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

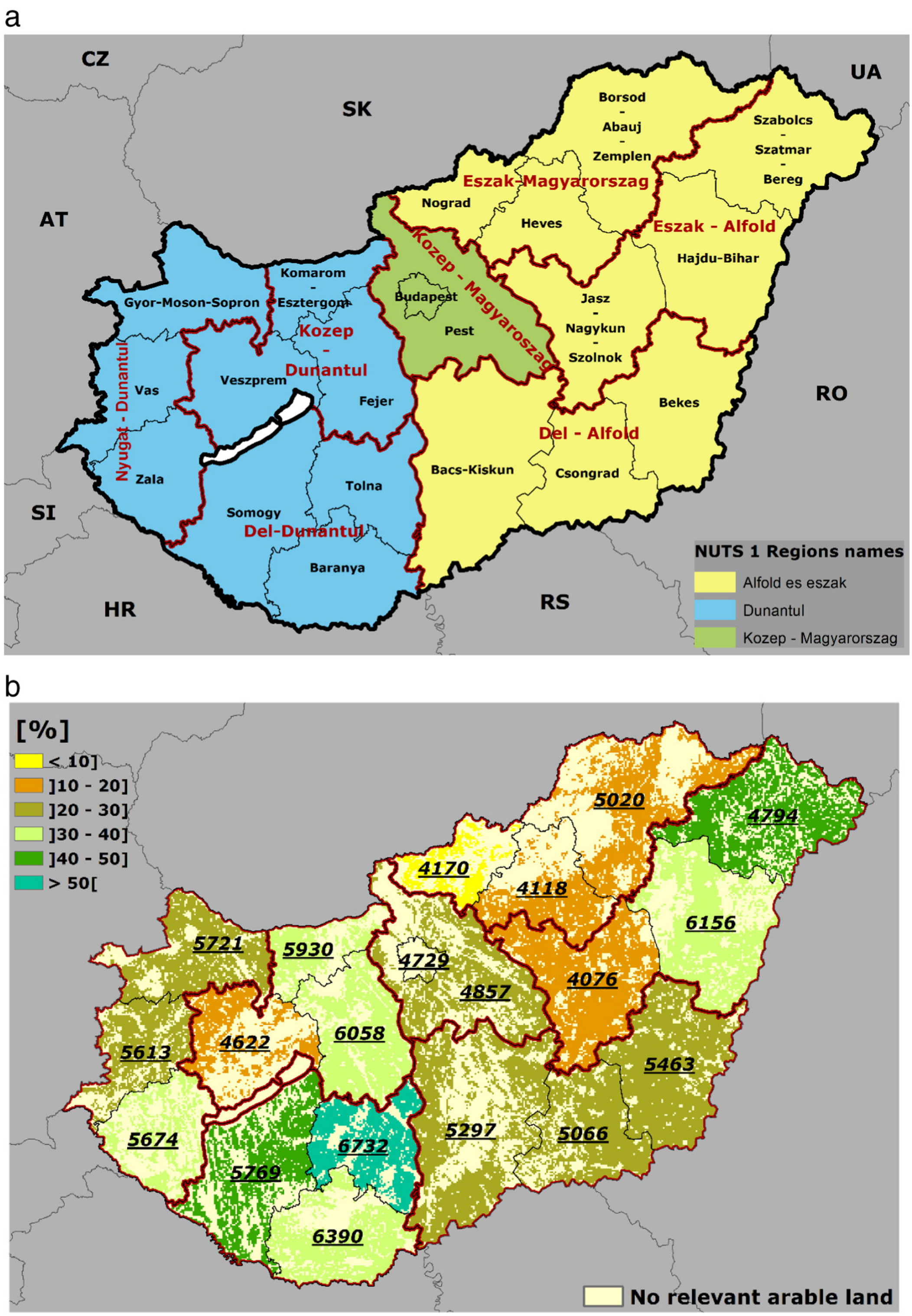

Fig. 1. a. Hungarian administrative divisions. The administrative units indicated are from the lowest spatial resolution to the highest: NUTS0 (national), NUTS1 (supra-regional; see legend), NUTS2 (regional; thick red lines) and NUTS3 (sub-regional; thin lines) levels. b. The percentage of the harvested maize area in arable land at sub-regional (NUTS3 level) based on the 2012 statistics. The numerical values depict 20 year average maize yields [kg ha^{-1}]. The "No relevant arable land" class defines pixels that have less than 40% of rainfed arable land, as defined in CORINE 2000 by class 2.1.1. The spatial resolution of the mask is 1 km × 1 km. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2004) as the smallest cartographic unit of the soil map. SMUs consist of homogeneous Soil Typological Units (STU) that have a set of different chemo-physical characteristics. Field capacity, permanent wilting point, available water capacity, rooting depth, drainage conditions, salinity and alkalinity were derived from basic soil properties using the pedotransfer rules from SINFO project (Baruth et al., 2006).

2.3. Aggregation of crop model output variables (predictors)

Crop model output variables were aggregated in space and time and used as predictors in the statistical analysis to establish relationships with yield statistics. The WOFOST model simulation was run on each STU with the weather conditions of the related grid cell. The aggregation of the model output variables followed three steps, a first one from STU to SMU, then from SMU to grid cell, and finally from grid to administrative level. The outputs simulated on individual suitable STU were aggregated to grid level as weighed by the share of the SMU in the grid cell considered. These were subsequently aggregated to the administrative units from NUTS3 to NUTS0 levels. Since the cultivated area of maize was not known for every single grid cell, simulated crop model output variables were aggregated to the different NUTS regions using the area of arable land derived from the CORINE Land Cover 2000 land cover database (Nunes de Lima, 2005) as a weighting factor. The WOFOST model output variables considered to be most relevant and important for crop yield prediction are listed in Table 1. The temporal sampling of these variables was performed on a weekly basis. The state of the cumulated variables DVS (−), WLB (kg ha^{-1}), WLSO (kg ha^{-1}), WLLAI (−) and TWC (mm), is thus reported on a weekly basis (thus accumulating impacts since emergence), while WLTR (mm $week^{-1}$) and RSM (−) are state variables produced on a weekly time step.

2.4. Regression-based crop yield forecasting

We wished to explain annual variability in yield due to climate variability as a function of a crop model-based predictor. Therefore, in a first step we have to estimate the trend in the reported crop yields which is usually associated with changes in agro-technology (Hafner, 2003)as these changes are not considered in the crop model simulation. This technological trend is not necessarily stable and smooth over a long period (De Wit et al., 2010).In our analysis we evaluated linear, quadratic and data-driven LOESS (locally weighted scatterplot smoothing; Cleveland and Devlin, 1988) trends to describe the multi-annual yield development. Once the trend was determined, simple linear regression was used to create relationships on a weekly basis between the yield residuals and crop model based predictors with output available at a weekly resolution, which can be accumulated over a period of several weeks, so that:

$$\hat{Y}^i_{res}|_k = f\left(P^i\right)|^j_k \tag{1}$$

where $\hat{Y}^i_{res}|_k$is the estimated yield residual for year i and week k and $f(P^i)|^j_k$ is the regression-based estimation using the crop model predictor P from weeks k to j. Finally, the yield was estimated in the following way:

$$\hat{Y}^i|_k = \hat{Y}^i_{trend} + \hat{Y}^i_{res}|_k \tag{2}$$

where $\hat{Y}^i|_k$ is the estimated yield, $\hat{Y}^i_{trend}$ is the trend and $\hat{Y}^i_{res}|_k$ is the estimated yield residual for year i in week k. The forecasts were thus calculated as the sum of the extrapolating yield value derived from the trend plus the estimated yield residual.

Table 1
Overview of the WOFOST crop model output variables, simulation periods, and the resulting total number of calculated regression equations during the yield forecast experiment. For each variable linear, quadratic and LOESS trends were determined for 1, 3, 7, and 20 administrative units corresponding to respectively NUTS0, 1, 2 and 3 regions.

WOFOST variables	Abbrev.	Period [week]	Number of calculated regression equations
Crop development stage	DVS	26 [18–43]	2418
Water limited biomass	WLB	26 [18–43]	2418
Water limited storage organ	WLSO	16 [28–43]	1488
Water limited lead area index	WLLAI	23 [18–40]	2139
Relative soil moisture	RSM	26 [18–43]	2418
Total water consumption	TWC	26 [18–43]	2418
Water limited transpiration	WLTR	26 [18–43]	2418

2.5. Crop yield forecast performance

Hungarian maize yield statistics were used to analyse the forecast error, the spatial and temporal behaviour of the forecasts, and to evaluate the different forecast configurations (see Table 1). The general aim of yield forecast was to maximize the coefficient of determination and minimize the root mean square error (RMSE). The RMSE of this national yield forecast was calculated to evaluate the accuracy for each weekly forecast against the yield statistics; see Eq. (3):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{n}} \tag{3}$$

where $\hat{Y}$ is the forecast and Y the statistical yield and n the number of the years.

