Agricultural Systems 141 (2015) 94–106

Contents lists available at ScienceDirect

Agricultural Systems

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

Improving operational maize yield forecasting in Hungary

Attila Bussay [*], Marijn van der Velde, Davide Fumagalli, Lorenzo Seguini

European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via E. Fermi 2749, I-21027 Ispra, VA, Italy

ARTICLE INFO

Article history:
Received 18 September 2014
Received in revised form 17 September 2015
Accepted 5 October 2015
Available online xxxx

Keywords:
Maize
Yield forecast
Hungary
Spatial aggregation
Trends
LOESS
MARS crop yield forecasting system
WOFOST

ABSTRACT

In most landscapes, accurate crop yield forecasting depends on a quantitative understanding of the relation between past weather, management and crop yield variability. We evaluated and improved the regression-based crop yield forecasting methodology currently employed in the MARS-Crop Yield Forecasting System (M-CYFS) for maize in Hungary. We quantified the effect of: 1) different statistical trends; 2) different crop growth simulation model outputs providing weekly predictors; 3) yield prediction lead times; and 4) spatial aggregation on the forecast accuracy as evaluated against statistical yield from 1993 to 2012. The LOESS (locally weighted scatterplot smoothing) trend provided the lowest root mean square error (RMSE) in describing the yield time-series compared to the quadratic and linear trend. Using the WOFOST crop model-based predictors to explain the yield residuals derived with each of the three trends, the lowest RMSEs were obtained with the Water Limited Leaf Area Index (WLLAI) and Water Limited Above Ground Biomass (WLB) predictors in combination with the LOESS trend. The LOESS trend was used to evaluate the effect of spatially aggregating subnational yield forecasts. During the first half of the crop cycle there are only marginal differences between the NUTS0 (national), NUTS1 (supra-regional), NUTS2 (regional), and NUTS3 (sub-regional) level. However, the NUTS0 forecast had a slightly lower accuracy from the start of flowering and onwards, indicating the possible benefit of maintaining spatial detail when aggregating data. The RMSE of the forecasts started to decrease in weeks 24 and 25. Even though the relative soil moisture decreased earliest, the best performing yield forecasts were associated with lead times of about 5–8 weeks before harvest and were obtained with the WLLAI and WLB as predictors. The best forecasts were associated with the critical phenological phases of flowering and grain-filling respectively occurring between weeks 27 to 30 and weeks 31 to 35. The best performing national forecast was based on NUTS1 level forecasts with an r^2 and a RMSE of respectively 0.8565 and 425.9 kg ha^{-1} using WLLAI as predictor. Finally, we compared the regression-based forecasts with operational forecasts performed by the Ministry of Agriculture of Hungary and the JRC-MARS forecasts from 2007 to 2012.

1. Introduction

Forecasting of crop yield throughout the growing season underpins estimates of agricultural production expected at the end of the growing season. As such, it can inform decisions of farmers, market analysts and policymakers alike. Providing benchmarks of the performance of crop yield forecasting is essential to ensure quality, but also to increase trust, uptake, and use of the forecasts. Accurate crop yield forecasting can also become of increasing importance in a changing climate characterized by increasingly variable weather (Salinger, 2005).

Globally, annual maize production averages about 861 million tons, contributing the largest share to total cereal production (FAO, 2012). Monitoring of growing conditions and yield forecasts are thus important for countries cultivating, exporting and importing maize. In Hungary, maize is the most commonly grown crop accounting for about 1.2 million hectares during the last years (KSH, 2013). The inter-annual variability of maize yields in Hungary is considerable, and therefore benefits from maize yield forecasting may be substantial.

In Hungary, the Ministry of Agriculture produces national yield estimates based on a countrywide collection of field observations including measurements to determine yield (e.g. length of maize cob, number of cobs, plant density). In the past (1997–2004), the Institute of Geodesy, Cartography and Remote Sensing used remote sensing for yield forecasting (Csornai et al., 2006). Since 2004, when Hungary joined the European Union, the Joint Research Centre of the European Commission has been operationally forecasting and publishing forecasts of end-of-season Hungarian crop yield using the MARS-Crop Yield Forecasting System (M-CYFS; see MARSWiki, 2015).

