Agricultural and Forest Meteorology 152 (2012) 223–232

Contents lists available at SciVerse ScienceDirect

Agricultural and Forest Meteorology

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

Efficient stabilization of crop yield prediction in the Canadian Prairies

Luke Bornn*, James V. Zidek

Department of Statistics, University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, Canada V6T 1Z2

ARTICLE INFO

Article history:
Received 11 February 2011
Received in revised form 13 July 2011
Accepted 16 September 2011

Keywords:
Crop yield prediction
Crop water stress index
Canadian Prairies
Bayesian
Smoothing
Spatial correlation

ABSTRACT

This paper describes how spatial dependence can be incorporated into statistical models for crop yield along with the dangers of ignoring it. In particular, approaches that ignore this dependence suffer in their ability to capture (and predict) the underlying phenomena. By judiciously selecting biophysically based explanatory variables and using spatially-determined prior probability distributions, a Bayesian model for crop yield is created that not only allows for increased modelling flexibility but also for improved prediction over existing least-squares methods. The model is focused on providing efficient predictions which stabilize the effects of noisy data. Prior distributions are developed to accommodate the spatial non-stationarity arising from distinct between-region differences in agricultural policy and practice. In addition, a range of possible dimension–reduction schemes and basis expansions are examined in the pursuit of improved prediction. As a result, the model developed has improved prediction performance relative to existing models, and allows for straightforward interpretation of climatic effects on the model's output.

1. Introduction

This paper presents a method for forecasting wheat crop yields in the Canadian Prairie Provinces—a challenging task due to dramatic variability in yield over space and time. Its importance, however, should not be understated: wheat is one of Canada's primary exports, accounting for 12% of wheat and barley traded in the world market. Thus variation in yield has considerable impact both within and beyond Canadian borders (Schmitz and Furtan, 2000). Enabling effective crop management, handling, and marketing thus requires accurate predictions of crop yield that account for and explain these variations. For example, these forecasts are helpful in setting insurance premiums and futures prices as well as in managing grain transport. Since spatial and temporal climate variability affect crop yields (Stone and Meinke, 2005; Potgieter et al., 2006), a crop yield forecasting method must include climate as an essential component if it is to be successful.

Several process-based models have been successfully used for crop yield prediction including the Agricultural Production Systems Simulator (APSIM) in Australia (Keating et al., 2003) as well as a web-based tool developed by the United States' Southeast Climate Consortium (Jagtap et al., 2002). These process-based models typically employ tunable and user adjustable deterministic and stochastic models to simulate biological and physical processes related to crop yield. While these models use knowledge pertaining to the individual processes, they often require significant input from the user, including a wide range of meteorological and environmental variables which may be difficult or expensive to obtain.

In contrast to the above, traditional statistical techniques are purely empirical. While these methods may result in accurate predictions, they typically lack the interpretability of process-based models (Barnett, 2004). As a result of this criticism, recent years have seen the development of statistical models that also provide interpretation of the underlying biophysical process (see, for example, Stephens (1995), Hansen et al. (2002)). One such process knowledge-based approach involves water stress indices (Potgieter et al., 2005, 2006; Qian et al., 2009a,b), the result of which has been of tremendous use and benefit to stakeholders, allowing for prediction and understanding of crop yield anomalies. While these models have improved the prediction of crop yield, there exists scope for improvement through (a) providing an efficient dimension reduction of explanatory variables; (b) accounting for uncertainty in the estimated technology trend; (c) modelling spatial correlation between regions.

This paper describes the results of a project coordinated by Agriculture and Agri-foods Canada to develop a model that explains and predicts wheat yield and its relation to climatic variables. With plans for an online implementation in the future, efficiency was required as a feature of the model, as was the ability to stabilize the effects of noisy measurements. Building on earlier work, we employ a crop water stress index (SI) to provide explanatory power for a new crop yield predictor (De Jong and Bootsma, 1996). To improve prediction over existing approaches, we extract

* Corresponding author. Tel.: +1 604 822 0570.
E-mail addresses: l.bornn@stat.ubc.ca (L. Bornn), jim@stat.ubc.ca (J.V. Zidek).

0168-1923/$ – see front matter
doi:10.1016/j.agrformet.2011.09.013

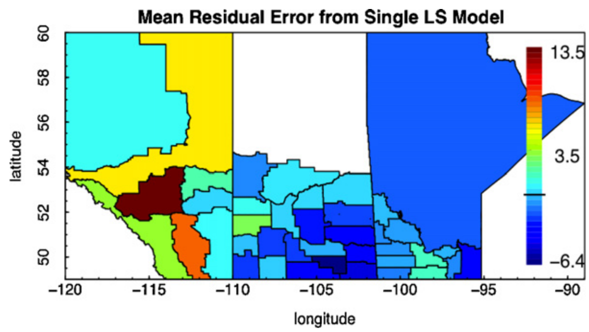

Fig. 1. Mean residuals from model (1). We observe that the model residuals are spatially correlated.

a sensitive yet low-dimensional summary of this stress index, comparing various alternatives and bases before ultimately selecting principal components. We then demonstrate its improved prediction performance compared to currently used windowed average approaches. In contrast to previous work which models each agricultural region separately, we create a unified model that allows strength to be borrowed from adjacent and nearby regions, thus stabilizing both inference and prediction. By employing a spatially-motivated context-specific prior distribution on the parameters of interest, we account for and use spatial correlation between sites while smoothing and consequently improving predictions.