To reach this goal we iteratively identified the optimal trend, the most reliable spatial aggregation level, and the best model-based predictor. To assess the performance of the crop yield forecasting methodology we hind-casted crop yield forecasts from 1993 to 2012 using the jack-knifing method (Abdi and Williams, 2010). For a selected year, the yield and model based predictor were left out one by one and the yield forecasts were performed for this year using the remaining data of 19 years.

2.5.1. Effect of trend

To determine whether the linear, quadratic or LOESS function led to the lowest RMSE we also used the jack-knifing methodology (Abdi and Williams, 2010). In the case of the LOESS method the smoothing factor was fixed at 1.000 thus considering all 19 annual yield values to provide the smoothest trend with minimal fluctuations. The trend calculation was performed for all administrative units with all three methods. In addition, to evaluate if the lowest RMSE would also results in the best temporal description of the yield residuals, we performed a forecasting experiment at NUTS0 level evaluating the different trend types in combination with the model predictors. The quality of NUTS0 yield forecast was evaluated as a function of the different trends used. Once the trend, with the lowest RMSE, was established, annual yield residuals were calculated as the difference of the reported statistical yield and the trend.

2.5.2. Regression

Regressions between the yield residuals and the predictors were evaluated from the first to the last week of the growing season, and changes in the regressions were evaluated for each week incrementally. Independent regression was performed for each sub-period (a continuous set of weeks at least containing the most recent week) and the sub-period with the highest coefficient of determination (r^2) was selected. This regression equation was used to determine the predicted deviation from the trend each week at each administrative level. The best forecast identified for each administrative unit at the different NUTS levels (NUTS3, NUTS2, NUTS1) were scaled up to make a national (NUTS0) level yield forecast. Scaling up was done through aggregation of the yield estimation taking into account the corresponding reported maize sowing areas at each administrative level. The summary of calculations is presented in Table 1.

2.5.3. Effect of spatial aggregation

The hypothesis was that using crop model outputs aggregated at a finer scale thus preserving more detail (i.e. NUTS3 vs. NUTS0), in combination with regional crop statistics to perform regional forecasts, would lead to higher forecasting accuracies at national level with longer lead times. To assess this the RMSE of the NUTS0 level forecasts based on respectively aggregated NUTS3, 2, and 1 level results was calculated. In addition, we counted the number of times the national yield forecast thus obtained outperformed a forecast solely based on the LOESS trend.

3. Results

3.1. Hungarian maize yield

The most important Hungarian maize producing areas have above average yield and lower yield variability compared to country level statistics. In fact, at sub-regional level, a strong relationship ($r^2 = 0.6345$; significant at $p < 0.001$ level) exists between the ratio of maize cultivated area (as percentage of the total arable land) and average regional yield. This suggests that, generally speaking, in regions with agro-ecological conditions favourable for maize, it is a widely grown crop (Fig. 1b). The maize yield series is highly variable but an increase in yields can be detected across all administrative regions as indicated by the positive slopes obtained when fitting linear trends through the data (1993–2012), even though this increase is hardly significant (Table 2, also see Fig. 2). Only 6 NUTS3 and 3 NUTS2 administrative units had linear trends significant at 5%. At national level the increase was 75.9 kg ha^{-1}. Year to year maize yields are highly variable reflecting crop growth processes dependent on weather variability. In relative terms (expressed as the relative deviation of the yield), the inter-annual yield variability is smaller in sub-regions (counties) with a higher average yield ($r^2 = 0.4058$). This probably relates to better risk management due to relatively advanced agro-technological practices (Table 2). There are considerable local differences in the variability of NUTS3 level maize yield (Table 2). The standard deviation of yield at sub-regional level was between 1238.6 and 1802.6 kg ha^{-1} and the relative standard deviation reached 21.5–34.1% (Table 2).

3.2. Effect of trend

The different trend functions fitted through national yield data are shown in Fig. 2. The variability in yield is much larger than the magnitude of the yield trend over the time period considered. The national yield residuals associated with each trend and obtained with the jack-knifing methodology are shown in Fig. 3. At national level the LOESS trend resulted in the lowest overall RMSE compared to the linear and quadratic trends (respectively 1115.9, vs. respectively 1215.95 and 1261.3.6 kg ha^{-1}). We also evaluated the RMSE at the NUTS1 to NUTS3 administrative levels. The LOESS trend led to the lowest RMSE for all administrative units. At NUTS3 level, the lowest, highest and median RMSEs that were obtained with the linear trend respectively equalled 909.5, 1643.5 and 1261.4 kg ha^{-1}, for the quadratic trend this equalled 951.8, 1715.5 and 1325.7 kg ha^{-1} while for the LOESS trend this equalled 846.4, 1538.0 and 1151.4 kg ha^{-1}.

In addition, a forecast experiment was performed at NUTS0 level to evaluate whether the LOESS trend in combination with the six model variables (WLB, WLSO, TWC, RSM, WLLAI, WLTR) indeed performs the best when forecasting yields. The results at NUTS0 level are shown in Table 3 and Fig. 4. The LOESS trend provided the lowest forecast error during the whole cropping season for all crop model variables in comparison to both the linear and quadratic trend. On average, the forecasting error, compared to the forecasts using linear and quadratic trend, was lowered by more than 100 kg ha^{-1} (Table 3). The forecasts with quadratic and linear trend generally have similar forecasting accuracy. Based on these results we selected the LOESS trend to evaluate the different crop model-based predictors and the effect of calculating forecasts at different spatial scales.

3.3. Crop model-based predictors

To analyse the importance of each individual crop model-based predictor (i.e. WOFOST model output variable) in terms of its possible use for yield forecasting as well as to evaluate lead times during the cropping season, correlation coefficients were calculated between yield residuals obtained with the LOESS trend and the predictors aggregated to national level with a weekly time step (Table 4).

Table 2
Main statistical characteristics of Hungarian maize yield and acreage statistics for the period 1993–2012 at NUTS 0, 1, 2, and 3 levels (* indicates the significance level at $p < 0.05$).

Administrative unit	NUTS level	Average yield [kg ha^{-1}]	Yield increase [kg ha^{-1} year^{-1}]	Linear trend r^2	Standard deviation of yield [kg ha^{-1}]	Relative deviation of yield [%]	Maize area (2012) in [1000 ha]	Arable land (2012) in [1000 ha]	Ratio [%]
Budapest	3	4729.0	124.1	0.207*	1612.3	34.1	10.1	40.5	24.9
Bacs-Kiskun	3	5297.3	49.3	0.036	1526.6	28.8	76.4	366.1	20.9
Baranya	3	6390.2	44.7	0.036	1390.6	21.8	87.2	225.6	38.7
Bekes	3	5463.3	74.8	0.084	1532.1	28.0	92.6	374.7	24.7
Borsod-Abauj-Zemplen	3	5020.2	159.1	0.447*	1407.7	28.0	46.2	250.4	18.4
Csongrad	3	5065.5	19.3	0.007	1331.5	26.3	53.6	255.2	21.0
Fejer	3	6058.1	106.6	0.122	1802.6	29.8	87.7	250.4	35.0
Gyor-Moson-Sopron	3	5721.2	127.0	0.327*	1313.7	23.0	50.8	225.9	22.5
Hajdu-Bihar	3	6155.6	93.2	0.164	1360.0	22.1	116.7	317.2	36.8
Heves	3	4118.4	80.4	0.142	1261.0	30.6	16.6	147.6	11.3
Jasz-Nagykun-Szolnok	3	4076.3	43.9	0.038	1340.3	32.9	36.0	329.4	10.9
Komarom-Esztergom	3	5929.9	96.9	0.157	1446.8	24.4	33.9	102.9	32.9
Nograd	3	4170.2	96.0	0.307*	1026.1	24.6	5.5	62.1	8.9
Pest	3	4857.0	108.4	0.150	1657.6	34.1	50.9	250.7	20.3
Somogy	3	5768.6	38.4	0.026	1415.0	24.5	109.1	249.8	43.7
Szabolcs-Szatmar-Bereg	3	4793.8	124.9	0.284*	1387.3	28.9	108.9	263.2	41.4
Tolna	3	6731.9	1.6	0.000	1668.9	24.8	109.5	214.2	51.1
Vas	3	5613.5	116.6	0.283*	1297.2	23.1	30.4	146.4	20.8
Veszprem	3	4622.3	80.6	0.148	1238.6	26.8	24.1	138.2	17.4
Zala	3	5673.7	72.2	0.123	1219.1	21.5	45.0	113.4	39.7
Kozep-Dunantul	2	5766.0	100.0	0.142	1568.9	27.2	145.7	491.4	29.6
Kozep-Magyarorszag	2	4842.5	110.8	0.164	1621.2	33.5	61.0	291.1	21.0
Nyugat-Dunantul	2	5691.9	101.6	0.245*	1215.2	21.4	126.3	485.6	26.0
Del-Dunantul	2	6303.5	22.8	0.009	1463.5	23.2	305.8	689.6	44.3
Eszak-Magyarorszag	2	4725.9	130.1	0.361*	1280.8	27.1	68.3	460.0	14.8
Eszak-Alfold	2	5258.9	102.4	0.208*	1327.5	25.2	261.6	909.8	28.8
Del-Alfold	2	5323.2	52.8	0.046	1457.0	27.4	222.6	996.0	22.4
Kozep-Magyarorszag	1	4842.5	110.8	0.164	1621.2	33.5	61.0	291.1	21.0
Dunantul	1	6035.1	60.6	0.069	1366.8	22.6	577.8	1666.7	34.7
Alfold es Eszak	1	5252.9	83.1	0.137	1328.8	25.3	552.5	2365.8	23.4
Magyarorszag	0	5597.8	75.9	0.113	1336.3	23.9	1191.3	4323.6	27.6

