European Journal of Agronomy 95 (2018) 33–44

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Data requirements for crop modelling—Applying the learning curve approach to the simulation of winter wheat flowering time under climate change

M. Montesino-San Martin[a,*], D Wallach[b], J.E. Olesen[c], A.J. Challinor[d], M.P Hoffman[e], A.K. Koehler[d], R.P Rötter[e,f], J.R. Porter[a,g]

[a] *Department of Plant and Environmental Science, University of Copenhagen, Højbakkegård Allé 30, 2630 Taastrup, Denmark*
[b] *National Institute for Agricultural Research (INRA), UMR AGIR, Toulouse, France*
[c] *Department of Agroecology, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark*
[d] *Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK*
[e] *University of Göttingen, Tropical Plant Production and Agricultural Systems Modelling (TROPAGS), Grisebachstraße 6, 37077 Göttingen, Germany*
[f] *University of Göttingen, Centre for Biodiversity and Sustainable Land Use, Büsgenweg 1, 37077 Göttingen, Germany*
[g] *System Montpellier SupAgro, INRA, CIHEAM-IAMM, CIRAD, Univ Montpellier, 34060 Montpellier, France*

ARTICLE INFO

Keywords:
Learning curve
Anthesis
Triticum aestivum
Dataset
Climate change

ABSTRACT

A prerequisite for application of crop models is a careful parameterization based on observational data. However, there are limited studies investigating the link between quality and quantity of observed data and its suitability for model parameterization. Here, we explore the interactions between number of measurements, noise and model predictive skills to simulate the impact of 2050′s climate change (RCP8.5) on winter wheat flowering time. The learning curve of two winter wheat phenology models is analysed under different assumptions about the size of the calibration dataset, the measurement error and the accuracy of the model structure. Our assessment confirms that prediction skills improve asymptotically with the size of the calibration dataset, as with statistical models. Results suggest that less precise but larger training datasets can improve the predictive abilities of models. However, the non-linear relationship between number of measurements, measurement error, and prediction skills limit the compensation between data quality and quantity. We find that the model performance does not improve significantly with a theoretical minimum size of 7–9 observations when the model structure is approximate. While simulation of crop phenology is critical to crop model simulation, more studies are needed to explore data needs for assessing entire crop models.

1. Introduction

Models are increasingly used in impact assessments of climate change on crop production and food security (Ruane et al., 2017). Models intended for these applications require suitable datasets to minimize the error in the projections (Wallach, 2011). The crop modelling community has repeatedly addressed and improved the definition of suitable datasets (Nix, 1983; Boote, 1999; Hunt et al., 2001; White et al., 2013). The latest efforts have been made in the context of AgMIP (Rosenzweig et al., 2013) and MACSUR (Rötter et al., 2013) projects. Boote et al. (2016) developed a generic qualitative method that ranks datasets based on the presence or absence of input and state variables. Kersebaum et al. (2015) designed a numerical classification approach where rules based on expert opinion provide scores for several desirable features. The total quality score of a dataset is the summation of scores from each feature. Further contributions to the definition of suitable datasets go through replacing expert opinion by empirically based rules. Hence, further research is needed assessing the impacts of dataset features on simulations and model performance. Confalonieri et al. (2016) worked in this direction by introducing a method for assessing changes in model performance depending on measurement errors. He et al. (2017) quantified the repercussions of the number of seasons and state variables on their effectiveness to calibrate a crop model. The results of these studies are key to elucidate the interactions between data and crop model but their comparison with the rules in Kersebaum et al. (2015) is not straightforward. In order to favour this comparison, features of datasets should be changed and assessed in a progressive and comprehensive manner.

* Corresponding author.
E-mail address: manuelmontesino@plen.ku.dk (M. Montesino-San Martin).

https://doi.org/10.1016/j.eja.2018.02.003
Received 31 October 2017; Received in revised form 2 February 2018; Accepted 14 February 2018
1161-0301/

The number of observations and the measurement error (as a proxy for number of replicates) are two essential features of datasets in the scoring system by Kersebaum et al. (2015). This is due to their critical role in estimating model parameters and their uncertainty (Wallach et al., 2011; Confalonieri et al., 2016) and the relevance of parameter uncertainty in impact assessments of climate change (Wallach et al., 2011; Wallach et al., 2017). Large and accurate datasets could reduce parameter uncertainty but the crop modelling community has suffered from chronic data scarcity exacerbated by ensemble modelling (Rötter et al., 2011; Jones et al., 2017). The maturation of new information technologies, namely mobile technology and remote sensing, and the implementation of new initiatives, such as crowdsourcing, could help solving this situation (Janssen et al., 2017) at the cost of accuracy. An assessment of suitable datasets for crop modelling in terms of number of observations and measurement error may bring light to the potential benefits of these technologies to improve crop impact projection performance.

