Agricultural and Forest Meteorology 221 (2016) 219–229

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Estimation of ecotype-specific cultivar parameters in a wheat phenology model and uncertainty analysis

Zunfu Lv[a,b,1], Xiaojun Liu[a,1], Liang Tang[a], Leilei Liu[a], Weixing Cao[a], Yan Zhu[a,*]

[a] National Engineering and Technology Center for Information Agriculture, Jiangsu Key Laboratory for Information Agriculture, Nanjing Agricultural University, Jiangsu Collaborative Innovation Center for Modern Crop Production, 1 Weigang Road, Nanjing, Jiangsu 210095, PR China
[b] Department of Agronomy, The Key Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, College of Agriculture and Food Science, Zhejiang A & F University, Lin'an, Hangzhou 311300, PR China

ARTICLE INFO

Article history:
Received 21 July 2015
Received in revised form 17 February 2016
Accepted 21 February 2016
Available online 1 March 2016

Keywords:
MCMC
Phenology model
Ecotype-specific cultivar parameter
Wheat

ABSTRACT

The objective of this study was to develop an effective method for predicting ecotype-specific cultivar parameters for the large-scale application of wheat crop simulation models. The Markov Chain Monte Carlo (MCMC) technique was explored for parameter calibration based on an existing model. We estimated the posterior probability distribution of ecotype-specific cultivar parameters at four ecosites (Huai'an, Zhengzhou, Weifang, and Shijiazhuang in China) by using the historical phenological stages of wheat and daily weather data from 1980 to 1995. 1000 sets of cultivar parameters which were randomly sampled from the posterior probability distribution at each ecosite from 1996 to 2005 were used to evaluate the MCMC-based method. Optimal 50 sets of parameters were chosen from the 1000 sets of parameters at each ecosite to represent the ecotype-specific cultivar parameters of the Jiangsu, Henan, Shandong and Hebei provinces to evaluate the MCMC-based method for the year 2005 at the regional scale. The results showed that the coefficients of determination (R^2) between the observed and estimated phenological stages ranged from 0.61 to 0.72, with a root mean square error (RMSE) of less than 3.6 days and a root mean square deviation $\left(\overline{\text{RMSD}}\right)$ of less than 3.7 days at the site scale. All of the RMSE and $\overline{\text{RMSD}}$ values for the three phenological stages obtained using the posterior probability distribution at the four ecosites were significantly lower than those based on the prior probability distribution. At the regional scale, R^2 between the observed and estimated phenological stages was greater than 0.86, with RMSE less than 3.4 days and $\overline{\text{RMSD}}$ less than 4.0 days in most of the grids for year 2005. The estimated phenological stages agreed well with the observations, suggesting that the MCMC technique has high reliability of for estimating multiple parameter combinations in a wheat phenology model. The combination of the present MCMC technique and a phenology model could be used for estimating the ecotype-specific cultivar parameters for the main wheat growing regions of China, which can be used to predict progress of wheat development at the regional scale.

1. Introduction

Crop simulation models have been widely used to predict grain productivity, precision crop management, regional agricultural design, and are considered as an important method for quantitative evaluation of agricultural production systems (Liu et al., 2002). Many of these simulation analyses are based on site-specific predictions. At the regional scale, however, significant spatial variations exist in agricultural production environments (climate conditions and soil characteristics) and management systems (Priya and Shibasaki, 2001; Shi et al., 2009). In turn, accurately applying a crop simulation model at the regional scale remains a challenging problem in the field of agricultural system analysis.

In recent years, studies have been carried out on the regional application of crop models. Priya and Shibasaki (2001) integrated environmental policy integrated climate (EPIC) model with a geographic information system (GIS) and established a spatial crop yield estimating system (Spatial EPIC). Shi et al. (2009) constructed a system for predicting and evaluating regional wheat productivity based on a wheat simulation model (WheatGrow) and GIS. In general, these models require input of various regional parameters, including regional cultivar parameters, management practices, as well as weather and soil data. Yet determination of cultivar parame-

* Corresponding author.
E-mail addresses: yanzhunjaueducn@163.com, yanzhu@njau.edu.cn (Y. Zhu).
[1] The first two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.agrformet.2016.02.016
0168-1923/

ters at the regional scale is a complex process, limiting the effective application of crop simulation models to regional scale predictions (Iizumi et al., 2009).

When a crop simulation model is extended from a site to a regional scale, the performance of a specific cultivar becomes less important. Instead, an integration of the main cultivars in the region is considered representative of an ecotype cultivar needed for predicting the regional productivity (Iizumi et al., 2009; Shi et al., 2009; Tao et al., 2009). The climatic ecotype is one type of ecotype. This climatic ecotype cultivar should have common characteristics of main cultivars grown in the region, such as regional climate adaptation after long-term crop selection. Therefore, the climate ecotype-specific cultivar parameters are related to regional climate and are considered constant within an ecological zone. For example, the parameters that determine the phenological characteristics have shown no clear scale dependency or, at most, weak scale dependency (Iizumi et al., 2014). Other parameters are cultivar-specific and differ across cultivars within an ecological zone. Parameters related to phenology are considered climate ecotype-specific cultivar parameters and are used for regional simulation in a wider region.

