Agricultural and Forest Meteorology 204 (2015) 10–21

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

A comparison of within-season yield prediction algorithms based on crop model behaviour analysis

B. Dumont[a,b,*], B. Basso[b], V. Leemans[a], B. Bodson[c], J.-P. Destain[c], M.-F. Destain[a]

[a] ULg Gembloux Agro-Bio Tech, Department Environmental Sciences and Technologies, 5030, Gembloux, Belgium
[b] Deptartment Geological Sciences, Michigan State University, East Lansing, MI 48823, USA
[c] ULg Gembloux Agro-Bio Tech, Department Agronomical Sciences, 5030 Gembloux, Belgium

ARTICLE INFO

Article history:
Received 3 April 2014
Received in revised form 13 January 2015
Accepted 21 January 2015
Available online 31 January 2015

Keywords:
STICS crop model
Climate variability
LARS-WG
Yield prediction
Log-normal distribution
Convergence in Law Theorem
Central Limit Theorem

ABSTRACT

The development of methodologies for predicting crop yield, in real-time and in response to different agro-climatic conditions, could help to improve the farm management decision process by providing an analysis of expected yields in relation to the costs of investment in particular practices. Based on the use of crop models, this paper compares the ability of two methodologies to predict wheat yield (*Triticum aestivum* L.), one based on stochastically generated climatic data and the other on mean climate data. It was shown that the numerical experimental yield distribution could be considered as a log-normal distribution. This function is representative of the overall model behaviour. The lack of statistical differences between the numerical realisations and the logistic curve showed in turn that the Generalised Central Limit Theorem (GCLT) was applicable to our case study. In addition, the predictions obtained using both climatic inputs were found to be similar at the inter and intra-annual time-steps, with the root mean square and normalised deviation values below an acceptable level of 10% in 90% of the climatic situations. The predictive observed lead-times were also similar for both approaches. Given (i) the mathematical formulation of crop models, (ii) the applicability of the CLT and GLTC to the climatic inputs and model outputs, respectively, and (iii) the equivalence of the predictive abilities, it could be concluded that the two methodologies were equally valid in terms of yield prediction. These observations indicated that the Convergence in Law Theorem was applicable in this case study. For purely predictive purposes, the findings favoured an algorithm based on a mean climate approach, which needed far less time (by 300-fold) to run and converge on same predictive lead time than the stochastic approach.

1. Introduction

Agricultural production is greatly affected by variability in weather Semenov et al., 2009; Supit et al., 2012. Providing an opportunity to study the effects of variable inputs, such as weather events on harvestable crop parts, crop models have been used successfully to support the decision-making process in agriculture Basso et al., 2011; Ewert et al., 2011; Thorp et al., 2008. The development of methodologies for predicting grain yield, in real time and in response to different agro-climatic conditions Dumont et al., 2014b; Lawless and Semenov, 2005, would further improve farm management decisions by providing an analysis of the trade-off between the value of expected crop yields and the cost of inputs.

Plant growth and development can be seen as systems linked to the environment in linear and non-linear ways (Campbell and Norman, 1989; Semenov and Porter, 1995). Many of the links between crop dynamics and atmospheric variables are non-linear and interdependent. Crop models were developed about 40 years ago as an effective substitute for ambiguous and cumbersome field experimentation (Sinclair and Seligman, 1996). The greater expectations from modelling rapidly led to increasingly detailed descriptions of the functioning of the biotic and abiotic components of cropping systems, leading to an increase in complexity and computer sophistication. Crop models provide the best-known approach for improving our understanding of complex plant processes as influenced by pedo-climatic and management conditions (Semenov et al., 2007), and they have proved to be more heuristic tools than simply a substitute for reality (Sinclair and Seligman, 1996). Most physically based soil-crop models operate on a daily time basis and simulate the evolution of variables of interest through daily dynamic accumulation.

* Corresponding author at: 2 Passage des Déportés, 5030 Gembloux, Belgium. Tel.: +32 81 622163; fax: +32 81 622167.
E-mail address: benjamin.dumont@ulg.ac.be (B. Dumont).

http://dx.doi.org/10.1016/j.agrformet.2015.01.014

In crop models, weather conditions need to be described as accurately as possible. Weather data are the input data that drive the model and daily crop growth. It has been shown that weather data have a greater effect on yield than technical data and soil parameterisation (Nonhebel, 1994). In addition, crop model predictions (such as phenological development, biomass growth, or yield elaboration) are affected by temporal fluctuations in temperature and/or precipitation, even when the mean values remain similar (Semenov and Porter, 1995). It has been demonstrated that historical mean weather data might be inappropriate for predicting crop growth because of the non-linear response of crops to agro-environmental conditions (Porter and Semenov, 1999, 2005; Semenov and Porter, 1995). The sequencing of weather events greatly affects dynamic crop simulations; interactive stresses might have a greater impact on the final value of crop characteristics of interest (such as grain yield) than individual stresses (Riha et al., 1996).

Important research has been done on estimating the form of historical crop yield distributions. Day (1965) analysed crop yield distributions using the Pearson System and found that: (i) crop yield distribution is generally non-normal and non-log-normal, whereas (ii) the skewness and kurtosis of yield distribution (the mathematical third and fourth central moment, respectively) depend on the specific crop and the amount of available nutrients. His conclusions were corroborated by Du et al. (2012), who considered that the development of a complete theory on the effect of input constraints on yield skewness required empirical studies on diverse crops grown in different production environments. Several authors (Just and Weninger, 1999; Ramirez et al., 2001) have tried to assess the normality of crop yield distribution, but have not been able to do so. Just and Weninger (1999) identified three specific reasons for this: (i) the misspecification of the non-random components of yield distributions, (ii) the misreporting of statistical significance and (iii) the use of aggregate time-series data to represent farm-level yield distributions. Numerous works have referred to the 'usual left-tail problem', which deals with the low probability of occurrence of some very low yields, characterised by particularly poor climate conditions (Hennessy, 2009a). More recently, Hennessy (2009a,b, 2011),),) analysed crop yield expectations with reference to the Law of the Minimum Technology and the Law of Large Number.

Within the context of yield prediction, there is a distinction between statistical models and process-based models. In the early 1960s, the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) developed a method for assessing crop yield based on several sources of information, including various types of surveys and field-level measurements. These yield forecasting models are based on analysing relationships of samples at the same stage of maturity in comparable months over the preceding 4 years (Allen et al., 1994; Keller and Wigton, 2003). More recently, the statistical models have been coupled with remote data and recorded climatic measurements covering a preliminary period of a few months (Doraiswamy et al., 2007). As the yield prediction model is empirical and not physically based, this approach has serious limitations: (i) the future impact of past stress effects is not integrated into the physiological plant growth and (ii) the compensation mechanisms of crop management are not fully considered.

Process-based crop model approaches appear to be better alternatives for yield prediction, but crop models should rely on data that reflect hypothetical future scenarios. An appropriate and sophisticated approach for predicting grain yield with incomplete weather data was described by Lawless and Semenov (2005). It is based on the use of the Sirius crop simulation model (Jamieson et al., 1998; Semenov et al., 2007, 2009) and the LARS-WG stochastic weather generator (WG) (Racsko et al., 1991; Semenov and Barrow, 1997). The methodology for predicting grain yield with incomplete weather data was related to the crop's life cycle: based on observed weather for the first part of the growing season, the authors used a stochastic WG to produce a probabilistic ensemble of synthetic weather time-series for the remainder of the season. WGs can be used to generate multiple stochastic realizations of extended sequences of real historical weather data (Lawless and Semenov, 2005; Mavromatis and Hansen, 2001; Mavromatis and Jones, 1998; Singh and Thornton, 1992), allowing risk assessment studies to be performed. The weather time-series built in this way were then used as an input in a crop simulation model to generate distributions of crop characteristics (such as phenological stages, end-season grain yields). As the season progressed, the uncertainty of the crop simulations decreased. This approach is interesting, but time-consuming and machine intensive.

Another method would involve replacing future data by forecasted weather. The initial problems here, though, are that forecasting has a time limit and that forecast accuracy diminishes with the long-time predictions. An added problem is the need to downscale data from a Global or Regional Climate Model (GCM/RCM) to local conditions at a resolution suitable for crop simulation models. The EU-funded DEMETER and ENSEMBLES projects are probably the two most representative examples of this application in Europe (Cantelaube and Terres, 2005; Challinor et al., 2005; Hewitt, 2004; Palmer et al., 2005). It is worth mentioning that GCM/RCM downscaling can be achieved by linking a seasonal forecast with a WG (Semenov and Doblas-Reyes, 2007), which allows yield prediction to be performed. It has been shown, however, that this approach is not any better at yield prediction than the approach based on historical climatology (Semenov and Doblas-Reyes, 2007).

Dumont et al. (2014b) have developed a similar approach. They assessed the potential of overcoming the lack of future weather data by using seasonal averages. For each of the climatic variables necessary to run the crop model (temperature, precipitation, solar radiation, vapour pressure, wind speed), they computed the seasonal averages as the daily mean values calculated from a 30-year historical weather database. Being based on only one future projection, it was very light in terms of computational requirement.

The aim of our study was to compare the efficiency of two crop yield prediction methodologies that are based only on historical records. To make the yield predictions, the Lawless and Semenov (2005) approach, based on using a high number of stochastically generated climate data, and the Dumont et al. (2014b) methodology, based on using seasonal averages, were selected. Both approaches benefit from the same amount of realised information. In each of the studies, relevant yield predictions could be made only at a late stage, but no research had ever compared the methodologies in an identical case study or using the same crop model. Comparing the efficiency of the two methodologies relied on an in-depth analysis of crop model behaviour based on a sound statistical foundation. The research findings reported by Day (1965) and Hennessy, 2009a,b, 2011 were applied to our study of crop model behaviour and the mathematical nature of the computed weather time-series is discussed in relation to the Convergence in Law Theorem and Central Limit Theorem (CLT).

2. Material and methods

2.1. Overview of the procedure

To answer the question of whether the predictive approaches have equal potential in terms of their ability to predict yield with the same accuracy and lead-time, we developed a four-step procedure (see Fig. 1). The first step focused on the applicability of the CLT to the weather input generation. In other words, it has to be