Agricultural and Forest Meteorology 383 (2026) 111139

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Evaluating the added value of subseasonal weather forecasts for EU national wheat yield forecasts

Maximilian Zachow [a], Ivana Aleksovska [b], Riccardo Henin [b], Harald Kunstmann [c,d], Michele Meroni [e,1], Lorenzo Seguini [b,f], Elena Tarnavsky [b], Senthold Asseng [a,*]

[a] Technical University of Munich, Department of Life Science Engineering, Digital Agriculture, HEF World Agricultural Systems Center, Freising, Germany
[b] European Commission, Joint Research Centre (JRC), Ispra, Italy
[c] Institute of Geography, University of Augsburg, Augsburg, Germany
[d] Institute of Meteorology and Climate Research (IMK-IFU), Karlsruhe Institute of Technology, Campus Alpin, Garmisch-Partenkirchen, Germany
[e] Seidor Consulting, Barcelona, Spain
[f] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, the Netherlands

ARTICLE INFO

Keywords:
Subseasonal weather forecast
Wheat yield
Gaussian process regression
Yield variability
MARS crop yield forecasting system

ABSTRACT

The MARS Crop Yield Forecasting System combines process-based crop model outputs, satellite vegetation indicators and gridded meteorological data within an analyst-guided statistical modelling approach. In May, a critical stage of wheat development, MARS publishes its first forecast based on observed data. While 10-day weather forecasts are often considered at the analyst's discretion, our study goes further by quantitatively incorporating four-week forecasts. To evaluate their added value, we introduce a new framework, MARS+Forecast, which extends MARS yield predictions using ECMWF four-week temperature and precipitation forecasts following the MARS publication date. This framework employs a data-driven yield model for 24 EU countries and is assessed for the harvest years 2007–2024. When assuming a perfect four-week weather forecast based on reanalysis data (MARS+Perfect), MARS forecasts would theoretically be improved in 16 countries, covering 60% of EU wheat area and 55% of wheat production. When adding an actual four-week forecast to MARS (MARS+Forecast), MARS forecasts were modestly improved in 8 countries, representing 39% of wheat area and 31% of wheat production. Extreme yield losses, such as in France in 2016 or Germany in 2018, are not captured by our model, because of the relatively short training datasets with a limited number of extreme years. MARS+Forecast extends the MARS system in May with a four-week weather forecast and demonstrates a novel approach with moderate improvement of the current operational MARS system. More training data, more accurate MARS predictions and improved four-week weather forecasts are needed for future improvements.

1. Introduction

Crop yield variability is projected to increase under climate change with serious risks for food security (Rosenzweig et al., 2014). To mitigate the consequences of crop failures, public institutions closely monitor yields. In the European Union, this task is carried out by the Monitoring Agricultural Resources (MARS) Crop Yield Forecasting System, which provides monthly updates and outlooks for major summer and winter crops across EU member states and neighboring countries (van der Velde et al., 2019). The primary stakeholder of MARS forecasts is the European Commission's Directorate-General for Agriculture and Rural Development (DG AGRI) to support agricultural policy decisions. Additionally, MARS forecasts contribute to the GEO-GLAM Agricultural Market Information System (AMIS), which provides global market intelligence on wheat, maize, rice, and soybeans. Beyond these institutional uses, MARS forecasts are publicly available and utilized by a wide range of stakeholders across industries. Among the crops monitored, winter wheat plays a particularly central role. With an annual EU production of 120–130 million tonnes, of which 20–30 million tonnes are exported (FAOStat, 2023), DG AGRI is interested in timely yield forecasts to create the agricultural balance sheet, guide export licensing, stabilize commodity markets and support food security

* Corresponding author at: Liesel-Beckmann-Str. 2 85354 Freising, Germany.
E-mail address: senthold.asseng@tum.de (S. Asseng).
[1] Under contract with the European Commission, JRC, Ispra, Italy.

https://doi.org/10.1016/j.agrformet.2026.111139
Received 2 November 2025; Received in revised form 13 February 2026; Accepted 19 March 2026
Available online 24 March 2026

in wheat-importing countries (van der Velde et al., 2019). Across the EU countries, winter wheat is typically sown in late autumn and harvested between June and early August. Early in the season (January-April), MARS yield forecasts rely mainly on historical yield trends or moving averages, as current-season crop conditions have yet to establish a clear influence. From May onward, when wheat reaches flowering and the reproductive stages, in-season information about growth conditions is explicitly incorporated. At this phenological phase, wheat has little capacity to recover from stress, so any adverse conditions directly reduce grain formation and final yields, making May a critical turning point for forecast accuracy. Nonetheless, the crop is still developing across Europe at this time, and yield outcomes remain uncertain. MARS forecasts are generated from a combination of process-based crop model outputs, weather observations, and remote sensing indicators, which are integrated into a statistical framework and evaluated by country analysts (van der Velde et al., 2019). 10-day weather forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) are provided to the analysts, but their use and interpretation are left to the analysts' judgment and may range from no consideration to partial subjective inclusion in the final yield forecast (van der Velde et al., 2019). Whether and how forecasts are considered depends on the analyst's assessment of soil conditions or mitigation strategies. For example, an analyst may refrain from lowering a yield outlook if 10-day forecasts indicate a relief from an ongoing drought. However, this approach lacks a standardized mechanism for systematically incorporating weather forecast information across countries and crops. Weather forecasts are produced by the ECMWF at three horizons: medium-range (up to 15 days), subseasonal (up to 46 days) and seasonal (up to 13 months). Seasonal climate models are designed to capture slow-moving weather drivers, which shape climate at monthly to seasonal scales. Their skill over Europe, however, typically vanishes beyond one month (ECMWF, 2024). Subseasonal models, by contrast, target lead times of up to 6 weeks and exploit different sources of predictability that are relevant for weather patterns on these shorter time scales. Given our interest in four-week horizons, the natural choice was to rely on subseasonal forecasts. Importantly, in May wheat across much of the EU begins its reproductive phase, a critical growth stage when stress directly affects grain formation and yield potential. At this point, even one additional month of reliable weather information can be decisive for anticipating yield variability. Building on this opportunity, our study introduces a framework, MARS+Forecast, that investigates how subseasonal weather forecasts can improve MARS yield forecasts without replacing them. Since MARS already integrates a wide range of information (historical trends, crop model outputs such as biomass indicators, remote sensing and weather observations), our aim is not to overwrite these forecasts but to adjust them based on forthcoming weather conditions. We therefore adopt a stacked regression approach: historic residuals from MARS forecasts (i.e. errors computed with final yield published by EUROSTAT) are modeled using a Gaussian Process Regression (GPR), with the expectation that part of these residuals has a systematic component explainable by future weather conditions. Estimated residuals are then added back to the original MARS forecast to generate an adjusted yield forecast. As input features to MARS+Forecast, we use the MARS May forecast, four-week outlooks of temperature and precipitation and the Standardized Precipitation Evapotranspiration Index (SPEI) for the two months preceding the forecast. The SPEI provides context by indicating whether projected weather is likely to mitigate or exacerbate ongoing stress (e.g., emerging drought) (Santini et al., 2022). To evaluate our approach, we trained, validated and tested the model for 24 EU countries over the period 2007–2024 using nested cross-validation. This allowed us to assess to what extent subseasonal forecasts can reduce structural errors in MARS predictions and enhance the reliability of EU wheat yield estimates.

2. Materials and methods

2.1. Yield data, MARS forecasts and study area

MARS relies on an integrated approach that combines outputs from a process-based crop model with vegetation indices derived from remote sensing and meteorological data. These inputs feed into a statistical model and the resulting forecasts are further reviewed by country analysts. Notably, forecasted weather data is not yet included in the quantitative MARS workflow. For this study, we extracted all available historical MARS forecasts from the system's archive (MARS, 2026), covering monthly national wheat yield forecasts from 2007 to 2024. The EU member states Croatia, Malta and Cyprus were excluded: Croatia joined the EU only in 2013 and therefore lacked a complete forecast record, while Malta and Cyprus have negligible wheat cultivation and are not covered by the MARS system. Hence, homogeneous data was available for the forecast months from May to July and for 24 EU countries. The final study area is shown in Fig. 1a. Across this study area, wheat harvest occurs between late June and early August. This study concentrates on forecasts issued in May with four-week outlooks towards the beginning or middle of June. This period is particularly relevant, as it coincides with the reproductive stage, when weather conditions are more relevant in determining final yield realization. In contrast, one-month weather forecasts issued in July have less influence on yield formation, since the crop is already in or approaching the harvest stage and future weather is therefore of limited relevance. May is also the point when MARS forecasts begin to diverge from historical trends, providing a meaningful testbed for early yield assessment. The historical MARS forecasts from May were published between the 9th and 23rd of May (MARS, 2026). In addition to the historical MARS forecasts, historical end-of-season yield statistics were obtained from EUROSTAT (EU, 2023). Previous studies have assessed the accuracy of MARS forecasts against official yields (van der Velde and Nisini, 2019). For the countries and years considered here, forecast errors vary across both space and time, with differences in spread (boxplot width) and systematic tendencies (bias, as indicated by median above or below zero), as illustrated in Figs. 1b and 1c.

2.2. Meteorological data

We propose a new framework, MARS+Forecast, where MARS is combined with weather forecast data. The data come from the subseasonal forecast model from the European Centre for Medium-Range Weather Forecasts (ECMWF). Subseasonal forecasts, up to 46 days ahead, fill the gap between medium-range (up to 15 days) and seasonal forecasts (up to 13 months). Seasonal forecast models have been used in a wide range of use cases (Bento et al., 2022; Brown et al., 2018; Zachow et al., 2024), while subseasonal forecast models are still understudied, possibly because of their shorter lead times. We obtained weekly subseasonal hindcast data from 2007 to 2024, initialized in May within three days after the publication of the corresponding MARS May forecast of the same year (MARS dates varied from the 9th to the 23rd of May). The hindcasts are represented by 10 ensemble members that allow probabilistic forecasts that quantify uncertainty to initial conditions. The outputs of these ensemble members were averaged to obtain one quasi-deterministic weather forecast for mean temperature and precipitation. Furthermore, hindcast data were aggregated from the native 0.5° grid resolution to the national level using a static cropland mask representing non-irrigated arable land. To achieve this, the Corine Land Cover dataset (100 m resolution) was used to identify arable land pixels (CORINE Land Cover, 2018). For each 0.5° grid cell, the number of underlying arable land pixels was counted, giving a measure of the proportion of arable land within that cell. These counts were then used as weights to compute a national average, ensuring that grid cells contributed in proportion to their arable land area. No bias correction was applied to subseasonal hindcasts, as our data-driven model uses

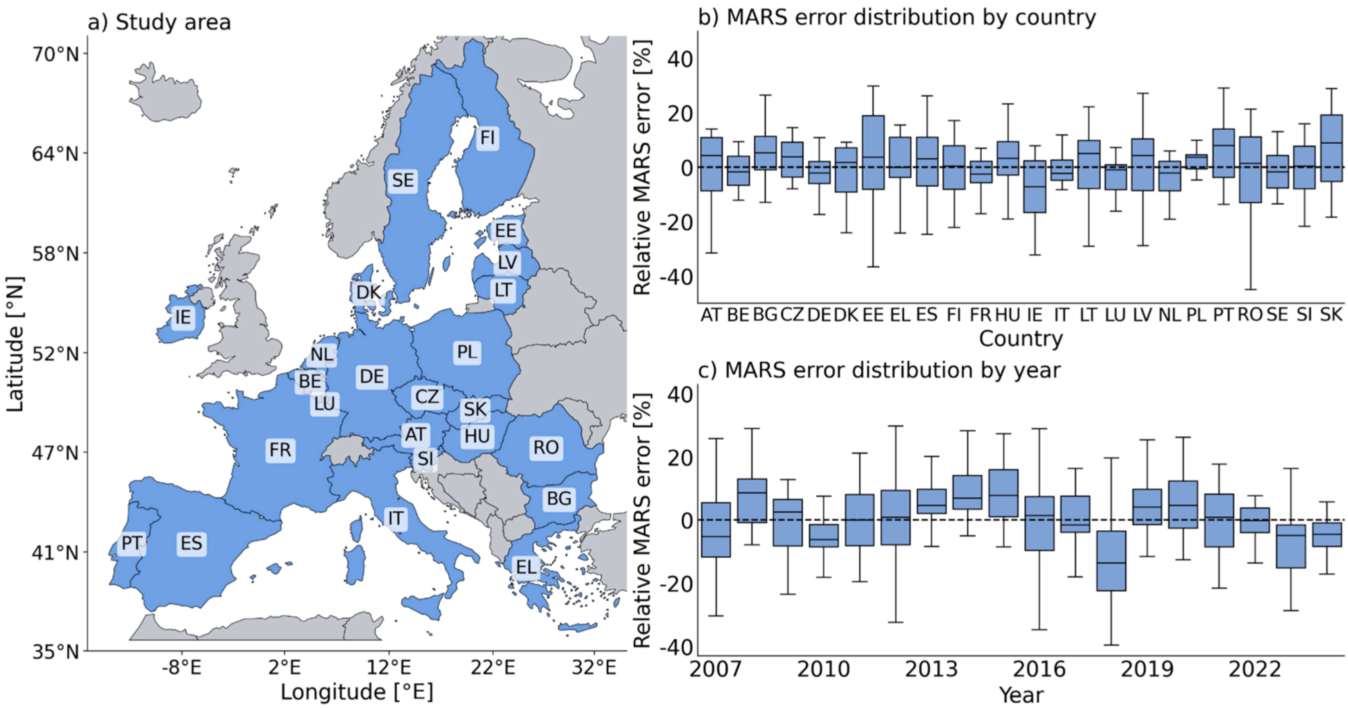

Fig. 1. Study area and MARS error distribution in May. (a) Countries considered in the study are shown in blue with two-letter country codes based on the nomenclature of the European Commission (Glossary, 2026). (b) Distribution of MARS errors by country. Errors are relative, expressed as percentage, and obtained by subtracting the MARS forecast from true yield and dividing the difference by true yield to harmonize different yield levels across the countries. (c) Distribution of these relative errors by year. In (b) and (c), outliers are not shown.

standardized input features and forecasted data is not mixed with observed data (see Section 2.3). Although subseasonal hindcasts are available for up to 46 days, we only used the first four weeks, as there is no forecast skill in Europe beyond one month (ECMWF, 2024). To avoid increasing the number of predictors and the so-called curse of dimensionality we opted for the parsimonious approach of computing the mean over the entire forecast period to have one deterministic precipitation and one deterministic temperature forecast per country and year, representing mean conditions over the four weeks following the MARS publication date. As a preliminary analysis, to assess the potential value of knowing future weather conditions, we first tested the MARS+Perfect approach, in which the four-week weather outlooks were based on ERA5 reanalysis data (Hersbach et al., 2020). Data processing was identical to the subseasonal hindcast data, aggregated from grid cells to national-level averages using the same land mask and converted into four-week means.

To contextualize the four-week outlook relative to the current crop season (for both reanalysis and subseasonal hindcast data), we incorporated the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010), calculated over the two months preceding the MARS forecast date. A two-month SPEI window was chosen to capture droughts that are still developing but have not yet caused irreversible damage and that are not fully reflected in MARS forecasts, which tend to remain conservative in lowering yield estimates as long as recovery remains possible. By incorporating a preceding SPEI as an additional predictor, we can determine whether e.g., future precipitation will positively impact crop yields, particularly when the preceding SPEI indicates dry-warm conditions. In contrast, additional rainfall has a limited effect under already wet-cold conditions. SPEI is derived from the climatic water balance (precipitation minus potential evapotranspiration, PET), where PET was estimated using the Hargreaves method (Hargreaves and Allen, 2003). The resulting water balance was standardized against a 27-year reference period (1980–2006) for each country. Historic temperature and precipitation data for SPEI calculation were taken from the MARS weather database, which are based on interpolated weather observations.

2.3. Modeling workflow and performance assessment

MARS+Forecast was designed to forecast MARS residuals, defined as the difference between final reported yields and MARS forecasts, rather than yields directly. This choice reflects the assumption that MARS residuals contain both structural and non-structural errors. Structural errors may stem from the absence of future weather information in the MARS workflow, which we explicitly incorporate here. By modeling residuals instead of absolute yields, we ensure that the core MARS signal is preserved while systematic shortcomings are corrected. The final predictor set comprised: four-week weather outlooks of precipitation and temperature after the MARS publication date, two-month SPEI prior to the MARS forecast date and the original MARS yield forecast. Predictors were standardized prior to training. Our modeling workflow for each country followed three steps:

1 Compute residuals by subtracting the MARS forecasts from final yields.
2 Train a model to learn the relationship between residuals and predictors.
3 Obtain final yield forecasts as addition of MARS forecasts plus estimated MARS residuals.

Gaussian Process Regression (GPR) was selected as the residual-learning model because it offers a flexible, non-parametric way to capture relationships without assuming a fixed functional form (e.g. linear or quadratic) (Rasmussen and Williams, 2008). Unlike many regression methods, GPR does not deliver a single fitted curve but a distribution over possible functions that could explain the data. Which functions are considered possible is determined by the choice of kernel. Here, we used an additive kernel composed of a radial basis function (RBF) kernel and a white-noise kernel. The RBF kernel encodes the assumption that residuals vary smoothly with the predictors, while the white kernel captures random variation that should not be explained by the model. Together, these kernels balance the signal relevant for improving MARS forecasts with the noise that inevitably exists in agricultural data. Each kernel has hyperparameters that control the behavior of the model. For the RBF kernel, the key parameter is the length scale, which determines how strongly nearby points influence each other: small values allow the function to follow short-term, noisy fluctuations, while larger values enforce smoother, more rigid responses. For the white kernel, the main parameter is the noise level, which defines how much of the variation is attributed to noise rather than explained structure; higher values reduce overfitting by letting the model ignore small, erratic deviations. These hyperparameters are optimized internally by GPR through maximizing the marginal likelihood of the observed training data. The final kernel parameters are those that maximize this likelihood, but because the likelihood surface can have multiple local optima, we restarted the search five times to improve robustness. The outcome is a distribution of functions consistent with the data, from which we used the mean function as the final yield adjustment. Although the optimization is handled automatically, it is still advisable to constrain the search space with reasonable bounds to avoid unrealistic solutions. Based on the literature (Martínez-Ferrer et al., 2021), we restricted the RBF length scale to (0.1, 1) and validated three candidate bounds for the white-kernel noise level: (0.001–0.3), (0.1–0.3), and (0.2–0.3). Finally, for model training, validation and testing, we employed nested cross-validation (Meroni et al., 2021). In this procedure, an inner loop is used to optimize hyperparameters (here, the noise-level bounds), while an outer loop evaluates the predictive skill of the selected model on unseen data. This ensures that model selection is unbiased with respect to test performance. The full modeling pipeline was trained, validated and tested independently for each of the 24 EU countries considered. Specifically, we used a leave-one-year-out scheme in the outer loop, where each year served as a single holdout sample while the remaining years (17) were used for training and hyperparameter optimization in the inner loop. Over the period 2007–2024, this results in 18 folds per country, each with 17 training years and one test year. To assess the added value of MARS+Forecast (and MARS+Perfect) relative to MARS, we compared absolute forecast errors. For each country c and year t, the absolute errors of the MARS forecast $|e_{MARS,c,t}|$ and MARS+Forecast $|e_{MARS+Forecast,c,t}|$ were computed. Relative skill was then expressed as:

$$Skill_{c,t} = \frac{|e_{MARS+Forecast,c,t}| - |e_{MARS,c,t}|}{|e_{MARS,c,t}|}$$

The skill was assessed equally for MARS+Perfect using its corresponding error $|e_{MARS+Perfect,c,t}|$. A negative value indicates that our approach reduced forecast error compared to MARS, while positive values indicate the opposite. Both yield overestimation and underestimation can have significant impacts on EU agricultural policy and global market stakeholders. MARS analysts are therefore tasked with producing the most accurate forecast possible based on a thorough interpretation of available data, without being guided by an asymmetric objective that penalizes one type of forecast error more strongly than the other. We therefore chose absolute errors as the basis for our skill metric, as these are assumption-free and directly reflect forecast deviation from observed yields regardless of direction. In addition, we evaluated a probabilistic forecasting setup by running the yield model with each of the 10 ensemble members of the weather forecast and assessed skill using the Continuous Ranked Probability Skill Score (CRPSS), which generalizes the mean absolute error (MAE) to probabilistic forecasts (Hersbach, 2000). Here, the reference is the deterministic MARS+Forecast approach using the ensemble mean weather forecast, so the CRPSS is calculated as 1 minus the ratio of the CRPS of the ensemble forecast to the MAE of the deterministic ensemble-mean forecast. Positive values indicate that the probabilistic approach is better with an upper limit of 1, while CRPSS values smaller or equal 0 indicate lack of skill over the reference forecast.

3. Results

An initial analysis of the skill of MARS wheat yield forecasts in May is given in Fig. 2. Over the study period from 2007 to 2024, MARS achieved skillful forecasts ($R^2 > 0$) over 10 out of 24 countries. The highest skill was achieved in Poland (PL, 0.65 R^2), Romania (RO, 0.63 R^2), Bulgaria (BG, 0.62 R^2), Spain (ES, 0.42 R^2), Portugal (PT, 0.32 R^2), Hungary (HU, 0.28 R^2) and Italy (IT, 0.20 R^2). The lowest skill was observed in Greece (EL, −0.69 R^2), Ireland (IE, −0.40 R^2), Czechia (CZ, −0.39 R^2), France (FR, −0.31 R^2), Austria (AT, −0.29 R^2), Denmark (DK, −0.25 R^2), Luxembourg (LU, −0.22 R^2), Sweden (SE, −0.16 R^2), Belgium (BE, −0.14 R^2), Germany (DE, −0.13 R^2) and Netherland (NL, −0.11 R^2). Countries with R^2 values around 0 where skill could be attributed to noise in the data and no general assessment can be made were Latvia (LV, 0.09 R^2), Slovenia (SI, 0.04 R^2), Estonia (EE, 0.01 R^2), Slovakia (SK, −0.02 R^2) and Lithuania (LT, −0.07 R^2). MARS forecast skill in June was comparable to May, while significant improvements were observed in July (Suppl. Mat. Fig. S1–2).

Fig. 3 indicates the theoretical performance change in wheat yield forecasts for May when MARS forecasts are combined with perfect four-week weather outlooks (MARS+Perfect). MARS+Perfect would improve MARS forecasts (Error difference to MARS < 0) over 16 out of 24 countries, corresponding to 60% of wheat area and 55% of wheat production. The largest error reduction of MARS was achieved for Slovakia (SK, −5.3%) and Spain (ES, −3.7%). Countries in which no improvement was possible (Error difference to MARS ≥ 0) were Sweden (SE, +1.9%), Luxembourg (LU, +1.9%), Czechia (CZ, +1.4%), Hungary (HU, +1%), Estonia (EE, +1%), Belgium, France and Bulgaria (BE, FR, BG, respectively, all around 0% error difference to MARS). Average performance of MARS+Perfect was best in May when compared to forecast months June and July (Suppl. Mat. Fig. S3–4).

In an operational setting, perfect four-week outlooks based on reanalysis data that are available a-posteriori, will be replaced with subseasonal weather forecasts, with skills shown in Fig. 4. Aggregated four-week precipitation forecasts at the country-level are skillful ($R^2 > 0$) over 17 out of 24 countries. The countries with the highest precipitation forecast skill are Portugal (PT, 0.78 R^2), Spain (ES, 0.72 R^2), Slovakia (SK, 0.53 R^2) and Italy (IT, 0.51 R^2). Countries with no skill ($R^2 \leq 0$) for precipitation forecasts are Ireland (IE, −0.96 R^2), Slovenia (SI, −0.31 R^2), Austria (AT, - 0.23 R^2), Belgium (BE, −0.13 R^2), Luxembourg (LU, −0.09 R^2), Sweden (SE, −0.06 R^2) and Estonia (EE, - 0.04 R^2). Temperature forecasts are skillful over all 24 countries of the study

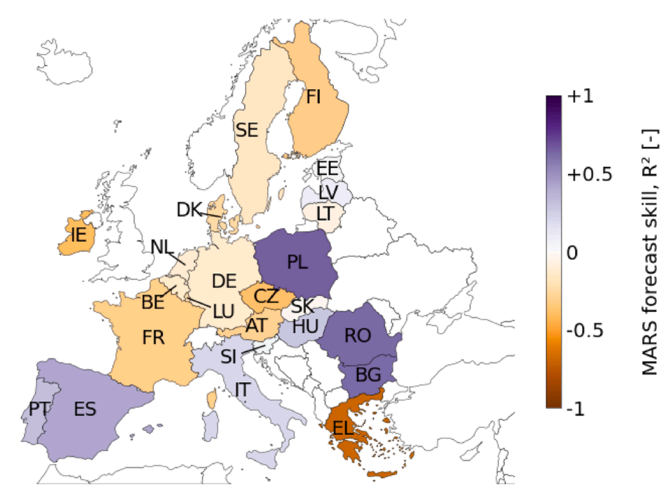

Fig. 2. Skill of MARS wheat yield forecasts in May (2007–2024). Forecast skill for MARS is calculated for each country using the R^2 metric, with negative values (orange shades) indicating no skill and positive values (purple shades) indicating skillful forecasts. MARS forecasts were published between 9 and 23 May each year (MARS, 2026).

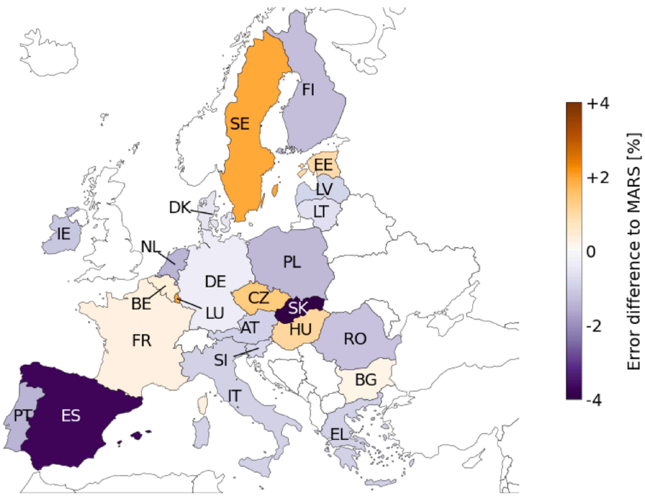

Fig. 3. Mean performance of MARS+Perfect wheat yield forecasts in May (2007–2024). MARS+Perfect integrates reanalysis data to simulate a perfect four-week weather outlook following the MARS publication date, where future conditions are known. The map shows mean performance relative to MARS at the country level. Negative values (purple shades) indicate improved forecasts compared to MARS, while positive values (orange shades) indicate that MARS performed better. Countries included in the analysis are labeled with two-letter codes, other countries are shown in white without labels. MARS+Perfect forecasts correspond to the same publication dates as MARS, which varied between 9 and 23 May each year (MARS, 2026).

region, with the highest skill in Spain (ES, 0.72 R^2), Portugal (PT, 0.61 R^2), Italy (IT, 0.59 R^2), Ireland (IE, 0.55 R^2) and France (FR, 0.55 R^2). The skill of weather forecasts was much worse in June, especially for precipitation, before it recovered again in July (Suppl. Mat. Fig. S5–6).

The skill of using subseasonal weather forecasts to forecast wheat yield within the MARS+Forecast approach is shown in Fig. 5. Only countries with skillful weather forecasts (Fig. 4) and skillful MARS+Perfect forecasts (Fig. 3) are included. MARS+Forecast improves MARS (mean error difference to MARS < 0) over 10 out of 24 countries, corresponding to 39% of wheat area and 31% of wheat production. The highest mean error difference was achieved for Spain (ES, −2.7%), Slovakia (SK, −2.3%), Romania (RO, −1.6%), Italy (IT, −1.2%) and Netherlands (NL, −1%). Countries in which no improvement was possible (mean error difference to MAR ≥ 0) were Latvia (LV, +0.3%), Germany (DE, +1%) and Lithuania (LT, +2.2%). In addition to mean error differences to MARS, the distribution of annual error differences is shown in Fig. 5b Improvements over MARS in more than half of the years (negative median) are achieved for 9 countries. Denmark, Latvia, Greece and Lithuania have positive median added values, meaning that MARS+Forecast underperformed MARS in more than half of the years. Larger skill variability was observed for Romania and Slovakia with wider boxplots. Smaller variability in annual performance was present in Italy, Netherlands, Poland and Denmark.

In Fig. 6, absolute forecast errors for MARS and MARS+Forecast are shown, for countries where mean and median error difference over MARS was negative (Fig. 5). In general, errors between MARS and MARS+Forecast are highly correlated. The highest correlation occurs in the Netherlands (Fig. 6d, 0.94) and in Portugal (Fig. 6f, 0.91) and the lowest correlation occurs in Spain (Fig. 6a, 0.77) and Italy (Fig. 6c, 0.82). In addition to the correlation, the mean absolute error (MAE) for both MARS and MARS+Forecast is reported in Fig. 6. The largest absolute difference occurs in Slovakia (Fig. 6h), with 0.64 and 0.74 t/ha MAE for MARS+Forecast and MARS, respectively. Other countries with large improvements were the Netherlands (0.46 versus 0.55 t/ha MAE), Spain (0.30 versus 0.39 t/ha MAE) and Italy (0.18 versus 0.25 t/ha MAE). There was no apparent relationship between the degree of correlation and the MAE difference.

An additional analysis where each weather forecast ensemble member ran its own wheat yield forecast to obtain a range of scenarios revealed little improvement over the deterministic approach (Suppl. Mat. Fig. S7).

Fig. 7 presents a forecast and predictor analysis for France and Germany, the two primary EU wheat producers. This analysis is based on MARS+Perfect using perfect four-week weather outlooks. It appears that wheat yield variability over the study period from 2007–2024 was not well captured by either approach. Notably, neither MARS nor MARS+Perfect successfully forecasted the extreme yield losses

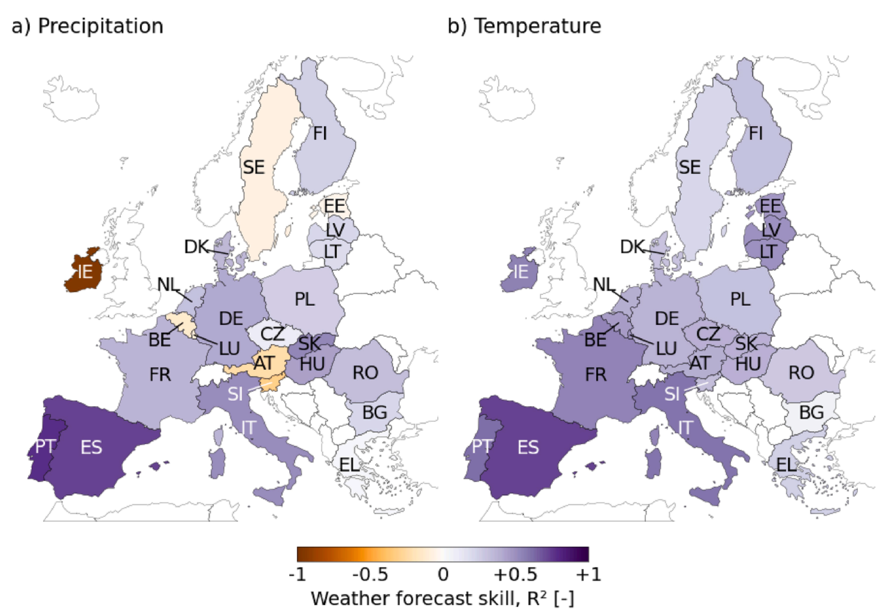

Fig. 4. Skill of four-week weather forecasts in May (2007–2024). Forecasted standardized (a) precipitation and (b) temperature are compared against standardized reanalysis data at the country level using the R^2 metric. Negative values (orange shades) indicate no forecast skill, while positive values (purple shades) indicate skillful forecasts. The analysis covers the study period 2007–2024. Countries included in the study are labeled with two-letter codes, all others are shown in white without labels. The weather forecasts were initialized within three days after the MARS publication date, which varied between 9 and 23 May each year (MARS, 2026).

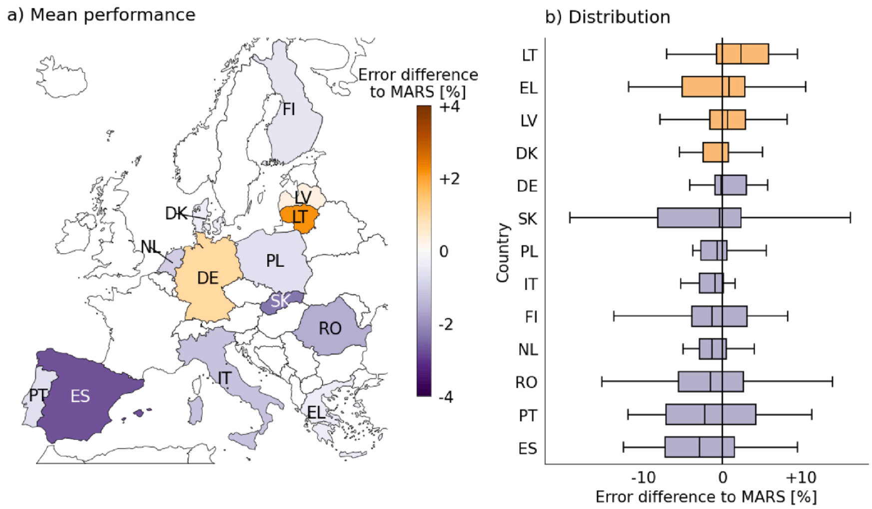

Fig. 5. Performance of MARS+Forecast wheat yield forecasts in May (2007–2024). The analysis includes only countries with skillful weather forecasts and skillful wheat yield forecasts under MARS+Perfect. (a) Country-level comparison of mean performance relative to MARS. Purple shades indicate countries where MARS+Forecast outperformed MARS on average with lower mean absolute error, while orange shades indicate the opposite. Countries included in the analysis are labeled with two-letter codes, others are shown in white without labels. (b) Distribution of annual performance differences between MARS+Forecast and MARS at the country level. Each boxplot shows the range and median of yearly error differences, with a vertical line at zero indicating equal performance. Purple boxplots denote countries where MARS+Forecast outperformed MARS in more than half of the years (median < 0), while orange boxplots denote the opposite (median ≥ 0). The weather forecasts used for the MARS+Forecast model were initialized within three days after the MARS publication date, which varied between 9 and 23 May each year (MARS, 2026).

experienced in France in 2016 and in Germany in 2018 (Fig. 7a-b). The MARS forecasts (which are used as an input to our approach) indicate above-normal wheat yield. MARS+Perfect failed even if anomalous weather conditions were captured in the four-week weather outlooks. For France, the four-week weather outlook from mid-May to mid-June 2016 indicated unprecedented wet conditions, with a standardized precipitation value of >2, and below-normal standardized temperatures (<0) (Fig. 7c, e). In Germany, our predictors showed slightly below-normal standardized precipitation (<0) and exceptionally warm conditions, with standardized mean temperature >2, for the 4 weeks after the forecast date in 2018 (Fig. 7d, f).

4. Discussion

We have proposed a data-driven model to forecast wheat yield across 24 EU countries in May, 1–3 months before harvest using as predictors the MARS forecast, the 2-month SPEI prior to the MARS forecast date and four-week weather outlooks of mean temperature and precipitation after the MARS forecast date. We tested a perfect weather outlook using reanalysis data (MARS+Perfect), which represents a hypothetical, always-correct forecast to assess the theoretical potential of the approach. In addition, we tested subseasonal weather forecasts from ECMWF (MARS+Forecast) to evaluate the operational accuracy when future weather conditions are unknown and must be derived from forecasts. Using an operational and analyst revised forecast (MARS) as an input to a new forecast approach is a novelty. In addition, this was the first study to assess the added value of subseasonal weather forecast data for wheat in 24 diverse wheat-growing countries. MARS+Perfect improved MARS over 16 countries, corresponding to 60% of wheat area and 55% of wheat production. The largest theoretical improvements, measured as relative MARS error reduction, were found in Slovakia (+5.3%) and Spain (+3.7%). MARS+Forecast improved MARS over 8 countries, representing 39% of wheat area and 31% of wheat production. Again, the largest improvement was found in Spain (+2.7%) and Slovakia (+2.3%). For France and Germany, the two largest wheat producers in the EU, both MARS and our frameworks were less successful, failing to capture their respective extreme yield losses in 2016 and 2018.

Before May, MARS forecasts rely mostly on yield trends. May marks the first month when forecasts incorporate growth conditions from the start of the season up to the forecast date. With winter wheat harvest typically occurring between late June and early August across the EU countries, the May forecasts are about 1–3 months away from the end of the season. An analysis of May MARS forecast skill showed varying performance across the 24 study countries (Fig. 2) with more skillful MARS forecasts for southern and eastern European countries and less skill for central and northern European countries. While a comprehensive interpretation of the MARS capabilities is not at the core of our study, we can think of three explanations for the spatial performance patterns that also affected the performance of our approach. First, wheat yield variability in the southern EU countries is mostly driven by water availability, thus more sensitive to precipitation and temperature. Second, wheat yield in the eastern European countries has less annual variability and is dominated by a long-term rising trend, correctly considered by MARS analysts. Third, central and northern European countries experience stagnating and high yield levels. The variability in these countries does not only depend on water availability, unlike in the South, which makes forecasting more challenging (Ben-Ari et al., 2018).

The first part of our analysis aimed at understanding the potential value of having weather outlooks explicitly incorporated in the May forecast. MARS+Perfect improved MARS over 18 out of 24 countries (Fig. 3). Regional performance differences may come from the limitations of our predictors to explain annual wheat yield. For example, the four-week outlook roughly covers the period from mid-May to mid-June. For countries with relatively late harvest (between July and August), weather conditions from mid-June until harvest remain unknown. Furthermore, the MARS forecast does not indicate the status of the crop at the forecast time but tries to anticipate yield development trajectories until harvest. If a crop is underdeveloped or stress conditions are observed by May, but recovery is possible because key grain formation stages are still to come, MARS usually remains cautious and does not lower yield expectations yet. This makes it difficult for our data-driven model to put the information of the four-week outlook into the right context and may lead to cases in which MARS cannot be improved. In some countries, MARS+Perfect even resulted in worse forecasts than MARS (Sweden, Estonia, Czechia, Luxembourg or Hungary). This may

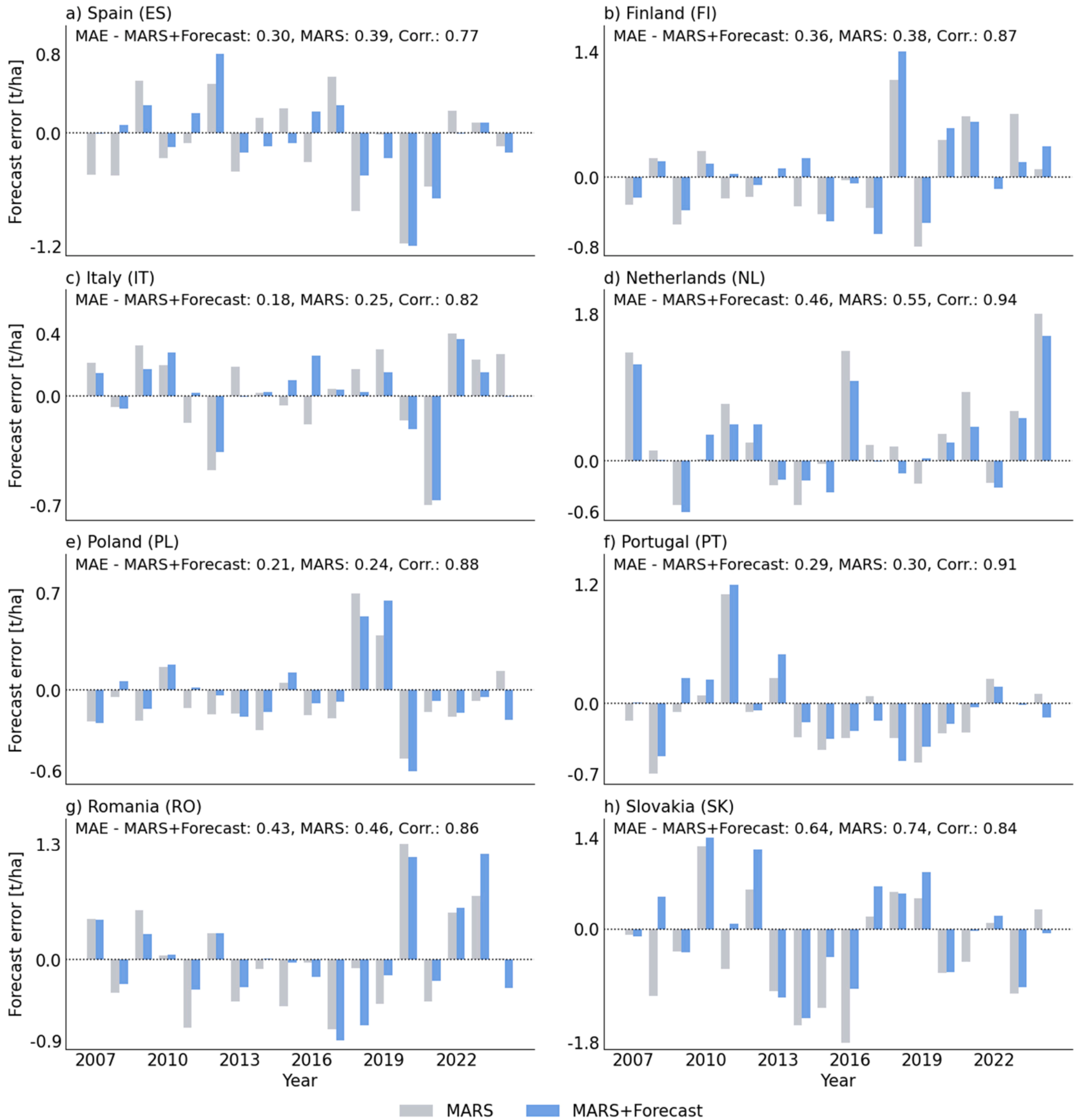

Fig. 6. Comparison of May forecast errors between MARS and MARS+Forecast. The analysis includes only countries with skillful yield forecasts under MARS+Forecast. Each panel corresponds to a country, with the mean absolute error (MAE) of each model and the correlation between their forecast errors reported. The weather forecasts used for the MARS+Forecast model were initialized within three days after the MARS publication date, which varied between 9 and 23 May each year (MARS, 2026).

be a drawback of our purely data-driven approach with only 17 years for training to make a forecast on the 18th year. If the 17 years does not sample enough variability for the model to learn the relationship between predictors and yield, it will possibly produce poor forecasts that are worse than MARS. For operational use, reanalysis data is replaced with subseasonal weather forecasts. Our analysis shows that the May subseasonal hindcast data used here are skillful for weather conditions up to four weeks in advance (Fig. 4). This skill holds true for most countries, except for precipitation forecasts in seven countries. For temperature, a gradient of decreasing skill is observed from the southwest to the northeast, which is consistent with verification metrics from the ECMWF (Büeler et al., 2020). This overall high skill of subseasonal hindcast data may come from our spatial and temporal aggregation of data from grid cells to national-level four-week averages (Monhart et al., 2019). A potential drawback of this aggregation is a loss of signal, as localized extreme events are smoothed out and not accounted for in the yield forecast. Yield forecasts based on subseasonal hindcast data (MARS+Forecast) were skillful over 10 out of 24 countries (Fig. 5a). As

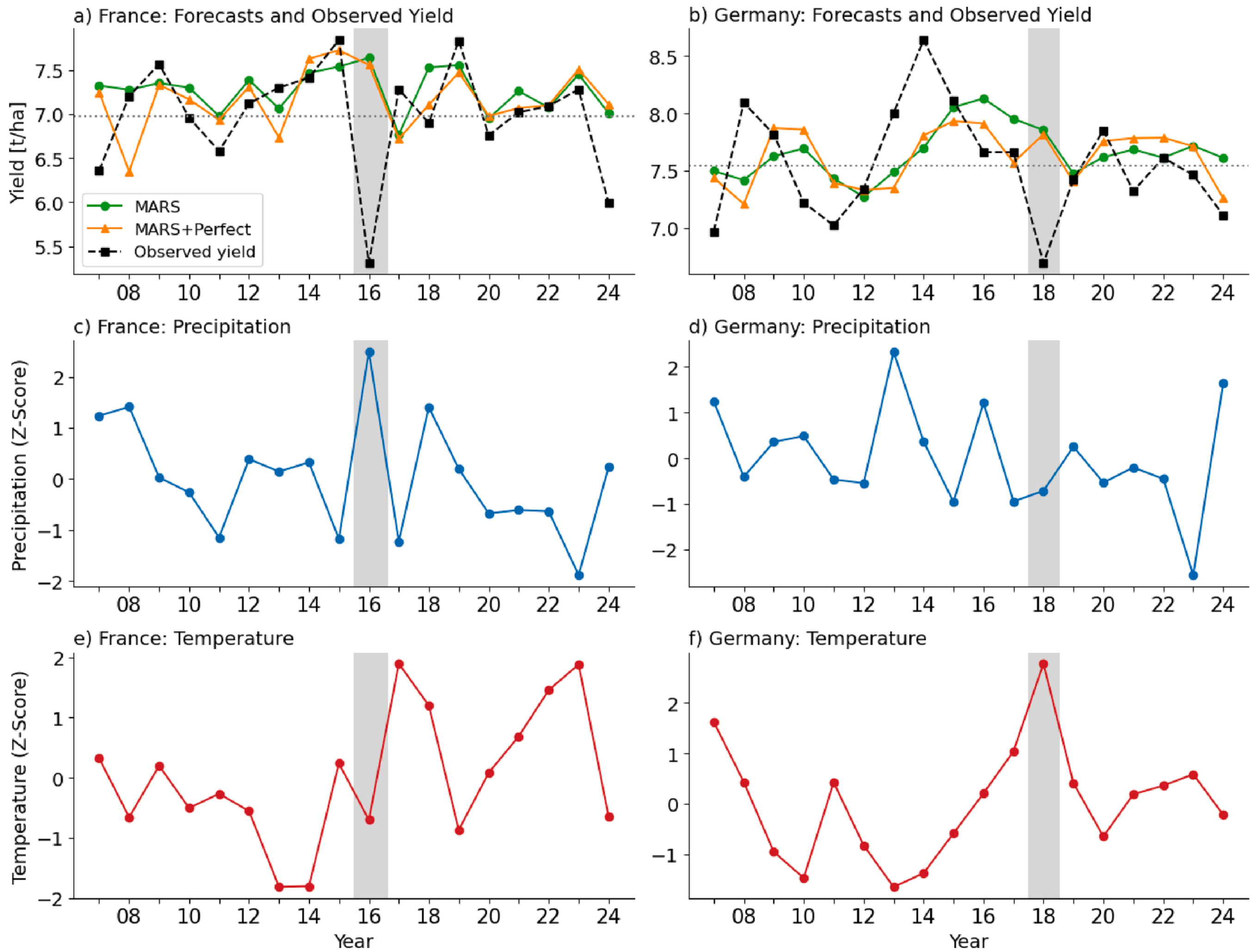

Fig. 7. Annual wheat yield forecasts and predictors for France and Germany (2007–2024).
(a-b) Comparison of observed yields (black dashed lines with square markers) with forecasts from MARS (green line with circular markers) and MARS+Perfect (orange line with triangular markers). Horizontal dotted lines indicate each country's average yield. (c-d) Standardized precipitation anomalies and (e-f) standardized mean temperature anomalies for the four-week outlook period using reanalysis data. The unit of standardized predictors Z-Score). Grey shaded areas highlight the extreme yield loss years in France (2016) and Germany (2018). The weather outlook corresponds to the four weeks following the MARS publication date, which varied between 9 and 23 May each year (MARS, 2026).

expected, a few countries that were previously skillful within MARS+Perfect now fail to improve MARS (Germany, Latvia, Lithuania), possibly because of skillful but imperfect weather forecasts. In addition to the mean performance, we also show the distribution of annual performance change by country (Fig. 5b). The large difference in variability is likely due to the performance variability of the underlying MARS forecast. As we showed in Fig. 6, our errors are highly correlated to the MARS errors. This is because we chose to forecast wheat yield by learning to forecast the residuals of MARS forecasts to true yield. This choice led to a model whose outputs are closely linked to the MARS forecasts. In line with this behavior is the tendency that our approach performed better in countries where MARS was already skillful (Fig. 2–3) and, in turn, performed worse in countries where historical performance of MARS was less good, such as in France and Germany (Fig. 7). These two major wheat producers are suitable examples to demonstrate the limitations of a purely data-driven crop yield modeling approach. Both extreme yield losses in 2016 in France and in 2018 in Germany were not forecasted, although the predictors suggested potential sources of risk (extreme wetness in France and extreme heat in Germany). However, both weather conditions were highly unusual and too extreme to be derived from the training data. If such conditions are not represented in the training data, the models fail to recognize and apply them during testing or validation. The 2016 French yield failure highlights additional complexities. Yield losses were strongly amplified by plant diseases, which are not explicitly represented in our model and are difficult to attribute to single weather predictors (Ben-Ari et al., 2018).

Future improvements to our approach could focus on addressing limitations related to both input data and model design. To provide context for the weather outlooks, we used the 2-month SPEI prior to the forecast date. However, SPEI does not represent information about the impact on the crop conditions. Incorporating crop condition indicators available at the forecast date, such as satellite-derived vegetation indices, could help the model contextualize the four-week weather outlook, particularly in situations where crop recovery or later stress is possible. Further improvements could also be achieved by using crop-specific land masks when aggregating weather data. In this study, we used a non-irrigated arable land mask (CORINE Land Cover, 2018), but in some countries winter wheat is cultivated only on a small fraction of total arable land (e.g. Italy). As a result, the aggregated weather predictors may dilute or omit the true weather signal affecting the crop. Incorporating high-resolution, crop-specific land masks could help ensure that the aggregated weather variables better reflect the conditions experienced by winter wheat. A further limitation lies in the purely data-driven nature of our approach, which constrains its ability to extrapolate to conditions not represented in the training data. This restricts forecast skill in years with unprecedented weather extremes or complex yield responses. Since additional historical data from the MARS archive is not available, future work could explore hybrid modeling frameworks that integrate expert knowledge into the data-driven framework. Such an approach could help the model recognize and adjust for rare but agronomically plausible scenarios, thereby improving robustness in the face of extreme or novel conditions. An additional factor not considered in this study is the role of farmer behavior. Evidence from the U.S. Great Plains suggests that winter wheat producers use climate forecasts for short-term management decisions with lead times of 0–2.5 months (Klemm and McPherson, 2018). However, these findings are difficult to transfer to European wheat production, where policy environments and management practices vary across and within member states. Since our framework targets the month of May, key management decisions such as planting, variety selection, and early-season fertilization have already been completed. The degree to which European farmers incorporate weather forecasts into their decisions earlier in the growing season, and how this might feed back into yield outcomes and prediction models, should be a direction for future research.

5. Conclusion

This study showed that extending the MARS system with subseasonal weather forecasts can enhance early-season wheat yield forecasts across several EU countries. The MARS+Forecast framework incorporates four-week rainfall and temperature outlooks into operational forecasting and delivers measurable improvements in areas covering over one-third of EU wheat production. Although extreme yield failures remain challenging to forecast, progress can be made through longer training records, continued improvements in MARS and advances in subseasonal rainfall and temperature forecasting.

CRediT authorship contribution statement

Maximilian Zachow: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ivana Aleksovska:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Riccardo Henin:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Harald Kunstmann:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Michele Meroni:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Lorenzo Seguini:** Validation, Resources, Methodology, Data curation, Conceptualization. **Elena Tarnavsky:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Senthold Asseng:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2026.111139.

Data availability

Data will be made available on request.

References

Büeler, D., Beerli, R., Wernli, H., Grams, C.M., 2020. Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. Q. J. R. Meteorol. Soc. 146, 3675–3694. https://doi.org/10.1002/qj.3866.

Ben-Ari, T., Boé, J., Ciais, P., Lecerf, R., Van der Velde, M., Makowski, D., 2018. Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France. Nat. Commun. 9, 1627. https://doi.org/10.1038/s41467-018-04087-x.

Bento, V.A., Russo, A., Dutra, E., Ribeiro, A.F.S., Gouveia, C.M., Trigo, R.M., 2022. Persistence versus dynamical seasonal forecasts of cereal crop yields. Sci. Rep. 12, 7422. https://doi.org/10.1038/s41598-022-11228-2.

Brown, J.N., Hochman, Z., Holzworth, D., Horan, H., 2018. Seasonal climate forecasts provide more definitive and accurate crop yield predictions. Agric. For. Meteorol. 260–261, 247–254. https://doi.org/10.1016/j.agrformet.2018.06.001.

ECMWF, 2024. ECMWF verification charts [WWWDocument]. URL https://charts.ecmwf.int/(accessed 11.29.24).

EU, 2023. Eurostat [WWWDocument]. URL https://ec.europa.eu/eurostat/data/database (accessed 9.14.23).

FAOStat, 2023. FAOSTAT [WWWDocument]. URL https://www.fao.org/faostat/en/#data/QCL (accessed 4.27.23).

Glossary, 2026. Country codes [WWWDocument]. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Country_codes (accessed 9.22.25).

Hargreaves, G.H., Allen, R.G., 2003. History and evaluation of Hargreaves evapotranspiration equation. J. Irrig. Drain. Eng. 129, 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53).

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems.

Klemm, T., McPherson, R., 2018. Assessing decision timing and seasonal climate forecast needs of winter wheat producers in the south-Central United States. J. Appl. Meteorol. Climatol. 57, 2129–2140. https://doi.org/10.1175/JAMC-D-17-0246.1.

CORINE Land Cover, 2018 (vector/raster 100 m), Europe, 6-yearly [WWW Document], 2020. URL https://land.copernicus.eu/en/products/corine-land-cover/clc2018 (accessed 9.22.25).

MARS, 2026. JRC MARS Bulletin Archive [WWWDocument]. URL https://agri4cast.jrc.ec.europa.eu/bulletinsarchive (accessed 9.22.25).

Martínez-Ferrer, L., Piles, M., Camps-Valls, G., 2021. Crop yield estimation and interpretability with gaussian processes. IEEE Geosci. Remote Sens. Lett. 18, 2043–2047. https://doi.org/10.1109/LGRS.2020.3016140.

Meroni, M., Waldner, F., Seguini, L., Kerdiles, H., Rembold, F., 2021. Yield forecasting with machine learning and small data: what gains for grains? Agric. For. Meteorol. 308–309, 108555. https://doi.org/10.1016/j.agrformet.2021.108555.

Monhart, S., Zappa, M., Spirig, C., Schär, C., Bogner, K., 2019. Subseasonal hydrometeorological ensemble predictions in small- and medium-sized mountainous catchments: benefits of the NWP approach. Hydrol. Earth Syst. Sci. 23, 493–513. https://doi.org/10.5194/hess-23-493-2019.

Rasmussen, C.E., Williams, C.K.I., 2008. Gaussian processes for machine learning, 3. print. ed Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. 111, 3268–3273. https://doi.org/10.1073/pnas.1222463110.

Santini, M., Noce, S., Antonelli, M., Caporaso, L., 2022. Complex drought patterns robustly explain global yield loss for major crops. Sci. Rep. 12, 5792. https://doi.org/10.1038/s41598-022-09611-0.

van der Velde, M., Nisini, L., 2019. 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. Agric. Syst. 168, 203–212. https://doi.org/10.1016/j.agsy.2018.06.009.

van der Velde, M., van Diepen, C.A., Baruth, B., 2019. The European crop monitoring and yield forecasting system: celebrating 25 years of JRC MARS bulletins. Agric. Syst. 168, 56–57. https://doi.org/10.1016/j.agsy.2018.10.003.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. https://doi.org/10.1175/2009JCLI2909.1.

Zachow, M., Kunstmann, H., Miralles, D.J., Asseng, S., 2024. Multi-model ensembles for regional and national wheat yield forecasts in Argentina. Environ. Res. Lett. 19, 084037. https://doi.org/10.1088/1748-9326/ad627c.