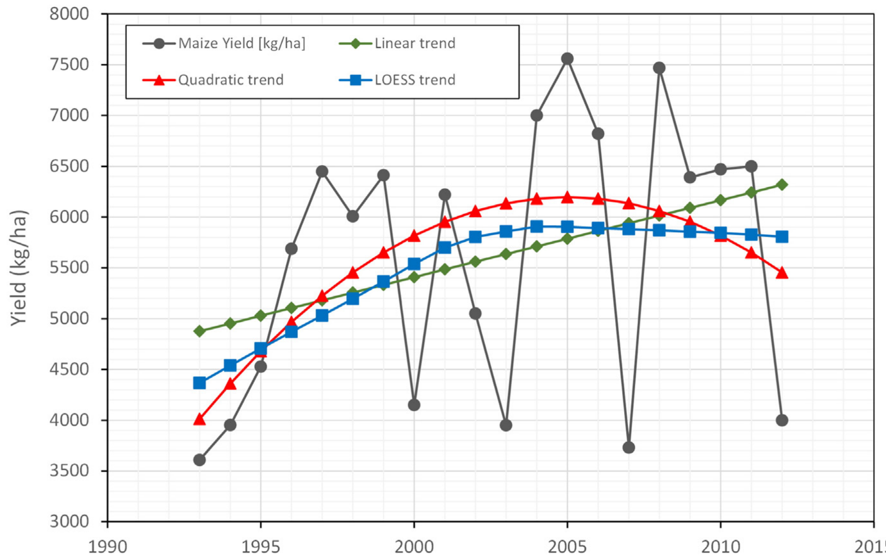

Fig. 2. Hungarian national maize yield (1993–2012) with linear, quadratic and LOESS trends.

Weeks exhibiting the highest correlations could provide information about the most significant and sensitive growth stages related to maize yield formation. The crop development stage (DVS) plays a key role in the synchronization of the simulation of different physiological processes. It was continuously negatively correlated with annual yield variability. A possible explanation is that accelerated crop development (due to high temperature) leads to a shortening of the crop cycle leaving less time for biomass and yield accumulation. More over periods with high temperatures often coincide with dry conditions also negatively affecting biomass and yield. Results obtained for DVS are included in Supplementary Fig. 1.

Simulated water limited above ground biomass (WLB) reveals a strong correlation with the yield residuals (Table 4). During the first 8 weeks of the crop cycle, the slightly negative correlation would indicate that early disadvantageous growing conditions have a positive effect on final yield. This paradox could possibly reflect drier weather that in reality leads to denser and deeper rooting systems (Comas et al., 2013) which may be advantageous during drier summer months. The sign of correlation flips in the first half of July (between week 26 and 27, when anthesis generally commences), indicating the significant change in the plant physiology and the possible problems in the usability of WLB around this period for yield prediction. The correlation coefficient became increasingly higher reaching $r > 0.90$ by mid-August (week 32). Theoretically, the water limited storage organs (WLSO) i.e. rainfed yield, should exhibit the strongest relation with yield residuals, but this was not the case [maximum $r = 0.818$]. The highest correlation coefficient of WLSO with the yield residuals occurred in the later stages of ripening and the maturity phenophase [weeks 40–43].

Relative soil moisture (RSM) and transpiration (WLTR) are reported on a weekly basis. Therefore, they do not summarize the growing conditions from the start of the season. However, for certain weeks they show good correlations with yield due to their importance for yield formation. RSM typically displays a medium to strong relationship ($r > 0.60$) shortly before flowering until the second phase of grain filling [weeks 25–33], but in week 31 the correlation reaches $r = 0.829$ (Table 4). During the second half of grain filling and the ripening stage, the correlation with

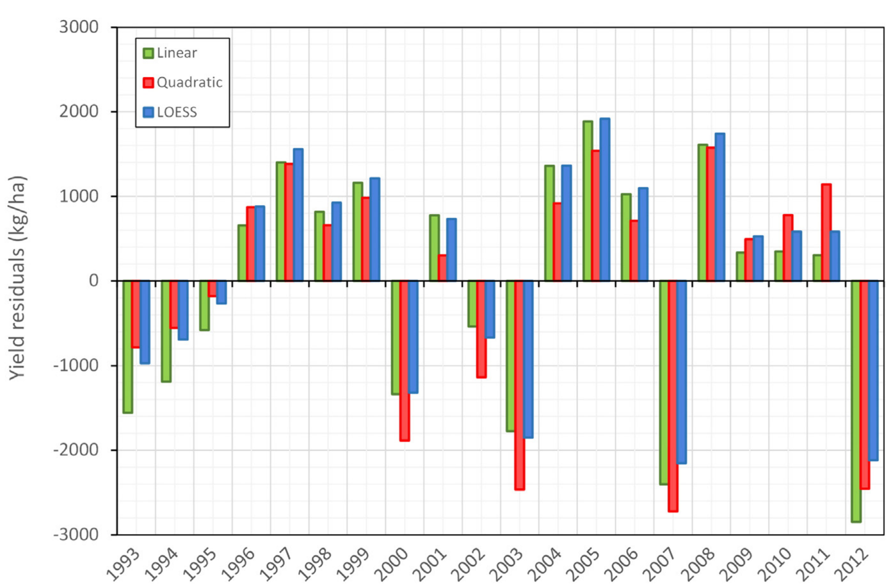

Fig. 3. National maize yield residuals obtained with the jack-knife approach for the linear, quadratic and LOESS yield trend functions (see Fig. 2).

Table 3
Average difference of RMSE of yield forecasts calculated over all weeks in the crop cycle using different trends (linear and quadratic minus LOESS trend) using seven WOFOST model output variables to explain the yield residuals.

Output variables	RMSE difference: linear trend − LOESS [kg ha^{-1}]	RMSE difference: quadratic trend − LOESS [kg ha^{-1}]
DVS	144.2	84.3
WLB	120.7	122.4
WLSO	94.8	121.3
TWC	120.5	137.5
RSM	126.5	129.1
WLLAI	151.0	113.3
WLTR	114.3	119.0

RSM is lower. This is probably because during late grain filling, senescence of the crop canopy, and the ripening stage, water supply has a limited effect on the final yield (Lobell et al., 2014). WLTR shows a negative correlation ($-0.223 < r < -0.595$) during most of the vegetative phase, because early intensive transpiration and associated soil water loss are disadvantageous for yield formation. WLTR becomes significant after week 26 until week 35, typically reaching its maximum in week 32 with $r = 0.892$. Total water consumption (TWC) is a cumulative variable depicting a similar correlation as WLB. The highest correlation ($r = 0.944$) is obtained with the WLLAI in week 34. The correlation becomes insignificant after the 35th week due to the senescence of the canopy. Regarding the early forecast skill of the predictors, the WLLAI shows higher correlation coefficients from the 26th week ($r > 0.724$). Similarly, the RSM and DVS have relatively better prediction capability from the 25th week with $r > 0.689$ and $r < -0.688$, respectively.

