G Model
EURAGR-25591; No. of Pages 10

ARTICLE IN PRESS

Europ. J. Agronomy xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark

Behzad Sharif[a,*], David Makowski[b], Finn Plauborg[a], Jørgen E. Olesen[a]

[a] Department of Agroecology, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark
[b] Unité Mixte de Recherche 211 Agronomie, INRA, AgroParisTech, Université Paris-Saclay, Thiverval-Grignon, France

ARTICLE INFO

Article history:
Received 2 May 2016
Received in revised form 15 September 2016
Accepted 19 September 2016
Available online xxx

Keywords:
Winter oilseed rape
Statistical models
Yield
Climate
Regression

ABSTRACT

Statistical regression models represent alternatives to process-based dynamic models for predicting the response of crop yields to variation in climatic conditions. Regression models can be used to quantify the effect of change in temperature and precipitation on yields. However, it is difficult to identify the most relevant input variables that should be included in regression models due to the high number of candidate variables and to their correlations. This paper compares several regression techniques for modeling response of winter oilseed rape yield to a high number of correlated input variables.

Several statistical regression methods were fitted to a dataset including 689 observations of winter oilseed rape yield from replicated field experiments conducted in 239 sites in Denmark, covering nearly all regions of the country from 1992 to 2013. Regression methods were compared by cross-validation.

The regression methods leading to the most accurate yield predictions were Lasso and Elastic Net, and the least accurate methods were ordinary least squares and stepwise regression. Partial least squares and ridge regression methods gave intermediate results. The estimated relative yield change for a +1°C temperature increase during flowering was estimated to range between 0 and +6 %, depending on choice of regression method. Precipitation was found to have an adverse effect on yield during autumn and winter. It was estimated that an increase in precipitation of +1 mm/day would result in a relative yield change ranging from 0 to −4 %. Soil type was also important for crop yields with lower yields on sandy soils compared to loamy soils. Later sowing was found to result in increased crop yield.

The estimated effect of climate on yield was highly sensitive to the chosen regression method. Regression models showing similar performance led in some cases to different conclusions with respect to effect of temperature and precipitation. Hence, it is recommended to apply an ensemble of regression models, in order to account for the sensitivity of the data driven models for projecting crop yield under climate change.

1. Introduction

Winter oilseed rape (*Brassica napus* L.) is one of the most important oilseed crops throughout the world. It is also widely cultivated in Denmark, currently mostly for biodiesel production. Quantifying the effect of different factors contributing to the yield of this crop could help in improving productivity, especially concerning responses to climate change.

There is little doubt that climate change will affect the yield of oilseed rape as well as of most other crops (Peltonen-Sainio et al., 2010). There have been few studies dealing with the effect of climate change on winter oilseed rape production in Europe. Tuck et al. (2006) estimated that the potentially suitable areas for oilseed rape production will increase in northern latitudes of Europe. Also, there have been several studies on effects of individual climatic factors affecting winter oilseed rape production (Diepenbrock, 2000; Rathke et al., 2006).

Temperature is an important factor influencing both development and growth throughout the growth phases of oilseed rape (Rathke et al., 2006). In cooler regions, resistance to low temperatures during winter is a determining factor. In such areas, rapid early growth of the crop due to high N availability, early sowing and warm temperatures might result in longer stems that enhance susceptibility to frost damages. There is a strong correlation between

* Corresponding author.
E-mail address: behzadsharif@gmail.com (B. Sharif).

http://dx.doi.org/10.1016/j.eja.2016.09.015
1161-0301/

intercepted radiation and plant growth during spring and full flowering (Mendham et al., 1981), and intercepted photosynthetically active radiation correlates well with crop yield (Rathke et al., 2006).

Drought stress during germination negatively affects crop establishment and subsequent growth (Rathke et al., 2006), but long wet periods may lead to less favorable seed beds and thus reduce crop establishment. Evans et al. (2008) investigated the relation between phoma stem canker (caused by *Leptosphaeria maculans* and *L. biglobosa*) and climatic variables in the UK. Their model could predict the onset and severity of the disease. Similarly, a study on light leaf spot disease (*Pyrenopeziza brassicae*) of winter oilseed rape in England and Wales showed the importance of climatic conditions also for this disease (Welham et al., 2004).

Many of the above-mentioned factors might be taken into account either by empirical models or by process-based models. However, some of the biotic interactions in particular with pests and diseases are notoriously difficult to handle in process-based simulation models (Ewert et al., 2015). In addition, parameterization of process-based crop models is difficult and the use of inaccurate parameter estimates can increase prediction errors (Wallach, 2011). Statistical models developed from large datasets could, therefore, offer useful alternatives to process-based models for predicting crop yield. However, most published studies using statistical models of crop-climate relationships are so far limited to annual averages of temperature and precipitation which may be useful for predictions, but would not take into account the effects of climate at key periods of time during the growing season (Lobell and Burke, 2010).

