and temporal evapotranspiration estimations. *Remote Sensing* 13(11), Article 2138. https://doi.org/10.3390/rs13112138.

Skovdal Christiansen, J., Thorsen, M., Clausen, T., Hansen, S. and Christian Refsgaard, J. (2004). Modelling of macropore flow and transport processes at catchment scale. *Journal of Hydrology* 299(1–2), 136–158. https://doi.org/10.1016/j.jhydrol.2004.04.029.

Steffen, W., Richardson, K., Rockstrom, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., de Vries, W., de Wit, C. A., Folke, C., Gerten, D., Heinke, J., Mace, G. M., Persson, L. M., Ramanathan, V., Reyers, B. and Sorlin, S. (2015). Sustainability. Planetary boundaries: Guiding human development on a changing planet. *Science* 347(6223), Article 1259855. https://doi.org/10.1126/science.1259855.

Tabbagh, A., Seger, M. and Cousin, I. (2022). Using proximal electromagnetic/electrical resistivity/electrical sensors to assess soil health. In: Lobsey, C. and Biswas, A. (Eds.), *Advances in Sensor Technology for Sustainable Crop Production*. Burleigh Dodds Science Publishing, Cambridge, UK.

Tang, Y. T., Marshall, L., Sharma, A. and Ajami, H. (2018). A Bayesian alternative for multi-objective ecohydrological model specification. *Journal of Hydrology* 556, 25–38. https://doi.org/10.1016/j.jhydrol.2017.07.040.

Tucker, G. E., Hutton, E. W. H., Piper, M. D., Campforts, B., Gan, T., Barnhart, K. R., Kettner, A. J., Overeem, I., Peckham, S. D., McCready, L. and Syvitski, J. (2022). CSDMS: A community platform for numerical modeling of Earth surface processes. *Geoscientific Model Development* 15(4), 1413–1439. https://doi.org/10.5194/gmd-15-1413-2022.

Van Evert, F. K. (2019). Data for developing, testing and applying crop and farm models. In: *Boote, Advances in Crop Modelling for a Sustainable Agriculture* (pp. 385–418). Burleigh Dodds Science Publishing, Cambridge, UK.

van Ittersum, M. K., Ewert, F., Heckelei, T., Wery, J., Olsson, J. A., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., Olsson, L., Rizzoli, A. E., van der Wal, T., Wien, J. E. and Wolf, J. (2008). Integrated assessment of agricultural systems - A component-based framework for the European Union (SEAMLESS). *Agricultural Systems* 96(1–3), 150–165. https://doi.org/10.1016/j.agsy.2007.07.009.

Published by Burleigh Dodds Science Publishing Limited, 2023.

van Klompenburg, T., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* 177, Article 105709. https://doi.org/10.1016/j.compag.2020.105709.

van Leeuwen, P. J., Kunsch, H. R., Nerger, L., Potthast, R. and Reich, S. (2019). Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society. Royal Meteorological Society* 145(723), 2335–2365. https://doi.org/10.1002/qj.3551.

Wallach, D., Makowski, D., Jones, J. W. and Brun, F. (2018). Data assimilation for dynamic models. In: Wallach, D., Makowski, D., Jones, J. W. and Brun, F. (Eds.), *Working With Dynamic Crop Models* (3rd edn., pp. 487–518). Academic Press.

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

Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R. P., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A., Liu, L., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ripoche, D., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y. and Asseng, S. (2017). The uncertainty of crop yield projections is reduced by improved temperature response functions. *Nature Plants* 3(8), Article 17102. https://doi.org/10.1038/nplants.2017.102.

Wang, E., Robertson, M. J., Hammer, G. L., Carberry, P. S., Holzworth, D., Meinke, H., Chapman, S. C., Hargreaves, J. N. G., Huth, N. I. and McLean, G. (2002). Development of a generic crop model template in the cropping system model APSIM. *European Journal of Agronomy* 18(1–2), 121–140. https://doi.org/10.1016/S1161-0301(02)00100-4.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. and Schewe, J. (2014). The inter-sectoral impact model intercomparison project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences of the United States of America* 111(9), 3228–3232. https://doi.org/10.1073/pnas.1312330110.

Webber, H., Lischeid, G., Sommer, M., Finger, R., Nendel, C., Gaiser, T. and Ewert, F. (2020). No perfect storm for crop yield failure in Germany. *Environmental Research Letters* 15(10), Article 104012. https://doi.org/10.1088/1748-9326/aba2a4.