3.4. National yield forecast based on regressions at different administrative levels

Considerable differences are related to NUTS0 level forecast in the 27th and 28th weeks when the regression resulted in a higher RMSE for TWC, WLB and between weeks 28 to 31 for WLSO indicating a temporarily decreased forecast skill (Fig. 5). The highest forecast skill was obtained from weeks 28 to 37 during which the best predictors were WLB, TWC, RSM, WLLAI and WLTR. The best performing forecast occurred at NUTS1 level with an r^2 and a RMSE of respectively 0.8565 and 425.9 kg ha^{-1} using WLLAI as predictor in week 34. Similar results at NUTS1 level were obtained with TWC (RMSE of 470.0 kg ha^{-1}) and WLB (RMSE of 463.1 kg ha^{-1}). Even though the best results were obtained aggregating at the NUTS1 level, calculations indicate only minimal differences between the NUTS levels, except during short periods.

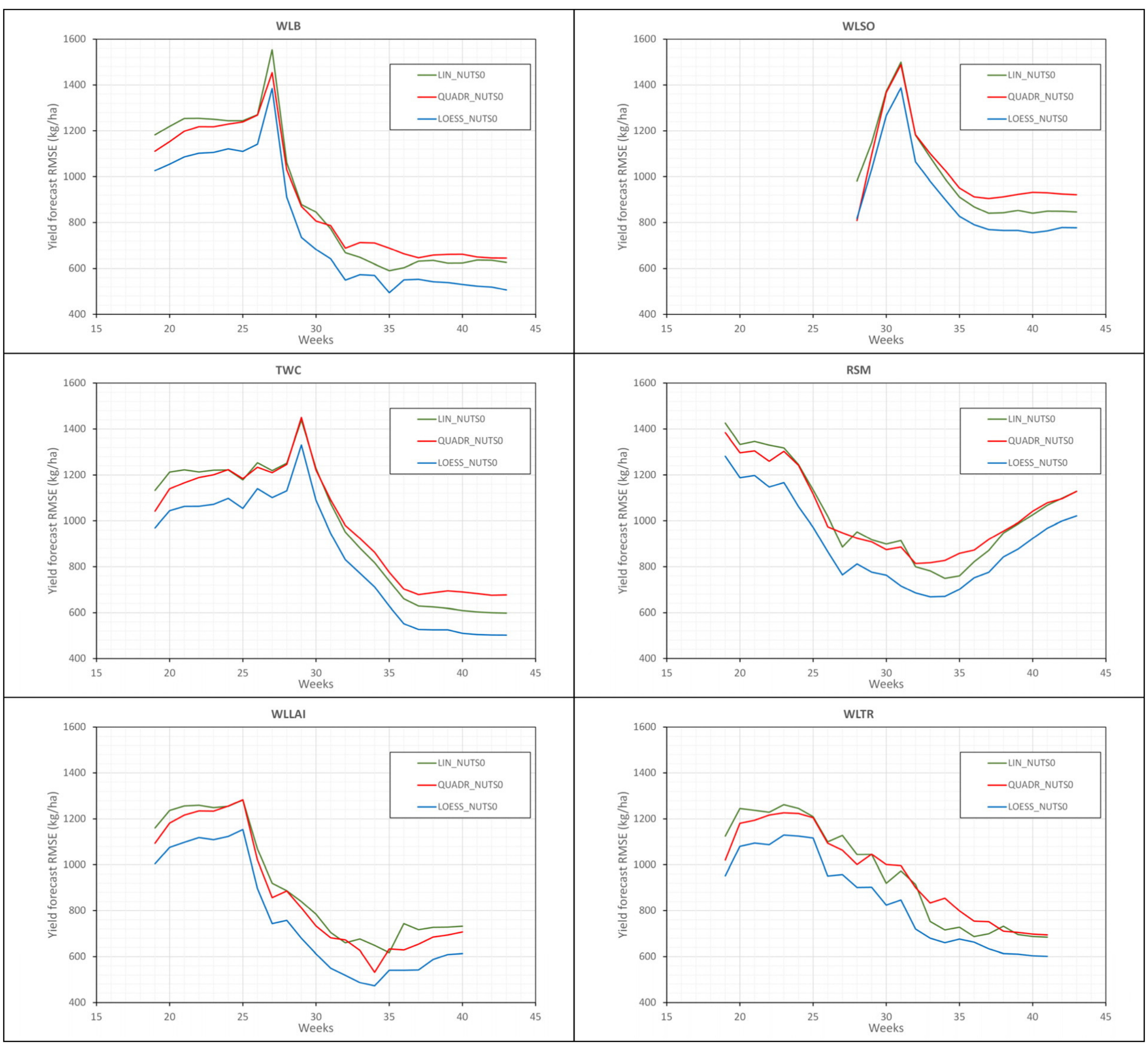

Fig. 4. RMSE of the national yield forecast by using the three different trends in combination with different WOFOST crop model-based predictors during the cropping cycle (weeks 18–43).

Table 4
Correlation coefficients calculated between weekly WOFOST model output variables (see Table 1 for abbreviations) and yield residuals as deviating from the LOESS trend at national level.

Week	DVS	WLB	WLSO	WLLAI	RSM	TWC	WLTR
18	−0.322	−0.349		−0.352	0.268	−0.494	−0.442
19	−0.519	−0.614		−0.634	0.187	−0.653	−0.595
20	−0.441	−0.549		−0.500	0.408	−0.556	−0.509
21	−0.561	−0.524		−0.512	0.435	−0.586	−0.532
22	−0.628	−0.547		−0.542	0.430	−0.545	−0.471
23	−0.586	−0.540		−0.511	0.533	−0.554	−0.506
24	−0.609	−0.526		−0.426	0.595	−0.565	−0.223
25	−0.688	−0.479		−0.028	0.689	−0.605	0.178
26	−0.713	−0.199	−0.383	0.724	0.753	−0.522	0.666
27	−0.758	0.447	−0.545	0.822	0.807	−0.324	0.703
28	−0.734	0.707	−0.791	0.828	0.675	0.055	0.747
29	−0.764	0.820	−0.514	0.849	0.704	0.413	0.667
30	−0.789	0.848	−0.171	0.882	0.700	0.561	0.680
31	−0.781	0.879	0.227	0.908	0.829	0.682	0.804
32	−0.808	0.908	0.573	0.922	0.656	0.764	0.892
33	−0.817	0.911	0.659	0.933	0.615	0.800	0.795
34	−0.849	0.914	0.721	0.944	0.597	0.831	0.766
35	−0.809	0.923	0.776	0.898	0.498	0.871	0.663
36	−0.696	0.920	0.800	0.767	0.336	0.900	0.582
37	−0.692	0.915	0.811	0.664	0.247	0.909	0.478
38	−0.673	0.910	0.813	0.599	0.202	0.910	0.463
39	−0.665	0.909	0.814	0.474	0.184	0.912	0.424
40	−0.676	0.910	0.817	0.345	0.053	0.917	0.274
41	−0.689	0.910	0.818	0.277	0.081	0.919	0.200
42	−0.720	0.910	0.818	0.313	0.073	0.920	0.178
43	−0.667	0.910	0.818	0.222	−0.103	0.920	0.231

The difference between the national forecasts based on information from the different NUTS levels was below 70 kg ha^{-1} for 87.6% of the forecasts and exceeded 100 kg ha^{-1} only in 3.7% of the forecasts. This means that in most cases only small differences in the yield forecasts occur, independent from whether the national yield forecast is based on 1 (NUTS0), 3 (NUTS1), 7 (NUTS2) or 20 (NUTS3) regression equations. Additionally, the usage of NUTS1-NUTS3 administrative level data during the yield forecast did not result in significantly better lead times of the national yield forecast.

3.5. Temporal development of forecast skill

The forecast skill of the regression method was also compared to a forecast solely based on the LOESS trend by assessing the number of years one or the other method performed better, in terms of the absolute difference with the reported yield, on a weekly basis (Fig. 6). The regression method based on the WLB and WLLAI predictors were moderately better than the trend even in the first part of the vegetative phase, with respectively only 12 and 14 years out of 20 exceeding the accuracy of the 'benchmark' forecast. From the end of the vegetative and beginning of the reproductive phase (week 26–28) the regression improved, and between week 30 and 35 reached the best performance, outperforming the LOESS-trend yield forecast 18 to 19 times out of 20. RSM did not lead to an advantage in the very early part of the considered period. However, the forecast with RSM improved earlier in the season compared to WLB and WLLAI, outperforming the benchmark 17 to 18 times out of 20 years typically during the 27th and 35th weeks. The TWC resulted in a low forecast skill until the beginning of grain filling, but improved quickly and towards maturity 18 to 19 times out of the 20 forecasts were better than the LOESS-trend yield forecast. Regarding the different NUTS levels, during the first half of the crop cycle there were only marginal differences. From the start of flowering and onwards the forecast based on NUTS0 level performed worst compared to forecasts aggregated from sub national level (NUTS3 to NUTS1).

