Performance of the MARS-crop yield forecasting system for the European Union: Assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015

M. van der Velde*, L. Nisini

European Commission, Joint Research Centre (JRC), Ispra 21027, Italy

ARTICLE INFO

Keywords:
Crop yield forecasting
Climate change
Crop production
AMIS

ABSTRACT

19,980 crop yield forecasts have been published for the European Union (EU) Member States (MS) during 1993–2015 using the MARS-Crop Yield Forecasting System (MCYFS). We assess the performance of these forecasts for soft wheat, durum wheat, grain maize, rapeseed, sunflower, potato and sugar beet, and sought to answer three questions. First, how good has the system performed? This was investigated by calculating several accuracy indicators (e.g. the mean absolute percentage error, MAPE) for the first forecasts during a season, forecasts one month pre-harvest, and the end-of-campaign (EOC) forecasts during 2006–2015 using reported yields. Second, do forecasts improve during the season? This was evaluated by comparing the accuracy of the first, the pre-harvest, and the EOC forecasts. Third, have forecasts systematically improved year-to-year? This was quantified by testing whether linear models fitted to the median of the national level absolute relative forecast errors for each crop at EU-12 (EU-27) level from 1993 to 2015 (2006–2015) were characterized by significant negative slopes. Encouragingly, the lowest median MAPE across all crops is obtained for Europe's largest producer, France, equalling 3.73%. Similarly, the highest median MAPE is obtained for Portugal, at 14.37%. Forecasts generally underestimated reported yields, with a systematic underestimation across all MS for soft wheat, rapeseed and sugar beet forecasts. Forecasts generally improve during the growing season; both the forecast error and its variability tend to progressively decrease. This is the case for the cereals, and to a lesser extent for the tuber crops, while seasonal forecast improvements are lower for the oilseed crops. The median EU-12 yield forecasts for rapeseed, potato and sugar beet have significantly (p-value < 0.05) improved from 1993 to 2015. No evidence was found for improvements for the other crops, neither was there any significant improvement in any of the crop forecasts from 2006 to 2015, evaluated at EU-27 level. In the early years of the MCYFS, most of the yield time series were characterized by strong trends; nowadays yield growth has slowed or even plateaued in several MS. In addition, an increased volatility in yield statistics is observed, and while crop yield forecasts tend to improve in a given year, in recent years, there is no evidence of structural improvements that carry-over from year-to-year. This underlines that renewed efforts to improve operational crop yield forecasting are needed, especially in the light of the increasingly variable and occasionally unprecedented climatic conditions impacting the EU's crop production systems.

1. Introduction

Fluctuations in European Union (EU) production levels can affect many people, organizations and companies operating in the food-production-chain. In 2012 for instance, EU crop production accounted for about 20%, 7%, 41% and 30% of the global wheat, maize, barley, rapeseed production respectively (data from FAO, www.fao.org). Early and transparent information on expected crop production levels in Europe can aid in global market stability, for instance through the G20 AMIS (Agricultural Market Information System) initiative (http://www.amis-outlook.org/). In Europe, the European Commission’s (EC) Joint Research Centre (JRC) utilizes the MARS Crop Yield Forecasting System (MCYFS, see MARSWiki (2017), and e.g. Bussay et al., 2015; Lecerf et al., 2018; Ceglar et al., 2018, and other articles in this Virtual Special Issue) to provide timely forecasts of crop yields for the European Union. To contribute to the management of the agricultural market of the EU, the JRC has been making in-season forecasts of expected crop yield for the major crops in the European
Union (EU) Member States (MS) since 1993. The crop yield forecasts are performed on a monthly basis during the growing season at EU and country level for all major crops (with a total national cultivated area above 10,000 ha). The forecasts are published in the so-called MARS Bulletin, which contains agro-meteorological analysis, forecasts, and thematic maps on crop monitoring, weather conditions, and yield expectations (available via https://ec.europa.eu/jrc/en/research-topic/crop-yield-forecasting).

To forecast crop production, forecasts of the end-of-season crop yield are multiplied by sown areas as provided by the MS to Eurostat, the statistical office of the EU. The crop production forecasts are communicated to DG-AGRI (the Directorate-General for Agriculture and Rural Development) of the EC, which uses these to assess expected crop production levels within the EU, to estimate changes in stocks across the EU, and to inform decisions regarding markets of cereals, tuber crops, oilseed crops, etc. MS usually report national crop data with a time lag of several months after the crop harvest. Therefore, the regular supply of information on crop growth and expected production levels during the growing season is essential to allow for timely decision making at EU level (see Van der Velde et al., 2018).

The ultimate objective of the MCYFS is thus to predict end-of-season crop production levels as the growing season progresses. In Fig. 1, the end-of-campaign EU soft wheat production forecast, along with the under- or overestimation with respect to reported data, is shown. From 1993 to 2013, the EU had expanded from 12 MS to 28 MS. During this period, the total EU production of wheat doubled from about 70 to 140 Mt. (Fig. 1). The accession of Member States to the EU, for instance the 2004 enlargement, is clearly visible in the increase of the total EU soft wheat production. Low overall EU soft wheat production levels can be related to adverse weather conditions. This was the case in 2007, when heavy rainfall lowered production in France, Germany, Belgium and the Netherlands, while at the same time, production in Hungary and Bulgaria was suffering from drought. In 2007, total EU soft wheat production was overestimated. Conversely, generally positive weather conditions in 2008 led to high production levels, which were underestimated by the MCYFS (Fig. 1).

One can imagine that maintaining operational forecasting activities, and the regular publication of forecasts over such a long period, requires solving of a host of political, bureaucratic, technical, and scientific challenges. All these factors will have affected the performance of the MCYFS in various ways. In this paper, we assess the performance of the MCYFS to forecast crop yield (we do not consider the area component of the production forecast). To ensure the quality of these forecasts, and identify ways to improve them, quality benchmarks are essential. Importantly, successfully benchmarking the performance of operational forecasts will also contribute to a better appreciation of whether they are fit-for-purpose, and thus guarantee their continued and appropriate use. Indeed, for operational seasonal crop yield forecasts to affect decisions by e.g. policy makers they should be relevant, reliable, be based on a systems approach, ensure stakeholder engagement, and – last but not least – the forecasts should have the highest possible accuracy (see e.g. Challinor, 2009).

Several governmental and private sector institutes across the world perform national level crop yield and production estimates (see e.g. Chipanshi et al., 2015). However, despite the importance of such forecasts, and the existence of several organizations that regularly provide them, with a large and diverse client-base that use them; there are surprisingly few studies on the performance of past forecasts done by these (large-scale) crop monitoring and yield forecasting systems (with the exception of e.g. Egelkraut et al., 2011).

The main objective of this manuscript is to assess the performance of the crop yield forecasts done by the MCYFS as published in the MARS Bulletin for EU Member States in the period from 1993 to 2015. We seek to address three questions. First, how good has the system performed? Second, have the forecasts improve during the growing season? Third, have the forecasts systematically improved from year-to-year?

2. Materials and methods

2.1. The MARS crop yield forecasting system

The MCYFS is used to monitor crop growth development, evaluate short-term effects of anomalous meteorological events, and provide monthly forecasts of crop yield and production. The MCYFS functions as an elaborate decision support system, which combined with expert knowledge, allows producing crop yield forecasts. Several interconnected software tools are used in the MCYFS including the CGMS (Crop Growth Monitoring System, Supit and Van der Goot, 2003) which contains a suite of crop models, and CoBo (the statistical Control Board, Genovese and Bettio, 2004) which allows the analyst to perform statistical analysis and crop yield forecasts.

2.1.1. CGMS

The main component of the CGMS is the World Food Studies (WOFOST) crop model (Boogaard et al., 2014; Supit, 1997; Van Diepen et al., 1989; De Wit et al., this Virtual Special Issue). WOFOST outputs are used for the forecasts of soft wheat, durum wheat, barley, rye, rapeseed, potatoes, and sugar beet. WOFOST is a biophysically based, dynamic, and explanatory point model performant across a range of meteorological, soil and agro-management conditions (De Wit et al., 2018). WOFOST simulates crop growth as the difference between assimilates produced by photosynthesis and consumed by respiration. Potential yield as well as water-limited yield can be simulated by WOFOST. Potential yield is determined by the defining factors CO₂, temperature, solar radiation, and crop characteristics. In addition to these factors, water limited yield is limited by water availability. Output variables of the model include development stage (DVS (~)), water limited biomass (WLB (kg ha⁻¹)), water limited storage organ yield (WLSO (kg ha⁻¹)), water limited leaf area index (WLLAI (~)),
total water consumption (TWC (mm)), water limited transpiration (WLTR \( \text{mm week}^{-1} \)), and root zone soil moisture (RSM (−)), also see Lecerf et al. (2018; this Virtual Special Issue) and De Wit et al. (2018; this Virtual Special Issue). For an investigation of the variance in yields explained by these crop model variables see Lecerf et al. (2018; this Virtual Special Issue).

Observe meteorological data is interpolated on a regular 25 km grid with a method based on the distance, altitude and climatic region similarity between the center of grid cells and weather stations, as described by Van der Goot (1998), and used as input to WOFOST. WOFOST runs on the intersection between the 25 km meteorological grid and units based on the European soil map (http://esdac.jrc.ec.europa.eu/). At the time of the yield forecast, the WOFOST is run with the interpolated and observed meteorological data that is extended by the 10-day ECMWF forecast (www.ecmwf.int). Model outputs are aggregated to national level and used as decadal predictors in the statistical analysis.

2.1.2. CoBo

CoBo (Control Board) is a software tool facilitating the statistical crop yield forecasts. Statistical analysis is used to link past yield variability with predictors derived from the interpolated meteorological observations, derived gridded agro-meteorological indicators, remotely sensed variables (e.g. López-Lozano et al., 2015), and the gridded crop model outputs, aggregated at national level (e.g. Supit, 1997). CoBo allows the analysts to perform a quantitative crop yield forecast using a set of statistical methodologies. Statistical methodologies that are available to the analysts and which can be used for the forecast are trend analysis, regression analysis (e.g. Bussay et al., 2015), and a similarity analysis (based on Principal Component Analysis (PCA) and Cluster Analysis, see Genovese and Bettio, 2004). Trends in the reported crop yields can be calculated using a variety of functions (e.g. linear, exponential, etc.). The trend generally provides the first crop yield forecast of the growing season.

The multiple-linear regression analysis requires the analyst to define a trend and identify predictors. The trend and predictors are input into the regression model to explain past yield variability. The derived regression model is then used with the purpose to forecast the end-of-season yield, by feeding the regression model with the predictor values of the ongoing year. By contrast, the similarity analysis relies on the identification of years where meteorological conditions and simulated crop model variables were most similar to those experienced in the current year. The PCA and cluster analysis uses the predictors of all available years to establish a similarity matrix among the years. Yields that were reported in these similar years are then weighed by the distance in component-space to come to a yield prediction (Genovese and Bettio, 2004).

2.2. MARS bulletin

At the start of each year, a roster of analysts working in the MARS forecast project is assembled. Following this, forecasting responsibilities for crop-country combinations (e.g. soft wheat in France) are distributed and assigned to the analysts. Once a crop-country combination has been assigned to an analyst, they will remain with the analyst for as long as the analyst works in the Unit to guarantee development of expertise. The analysts’ use the statistical framework of the MCYFS to forecast crop yields on a monthly basis during the growing season in combination with secondary information. Secondary information includes information from news sources and agricultural organizations that helps to understand what the critical periods were during the growing period. Currently, the MARS Bulletin is published each month during the growing season. The crop yield forecasts are published in the MARS Bulletins (https://ec.europa.eu/jrc/en/research-topic/crop-yield-forecasting). The MARS Bulletin archive (http://ies-webarchive-ext.jrc.it/mars/Bulletins-Publications.html) gathers all Bulletins published. The crop yield forecasts evaluated in this paper were collected from all the MARS Bulletins that were previously published form 1993 to 2015.

2.3. Analysis

The three questions and objectives posed in the introduction were investigated as described in the three subsections below. The calculation of the accuracy indicators and the assessment of the in-season improvement were done for the 2006–2015 period when forecasts were performed for all MS part of the EU-27 (Croatia, a Member State since 2013 is not included because there was too little data). Structural improvements were analysed for the periods 1993–2015 and 2006–2015 for respectively the EU-12 and the EU-27.

2.3.1. Accuracy

The accuracy of the past MCYFS forecasts was investigated by calculating the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), the Root Mean Square Error (RMSE), and the Coefficient of Residual Mass (CRM) of the first forecasts during the season, the forecasts a month before harvest (the pre-harvest forecast), and the EOC forecasts for soft wheat, durum wheat, grain maize, rapeseed, sunflower, potato and sugar beet for the period from 2006 to 2015, as compared to reported yields. For instance, the MAE was defined as:

$$\text{MAE} = \frac{\sum_{i=1}^{n}|O_i - F_i|}{n}$$  \hspace{1cm} (1)

Where \( n \) is the number of cases, \( O_i \) is the observed value and \( F_i \) is the forecasted value (which can be any one of the three different forecasts distinguished). The MAE represents the overall error and compared to the RMSE is less sensitive to errors in large deviations from the mean (Willett and Matsuura, 2005; Hyndman and Koehler, 2006). Similarly, the scale-independent MAPE was calculated as the average of the Absolute Percentage Error (APE) for each pair of yield forecasts and reported data, with APE defined as:

$$\text{APE} = \left| \frac{(O_i - F_i)}{O_i} \right| \times 100$$  \hspace{1cm} (2)

An additional indicator is the Coefficient of Residual Mass (CRM, Loague and Green, 1991) defined as:

$$\text{CRM} = 1 - \frac{\sum_{i=1}^{n} F_i}{\sum_{i=1}^{n} O_i}$$  \hspace{1cm} (3)

The CRM provides information on the tendency of the model to overestimate (when negative) or to underestimate (when positive) the reported data. The minimum is \(-\infty\), the maximum is \(+\infty\), while the optimal value is 0.

2.3.2. Improvements during growing season

The improvement during the growing season was evaluated by comparing the EOC forecast with the pre-harvest forecast and with the first forecast of the season for all crops considered using the PE with respect to the reported yield. Usually, the first forecast is done in March or April and generally is a trend forecast, as selected by the analyst. The harvest month for each crop was determined as per Supplementary Table 1. Results were computed for the 2006–2015 period and are summarized at EU-27 level.

2.3.3. Improvements from 1993 to 2015

Structural improvements in the performance of the MCYFS to forecast yields were assessed for the periods from 1993 to 2015 and 2006–2015 for the EOC forecast. In order to test whether forecast errors converged towards zero over time using a fitter linear model, absolute PE values at national level were used for the calculation. Tests were performed to quantify whether the median absolute relative forecast
error (%) for each year had decreased from 1993 to 2015 at EU-12 level, and had decreased from 2006 to 2015 at EU-27 level, by evaluat-
error (%) for each year had decreased from 1993 to 2015 at EU-12 level, and had decreased from 2006 to 2015 at EU-27 level, by evaluat-
error (%) for each year had decreased from 1993 to 2015 at EU-12 level, and had decreased from 2006 to 2015 at EU-27 level, by evaluat-
error (%) for each year had decreased from 1993 to 2015 at EU-12 level, and had decreased from 2006 to 2015 at EU-27 level, by evaluat-

3. Results

3.1. Forecasts in numbers

In the period from 1993 to 2015, a total of 19,980 yield forecasts were published for crops grown in the EU Member States. In this period, the number of forecasts published each year for soft wheat increased from 30 in 1993, to 91 in 1994, and to 234 in 2015 (Fig. 2). Similarly, the number of forecasts published for potatoes increased from seven in 1993 to 197 in 2015. The number of published yield forecasts for each crop in this period is shown for France, Italy, Austria and Romania (Fig. 2), with a large share of soft wheat and barley forecasts (including spring and winter barley, not evaluated here).

3.2. Accuracy

Fig. 3 illustrates, for soft wheat in Germany, sugar beet in France, grain maize in Romania, and sunflower in Spain, the values of reported yields, the EOC forecasts, the range of forecasts during one season, and the associated relative percentage error from 1993 to 2015. Time series are characterized by more (e.g. grain maize Romania) or less variability, and the lack, or occurrence (e.g. sugar beets in France), of a trend in the yields. The variability in German soft wheat yields, and the range of forecast values during a season, seems to have increased after 2000. A tendency to underestimate sugar beet yields in France can be observed. An extremely low sunflower yield was forecasts for Spain in 2003, while the reported yield was at average levels (~1 t ha⁻¹⁻¹). The relative error varies substantially between crops, and years. In a first section, the accuracy results will be discussed with reference to crops and Member States, while the assessment of the in-season and structural performance will be discussed in a second section.

3.2.1. Member States

Table 1 summarizes the values of the accuracy indicators obtained for each crop across all MS for the period from 2006 to 2015 in terms of minimum, median and maximum values, across years, and Member States. The MAE of the JRC-MARS forecasts for soft wheat ranged from 0.156 to 0.679 t ha⁻¹ with a median of 0.379 t ha⁻¹ across all MS (Table 1). The value for the median MAE for durum wheat is similar to soft wheat, but the minimum value is lower, and the maximum is higher. Similar values are obtained for durum wheat. The largest MAEs are obtained for the tuber crops with median values of 1.56 and 3.86 t ha⁻¹ respectively for potato and sugar beet. Since it does not scale with the magnitude of the average crop yield, the MAE allows making comparisons between forecasts of different crops. The median MAE across all crops, and all MS, evaluated for all years from 2006 to 2015, ranged from 6.31 to 10.15%, respectively for potato and durum wheat. Maximum MAPEs ranged from 15.81 to 28.86% for respectively sugar beet and durum wheat. The MAE of soft wheat forecasts ranged from 0.16 to 0.68 t ha⁻¹ with a median of 0.38 t ha⁻¹ across all MS. For soft wheat, the lowest MAPE was obtained for France at 3.11% (MAE of 0.22 t ha⁻¹), the median value was 7.18% (0.38 t ha⁻¹) for Bulgaria, while Portugal had the highest MAPE at 21.06% (0.38 t ha⁻¹). The median values for the CRM are slightly above zero, indicating a general tendency of the forecasts to underestimate reported yield. The median value of CRM across all MS and years closest to zero is obtained for potato (0.004) followed by soft wheat (0.013). The highest median value for the CRM – and thus the crop for which yields are most under-
derestimated is rapeseed (0.058).

Fig. 4 displays the MAPE and CRM of the EOC forecasts grouped by crop, for all MS (for MS codes, see Suppl. Table 1). The MAPE for a large share of crop-MS combinations is below or around 5%. The MAPE varies considerable within each crop, with notable outliers visible (e.g. durum wheat for SK and BG). The graph for CRM indicates a systematic underestimation of forecasts for soft wheat, rapeseed and sugar beet yields. Although yield underestimations also dominate the pattern for durum wheat, grain maize, sunflower, and to a lesser extent potato, general forecast overestimations of reported yields also occur, with notable outliers such as durum wheat in BG, and sugar beet in LV.

3.2.2. In-season improvements

We consider improvements within one season by evaluating the first, pre-harvest, and EOC forecast for all crops for all MS of the EU from 2006 to 2015 (Fig. 5). Both the forecast error (MAPE) and variability tend to progressively decrease for the cereal crops. The median maize forecast error decreases from 12 to 8% from the first to the pre-harvest forecast, while the variability in durum wheat forecast errors dramatically decreases from the first to the pre-harvest forecast. In-season forecast improvements are lower for both oilseed crops, rape
derseed and sunflower. In rapeseed, the distribution of the pre-harvest and EOC forecast is characterized by relatively more forecasts with lower errors, compared to the first forecast of the year, while the median forecast error only decreases slightly, from 9.9 to 8.9% for respec-
tively the first and EOC forecast. The median in-season forecast error does not decrease for sunflower, although the variability de-
creases. Mixed results are obtained for the tuber crops, although the relative forecast errors tend to be lower compared to the other crops.
Lowest potato forecast errors are obtained for the EOC forecast, with a slight reduction in variability compared to the first and pre-harvest forecasts. The pre-harvest and EOC forecasts for sugar beet are nearly comparable, with the pre-harvest forecast characterized by a slightly lower median forecast error compared to the EOC forecast.

3.2.3. Year-to-year improvements

Table 2 summarizes the results of the assessment of structural improvements in forecasts. From 1993 to 2015, the EU-12 median absolute relative forecast error significantly decreased for rapeseed (see Fig. 6), potato, and sugar beet ($p$-value < 0.05). In other words, yield forecasts for these crops have improved structurally. No evidence was found for improvements for the other crops. In the period from 2006 to 2015, no evidence was found for structural improvements in the EU-27 forecast accuracy for any of the crops. If we evaluate the year-to-year tendencies at MS level, for instance for the 25 MS EOC soft wheat forecasts, we find that four significantly improved (BG, ES, IT, RO), as testified by a significant linear reduction in the absolute percentage

Fig. 3. Reported crop yields, end-of-campaign forecasts, and range of forecasts made (upper panels) and relative percentage errors (lower panels) for soft wheat in Germany (upper left), sugar beet in France (upper right), grain maize in Romania (lower left) and sunflower in Spain (lower right). N.B.: the Error (%) bar graphs have Y-axes with different ranges.
from 2006 to 2015 (p-value < 0.05), while two (LV, SK) significantly deteriorated (see Supplementary Table 6). The remaining MS showed no significant changes either way.

### Table 1

|                  | Soft wheat | Durum wheat | Grain maize | Potato | Rapeseed | Sugar beet | Sunflower |
|------------------|------------|-------------|-------------|--------|----------|------------|-----------|
| MAE (t ha⁻¹)     | 0.156      | 0.11        | 0.279       | 0.929  | 0.05     | 1.948      | 0.111     |
| median           | 0.379      | 0.368       | 0.5695      | 1.563  | 0.2455   | 3.858      | 0.1615    |
| max              | 0.679      | 0.923       | 1.388       | 3.878  | 0.464    | 16.03      | 5.541     |
| MAPE (%)         | 3.105      | 3.662       | 2.542       | 2.538  | 1.389    | 3.314      | 5.369     |
| median           | 7.181      | 10.1505     | 7.9155      | 6.309  | 8.918    | 6.8055     | 8.933     |
| max              | 21.056     | 28.859      | 17.289      | 16.692 | 16.775   | 15.808     | 26.019    |
| RMSE (t ha⁻¹)    | 0.205      | 0.11        | 0.334       | 1.075  | 0.05     | 2.75       | 0.137     |
| median           | 0.473      | 0.4585      | 0.6545      | 1.848  | 0.314    | 4.89       | 0.196     |
| max              | 0.808      | 1.136       | 1.532       | 4.585  | 0.601    | 22.909     | 0.684     |
| CRM (−)          | −0.019     | −0.26       | −0.101      | −0.058 | 0.008    | −0.453     | −0.059    |
| median           | 0.013      | 0.032       | 0.014       | 0.01   | 0.058    | 0.024      | 0.037     |
| max              | 0.104      | 0.064       | 0.061       | 0.073  | 0.117    | 0.071      | 0.151     |

Fig. 4. The mean absolute percentage error MAPE (%, left panel) and the coefficient of residual mass CRM (−, right panel) of the end-of-campaign forecast, grouped by crop, with bars indicating Member States, evaluated for the period from 2006 to 2015. For MS codes, see Suppl. Table 1.

4. Discussion

4.1. Yield data

One major assumption in the forecast process is that the yield statistics as provided by MS to Eurostat are objective, reflect reality, and
are coherent in time and space. Our experience has been that these ‘official’ yield statistics are sometimes still altered several years after their first publication. For instance, in August 2017, 2008 and 2009 Irish oilseed rape yields reported in November 2015 were changed from 3.49 to 3.60 and from 3.39 to 3.70 t ha\(^{-1}\). Other examples include changes of > 25%. Obviously, this affects the assessment of past forecast performance, indeed past forecasts may have been good, given that their purpose was to – to put it bluntly – forecast reported statistics.

4.2. Evaluating forecast performance

Reasons for structural improvements in forecasts may also relate to improved quality of the reported yield statistics. Logically, analysing relationships between meteorological drivers and yield anomalies using higher quality statistics will result in better statistical models. We suspect that this may be one of the reasons for the improved forecasts of rapeseed yields. ‘Rape and turnip rape seeds’ is a combined statistical class for reporting to Eurostat (I1110), and although both crops are rather similar, varying shares throughout the years will result in less-coherent statistical time series. The tendency to underestimate systematically crop yields partially relates to the statistical methods employed, and may partially relate to a conservative attitude of analysts. Indeed, the MCYFS is a quantitative forecasting system, but is strongly analyst driven. As such, the MCYFS approach is different compared to, for instance, the forecasting methodology employed in Canada. The Integrated Canadian Crop Yield Forecaster (Chipanshi et al., 2015) makes a clearer distinction between automated model-based forecasts, and analyst interpretation. In the case of the MCYFS, it would be useful to disentangle the contribution of analyst decisions to past and present forecast performance.

Fig. 5. Boxplots of the relative percentage error (%) of the first (generally March), the pre-harvest (one month before harvest) and end-of-campaign forecasts from 2006 to 2015 for soft wheat, durum wheat, grain maize, rapeseed, sunflower, potato and sugar beet calculated across all European Union Member States (MS). See Supplementary Table 1 for the harvest month for all crops and MS. The pre-harvest month is defined as the month before harvest and can differ between crops and MS. N.B.: the boxplots have Y-axes with different ranges.
Improving operational crop yield forecasts

4.3.1. Forecasting methodologies

In the MCYFS, the statistical methods used have inherent flaws, for instance when it comes to unprecedented conditions or yield levels that are out of the usual range of expectations. In recent years an increasing amount of publications have appeared on crop forecasting using a variety of novel methods and approaches (Becker-Reshef et al., 2010; Ceglar et al., 2016; Iizumi et al., 2013; Ferisse et al., 2015; Schaubberger et al., 2017; Sharif et al., 2017) including probabilistic approaches (e.g. Ben-Ari et al., 2016), random forest (e.g. Jeong et al., 2016), time series techniques such as autoregressive and moving average models (e.g. Choudhury and Jones, 2014). Some of these methods, particularly those based on regression should be implemented in the MCYFS in order to obtain more robust (in terms of sensitivity to the outliers) and accurate forecasts. Opposed to these ‘proof-of-concept’ crop yield forecasting methodologies published in the scientific literature (e.g. Schaubberger et al., 2017), with considerable performance levels (e.g. $R^2 > 0.6$), there is little quantitative data on the performance of operational crop yield forecasting systems. Yet, it is worth noticing that, at first sight, there seems to be a discrepancy between the performance of operational yield forecasting systems and recent methods published in the literature. Therefore, better mechanisms that lead to quicker implementation of scientific advances in operational crop yield forecasting are needed.

4.3.2. Crop models

Improving crop models is a challenging task. Over the last years, research networks such as the global Agricultural Model Inter-comparison and Improvement Project (AgMIP; www.agmip.org), and the European Knowledge Hub “Modelling European Agriculture with Climate Change for Food Security” (MACSUR; www.macsur.eu), have been exemplary in bringing this challenge forward across many fronts (e.g. Nendel et al., 2018 and references therein). Notwithstanding recent progress in the development, use, and combinations of crop models (e.g. Balkovič et al., 2013; Maiorano et al., 2017; Challinor et al., 2018), limitations to simulate certain impacts, such as over-wet conditions, or excessive heat during flowering, still provide a challenge to operational crop yield forecasting. For instance, high precipitation amounts during the French soft wheat harvest in 2007 impacted yields considerably and crop models struggled to reproduce these impacts (Van der Velde et al., 2012). Indeed, it sometimes remains difficult to prove the impact of a single weather event on crop yield, as yield integrates the cumulative effect of weather variability throughout the season, and crops can recover from impacts (e.g. Gobin, 2018). Disentangling different drivers is not straightforward, especially when considering that meteorological events can trigger a variety of indirect processes that may operate at longer time scales, such as the development of diseases, and which so far prove hard to predict (see e.g. Ben-Ari et al., 2018). Further exacerbating the forecasting challenge are changes in phenology trends of crops, driven by the interaction of climate change and management (Eyshi Rezaei et al., 2017). For instance, this may alter the exposure of sensitive phenophases to extreme weather events, affecting crop yield. Improved process understanding is needed (e.g. Marte and Dambreville, 2018). Better data and spatial allocation of information related to where crops are grown and on which soils (e.g. Folberth et al., 2016), and how they are managed, are also needed to improve forecasting of crop yield. Benefits will also be gained from continuously improving skill in medium and extended-range (e.g. Buizza and Leutbecher, 2015), and seasonal forecasts (Doblas-Reyes et al., 2013; Soares et al., 2018), novel downscaling strategies, and further development of integrated approaches (e.g. probabilistic commodity forecasts integrated in a seasonal climate forecasting system, Potgieter et al., 2003; Stone and Meinke, 2005).

4.3.3. Remote sensing

Satellite remote sensing has been at the basis of crop monitoring and forecasting (e.g. MacDonald and Hall, 1980; Genovese et al., 2001). Recently, advances in the use of 1-km resolution satellite earth observation (EO) indicators for yield forecasting of cereals at sub-national level have been implemented in our operational system (see e.g. López-Lozano et al., 2015). Performances differ from one region to another, and improvements can be made, especially in the combined use with meteorological variables. Furthermore, the lack of dynamic crop masks has restricted the use of such 1-km EO products for operational yield forecasting.
forecasting.

The EU’s Copernicus program, and especially the suite of Sentinel satellites, providing freely available and high resolution satellite data with short revisit cycles, is enabling new opportunities for in-season crop monitoring and yield forecasting. For example, the ESA Sen2Agri project (http://www.esa-sen2agri.org) aims to develop, demonstrate, and facilitate the Sentinel-2 time series contribution to the satellite EO component of agricultural monitoring across a range of crops and agricultural practices (Defourny et al., in review). In addition, and in combination with appropriate data storage and processing facilities, the availability of new ground-based data streams, e.g. publically available parcel level crop type declarations by farmers, have the capacity to drastically transform our capacity to monitor crops (D’Andrimont et al., 2017). A subsequent challenge is translating this improved capacity to monitor crops into better operational yield forecasts. Many challenges, including spatially and coherently aggregating this heterogeneous field-level information up across scales – remain (for example see Pagani et al., 2018).

4.4. Forecasting in an increasingly variable climate

In the early years of the MCYFS, most of the yield time series were characterized by strong trends; in recent years, yield growth has slowed, and yields seem to have plateaued in many countries (Grasstill et al., 2013). In addition, an increased volatility in yields is observed in several European countries (e.g. Hawkins et al., 2013; cf. Fig. 3). Therefore, one could argue that forecasting yields has become harder. Besides this, past improvements for rapsed, potato and sugar beet largely related to large gains during the initial years of forecasting. Indeed, while crop yield forecasts tend to improve in a given year, in recent years there is no evidence of systematic improvements that carry-over from year-to-year. With the prospect of an increasingly variable climate in Europe (Kovats et al., 2014), accompanied by the occasional occurrence of unprecedented conditions (e.g.Ben-Ari et al., 2018), this lack of progress needs to be reversed.

5. Conclusion

We have provided an assessment of the performance of the MCYFS for the EU. The accuracy differs by crop and MS, with the median MAPE across the EU-27 ranging from about 6.31 to 10.15% for respectively potato and durum wheat. Overall, crop yield forecasts tend to improve during the growing season. The pre-harvest forecast, most crucial to farmers by crop and MS, with the median MAPE for the EU. The accuracy diagnostics indicates for extreme wheat and maize yield losses. Agric. For. Meteorol. 220, 130–140. https://doi.org/10.1016/j.agrformet.2016.01.009.

Balković, J., Van der Velde, M., Schmid, E., Skalsky, R., Khabarov, N., Obsteiner, M., Stäumer, B., Xiong, W., 2013. Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. Agric. Syst. https://doi.org/10.1016/j.agsy.2017.06.009. Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens Environ 114 (6), 1312–1323. https://doi.org/10.1016/j.rse.2010.03.010.

Ben-Ari, T.J., Adrian, J., Klein, T., Calanca, P., Van der Velde, M., Makowski, D., 2016. Identifying indicators for extreme wheat and maize yield losses. Agric. For. Meteorol. 220, 130–140. https://doi.org/10.1016/j.agrformet.2016.01.009.

References

Balković, J., Van der Velde, M., Schmid, E., Skalsky, R., Khabarov, N., Obsteiner, M., Stäumer, B., Xiong, W., 2013. Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. Agric. Syst. https://doi.org/10.1016/j.agsy.2017.06.009. Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens Environ 114 (6), 1312–1323. https://doi.org/10.1016/j.rse.2010.03.010. Ben-Ari, T.J., Adrian, J., Klein, T., Calanca, P., Van der Velde, M., Makowski, D., 2016. Identifying indicators for extreme wheat and maize yield losses. Agric. For. Meteorol. 220, 130–140. https://doi.org/10.1016/j.agrformet.2016.01.009.

We gratefully acknowledge the multiple collaborators and analysts who have contributed to the development, maintenance, and use of the MARS-Crop Yield Forecasting System over the years since its conception.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.agsy.2018.06.009.
91–111.
Grassini, P., Eskridge, K.M., Cassman, K.G., 2013. Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat. Commun. 4, 2918. https://doi.org/10.1038/ncomms3918.
Hawkins, E., Fricker, T.E., Challinor, A.J., Ferro, C.A.T., Ho, C.K., Osborne, T.M., 2013. Increasing influence of heat stress on French maize yields from the 1960s to the 2000s. Glob. Chang. Biol. 19, 937–947.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688.
Izumi, T., Sakuma, H., Yokozawa, J., Luo, J.-J., Challinor, A.J., Brown, M.E., Sakurai, G., Yamagata, T., 2013. Prediction of season climate-induced variations in global food production. Nat. Clim. Chang. 3, 904–908. https://doi.org/10.1038/nclimate1945.
Jeong, J.H., Jonathan, P., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Ethan, E.B., Timlin, D.J., Shim, K.M., Gerber, J.G., Vangimalla, R.R., Kim, S.H., 2016. Random forests for global and regional crop yield predictions. PLoS ONE 11 (6), e0156571.
Kovats, R.S., Valentini, R., Roover, I.M., Georgopoulou, E., Jacob, D., Martin, E., Roussevelli, M., Sousana, J.F., 2014. In: Barros, V.R., Field, C.B., Dokken, D.J., Mastrandrea, M.D., Mach, K.J., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L. (Eds.), Europe: In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1267–1326.
Lecerf, R., Ceglar, A., López-Lozano, R., Van der Velde, M., Baruth, B., 2018. Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. Agric. Syst. (this Virtual Special Issue), https://doi.org/10.1016/j.agsy.2018.03.002.
Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. J. Contam. Hydrol. 7 (1–2), 51–73.
López-Lozano, R., Duveiller, G., Seguin, L., Meroni, M., García-Gondad Hooker, J., Lee, O., Baruth, B., 2015. Towards regional grain yield forecasting with 1km-resolution EO biophysical products: strengths and limitations at pan-European level. Agric. For. Meteorol. 202, 5–21. https://doi.org/10.1016/j.agrformet.2015.02.021.
MacDonald, R.B., Hall, F.G., 1980. Global crop forecasting. Science 16, 208(4445), 91–111.
Willmott, C.M., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Dyn. 23 (1–2), 7–21. https://doi.org/10.1007/s00382-004-0533-1.
Marte, P., Dambreville, A., 2018. Modelling the Coordination of Leaf Growth for Wheat. Plant physiology, pp. 00986.2017. https://doi.org/10.1104/pp.17.00986.
Nendel, C., Rötter, R.P., Thorburn, P.J., Boote, K.J., Ewert, F., 2018. Editorial introduction to the special issue ‘modeling cropping systems under climate variability and change: impacts, risk and adaptation’. Agric. Syst. 159, 139–143.
Pagan, V., Guarnieri, T., Busetto, L., Ranghetti, L., Boschetti, M., Movedi, E., Camposeo, T., García-Haro, F.J., Katzanou, D., Stravrakoudis, D., Ricciardelli, E., Romano, F., Holex, C., Collivignarelli, F., Granell, C., Castleyn, D., Conflanioni, R., 2018. A high resolution, integrated system for rice yield forecasts at district level. Agric. Syst. https://doi.org/10.1016/j.agri.2018.05.007. (this Virtual Special Issue, accepted).
Pottgier, A.B., Everingham, Y.L., Hammer, G.L., 2003. On measuring quality of a probabilistic commodity forecast for a system that incorporates season climate forecasts. Int. J. Climatol. 23 (10), 1195–1210.
Schauerbeke, B., Gornott, C., Wechung, F., 2017. Global evaluation of a semi-empirical model for yield anomalies and application to within-season yield forecasting. Glob. Chang. Biol. 23 (11), 1365–2486. https://doi.org/10.1111/gcb.13738.
Sharifi, B., Makowski, D., Plauborg, F., Olsen, J.E., 2017. Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark. Eur. J. Agron. 82, A, 11–20.
Stone, R.C., Meinke, H., 2005. Operational seasonal forecasting of crop performance. Philos. Trans. R. Soc. B 360, 2109–2124. https://doi.org/10.1098/rstb.2005.1753.
Soares, M.B., Alexander, M., Dessai, S., 2018. Sectoral use of climate information in Europe: a synoptic overview. Clim. Serv. 9 (2018), 5–20.
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. TreemailPubl, Heelsum.
Van der Goot, E., 1998. Spatial interpolation of daily meteorological data for the Crop Growth Monitoring System (CGMS). In: M. Bindi, B. Gozzini (eds). Proceedings of a 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, p 141-153.
Van der Goot, 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, TreemailPubl, Heelsum (2003)Van der Velde, M., Tubiello, F.N., Vriend, A., Bourouhi, F., 2012. Impacts of extreme weather on wheat and amazine in France: evaluating regional crop simulations against observed data. Clim. Chang. 113 (3–4), 751–765. https://doi.org/10.1007/s10584-011-0368-2.
Van der Velde, M., Bivetti, I., Ayad, M., Niemeyer, S., Santini, F., Van den Berg, M., 2018. Use and relevance of European Union crop monitoring and yield forecasts. Agric. Syst. https://doi.org/10.1016/j.agri.2018.05.001. (this Virtual Special Issue, accepted).
Van Diepen, C.A., Wolf, J., van Keulen, H., 1989. WOFOST: a simulation model of crop production. Soil Use Manag. 5, 16–24.
Willmott, C.M., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30, 79–82.