The learning curve approach evaluates in a progressive manner the impact of the size and measurement error of the calibration dataset on model performance. Learning curves are graphs displaying the evolution of simulation errors with the size of the training dataset (Perlich et al., 2003; Perlich, 2011). Errors usually evolve asymptotically with the size of the training dataset, increasing for the training dataset and decreasing for the testing dataset. The shape of the curves can reveal, for instance, when the model is considered to have a sufficiently large calibration dataset. The size is considered large enough when greater observations produce small changes in the simulation skills. However, defining when the changes are small enough depends on the model application. The learning curve approach has been used in the past with statistical models in the field of machine learning (e.g. Perlich, 2011 or Figueroa et al., 2012). To our knowledge, the method has not been applied yet for the assessment of dataset features in crop modelling.

Drawing the learning curves requires calibrating and evaluating the model repeatedly, changing the size of the calibration dataset. This makes the process computationally demanding and data intensive. Phenology combines its relevance for yield (Craufurd and Wheeler, 2009) with its simple mathematical formulation and fast execution (e.g. Ceglar et al., 2011). Within the phenology phases, flowering is particularly critical; it is a very sensitive phase to temperature extremes (Ugarte et al., 2007) and it defines the balance between source-sink organs. Therefore, the simulation of flowering time represents a practical starting point to introduce the learning curve approach into crop modelling. Phenology modelling offers several working solutions with different mathematical formulations (Ceglar et al., 2011; Alderman and Stanfill, 2017). Learning curves are likely influenced by model structures, since prediction skills of different modelling hypotheses vary due to specific error compensations forged during calibration (Wallach et al., 2011). Hence, robust conclusions about data-model interactions with the learning curves require the assessment of multiple structures.

Our study aims to analyse the influence of datasets on model simulation performance. More specifically, we seek to elucidate the impact of number and measurement error of crop state variables on the prediction skills of a phenology model intended for climate change applications. We apply the learning curve approach which allows the progressive assessment of properties of datasets and brings the opportunity to compare the evolution of model performance with the scoring rules specified in the data classification system. Additionally, we inspect possible compensations between size and measurement error thanks to their joint analysis.

2. Methods

The generation of learning curves is a two-step process repeated multiple times. The first step is the calibration and evaluation of the models against the training (or calibration) dataset. The second step is the evaluation of the predictive skills of the model against the testing (or evaluation) dataset. The training dataset varies in number of observations (quantity of observations) and levels of measurement error (quality of observations). Long series of records (greater than 10 seasons) of flowering dates required to construct the learning curves are scarce. Hence, data is replaced by the simulations of a "*perfect model*" with structure and parameter values considered to be true. The simulations from such perfect models are masked with different levels of noise. This perfect model approach gives us full control over the number of seasons and errors introduced in the datasets. In addition, it allows the evaluation of the simulation model predictive skills against the perfect model under climate change.

Two phenology models for simulating anthesis dates of winter wheat under climate change are considered; the Broken-Sticks (BS) and Continuous Curvilinear (CC) (Wang and Engel, 1998) models. The BS is a wide-spread practical model to simulate phenology whereas the CC model is considered a more realistic version from a biological perspective (Streck et al., 2008). Consequently, we assume that the CC model is the "*perfect model*" and the BS and the CC models are used as simulation models. Thus, two situations concerning model structures are assessed; (S1) the structure of the simulation model is an exact representation of reality (the simulation model and the "*perfect model*" are the same, both represented by the CC model), and (S2) the structure of the simulation model approximates the reality (the BS and the CC model correspond to the simulation model and the "*perfect model*" respectively). The results are used to analyse the shape of the learning curves and understand the relationships between measurements, errors and model structures.

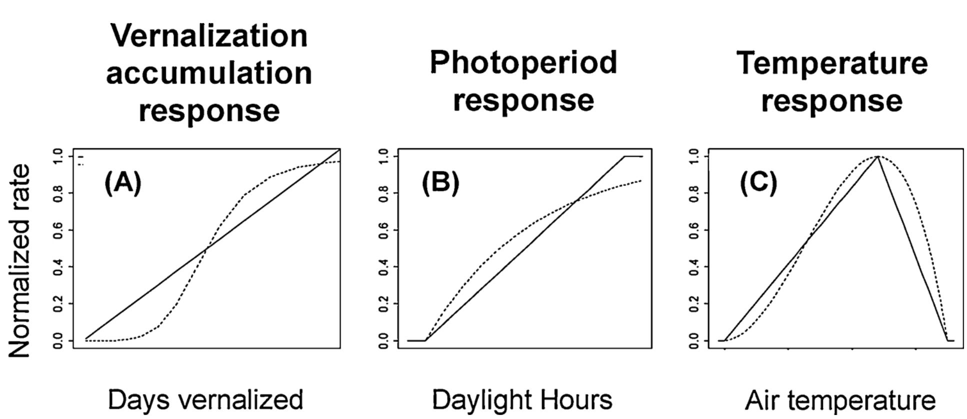

Fig. 1. Normalized responses of crop development to vernalization (A), photoperiod (B) and temperatures (C) simulated by the Broken-Sticks Model (solid line) and the Continuous Curvilinear Model (dashed line).