The temporal development of the RMSE of the forecasts using the LOESS trend was further analysed by focusing on the national yield forecasts aggregated from NUTS level 1 (Fig. 7) throughout the whole growing season. We considered the following model output variables: WLB, TWC, RSM, WLLAI, WLSO and WLTR. Different crop model variables exhibit the best forecast skill at different times in the growing season. At the start of the season none of the model variables have a significant forecasting capability, and the magnitude of the RMSE (>940 kg) approaches the deviation of yield around the LOESS trend. During the period of weeks 20 to 23, the RMSE was the lowest for the regression with TWC, but during the following 2 weeks, the DVS-based forecast performed best while RSM provided the best forecasts during week 26. The RMSE of the forecast performed with the WLLAI started to decrease significantly from week 25 and became the best forecast from week 27 (just before the start of flowering). The highest forecasting accuracy (RMSE < 470 kg) was reached and maintained between weeks 32 and 35. The minimum RMSE was observed during week 34, corresponding to 425.9 kg. After a sharp decrease in RMSE the WLB also was appropriate for yield forecasting from week 32 and onwards although the accuracy remained below the results obtained with the WLLAI. At the very late phase of the crop cycle the TWC led to a similar accuracy as the WLB. Surprisingly the WLSO performed weakly and only improved very late in the cropping season.

Overall, the maize crop yield forecasting method was able to explain most of the yield variability and thus performed satisfactorily. The best predictors (e.g. WLLAI and WLB) were able to detect the extreme years and the RMSE was reduced to 30% compared to the LOESS-trend forecast (Fig. 8). The regression between the real and estimated yield residuals by WLLAI and WLB was quite strong with $r^2 = 0.877$ and $r^2 = 0.857$ respectively (Fig. 8). The highest prediction errors occurred in wet years such as 1995, 1999 and 2010.

3.6. Multiple regression

We also evaluated the predictive capability of multiple-regression analysis (ignoring possible multi-collinearity between the crop model variable outputs) to assess whether a significant improvement in the forecast would be obtained. Therefore, when extrapolating this model, i.e. to use it for forecasting, there would be a risk of unreliable yield forecasts given the instability of the regression coefficients estimates. Multiple regression was calculated using the most logical crop model output combinations: WLB–RSM–WLLAI, WLB–TWC–WLLAI, WLB–RSM–DVS, RSM–WLLAI–DVS and TWC–WLLAI–DVS. The results are similar to the simple regression; showing the lowest RMSE in the range of 402.5–451.2 kg ha^{-1}. The lowest forecast error was obtained when aggregating at NUTS0 and NUTS1 level from week 32 to 34. The only exception was the WLB–RSM–WLLAI combination which resulted in the lowest RMSE with 379.5 kg ha^{-1}, but this occurred in the 38th week (15–23 September) which is late without any lead time as the harvest would already have occurred.

4. Discussion

Our yield forecast study showed that the highest coefficients of determination reached similar values exceeding ($r^2 > 0.8$) while the lowest RMSE were in the range of 400–500 kg ha^{-1}. In order to improve these forecasts significant improvement of the crop model (e.g. local calibration, a better description of crop phenology, accounting for additional factors like crop diseases, including negative effects of water excess) or inclusion of remote sensing data as statistical predictors would need to be applied. However, a limit in the forecasting ability will also be determined by the uncertainty in the statistics themselves.

4.1. Trends

The choice of the trend is very important for the accuracy of the forecast. Determining trend type is not always obvious if considerable variability or break points exist in the time series because of e.g. political, economic, ecological, or climatological reasons. The trend provides 1) yield residuals for the regression analysis and 2) the 'baseline' yield

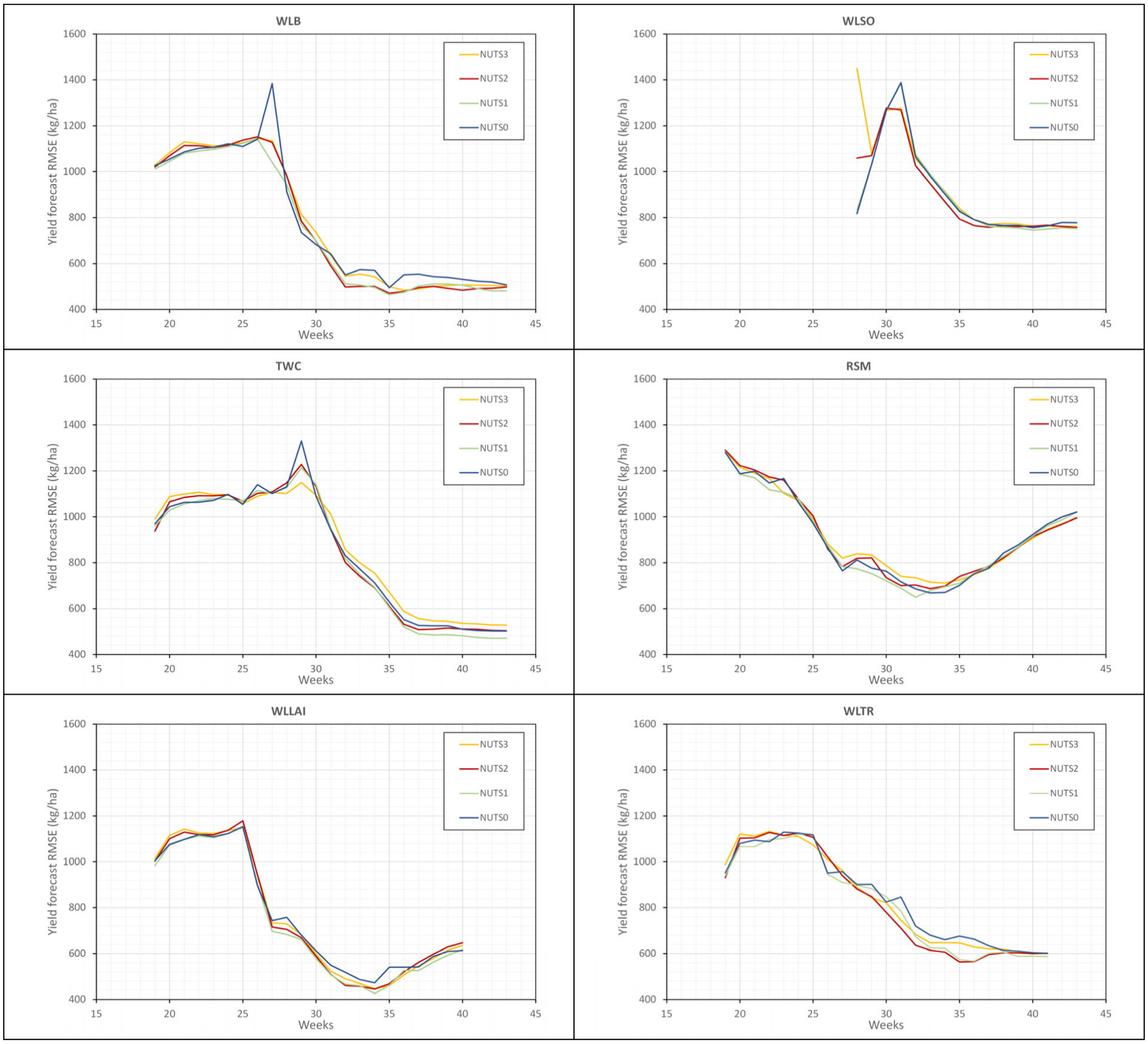

Fig. 5. Comparison of the RMSE of NUTS0 to NUTS3 level aggregation based national yield forecasts using different WOFOST crop model-based predictors in combination with the LOESS trend.

value for the target year. We concluded that the LOESS trend provided the best choice to determine yield residuals. However, due to the analytical construction of the LOESS trend, it may not be optimal in a real forecasting mode since it primarily is an interpolation technique. Therefore, to determine the baseline yield value of the forecast year, it may be used in combination with other techniques (e.g. a linear trend).

4.2. Crop model

Given the weak performance of the WLSO model variable, improvements could focus on the partitioning of biomass in the storage organs (harvest index) in the WOFOST model. The best result of WLLAI was unexpected, since the WLB or water supply parameters are usually thought to play a key role in the yield forecast. At the same time this result was not unprecedented (Baez-Gonzalez et al., 2005). WLLAI was related to the WLB since the biomass assimilate is partitioned mainly to the leaves during the vegetative phase. Additionally, the WLLAI determines the solar radiation interception and indirectly relates to transpiration (Shuttleworth and Wallace, 1985). The high LAI and consequently the high solar radiation interception during the flowering and grain filling growth stages is crucial to obtain a high yield. This also indicates that a remotes sensing based vegetation index could perform well in crop yield forecasting if the maize crop can be identified adequately.