White, J. W., Hunt, L. A., Boote, K. J., Jones, J. W., Koo, J., Kim, S., Porter, C. H., Wilkens, P. W. and Hoogenboom, G. (2013). *Integrated Description of Agricultural Field Experiments and Production: A ICASA Version 2.0 Data Standards. Computers and Electronics in Agriculture* (Vol. 96, pp. 1–12). https://doi.org/10.1016/j.compag.2013.04.003.

Wischmeier, W. C. and Smith, D. D. (1965). *Predicting Rainfall-Erosion Losses From Cropland East of the Rocky Mountains* (Vol. 282). United States Department of Agriculture.

Wöhling, T., Gayler, S., Priesack, E., Ingwersen, J., Wizemann, H. D., Hogy, P., Cuntz, M., Attinger, S., Wulfmeyer, V. and Streck, T. (2013). Multiresponse, multiobjective

calibration as a diagnostic tool to compare accuracy and structural limitations of five coupled soil-plant models and CLM3.5. *Water Resources Research* 49(12), 8200–8221. https://doi.org/10.1002/2013WR014536.

Wöhling, T., Schöniger, A., Gayler, S. and Nowak, W. (2015). Bayesian model averaging to explore the worth of data for soil-plant model selection and prediction. *Water Resources Research* 51(4), 2825–2846. https://doi.org/10.1002/2014WR016292.

Wolters, S., Soderstrom, M., Piikki, K., Reese, H. and Stenberg, M. (2021). Upscaling proximal sensor N-uptake predictions in winter wheat (*Triticum aestivum* L.) with Sentinel-2 satellite data for use in a decision support system. *Precision Agriculture* 22(4), 1263–1283. https://doi.org/10.1007/s11119-020-09783-7.

Yapo, P. O., Gupta, H. V. and Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology* 204(1–4), 83–97. https://doi.org/10.1016/S0022-1694(97)00107-8.

Younes, N., Joyce, K. E. and Maier, S. W. (2021). All models of satellite-derived phenology are wrong, but some are useful: A case study from northern Australia. *International Journal of Applied Earth Observation and Geoinformation* 97, Article 102285. https://doi.org/10.1016/j.jag.2020.102285.

Zappa, L., Schlaffer, S., Bauer-Marschallinger, B., Nendel, C., Zimmerman, B. and Dorigo, W. (2021). Detection and quantification of irrigation water amounts at 500 m using Sentinel-1 surface soil moisture. *Remote Sensing* 13(9), Article 1727. https://doi.org/10.3390/rs13091727.

Zare, H., Weber, T. K. D., Ingwersen, J., Nowak, W., Gayler, S. and Streck, T. (2022). Combining crop modelling with remote sensing data using a particle filtering technique to produce real-time forecasts of winter wheat yields under uncertain boundary conditions. *Remote Sensing* 14(6), 1360.

Zhao, G., Hoffmann, H., van Bussel, L. G. J., Enders, A., Specka, X., Sosa, C., Yeluripati, J., Tao, F. L., Constantin, J., Raynal, H., Teixeira, E., Grosz, B., Doro, L., Zhao, Z. G., Nendel, C., Kiese, R., Eckersten, H., Haas, E., Vanuytrecht, E., Wang, E., Kuhnert, M., Trombi, G., Moriondo, M., Bindi, M., Lewan, E., Bach, M., Kersebaum, K. C., Rotter, R., Roggero, P. P., Wallach, D., Cammarano, D., Asseng, S., Krauss, G., Siebert, S., Gaiser, T. and Ewert, F. (2015). Effect of weather data aggregation on regional crop simulation for different crops, production conditions, and response variables. *Climate Research* 65, 141–157. https://doi.org/10.3354/cr01301.

Zhao, G., Hoffmann, H., Yeluripati, J., Xenia, S., Nendel, C., Coucheney, E., Kuhnert, M., Tao, F. L., Constantin, J., Raynal, H., Teixeira, E., Grosz, B., Doro, L., Kiese, R., Eckersten, H., Haas, E., Cammarano, D., Kassie, B., Moriondo, M., Trombi, G., Bindi, M., Biernath, C., Heinlein, F., Klein, C., Priesack, E., Lewan, E., Kersebaum, K. C., Rotter, R., Roggero, P. P., Wallach, D., Asseng, S., Siebert, S., Gaiser, T. and Ewert, F. (2016). Evaluating the precision of eight spatial sampling schemes in estimating regional means of simulated yield for two crops. *Environmental Modelling and Software* 80, 100–112. https://doi.org/10.1016/j.envsoft.2016.02.022.

Published by Burleigh Dodds Science Publishing Limited, 2023.