Following this introduction, Section 2 describes the crop yield forecasting problem and available data. This section works through a series of successively improved models, eventually leading to a Bayesian model in Section 2.3 which jointly models all regions simultaneously. Model testing and diagnostics are explored in Section 3. Lastly, Section 4 concludes the work.

2. Materials and methods

This paper models crop yield in the Canadian Prairies as a function of climate-related explanatory variables. The data include annual wheat yields (in bushels per acre) along with associated measurements of a crop water stress index and growing degree day (both described later) for 40 agricultural regions (plotted in Fig. 1) across the Canadian Prairies from 1976 to 2006. The agricultural regions are those used in the 2006 Canadian Census of Agriculture, through which the data are also obtained, and are determined from climate and soil information. For each of the 31 years and 40 regions, yield is an aggregated average across the region. Likewise, stress index and growing degree day are calculated regionally, but on a daily basis throughout the growing season (April 1 to September 30).

2.1. Incorporating soil water

The well recognized influence of soil water on crop yields dictates its inclusion in any yield prediction model (De Jong and Bootsma, 1996). However, due to the time-consuming and costly process of measuring soil water content, in practice its effects must be inferred from more widely available environmental variables such as precipitation, temperature, and easily measured crop and soil-related factors. A suite of models have been developed which attempt to understand soil water availability in the context of these environmental variables. Beginning with simple water balance approaches that balance precipitation and soil water storage with evapotranspiration and water runoff, these models have increased in their complexity over the years (Thornthwaite, 1948; De Jong and Bootsma, 1996). For the reasons given below we focus on budget models, which build on the premise that above a certain threshold (called the ‘field capacity’), soil cannot absorb any more water and therefore any additional water is drained off through runoff or drainage. Also, if the soil water fails to be replenished through precipitation, irrigation, or other sources, the soil reaches a point where plant roots are no longer capable of uptaking water. This stage is known as the ‘wilting point’.

Evapotranspiration, which describes the sum of evaporation and plant transpiration, measures the water lost from plants, soil, and other land surfaces into the atmosphere. There are two key components in the budget model, potential evapotranspiration (PET) and actual evapotranspiration (AET). PET represents the atmospheric demand for evapotranspiration; specifically, it accounts for the energy available to evaporate water and transport it into the lower atmosphere. AET is the actual water content available for evaporation and transpiration, and relies on plant physiology and soil characteristics for its calculation. When the soil has ample water, the actual evapotranspiration (AET) can equal the PET. However when the soil is not at its field capacity, AET will be less than PET. More details on these concepts and soil science in general may be found in Brady et al. (1999).

Budget models are straightforward to implement since they require a minimum of meteorological data as well as soil field capacities and wilting points. While more advanced models have been built which include soil hydraulic characteristics and more complex relationships between soil, plant, and meteorological systems, these models require considerably more information from the user, including detailed soil and plant characteristics. Because of the additional variables required by these models, we employ a budget model in the remainder of this work. Our model uses crop water stress index (SI) over agricultural land, defined as 1-AET/PET (Qian et al., 2009a,b). This quantity will be near 0 when water is plentiful in the soil and near 1 when the plant is stressed by a lack of available moisture. Intuition might suggest directly including precipitation, temperature, soil and plant information into the model. However, doing so would add a large number of variables, especially considering that many of these variables are observed for every day of the growing season. Using the SI instead provides an economical reduction in the dimensionality of the description space in a way that respects the biophysical processes involved in soil water movement and availability.

2.1.1. Predicting yield with SI

We begin by detailing the process of fitting a regression model to crop yield using least squares (LS). First let $y_{j,t}$, $j = 1, \ldots, 40$ be the yield from region j for years $t = 1976, \ldots, 2006$. Since SI is a daily value, we create an annual average for each year and region; let $\overline{si}_{j,t}$ denote the vector of these means in year t for each region j. We begin by fitting a common regression model to all regions, specifically

$$y_{j,t} = \beta_0 + \beta_1 t + \beta_2 \overline{si}_{j,t} + \epsilon_{j,t}. \tag{1}$$

Here $\epsilon_{j,t}$ for year t and region j represents a combination of model and measurement error. While previously developed statistical models for crop yield account for a technology trend by first fitting a regression on time and then modelling the residuals, such approaches yield little understanding about the uncertainty associated with forecasting. In particular, while forecasts that use detrended data may be similar, their associated variances will be biased as uncertainty in the trend model is ignored. In fact, to properly account for all sources of variability the technology trend should be an integral part of any forecasting model.

To begin, note that the simple model in Eq. (1) relies on only 3 parameters—all regions are described by the same equation. The validity of inference for such a model relies on assumptions including for instance that the errors $\epsilon_{j,t}$ are stochastically independent