Results suggest that crop model improvements should focus on a better simulation of crop production under wet conditions. The WOFOST model implemented in the M-CYFS currently does not simulate the effect of water logging and over-wet conditions on maize growth and therefore systematically overestimated the maize yield during the presence of water surplus, as mentioned previously. At the same time the most severe yield losses were always related to drought events, given maize's sensitivity to water scarcity (e.g. Van der Velde et al., 2010). Nevertheless, the inaccurate yield estimation during the severe drought of 2007 (Fig. 8) cannot solely be evaluated as a model error, because farmers will harvest the most affected grain maize earlier as green maize to minimize crop losses. Indeed, the harvested area of green maize in 2007 was 140,958 ha while it was 94,359 ha in 2006 and 90,787 ha in 2008. In this case, expected yield loss leads to a decrease of harvested grain maize area thus amplifying the decrease in grain maize production. At the same time, it is well-known that ordinary regression cannot reproduce extremes accurately as the variance of the predicted values will always be lower than the variance in the observed data.

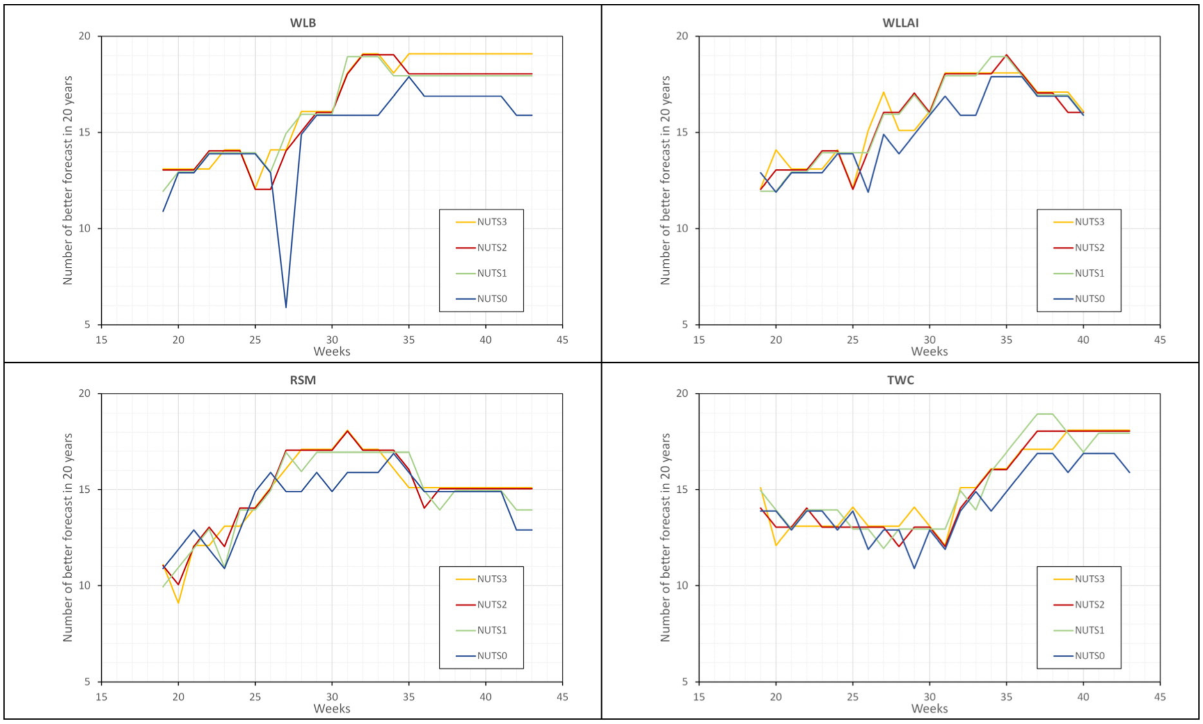

Fig. 6. The number of superior yield forecasts throughout the cropping cycle obtained with the regression based method using the LOESS trend in combination with four predictors (WLB, WLLAI, TWC and RSM) compared to a forecast solely based on the LOESS trend evaluated over the 20-years considered in terms of the absolute difference with the reported yield.

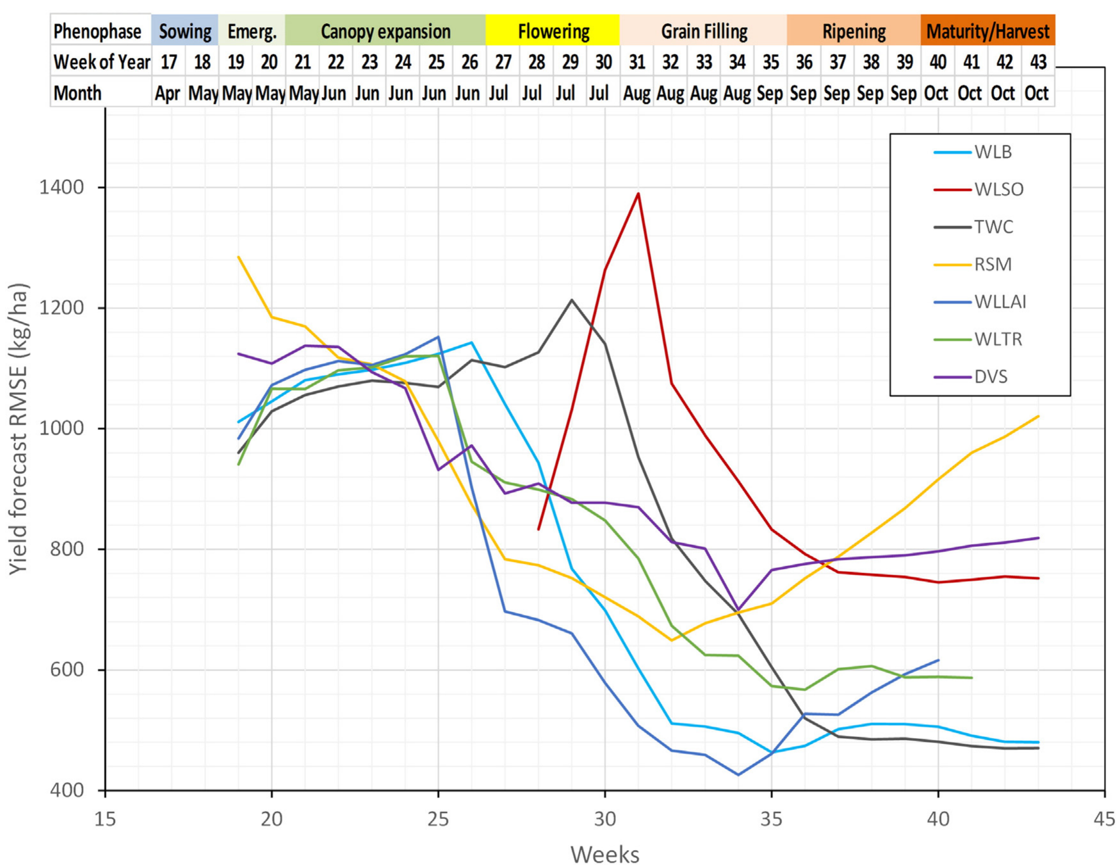

Fig. 7. The temporal development of the RMSE of national yield forecasts based on the LOESS trend in combination with the different WOFOST crop model-based predictors as aggregated from NUTS1 level. The crop development stages are indicated at the top of the Figure.

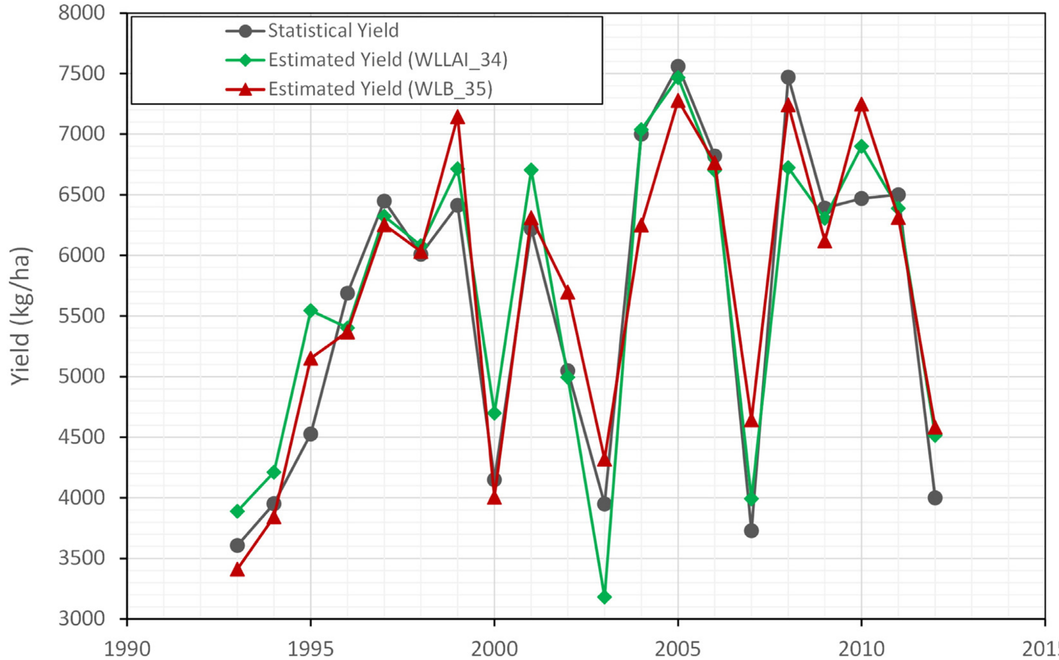

Fig. 8. National reported maize yield and the best estimated hind-casted yields(jack-knifing) using the regression with WLLAI (week 34) and WLB (week 35) at NUTS1 level aggregated to national level with r^2 of respectively 0.8772 and 0.8565.

4.3. Use of weather forecasts

The identification of good forecast skill obtained around the start of flowering from week 26 and onwards (with a maximum in week 34 and 35 with WLLAI or WLB) suggested a potential benefit from short-range weather forecasts. The lead-time of the yield forecast could be improved if extended-range weather forecasts (between 10 and 30 days) were used in the period prior to week 26. At the same time, the last 5–8 weeks of maize crop cycle have a limited effect on the yield forecast, since weather conditions generally have a limited impact on maize yield. Generally, maize is not that sensitive to the weather conditions during the harvest campaign in Hungary, since yield losses are generally negligible and unfavourable conditions usually only cause a delay of the harvest.

4.4. Benefits of regional forecasting

National yield forecasts aggregated from subnational level did not provide higher yield forecast accuracy. Benefits of regional forecasting are likely accrued when 1) countries, or higher level administrative areas, are divided by natural barriers such as mountains which lead to different climates, 2) the distribution of the crop is not homogeneous but concentrated in certain regions, 3) and on the model side, if regional calibration of the crop model is performed. In this case study of maize in Hungary, the best national yield forecast was performed with NUTS1 level data and 3 regression equations. Further increase to NUTS 2 level (7 equations) lead to better result only for WLTR, but this model variable had a lower forecast skill than WLLAI and WLB. Results at NUTS3 level (20 equations) did not yield any improvement. In most of the cases the differences are small between the results obtained at different NUTS levels. The limited improvements in Hungary can be attributed to 1) maize being homogenously distributed across the country and 2) the application of a crop model which was calibrated at national level. Improvements are likely to be obtained with regional model calibration and better crop masks. At the same time there could be limitations of crop model usage at finer administrative regions (e.g. NUTS3 level), because several factors not implemented in the crop model (like farm management, special agro-techniques, specific soils) can lead to extremes in yield, not simulated by the model, while at coarser levels these factors are averaged out and thus less relevant in the crop yield forecasting.

4.5. Comparison to operational crop yield forecasting.

We compared the regression-based forecasts to operational forecasts performed by the Hungarian Ministry of Agriculture, and those by MARS, JRC for the period 2007–2012 (during this period data was available for both datasets). During this period the Hungarian Ministry of Agriculture provided two *estimates* of yield based on observations during the last dekad of July and issued on the 1st of August and a final yield estimation done in the first dekad of September and issued on the 12th of September. The JRC provided monthly forecasts over the same time period and we used the forecasts made in August and September. The performance of the different approaches is illustrated in Fig. 9 and expressed as the ratio of the forecast made and the reported yield. The overall average RMSE equalled 492, 212, 428, 317 and 252 kg ha^{-1} for respectively the first and final estimate from the Ministry of Agriculture, the regression based model (using WLLAI in dekad 34), and the MARS Bulletin forecast reported in August and September. If we exclude 2008 which was a record year and which all approaches underestimated, the number change to 488, 165, 300, 252 and 227 kg ha^{-1}. The final September estimate by the Ministry of Agriculture is closest to the reported yield, followed by the forecast reported in the MARS Bulletin. Both perform better than the regression based forecast we evaluated here. This points to the value of field sampling for national-level yield estimates as well as to the importance of expert knowledge in the MARS crop yield forecasting process.

5. Conclusions.

Our results indicate that the Hungarian national maize yield forecast with the highest accuracy was done using data aggregated from the NUTS1 level using WLLAI as the predictor (r^2 and RMSE of respectively 0.8565 and 425.9 kg ha^{-1}). We further conclude that LOESS trend provided the lowest RMSE in describing the yield time-series compared to the quadratic and linear trend. The best performing yield forecasts with the largest lead times were obtained with the WLLAI and WLB as

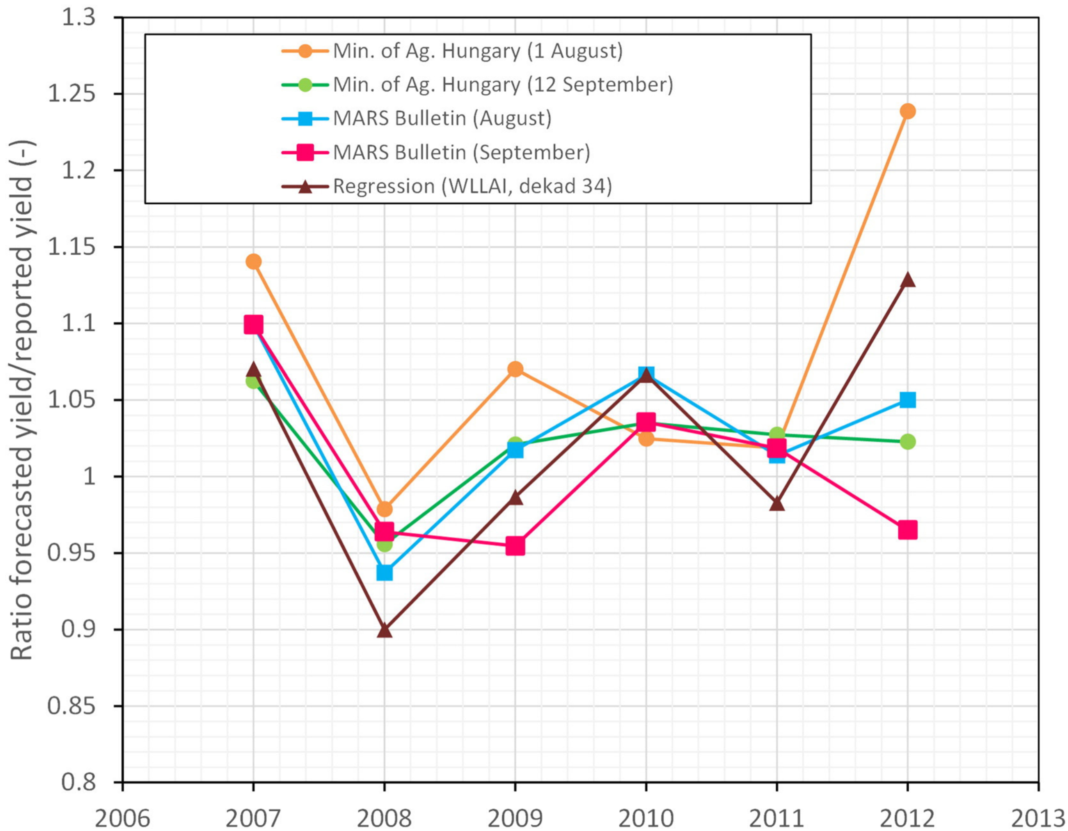

Fig. 9. Comparison of the operational national and regression-based Hungarian maize yield estimates and forecasts for the period 2007–2012. During this period the Hungarian Ministry of Agriculture provided estimates based on field-sampling in August and September, the JRC-MARS provided forecasts in the August and September Bulletin, and we included the best-performing regression-based forecast using the Water Limited Leaf Area Index (WLLAI) in dekad 34.

predictors. The best forecasts were associated with the critical phenological stages of flowering and grain-filling stages respectively occurring between weeks 27 to 30 and weeks 31 to 35. This corresponds to lead times of about 5–8 weeks before the start of maize harvest. The NUTS0 forecast had a slightly lower accuracy from the start of flowering and onwards. However, the benefits of using finer scale forecasts, aggregated to national scale, were minimal in this case-study. Prerequisites for forecasts to optimally benefit from NUTS2 and NUTS3 level approaches may be regionally calibrated crop models, spatially heterogeneous crop distributions, and the occurrence of a spatially variable climate. Hungarian yield estimates and the operational M-CYFS forecasts performed from 2007 to 2012 were significantly better than the regression-based method. This was attributed to the value of field-sampling in the Hungarian national estimates and the importance of expert knowledge in the M-CYFS forecasting process.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.agsy.2015.10.001.

Acknowledgements

We thank Tibor Tóth for providing the Hungarian National Statistical data as well as discussions related to the research presented in this paper. We also thank the three anonymous reviewers whose comments led to significant improvements of the paper.

References

Abdi, H., Williams, L.J., 2010. Jackknife. In: Salkind, N.J., Dougherty, D.M., Frey, B. (Eds.), Encyclopedia of Research Design. Sage, Thousand Oaks, pp. 655–660.
Abedinpour, M., Sarangi, A., Rajput, T.B.S., Singh, M., Pathak, H., Ahmad, T., 2012. Performance evaluation of AquaCrop model for maize crop in a semi-arid environment. Agric. Water Manag. 110, 55–66.
Baez-Gonzalez, A.D., Kiniry, J.R., Maas, S.J., Tiscareno, M.L., Macias, C.J.J.L., Mendoza, J.L., Richardson, C.W., Salinas, J.G., Manjarrez, J.R., 2005. Large-area maize yield forecasting using leaf area index based yield model. Agron. J. 97 (2), 418–425.
Baruth, B., Genovese, G., Leo, O., 2007. CGMS Version 9.2. User Manual and Technical Documentation. JRC Scientific and Technical ReportsEUR, p. 22936.
Baruth, B., Genovese, G., Montanarella, L., 2006. New Soil Information for the MARS Crop Yield Forecasting System. EUR Report EUR 22499, p. 95.
Boogaard, H.L., De Wit, A.J.W., te Roller, J.A., Van Diepen, C.A., 2014. WOFOST Control Centre 2.1; User's Guide for the WOFOST Control Centre 2.1 and the Crop Growth Simulation Model WOFOST 7.1.7. Wageningen University & Research Centre, Wageningen (Netherlands), Alterra.
Boons-Prins, E.R., De Koning, G.H.J., Van Diepen, C.A., Penning de Vries, F.W.T., 1993. Crop specific simulation parameters for yield forecasting across the European community. Simulation Reports CABO-TT no. 32CABO-DLO, Wageningen, The Netherlands (43 p. + Appendices).
Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83 (403), 596–610.
Comas, L.H., Becker, S.R., Von Mark, V.C., Byrne, P.F., Dierig, D.A., 2013. Root traits contributing to plant productivity under drought. Front. Plant Sci. 4, 442.
Csornai, G., Wirnhardt, C., Suba, Z., Nádor, G., Tikász, L., Martinovich, L., Bognár, E., 2006. CROPMON: Hungarian crop production forecast by remote sensing. ISPRS Archives 36 (8), W48.
De Wit, A., Baruth, B., Boogaard, H., Van Diepen, K., Van Kraalingen, D., Micale, F., te Roller, J., Supit, I., Van den Wijngaart, R., 2010. Using ERA-INTERIM for regional crop yield forecasting in Europe. Clim. Res. 44 (1), 41.
FAO, 2012. Statistical Yearbook 2013. World Food and Agriculture. FAO (Food and Agriculture Organization of the United Nations), Rome.
Genovese, G., Bettio, M., 2004. Methodology of the MARS Crop Yield Forecasting System, Statistical Data CollectionProcessing and Analysis Vol. 4. Office for Official Publications of the EU, Luxembourg, p. 92 (EUR 21291/EN/4).
Hafner, S., 2003. Trends in maize, rice, and wheat yields for 188 nations over the past 40 years: a prevalence of linear growth. Agric. Ecosyst. Environ. 97 (1), 275–283.
KSH, 1994. Hungarian Statistical Yearbook 1993. KSH, Budapest, p. 447.
KSH, 1995. Hungarian Statistical yearbook 1994. KSH, Budapest, p. 481.
KSH, 1996. Hungarian Statistical Yearbook 1995. KSH, Budapest, p. 545.
KSH, 1997. Hungarian Statistical Yearbook 1996. KSH, Budapest, p. 604.
KSH, 1998. Hungarian Statistical Yearbook 1997. KSH, Budapest, p. 673.
KSH, 2013. The Role of Agriculture in the Hungarian Economy (in Hungarian) http://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/mezoszerepe12.pdf (Accessed on 4th July, 2014).
KSH, 2014. Dissemination Database; Statistics by Subject (in Hungarian). http://statinfo.ksh.hu/Statinfo/themeSelector.jsp?&lang=hu (Accessed on 22nd March, 2014).

Kucera, L., Genovese, G. (Eds.), 2004. The MARS Crop Yield Forecasting SystemCrop Monographies on Central European Countries Vol. 3. EUR, Czech Republic, Slovakia, Hungary, Slovenia (21290/EN/3).
Lazar, C., Genovese, G., 2004. Methodology of the MARS Crop Yield Forecasting System, Agrometeorological ModellingProcessing and Analysis Vol. 2. Office for Official Publications of the EU, Luxembourg, pp. 10–31 (EUR 21291/EN/2).
Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M., Hammer, G.L., 2014. Greater sensitivity to drought accompanies maize yield increase in the US Midwest. Science 344 (6183), 516–519.
MARSWiki, 2015. http://marswiki.jrc.ec.europa.eu/agri4castwiki/index.php/Main_Page (Accessed on 11th March, 2015).
Micale, E., Genovese, G., 2004. Methodology of the MARS Crop Yield Forecasting System, Meteorological Data Collection Processing and Analysis Vol. 1. Office for Official Publications of the EU, Luxembourg, pp. 5–26 (EUR 21291/EN/1).
Nunes de Lima, M.V. (Ed.), 2005. CORINE land cover updating for the year 2000. IMAGE2000 and CLC2000. Products and Methods. EC, DG JRC, IES, ISBN 92–894-9862-5.
Salinger, M.J., 2005. Climate variability and change: past, present and future—an overview. Clim. Chang. 70 (1–2), 9–29.
Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops—an energy combination theory. Q. J. R. Meteorol. Soc. 111 (469), 839–855.
Soler, C.M.T., Sentelhas, P.C., Hoogenboom, G., 2007. Application of the CSM-CERES-Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment.European. J. Agron. 27 (2), 165–177.
Supit, I., 1997. Predicting national wheat yields using a crop simulation and trend models. Agric. For. Meteorol. 88 (1–4), 199–214.
Supit, I., Van der Goot, E., 2003. Updated System Description of the WOFOST Crop Growth Simulation Model as Implemented in the Crop Growth Monitoring System Applied by the European Commission. Treemail Publ, Heelsum.
Supit, I., Hooijer, A.A., Van Diepen, C.A., 1994. System Description of the WOFOST 6.0 Crop Simulation Model Implemented in CGMSTheory and Algorithms. EUR 15956 EN vol. 1. Joint Research Center, Commission of the European Communities, Luxembourg.
Van der Goot, E., 1998. Spatial interpolation of daily meteorological data for the Crop Growth Monitoring System (CGMS). In: Bindi, M., Gozzini, B. (Eds.), Proceedings of Seminar on Data Spatial Distribution in Meteorology and Climatology, 28 September–3 October 1997, Volterra, Italy. EUR 18472 EN. Office for Official Publications of the EU, Luxembourg, pp. 141–153.
Van der Velde, M., Wriedt, G., Bouraoui, F., 2010. Estimating irrigation use and effects on maize yield during the 2003 heatwave in France. Agric. Ecosyst. Environ. 135 (1), 90–97.
Van Diepen, C.A., De Koning, G.H.J., 1990. Crop Data Files for Use with WOFOST Version 5.0.
Van Diepen, C.A., Wolf, J., Van Keulen, H., 1989. WOFOST: a simulation model of crop production. Soil Use Manag. 5, 16–24.
Van Heemst, H.D.J., 1988. Plant data values required for simple crop growth simulation models: review and bibliography. Simulation Report CABO-TT 17. Centre for Agrobiological Research (CABO) and Department of Theoretical Production Ecology, Agricultural University Wageningen, The Netherlands.
Vizvári, B., Bacsi, Z., 2003. Structural problems in Hungarian agriculture after the political turnover. J. Cent. Eur. Agric. 4 (2), 175–184.