The M-CYFS is a crop yield forecasting system operational over Europe. Besides remotely sensed information, and statistical analysis, it benefits from crop growth modelling with the Crop Growth Monitoring System (CGMS; Supit and Van der Goot, 2003; for more details see below). Indeed, crop models have been evaluated for their use in yield forecasts, including CERES-Maize (Soler et al., 2007), WOFOST (Supit, 1997) and AquaCrop (Abedinpour et al., 2012). Generally, results

---

[*] Corresponding author.
E-mail address: attila.bussay@jrc.ec.europa.eu (A. Bussay).

http://dx.doi.org/10.1016/j.agsy.2015.10.001
0308-521X/© 2015 Published by Elsevier B.V.

indicated a good agreement between simulated and observed yields with coefficients of determination generally exceeding 0.80.

The general objective of this manuscript is to enhance the performance of the M-CYFS for operational maize yield forecasting in Hungary. We specifically aim to determine the importance of several factors affecting the operational M-CYFS yield forecasts including 1) the trend chosen to de-trend the maize yield time-series and determine the yield residuals, 2) the crop model output variable used as a predictor to explain the variability in yield residuals, 3) crop phenology in terms of the timing when the correlation between the crop model predictor and yield residuals is strongest, and 4) the effect of spatial aggregation from NUTS3/2/1 to NUTS0 level on national yield forecasts. From this analysis, we aim to determine the approach resulting in the highest possible accuracy to aid operational maize yield forecasts for Hungary with the M-CYFS.

2. Materials and methods

In the following sections we will describe the Hungarian national yield statistics collected for this study, introduce the M-CYFS, including the CGMS, and its data needs, then we will describe how the crop model simulations provide predictors used for the crop yield forecasting. We continue with presenting the regression analysis used to forecast the crop yields, and finally we will analyse the performance of the crop yield forecasts, with specific attention for the effects of spatial data aggregation, and the timing of the best forecasts with respect to crop growth stages.

2.1. Hungarian crop statistics

The political turnover in Hungary in 1990 caused significant changes in Hungarian agriculture, including ownership structure and profitability, as well as in the intensification levels of agro-techniques (e.g. fertilization, mechanization etc.). Therefore, the yields dropped dramatically during the first years of the 1990s (Vizvári and Bacsi, 2003). The yields started to recover in 1993, therefore this year was selected as the starting year of the analysis. The reported Hungarian maize yield, acreage and production time-series were obtained from the webpage of Hungarian Central Statistical Office (HCSO; KSH, 2014) and statistical yearbooks (KSH, 1994, 1995, 1996, 1997, 1998) over the period 1993–2012 at national, supra-regional, regional and sub-regional (respectively NUTS0 to NUTS3) levels (see Fig. 1a). The crop yield as reported by HCSO represents the best available and unbiased information on Hungarian crop yields. Maize is commonly cultivated in Hungary with 8.9 to 51.1% of the arable land area at sub-regional level grown with maize (Fig. 1b). The statistical data were plotted and visually inspected to detect outliers or obvious errors. A cross-check was performed to ensure that the NUTS0–NUTS3 level data were in correspondence with each other regarding the sums, totals and averages of area, yield and production values.

