Agricultural and Forest Meteorology 260–261 (2018) 300–320

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

# Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices

Anikó Kern[a], Zoltán Barcza[b,c,d,*], Hrvoje Marjanović[e], Tamás Árendás[f], Nándor Fodor[f], Péter Bónis[f], Péter Bognár[a], János Lichtenberger[a,g]

[a] Department of Geophysics and Space Science, Eötvös Loránd University, H-1117 Budapest, Pázmány P. st. 1/A, Hungary
[b] Department of Meteorology, Eötvös Loránd University, H-1117 Budapest, Pázmány P. sétány 1/A, Hungary
[c] Excellence Center, Faculty of Science, Eötvös Loránd University, H-2462 Martonvásár, Brunszvik u. 2., Hungary
[d] Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, 165 21 Prague 6, Kamýcká 129, Czech Republic
[e] Croatian Forest Research Institute, HR-10450 Jastrebarsko, Cvjetno naselje 41, Croatia
[f] Centre for Agricultural Research, Hungarian Academy of Sciences, H-2462 Martonvásár, Brunszvik u. 2., Hungary
[g] Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, H-9400 Sopron, Csatkai u. 68., Hungary

ARTICLE INFO

*Keywords:*
Crop yield
Statistical modelling
Yield forecast
Climate variability
Remote sensing
MODIS NDVI

ABSTRACT

In the present study, multiple linear regression models were constructed to simulate the yield of winter wheat, rapeseed, maize and sunflower in Hungary for the 2000–2016 time period. We used meteorological data and soil water content from meteorological reanalysis as predictors of the models in monthly resolution. We included annual fertilizer amount in the analysis to remove trend from the census data. We also used remote sensing based vegetation index to extend the approach for early crop yield forecast purposes and to study the added value of proxy data on the predictive power of the statistical models. Using a stepwise linear regression-like method the most appropriate models were selected based on the statistical evaluation of the model fitting. We provided simple equations with well interpretable coefficients that can estimate crop yield with high accuracy. Cross-validated explained variance were 67% for winter wheat, 76% for rapeseed, 81% for maize and 68.5% for sunflower. The modelling exercise showed that positive anomaly of minimum temperature in May has a substantial negative effect on the final crop yield for all four crops. For winter wheat increasing maximum temperature in May has a beneficial effect, while higher-than-usual vapour pressure deficit in May decreases yield. For maize soil water content in July and August is crucial in terms of the final yield. Incorporation of the vegetation index improved the predictive power of the models at country scale, with 10%, 2% and 4% for winter wheat, rapeseed and maize, respectively. At the county level, remote sensing data improved the overall predictive power of the models only for winter wheat. The results provide simple yet robust models for spatially explicit yield forecast as well as yield projection for the near future.

## 1. Introduction

Field crops are in the focus of plant research worldwide given their importance in human food production and animal husbandry (Goudriaan et al., 2001; Godfray et al., 2010; Asseng et al., 2013; Bassu et al., 2014; Ewert et al., 2015; see also the AgMIP[1] and MACSUR[2] initiatives). Agronomists follow and study the life cycle of cereal crops, including the timing of phenological events like emergence, tillering, leaf elongation, anthesis, grain filling, physiological maturity as well as final crop yield. Weather, soil type, cultivar selection, agrotechnology, air pollution, the presence of pathogens and weeds jointly affect plant production and final yield in a complex way (Berzsenyi and Győrffy, 1995; Kaufmann and Snell, 1997; Dobor et al., 2016; Parkes et al., 2017).

The effect of weather on crop growth is widely documented (Farooq et al., 2011; Hatfield et al., 2011; Eyshi Rezaei et al., 2015; Jin et al., 2016; Siebert et al., 2017). Long-term field experiments provide invaluable data on climate-crop interactions (Berzsenyi and Győrffy, 1995; Berzsenyi et al., 2000; Blair et al., 2006; Johnston et al., 2016; Wen et al., 2016). In most of the cases, lessons learned from these small-scale field experiments are used to spatially and temporally extrapolate crop response to weather variability and agromanagement, which may

* Corresponding author at: Department of Meteorology, Eötvös Loránd University, H-1117 Budapest, Pázmány P. st. 1/A, Hungary.
*E-mail address:* zoltan.barcza@ttk.elte.hu (Z. Barcza).

[1] http://www.agmip.org/
[2] https://macsur.eu/

https://doi.org/10.1016/j.agrformet.2018.06.009
Received 22 February 2018; Received in revised form 19 May 2018; Accepted 10 June 2018

neglect key elements like spatial variability of soil, cultivar selection, pest damage, heterogeneity of management, irrigation patterns and others (Nendel et al., 2013; Gornott and Wechsung, 2016). As a consequence, our knowledge is still incomplete regarding the effect of the key meteorological parameters on crop growth and yield in large spatial scales (e.g. Luo, 2011; Frieler et al., 2017). This is a critical issue which might hamper the quantification of future crop production under changing climate and cause substantial uncertainty in the projections (Asseng et al., 2013; Bassu et al., 2014; Hatfield et al., 2011; Wang et al., 2017).

Mathematical models of different complexity are widely used to study the effect of environmental conditions, management and cultivar selection on crop growth on the local, regional and global scale. These models can be used to predict the effect of climate change on crop growth. Process-based crop models like APSIM, DSSAT, CropSyst, STICS, WOFOST, 4 M (Asseng et al., 2013; Fodor et al., 2014), and also combined biogeochemical-crop models like Biome-BGCMuSo, ORCHIDEE-STICS, LPJmL, CLM-crop, JULES-SUCROS, SiBcrop (see Hidy et al., 2016 and references therein) have large input requirement which could prevent their spatial generalization and application. In contrast, statistical models of crop growth using regression models provide a simpler alternative for spatially explicit studies (Lobell and Field, 2007; Lobell and Burke, 2010; Shi et al., 2013; Gaál et al., 2013; Mirschel et al., 2014; Pinke and Lövei, 2017). These models have the advantage that they include the effect of weather-induced pathogens, air pollution and other effects that are typically neglected by process-based models (Lobell et al., 2006; Shi et al., 2013; Gornott and Wechsung, 2016). Statistical models are not free from problems. Co-linearity of the driving meteorological parameters might complicate the interpretation of model results (Nicholls, 1997; Lobell and Burke, 2010; Quinn and Keough, 2002; Shi et al., 2013). Nevertheless, improvement and application of statistical models might be essential for near-future (e.g. up to 2050) prediction of effects of climate change on crop yield.

From the beginning of the 1980 s, remote sensing acts as another major data source for monitoring crop development and forecast the final yields at large spatial context. The derived spectral vegetation indices, like the Normalized Difference Vegetation Index (NDVI), and biophysical parameters such as the Leaf Area Index (LAI) or the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) are all in relation with plant status and productivity, therefore they are widely used to estimate crop yield (Maselli et al., 1992; Rasmussen, 1992; Benedetti and Rossini, 1993; Doraiswamy et al., 2003; Ferencz et al., 2004; Balaghi et al., 2008; Mkhabela et al., 2005, 2011; Bolton and Friedl, 2013; Sakamoto et al., 2013; Kowalik et al., 2014; Huang et al., 2015; López-Lozano et al., 2015). NDVI is one of the most frequently used vegetation indices for this purpose. Since 2000 the measurements of the Moderate Resolution Imaging Spectroradiometer (MODIS) became one of the most popular data sources given its sophisticated on-board calibration and high radiometric and spatial accuracy (Justice et al., 1998) which makes it suitable for improved yield estimation. Other remote sensing based indices were also proposed to support crop yield estimations, like the Normalized Difference Water Index (NDWI) (Gu et al., 2007; Bolton and Friedl, 2013). NDWI is sensitive to soil water content which is directly related to plant status and production.

Joint application of meteorological data and remote sensing based vegetation indices is a relatively new approach, and seems to be promising based on the reported improvements in crop yield prediction (Prasad et al., 2006; Balaghi et al., 2008; Kogan et al., 2013; Huang et al., 2014b; Franch et al., 2015). Several other authors focusing on remote sensing based crop yield forecast concluded that additional use of meteorological predictors is a potential step towards improved methods. In this study, we explore this option with the combined use of MODIS NDVI and meteorological data.

Hungary is located in the Carpathian Basin in Central Europe and is known to have favourable conditions for crop growth (Trnka et al., 2016). In Hungary, the largest share of land use is agriculture (85%), where arable land occupies ~48% of the total land area (Barcza et al., 2010). Though the share of agriculture in the total GDP is relatively small, the sector provides income to several hundreds of thousand families (HCSO, 2017a), which means strong economic impact for the country. Climate anomalies in Hungary are characteristic for a wider Central European domain (Pinke and Lövei, 2017), which indicates that Hungary is a good candidate for studying the effects of the meteorological elements on crop production in Central Europe. Hungary is a drought-prone country, where the larger crop yield losses are typically attributed to dry spells (Szinell et al., 1998; Fiala et al., 2014; Gulácsi and Kovács, 2015; Pinke and Lövei, 2017). As climate change is expected to increase the number of extreme events worldwide including dry spells and heatwaves (Ewert et al., 2015; Troy et al., 2015; Mäkinen et al., 2017), the effect of climate on crop production, that is representative to present-day conditions in Hungary, might become typical in other geographical locations in the mid-latitudes (e.g. in Western Europe with more humid climate today). Therefore, understanding weather effect on crop yield in Hungary might support the understanding of the relationship between climate anomalies (represented by weather elements) and crop production in a wider spatial context.

The four most important field crops in Hungary (maize, winter wheat, rapeseed and sunflower) are the subject of the present study. Our hypothesis is that temporally and spatially aggregated meteorological variables have a distinct and detectable effect on the yields of the selected crop types both at country and county scale. We also hypothesize that using multiple linear regression modelling it is possible to predict crop yield with reasonable accuracy and predictive power. Remote sensing is expected to improve the predictive skill of the yield models. The aims of the present study are (i) to quantify the effect of weather (using monthly anomalies) and other environmental elements (most of all soil water content anomaly and fertilizer amount) on crop yield at county and country scale; (ii) to explore which climate elements are more important than the others in terms of relationship with crop yield; (iii) to construct a "best" model for all four crop types to simulate crop yield with multiple linear regression using minimal number of predictors; (iv) to extend the statistical modelling approach with remote sensing information via using vegetation indices. The inclusion of remote sensing data is performed to evaluate whether this additional proxy variable improves model performance both on country and county scale. The results can be used for early crop yield forecast (all models) for the ongoing year to support economic decisions, and for the projection of crop yields using climate model output (models based on environmental variables only) for the forthcoming few decades (e.g. up to 2050) assuming no major change in the driving variables of the yield variability.

The presented method can be adopted for other countries and can support the improved understanding of weather effect on crop production. A novel feature is the application of soil water content as a predictor in the statistical modelling.

## 2. Materials and methods

### 2.1. Target area

The target area of this study is Hungary (with an area of 93,030 $km^2$), located in Central Europe (Fig. 1). Based on data of the applied meteorological dataset (see Section 2.3) during 1981–2010 the mean annual temperature in Hungary was ~11.2 °C (with a ~8.6 °C minimum and 12.1 °C maximum values in Northern and in Southern Hungary, respectively), while mean annual precipitation was 600 mm (with a ~530 mm minimum and ~800 mm maximum value in the Central and Western part of the country, respectively).