Yet it is difficult to obtain ecotype-specific parameters with traditional field-testing procedures (He, 2008). Thus, an estimation method is needed for generating ecotype-specific parameters using regional historical data of weather conditions, crop phenology, and yield. Jin and Shi (2006) adopted a trial and error method to determine the ecotype-specific parameters using the time series yield data and weather conditions at a county scale. Xiong et al. (2007) also applied the trial and error method to calibrate ecotype-specific parameters for simulating regional maize production with the CERES-Maize model. Several other methods are used for estimating these parameters, such as genetic algorithms (Dai et al., 2009) and simulated annealing method (Ferreyra, 2004). However, these methods present a partially optimal rather than globally optimal combination of model parameters and do not reflect the characteristics of similar impacts of different parameters (i.e., different combinations of the parameters tend to produce similar simulation results) (He et al., 2009). To solve this problem, some researchers have used the Bayesian approach to obtain a posterior distribution of a parameter based on a prior distribution and an observed dataset for optimizing the parameters of a multivariate nonlinear model. This approach is becoming popular in regional simulation studies and has been used with a crop growth model (Iizumi et al., 2013), a canopy transpiration model (Samanta et al., 2007) and a terrestrial ecosystem model (Harmon and Challenor, 1997; Xu et al., 2006). Other researchers have confirmed the benefits of using this methodology for impact assessment in the fields of forest, biogeochemistry and groundwater flow (Van Oijen et al., 2005; Arhonditsis et al., 2008; Hassan et al., 2009).

The Bayes' approach (Gill 2002; Braswell et al., 2005; Raupach et al., 2005; Van Oijen et al., 2013) requires two types of information: prior information based on expert knowledge and data collected via experimentation. Its basic principle is to start with a prior probability distribution of crop model parameters, which describes our perception of the parameter values before experimental observations. Through updating prior parameter distributions and using the measurements, the Bayesian framework produces posterior parameter distributions (Makowski et al., 2006). Due to the complexity of crop models (non-linearity, high number of parameters), it is almost impossible to analytically calculate the posterior parameter distributions (Gelman et al., 2004). However, the growing power of computers and the development of new methods make the Bayesian approach more accessible even for complex models (Makowski et al., 2002). The Markov Chain Monte Carlo (MCMC) methods are effective in solving this problem (Hassan et al., 2009). The method is based on the Bayesian theoretical framework to establish a balanced distribution of the Markov chain and samples from its balanced distribution. By continuously updating the sample information, the chain fully searches parameter intervals and ultimately converges to the high probability density areas. This method helps to transform some complex high-dimensional problems into simple low-dimensional ones.

The objectives of this study were to find the posterior distribution of ecotype-specific cultivar parameters by combining the MCMC method with a process-based wheat model (WheatGrow) developed by the authors' group (Cao and Moss, 1997; Yan et al., 2000; Lv et al., 2013), to validate the effect of the MCMC-based method for parameter estimation at site and regional scales, and to quantify the parameter uncertainty on phenology estimation at site and regional scales. The expected results may provide a technical basis for estimating ecotype-specific cultivar parameters and developing regional applications of crop simulation models.

2. Materials and methods

2.1. Selection of study regions

Based on the wheat sowing area of each county in China obtained from the statistical yearbook for 2005, we used the sowing-area-weighted method (Jagtap and Jones, 2002) to aggregate the data of county-level wheat sowing areas in 2005 to $0.5^\circ \times 0.5^\circ$ resolution grid-level value. We then calculated the ratio of the wheat sowing area to the total area in each grid (Fig. 1).

Four major wheat production provinces, the Jiangsu, Hebei, Henan and Shandong provinces, which have the largest wheat sowing fraction of land in China, were chosen as study regions. Within the study region, we selected four representative ecosites (represented as green stars in Fig. 1), Huai'an (119.02°, 33.4°), Zhengzhou (113.4°, 34.49°), Weifang (119.05°, 36.42°) and Shijiazhuang (114.38°, 37.53°), to calibrate and validate the ecotype-specific cultivar parameters at the site scale. We used 62 other ecosites (represented as green triangles in Fig. 1) for validation at the regional scale.

2.2. Data collection and processing

Daily meteorological data from 1980 to 2005 for our four representative ecosites, and from the 62 other ecosites from 2004 to 2005, were provided by the Chinese Meteorological Administration. The data included daily maximum and minimum temperatures, precipitation, and sunshine hours. The sunshine hours were converted to solar radiation using Pohlert's method (2004). The National Meteorological Center Library of China provided the historical measurements of phenological stages, (which included sowing, jointing, anthesis, and maturity in wheat crops) at the four representative ecosites from 1980 to 2005, as well as measurements from the 62 other ecosites from 2004 to 2005. The different cultivars planted at each ecosite from 1980 to 2005 were cultivars recommended in each province.

Since wheat phenology prediction and uncertainty analysis at the regional scale required daily meteorological data at each $0.1^\circ \times 0.1^\circ$ resolution grid, the thin-plate smoothing splines of Hutchinson (TPS), described in the ANUSPLIN software package version 4.3 (Hutchinson, 1995, 2004), were used to generate the interpolated daily meteorological surfaces for the study area. This was based on the meteorological data from 62 ecosites in the four Chinese provinces from 2004 to 2005. The raw data from Digital Elevation Model (DEM) provided by the Consultative Group on International Agricultural Research (New et al., 2002) were used to help generate interpolated daily meteorological data. A second-order spline was fit by using latitude, longitude, and ele-