2.2. The MARS-Crop Yield Forecasting System

The M-CYFS is used to monitor weather conditions, crop growth and development, determine the impact of extreme meteorological events, and provide monthly forecasts of crop yield at national and European Union level. The crop yield forecasts are published in the MARS bulletins (http://mars.jrc.ec.europa.eu/mars/Bulletins-Publications/). The M-CYFS contains past and real-time meteorological observations, agro-meteorological and biophysical modelling with the Crop Growth Monitoring System (CGMS; Supit and Van der Goot, 2003), data derived from remote sensing, and statistical analyses performed within a dedicated software tool, the Control Board (CoBo; Genovese and Bettio, 2004). Analysts use CoBo to statistically link selected outputs from the archive of CGMS simulations to time-series of reported crop yield data using a variety of statistical methods. Subsequently, this analysis is used to perform the operational crop yield forecasts, first by feeding the CGMS with real-time and 10-day short term ECWMF weather forecasts, and secondly by using these crop model outputs with the previously established statistical relationships, in addition to the extrapolated trend. The regression-based crop yield forecasting methodology of the M-CYFS thus consists of the following components; 1) importing of maize yields statistics, 2) determining the trend in the statistics to calculate the yield residuals and 3) a statistical regression analysis is performed on the yield residuals and crop model-based predictors fed with past weather, and finally 4) forecasting an end-of-season crop yield using the current season's model outcome.

2.2.1. CGSM (WOFOST) simulations

The CGMS is a platform that includes several crop models that are spatially implemented over the regions for which regular crop yield forecasts are made. The core crop model of the CGMS is the World Food Studies (WOFOST) crop growth simulation model (Boogaard et al., 2014; Supit et al., 1994; Van Diepen et al., 1989). WOFOST is used here for the simulation of maize in Hungary. WOFOST is a biophysically-based, dynamic and explanatory point model that can be applied across a range of meteorological, soil and agro-management conditions (De Wit et al., 2010). WOFOST simulates crop growth as the difference between assimilates produced by photosynthesis and consumed by respiration. The main process controlling growth and partitioning of assimilates is crop development stage (DVS) which describes crop phenology (Boogaard et al., 2014). DVS is a dimensionless state variable being primarily a function of temperature and day length which varies between 0 (sowing) through 100 (anthesis) to 200 (maturity). Potential yield as well as water limited yield can be simulated by WOFOST. Potential yield is determined by the defining factors CO_2, temperature, solar radiation and crop characteristics. In addition to these factors water limited yield is constrained by water availability.

2.2.1.1. Crop model parameters and management. The crop model parameter settings were identical with the operational settings of the WOFOST model for Hungary, which came from the results of previous studies (Van Heemst, 1988; Van Diepen and De Koning, 1990; Boons-Prins et al., 1993). The calibration of crop phenology was based on crop monographs on Central European Countries (Kucera and Genovese, 2004). The sowing date was determined at regional level, and considered as a fixed Julian day during the period of 1993–2012, notwithstanding possible changes in maize varieties and their spatial distribution during the last 20 years, as well as the occurrences of early start of the growing season (favourable weather conditions) or delayed start (unfavourable weather conditions). Irrigation of maize was of limited importance in Hungary during the last 20 years. In 2012 for example, the irrigated land surface was only 112,669 ha (or 2.6%) from a total of 4,323,638 ha of Hungarian arable land (KSH, 2013). Therefore, the model simulations were performed for rain-fed (water limited) conditions. Bare soil conditions were simulated during winter and maize was cultivated during the summer cropping season.

2.2.1.2. Meteorological data. The meteorological data-base of the M-CYFS (Baruth et al., 2007) was used as input for the crop model simulations of this study. The database contains interpolated meteorological data on a regular grid with a mesh of 25 km using an interpolation method based on the distance, altitude and climatic region similarity between the centre of grid cells and weather stations as described by Van der Goot (1998). The interpolation was based on 19 weather stations which supplied continuous observations over Hungary as well as an additional 20 Hungarian weather stations with intermittent observations. The meteorological observations of surface weather stations were automatically quality checked using a routine data procedure (Micale and Genovese, 2004).

2.2.1.3. Soil. Data from the Soil Geographical Data Base of Europe (SGBDE) version 4.0 with a resolution of 1:1,000,000 were collected. Soil mapping units (SMU) were defined following Lazar and Genovese