Here we aim to analyze the effects of the regression technique, of the temporal resolution of the climatic input data, and of the level of model complexity on the estimated effect of climate on winter oilseed rape yield. Seven different statistical methods, two temporal resolutions, and three levels of model complexity are combined to develop 42 regression models. These models are fitted to a large dataset and are evaluated by cross-validation. Results are used to estimate effects of climate changes on oilseed rape yields and to analyze the associated uncertainties.

2. Materials and methods

2.1. Data

Experimental yield data was acquired from Danish Agricultural Advisory Service for winter oilseed rape. Every year replicated field experiments with a randomized block design (typically four replicates) were conducted at several sites in Denmark to test winter oilseed rape under different management practices, including comparison of varieties or use of different rates of fertilization or use of pesticides. Here, we only used yield data obtained under one specific type of management: the common (standard) management practice for winter oilseed rape, which applied recommended dosages of fertilizer and pesticides and used a reference variety. The final dataset used in this study included 689 winter oilseed rape yield data collected from 1992 to 2013 in 239 sites. The yield in the experiments was measured using a plot combine harvester, and yields were recorded with a moisture content of 8%. All considered trials were characterized by location, soil type, and sowing date.

In order to derive and assign weather data to each experiment, a 40 × 40 km climate grid (Kristensen et al., 2011) covering Denmark was used to obtain interpolated daily weather data. Number of yield observations over the applied grids is presented in Fig. 1. The climatic factors included were average temperature, global radiation and precipitation. Location of each experiment was assigned to one of the grid cells based on the postal address. Fig. 2 shows the yield of winter oilseed rape at all the sites used for this study, based on year of harvest. There is a clear increasing yield trend that continues throughout the period from 1992 to 2013.

In order to compare the effect of temporal resolution of climatic data on the accuracy of yield prediction, climate variables were averaged over both fortnightly and monthly temporal resolutions, and were then considered as two series of model inputs. Other temporal resolutions could have been used, but since the purpose of our study was to compare the performance of different regression techniques, we only composed two temporal resolutions.

For each observation, soil type was characterized based on the Danish soil classification system (Krogh et al., 2003), which uses topsoil texture as the classification criterion. Based on this we classified soils into two classes, sandy and loamy soils, and soils with clay content less than 10% in the topsoil were considered as sandy soils and the rest as loamy soils. Soil type was thus defined as a dummy variable that could be either loamy or sandy. There were in total 511 observations with loamy soils, and 175 observations with sandy soils.

The previous crop was also taken into account as a dummy variable, which was classified as cereals, grass, pea, or set-a-side with 621, 40, 10 and 18 observations, respectively.

Sowing date was defined as a continuous variable and was specified by its "day of year" value.

2.2. Statistical models

Statistical models of three levels of complexity were defined. The model with the lowest level of complexity was defined by:

$$\log\left(Yield_j\right) = b_0 + b_1 \times YEAR_j + \sum_{i=1}^{n} b_{2i} \times TEMP_{ij} + \sum_{i=1}^{n} b_{3i} \times RAD_{ij} + \sum_{i=1}^{n} b_{4i} \times PREC_{ij} + b_5 \times SOIL_j + b_6 \times Pre_{CROP\,j} + b_7 \times Sowing_j + \varepsilon_j \qquad (1)$$

where *TEMP*, *PREC* and *RAD* are average temperature, precipitation and global radiation, respectively, over the i-th time period of the agricultural year for the j-th observation, *YEAR, SOIL*, *Sowing* and Pre_{CROP} are year number, soil type, sowing date and previous crop, respectively, ε_j is a residual random error term, and b_x, x ϵ [0,7] are unknown parameters. Logarithm of yield was used as the response variable, which enabled the estimation of the relative yield change corresponding to changes in input parameters.

The model with intermediate level of complexity included quadratic terms for precipitation and sowing dates, but not for the other inputs, and was defined as:

$$\log\left(Yield_j\right) = b_0 + b_1 \times YEAR_j + \sum_{i=1}^{n} b_{2i} \times TEMP_{ij} + \sum_{i=1}^{n} b_{3i} \times PREC_{ij} + \sum_{i=1}^{n} b_{4i} \times PREC_{ij}^{\;2} + b_5 \times SOIL_j + b_6 \times Pre_{CROP\,j} + b_7 \times Sowing_j + b_8 \times Sowing_j^{\;2} + \varepsilon_j \qquad (2)$$

Note that the radiation was not included in Eq. (2), to test if reducing possible collinearities between radiation and temperature could increase the performance of the model.

The model with the high level of complexity included quadratic terms for all continuous input variables, except year: