

Multi-model ensembles for regional and national wheat yield forecasts in Argentina

Maximilian Zachow, Harald Kunstmann, Daniel Julio Miralles,
Senthold Asseng

Angaben zur Veröffentlichung / Publication details:

Zachow, Maximilian, Harald Kunstmann, Daniel Julio Miralles, and Senthold Asseng. 2024. "Multi-model ensembles for regional and national wheat yield forecasts in Argentina." *Environmental Research Letters* 19 (8): 084037.
<https://doi.org/10.1088/1748-9326/ad627c>.

LETTER • OPEN ACCESS

Multi-model ensembles for regional and national wheat yield forecasts in Argentina

To cite this article: Maximilian Zachow *et al* 2024 *Environ. Res. Lett.* **19** 084037

View the [article online](#) for updates and enhancements.

You may also like

- [Large-scale reflective optical Janus color materials](#)
Biao Wu, Zhengqi Liu, Xiaoshan Liu et al.
- [Multiband tangent space mapping and feature selection for classification of EEG during motor imagery](#)
Md Rabiul Islam, Toshihisa Tanaka and Md Khademul Islam Molla
- [The influence of different types of SCMs on microstructure and macroscopic properties of cementitious materials](#)
Katalin Kopecskó and Lilla Mlinárik

Breath Biopsy Conference

BREATH BIOPSY[®]

Join the conference to explore the **latest challenges** and advances in **breath research**, you could even **present your latest work!**



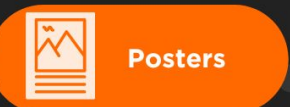
5th & 6th November
Online



Main talks



Early career sessions



Posters

Register now for free!

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Multi-model ensembles for regional and national wheat yield forecasts in Argentina

OPEN ACCESS

RECEIVED

28 February 2024

REVISED

24 June 2024

ACCEPTED FOR PUBLICATION


12 July 2024

PUBLISHED

23 July 2024

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Maximilian Zachow¹ , Harald Kunstmann^{2,3}, Daniel Julio Miralles⁴ and Senthold Asseng^{1,*}¹ Technical University of Munich, Department of Life Science Engineering, Digital Agriculture, HEF World Agricultural Systems Center, Freising, Germany² Institute of Geography, University of Augsburg, Augsburg, Germany³ Institute of Meteorology and Climate Research (IMK-IFU), Karlsruhe Institute of Technology, Campus Alpin, Garmisch-Partenkirchen, Germany⁴ School of Agriculture, Department Plant Production, IFEVA-CONICET, University of Buenos Aires, Buenos Aires, Argentina

* Author to whom any correspondence should be addressed.

E-mail: senthold.asseng@tum.de**Keywords:** seasonal climate model, agricultural monitoring systems, statistical model, multi-model ensemble, crop yieldSupplementary material for this article is available [online](#)**Abstract**

While multi-model ensembles (MMEs) of seasonal climate models (SCMs) have been used for crop yield forecasting, there has not been a systematic attempt to select the most skillful SCMs to optimize the performance of a MME and improve in-season yield forecasts. Here, we propose a statistical model to forecast regional and national wheat yield variability from 1993–2016 over the main wheat production area in Argentina. Monthly mean temperature and precipitation from the four months (August–November) before harvest were used as features. The model was validated for end-of-season estimation in December using reanalysis data (ERA) from the European Centre for Medium-Range Weather Forecasts (ECMWF) as well as for in-season forecasts from June to November using a MME of three SCMs from 10 SCMs analyzed. A benchmark model for end-of-season yield estimation using ERA data achieved a R^2 of 0.33, a root-mean-square error (RMSE) of 9.8% and a receiver operating characteristic (ROC) score of 0.8 on national level. On regional level, the model demonstrated the best estimation accuracy in the northern sub-humid Pampas with a R^2 of 0.5, a RMSE of 12.6% and a ROC score of 0.9. Across all months of initialization, SCMs from the National Centers for Environmental Prediction, the National Center for Atmospheric Research and the Geophysical Fluid Dynamics Laboratory had the highest mean absolute error of forecasted features compared to ERA data. The most skillful in-season wheat yield forecasts were possible with a 3-member-MME, combining data from the SCMs of the ECMWF, the National Aeronautics and Space Administration and the French national meteorological service. This MME forecasted wheat yield on national level at the beginning of November, one month before harvest, with a R^2 of 0.32, a RMSE of 9.9% and a ROC score of 0.7. This approach can be applied to other crops and regions.

1. Introduction

Crop yield is expected to become more variable [1, 2] under advancing climate change [3, 4] with possible consequences for food security [5, 6]. To counteract these disturbances, crop yield forecast models are used [7]. These models assist to prepare humanitarian aid efforts [8] or to negotiate trade agreements [9]. However, yield forecasts that are provided before harvest face unknown weather conditions between

forecast issue date and harvest. When meteorological and vegetation indices are available for machine learning models, this gap of unknown weather is usually ignored [10]. Alternatively, statistical [11, 12] or dynamical seasonal climate models (SCMs) [13–15] are used. SCMs with prediction ranges of several months have become increasingly skillful [16] and can help anticipating unexpected weather. In addition, Multi-Model-Ensembles (MMEs) of individual SCMs have improved weather forecasts [17, 18] and

crop yield forecast [15, 19–21]. The MMEs used are usually provided ‘out-of-the-box’ from climate centers such as the APCC-MME [20, 21] or the MMEs are built without analyzing the accuracy of the participating SCMs [15]. In the field of crop yield forecasting, no attempt has been made yet to systematically analyze SCMs to build a more accurate MME. SCM forecast skill varies by region, forecasted quantity, lead time, spatial and temporal aggregation and forecast issue month [22, 23]. Here, we propose a linear regression model to forecast regional and national crop yield variability with a MME that is built specifically for a model use case by analyzing ten SCMs [18]. Other assessment studies on the seasonal forecast performance over South America indicated moderate to poor skill for the Argentinian Pampas [23–25], one of the most important agricultural region in the world. Given that Argentina is a major wheat producer and exporter in South America, contributing around 60%–70% of total South American wheat grain output [26, 27], we use Argentina as a test country and wheat as the model crop to analyze if seasonal forecast performance can be improved and used within a crop yield model when a MME is built systematically. Like previous studies [6, 10, 28–30], we assess the approach on regional level across a wide range of climate conditions and on national level. Like several other countries in the Southern Cone, Argentina’s cereal crops have been negatively affected by climate change. The rise in average daily temperature, mostly due to higher nighttime temperatures, accelerates crop phenology and diminishes yields [31]. In that context, wheat yield forecasts are crucial for Argentina. In our proposed approach, we have validated both the end-of-season estimation case using the European Centre for Medium-Range Weather Forecasts Reanalysis data (ERA) and the in-season forecast case using a MME.

2. Material and methods

The data flow of this study is shown in figure 1, including indications on the corresponding sections with further information. Section 2.1 explains the preparation of historical wheat data from Argentina. Sections 2.2 and 2.3 summarize the preprocessing steps of ERA and SCM data. Finally, section 2.4 explains the underlying model training and the validation procedure for both end-of-season (ERA) and in-season (MME) applications.

2.1. Region of study and historical wheat data

In Argentina, wheat cultivation is mainly rainfed [32] and the season spans from June to December [33]. Around 95% of wheat is produced in the five provinces Santa Fe, Córdoba, Entre Ríos, Buenos Aires and La Pampa [34]. From these provinces, historical wheat data was collected from 182 municipalities through the Ministry of Agriculture of

Argentina [34]. During data preparation, municipalities with incomplete time series from 1993–2016 were dropped (142 municipalities remaining) and assigned to regions. The regions used in this study were obtained from the ‘Map of Argentinian wheat sub-regions and other winter cereals’ [35]. In this map, municipalities form a region based on similar soil properties as well as temperature and precipitation during the winter cropping season from April to December. The twelve regions analyzed in this study are depicted in figure 2(a). Municipality yield was aggregated to regional (supplementary figure S1) and national level (black line, figure 2(b)) using a weighted average with weights being the harvested area.

Yield anomalies for each region and on national level is computed as follows using absolute yield:

$$\text{anomaly}_t = \frac{\text{absolute}_t - \overline{\text{absolute}_{t-2:t+2}}}{\overline{\text{absolute}_{t-2:t+2}}} * 100. \quad (1)$$

Yield anomaly for year t (anomaly_t) was calculated by subtracting the 5 year-moving average of absolute yields from years $t - 2$ to $t + 2$ from the current year’s absolute yield (absolute_t) and dividing the difference by the 5 year moving average and multiply the result by 100 to obtain a percentage. The moving average approach to estimate expected yield has been proposed by Iizumi *et al* on crop yield variability forecasts [21]. National wheat yield anomalies are depicted in figure 2(c). Regional absolute yield as well as the trend and anomalies can be found in the supplementary materials (supplementary figures S1 and S2).

2.2. Temperature and precipitation data

Wheat yield (grains per unit area) is determined by two numerical components: Spikes per unit area and grains per spike. There is an important plasticity of these numerical yield components depending on the environmental conditions to which wheat is exposed. For example, a greater initiation of tillers prior to stem elongation (August–September) that results in greater establishment of spikes will in part be offset by a reduction in the number of fertile florets (grains) per spike at anthesis (October–November) [36]. To account for the climate impact on yield formation, we used monthly mean temperature and precipitation from the last four months before harvest (August–November) [31]. Higher resolution climate data were not available for SCMs coming from the North-American-Multi-Model Ensemble (NMME) (see section 2.3). In addition, Asseng *et al* [37] and Wall *et al* [38] have shown a robust linear relationship between wheat yield and seasonal mean temperature. Other crop yield models for Argentina used additional variables like soil moisture [30], large-scale climate indices like the El Niño–Southern Oscillation index [21, 39] or remotely sensed quantities like land

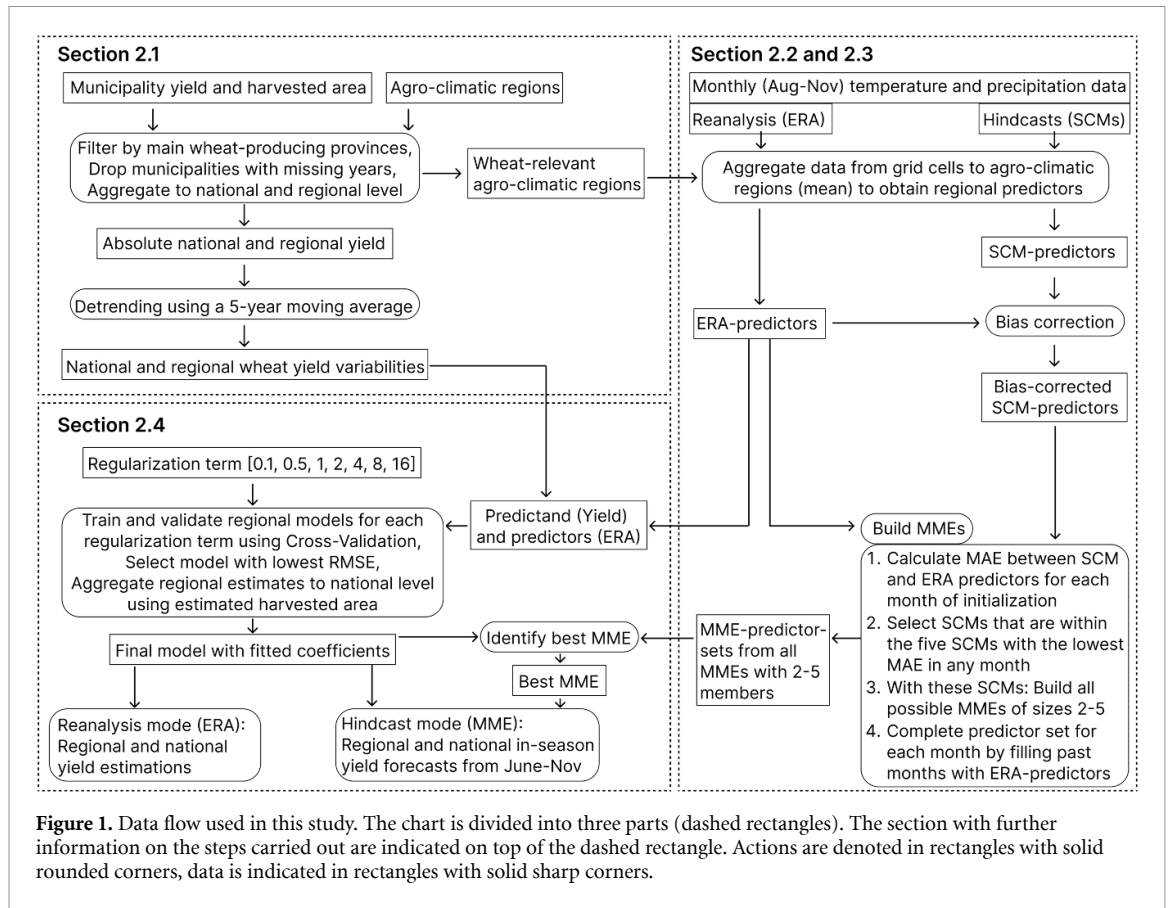


Figure 1. Data flow used in this study. The chart is divided into three parts (dashed rectangles). The section with further information on the steps carried out are indicated on top of the dashed rectangle. Actions are denoted in rectangles with solid rounded corners, data is indicated in rectangles with solid sharp corners.

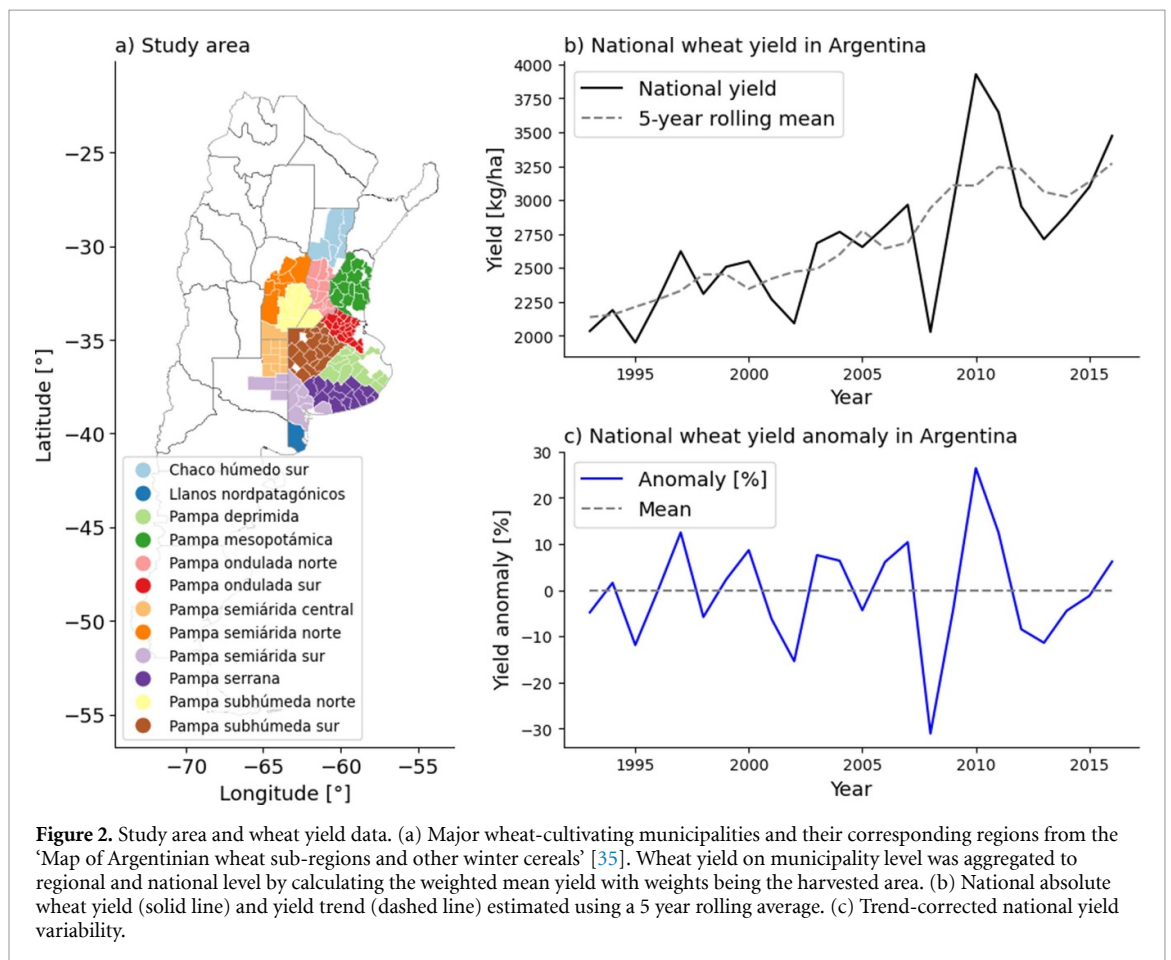
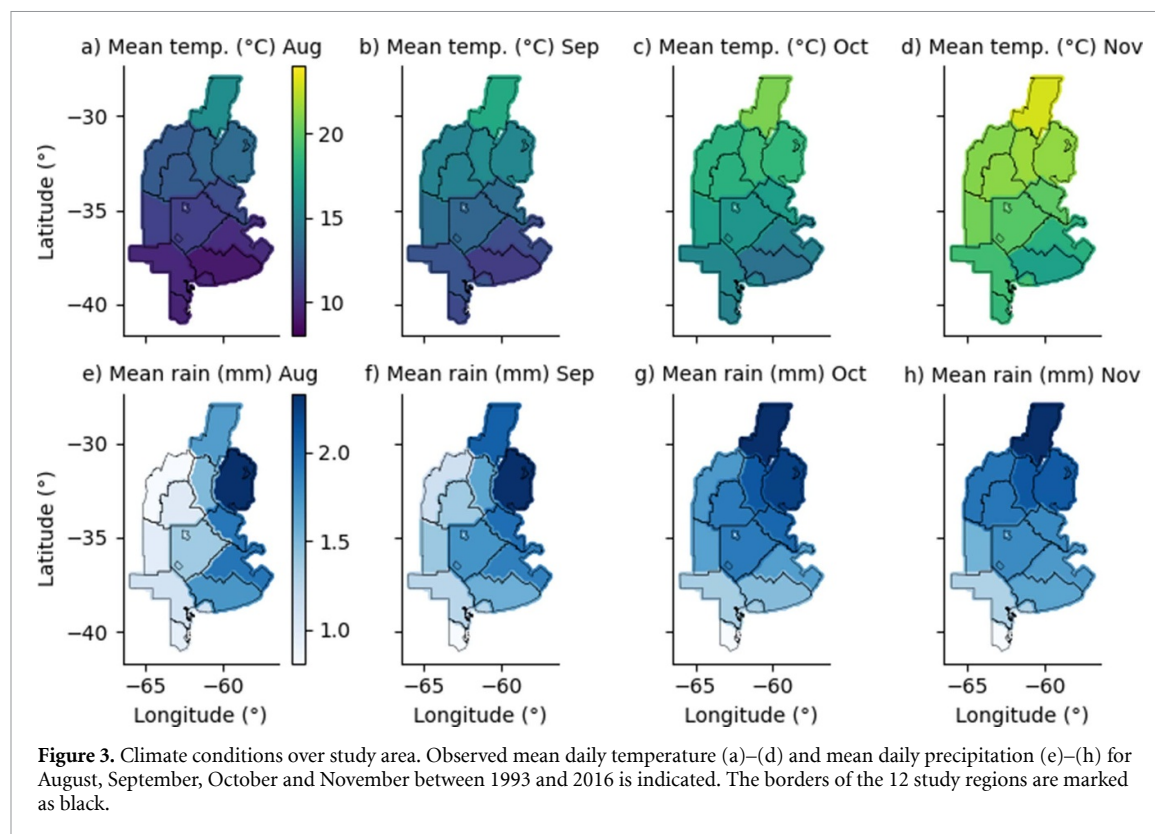


Figure 2. Study area and wheat yield data. (a) Major wheat-cultivating municipalities and their corresponding regions from the ‘Map of Argentinian wheat sub-regions and other winter cereals’ [35]. Wheat yield on municipality level was aggregated to regional and national level by calculating the weighted mean yield with weights being the harvested area. (b) National absolute wheat yield (solid line) and yield trend (dashed line) estimated using a 5 year rolling average. (c) Trend-corrected national yield variability.



surface temperature and vegetation indices [28, 40]. Our objective here is an isolated analysis of the in-season forecast accuracy coming from a MME (see section 2.3). Therefore, we decided to solely use meteorological indices [13, 20]. The in-season forecasts are benchmarked with the end-of season estimation, when all indices are available as observations and crop yield can be estimated under full knowledge. The data for the end-of-season estimation was gathered from the ERA5-Land reanalysis dataset (ERA) for August to November from 1993–2016 [41], which was available at 0.1° spatial resolution and was subsequently aggregated to our study regions. The monthly mean climate conditions from 1993–2016 per variable and region are indicated in figure 3. In addition to providing the end-of-season benchmark, ERA data was used for model training (finding the coefficients) for the in-season forecast validation (see section 2.4) using MME data.

2.3. Seasonal climate forecasts

To provide wheat yield forecasts before harvest, forecasted monthly temperature and precipitation data were obtained as hindcasts (retrospective forecasts) from SCMs. Three platforms that offer access to hindcasts of multiple SCMs are: The Copernicus Climate Data Store from the Copernicus Climate Change Service (C3S) [42], the Data Library of the NMME [43] and the Climate Information Toolkit of the Asia-Pacific Climate Center (APCC) [44]. Hindcasts from APCC were generally not considered, because of their

sparse spatial resolution of 2.5° . The NMME consists of six SCMs and the C3S offers access to eight SCMs, however some SCMs were available on both platforms and others had missing data for some years. Hence, we chose 10 SCMs (table 1). For each SCM, hindcasts from 1993–2016 across the study region at a resolution of 1.0° were collected. The study period was chosen, as it was the largest common period where hindcasts were available for all 10 SCMs. To validate the in-season performance of the wheat yield model, we used hindcasts that were initialized at the beginning (within the first week) of each month from June to November throughout the cropping season with forecasts for the months August–November. The highest temporal resolution across all SCMs were monthly mean data and therefore used across all applications. At each month of initialization (from June to November), hindcasts of monthly mean temperature and precipitation for August, September, October and November were obtained, when these months were in the future. Each SCM provides an ensemble of hindcasts that are run with different initial conditions [45]. Here, we took the ensemble mean to only have one output per SCM.

The gridded hindcasts were then aggregated to the twelve study regions. Next, regional hindcasts were bias-adjusted with ERA data. For temperature data, scaled distribution mapping [53] was used for bias-adjustment and for precipitation data, linear scaling was applied [54]. The structural errors of individual SCMs in forecasting weather conditions can be

Table 1. Seasonal climate models used in this study. From each model, monthly hindcasts were generated from June to November to forecast monthly mean temperature and rainfall for August, September, October and November during 1993–2016. In addition to these seasonal climate models, reanalysis data from the European for Medium-Range Weather Forecasts (ERA) provided by the Copernicus Climate Data Store was used for model training, end-of-season estimation and to complete the feature vector for in-season forecasts e.g. to provide August and September data for forecasts made in beginning of October. The seasonal climate models used in this study were all obtained in grid cell resolutions of 1° .

ID	Abbreviation	Country	System name	Source
1	CMCC	Italy	SPS3.5	C3S [46]
2	DWD	Germany	GCFS2.1	C3S [47]
3	ECCC	Canada	CANSIPsv2.1	C3S [48]
4	ECMWF	European Union	SEAS5	C3S [49]
5	METFR	France	SYS8	C3S [50]
6	UKMO	United Kingdom	GloSea6	C3S [51]
7	NCAR	United States of America	CCSM4	NMME [52]
8	GFDL	United States of America	CM2.5	NMME [52]
9	NASA	United States of America	GEOS5	NMME [52]
10	NCEP	United States of America	CFSv2	NMME [52]

reduced by combining their outputs to MMEs [17]. Smaller MMEs with data from the best (maximum five) available SCMs outperform larger MMEs [18]. To identify the best SCMs, we calculated the mean absolute error (MAE) between SCM and ERA predictors across all variables and regions, separately for each month of initialization. Then, the ten SCMs are filtered and only those are kept that are at least in one month within the top five SCMs [18] with the lowest MAE. Those SCMs that are not ranked within the top five SCMs in any month are discarded. With the remaining SCMs all possible MMEs of sizes 2–5 were built, e.g. for MME size 2, all possible SCM-pairs out of the best SCMs were built, for MME size 3, all possible MMEs with three SCMs were built and so on. To avoid spatial and temporal overfitting, we decided to validate the same MME for all regions and months of initialization. The final MME was selected based on the coefficient of determination (R^2) of national yield forecasts. If R^2 were the same, the selection was made using root-mean-square error (RMSE) (see section 2.4). Since hindcasts are retrospective forecasts and our ultimate technical objective are in-season yield forecasts, we will align the terminology and refer to hindcasts as forecasts from now on to make the text clearer and easier to read.

2.4. Statistical model and analysis

Regional wheat yield variability $\hat{y}_{\text{regional}}$ was modeled as a linear combination of the eight available predictors x_1, \dots, x_8 (monthly mean temperature and precipitation from August to November) multiplied with their corresponding learned coefficients β_1, \dots, β_8 [2] plus an intercept β_0 . We determined the coefficients using a ridge estimator [3] to avoid overfitting [55].

$$\hat{y}_{\text{regional}} = \beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8 \quad (2)$$

$$\operatorname{argmin}_{\beta} \sum_{i=1}^N (y_i - \beta' x_i)^2 + \alpha \sum_{j=1}^8 \beta_j^2. \quad (3)$$

In ridge regression, coefficients are found by minimizing the sum of squared errors between estimated and actual yield anomaly ($\sum_{i=1}^N (y_i - \beta' x_i)^2$) and the sum of squared coefficients ($\alpha \sum_{j=1}^8 \beta_j^2$), with α being a parameter that controls the strength of the coefficient regularization. To find the optimal α , we used Leave-One-Year-Out Cross-Validation (LOOCV). For each α , regional models were trained and validated (using ERA data) and the average performance metrics across all regions were calculated. The statistics used in this study are the RMSE, R^2 and the MAE [56]. The model with the lowest RMSE across all regions was selected as the final model. In addition to the regression analysis, we validated how well our model discriminates between positive and negative yield anomalies using the Receiver Operating Characteristic (ROC) [57]. ROC scores above 0.6 indicate skill that is better than random guessing with 1 being a perfect model [58]. For the in-season forecasts, the final model was validated with MME data. To complete the feature set for yield forecasts made at, e.g. the beginning of October, ERA data was used for August and September and MME data was used for October and November. Yield estimations and forecasts on regional level were aggregated to national level by using estimated harvested area as weights. Actual harvested area is unknown before harvest, so harvested area for each year was estimated as the mean harvested area from all other years.

3. Results

The best performing end-of-season model, trained and validated with ERA features, was found with a ridge regression regularization term α of 8. The average regional and national estimation performances of the final model are given in figure 4. The end-of-season model achieves a R^2 of 0.24, a RMSE of 17.7% and a ROC score of 0.65 on regional level (figure 4(a)), validated via LOOCV over all 24 years

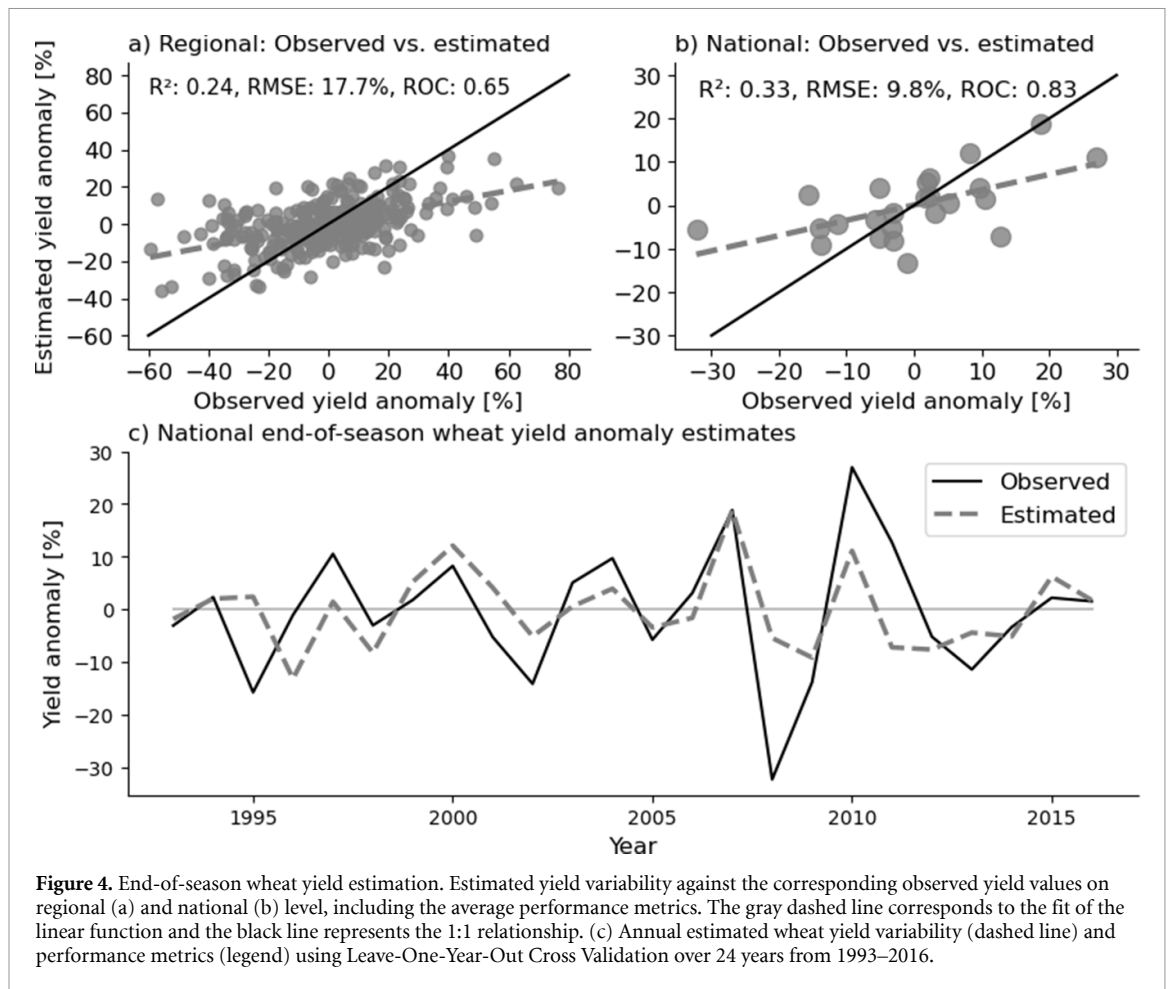


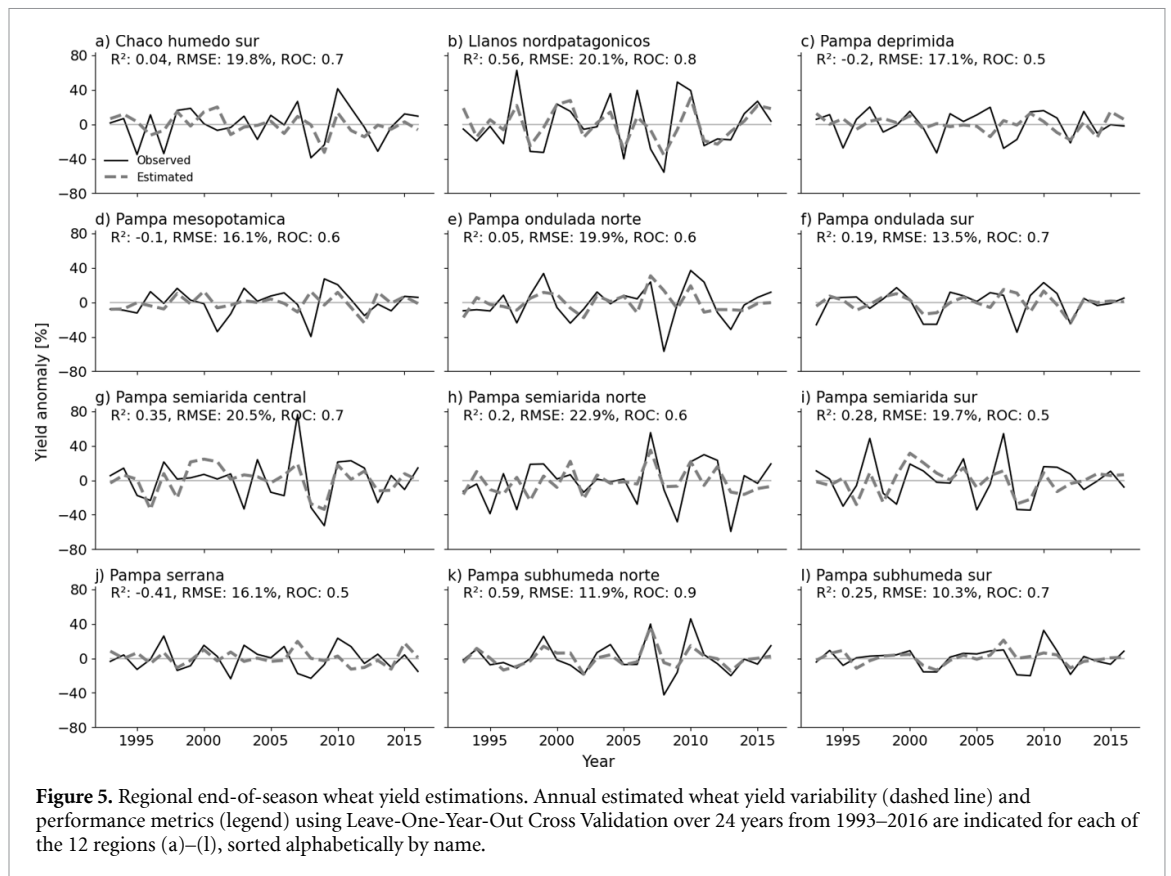
Figure 4. End-of-season wheat yield estimation. Estimated yield variability against the corresponding observed yield values on regional (a) and national (b) level, including the average performance metrics. The gray dashed line corresponds to the fit of the linear function and the black line represents the 1:1 relationship. (c) Annual estimated wheat yield variability (dashed line) and performance metrics (legend) using Leave-One-Year-Out Cross Validation over 24 years from 1993–2016.

from 1993 to 2016. National wheat yield estimations were possible with R^2 of 0.33, RMSE of 9.8% and a ROC score of 0.83 (figure 4(b)). The fitted line through the data points of annual observed yield variability and the corresponding estimated yield variability (gray dashed slope, figures 4(a) and (b)) is less steep than the black identity line. This indicates that both regional and national estimates have a tendency of underestimating the extent of extreme years (yield loss and yield gain). This tendency can also be seen in figure 4(c), where the annual yield anomaly estimates (gray dashed line) are plotted against the actual observed estimates (black solid line). The national estimation model discriminates well between positive and negative yield anomalies (ROC score of 0.8, figure 4(b)) but e.g. in the two most extreme year 2008 (yield loss) and 2010 (yield gain), it failed to estimate the magnitude (figure 4(c)).

A regional analysis of the end-of-season estimation model is given in figure 5. The twelve regions are sorted alphabetically and their annual wheat yield estimations (gray dashed line) are shown alongside the corresponding actual yield anomaly (black solid line) (figures 5(a)–(l)). In Pampa deprimida (figure 5(c)), Pampa mesopotamica (figure 5(d)) and Pampa serrana (figure 5(j)), skillful end-of-season

estimations are not possible, indicated with a negative R^2 (trivial approach of estimating mean yield would perform better). In addition, Pampa ondulada norte (figure 5(e)), Pampa semiardida norte (figure 5(h)) and Pampa semiardida sur (figure 5(i)) have ROC scores of 0.6 or lower, indicating poor skill in discriminating between positive and negative yield anomalies. These six regions with either $R^2 \leq 0$ or $ROC \leq 0.6$ are not considered for the subsequent analysis of the in-season yield forecast model using MME data, since they already failed to provide skillful estimates with ERA data. The best wheat yield estimations are achieved in Llanos nord-patagonicos (figure 5(b)) and in Pampas subhmeda norte (figure 5(k)) with R^2 of 0.56 and 0.59, RMSE of 20.1% and 11.9% and ROC scores of 0.8 and 0.9, respectively. In terms of spatial distribution, wheat yield estimates are generally more skillful in the western regions (supplementary figure S3).

To forecast wheat yield variability in-season before the predictors are available as observations, forecasted features derived from SCM data were used. For each month of initialization from June [6] to November [11], the MAE across all years and regions between forecasted and observed ERA features is shown in figure 6. A spatial analysis of the



regional forecast skill by SCM and month of initialization can be found in supplementary figure S4. Up until September, the performance among the SCMs is mixed with no clear trend of improvement (figure 6). In October and November, most SCMs become more accurate. In November, all SCMs perform best across all months of initialization. The five best SCMs with the lowest MAE per month of initialization are indicated with a superscript asterisk. The best SCMs in June are ECCC, NASA, CMCC, UKMO and DWD. From July to October, the five best SCMs are always CMCC, DWD, ECCC, ECMWF and ECMWF. In November, the best SCMs are CMCC, ECCC, ECMWF, METFR and UKMO. The count of how often each SCM is within the top five is indicated at the right next to the heatmap. The three SCMs from NCAR, GFDL and NCEP are never within the best SCMs and are not considered for the MME analysis. The remaining seven SCMs were used to build all possible MMEs of sizes 2–5. None of these 120 MMEs makes skillful yield forecasts on national level before November (supplementary table S1). The best forecasts (measured as R^2 and RMSE) were achieved with a 3-member-MME using the unweighted average of the forecasted features from ECMWF, NASA and METFR (supplementary table S1 and S2). This MME also outperforms the yield forecasts made with individual SCM data (supplementary table S1 and S2). Average regional in-season wheat yield forecast quality of this

3-member-MME is shown in figure 7. Although forecasts improve from June (figure 7(a)) to November (figure 7(f)) with lower RMSE, higher R^2 and higher ROC scores, skillful yield forecasts are never possible, indicated through a $R^2 \leq 0$ and a ROC score ≤ 0.6 for all months (figures 7(a)–(f)). Forecasts generally fail to anticipate the extent of extreme years, depicted through the gray dashed slopes that are more flat than the 1:1 identity lines. A regional analysis of the forecasts quality reveals that skillful forecasts are possible in November for the two regions Llanos nordpatagónicos and Pampa semiárida sur and in October for Pampa semiárida central (supplementary figure S5, supplementary table S3).

In-season wheat yield forecast quality on national level is shown in figure 8, where forecasts are compared with observed yield anomalies throughout the cropping season. From June (figure 8(a)) until October (figure 8(e)), national wheat yield forecasts are not skillful, depicted through a $R^2 \leq 0$ and a ROC score ≤ 0.6 . Both R^2 and RMSE improve throughout the season. In November, the 3-member MME (with ECMWF, NASA and METFR data) makes skillful national wheat yield forecasts with a R^2 of 0.32, RMSE of 9.9% and a ROC score of 0.7 (figure 8(f)). In November, forecast skill has approached the end-of-season estimation skill (figure 4(a)). The approach could not forecast the extent of extreme low yields (e.g. 2002, 2008) but anticipated correctly a negative yield event in these years.

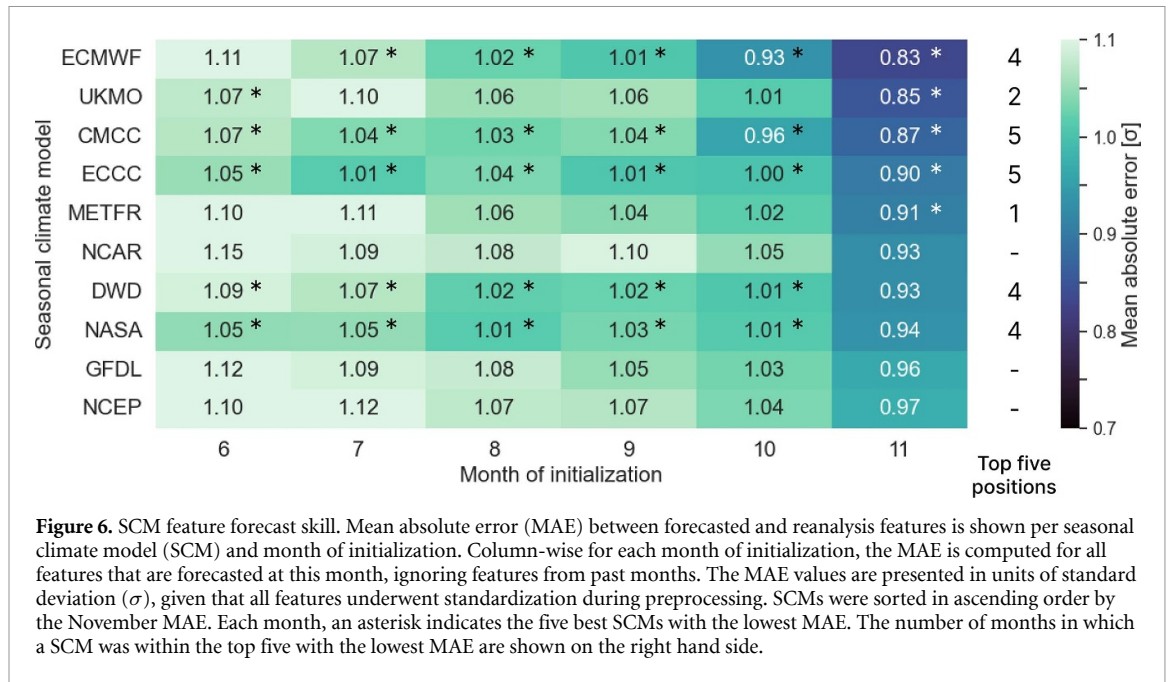


Figure 6. SCM feature forecast skill. Mean absolute error (MAE) between forecasted and reanalysis features is shown per seasonal climate model (SCM) and month of initialization. Column-wise for each month of initialization, the MAE is computed for all features that are forecasted at this month, ignoring features from past months. The MAE values are presented in units of standard deviation (σ), given that all features underwent standardization during preprocessing. SCMs were sorted in ascending order by the November MAE. Each month, an asterisk indicates the five best SCMs with the lowest MAE. The number of months in which a SCM was within the top five with the lowest MAE are shown on the right hand side.

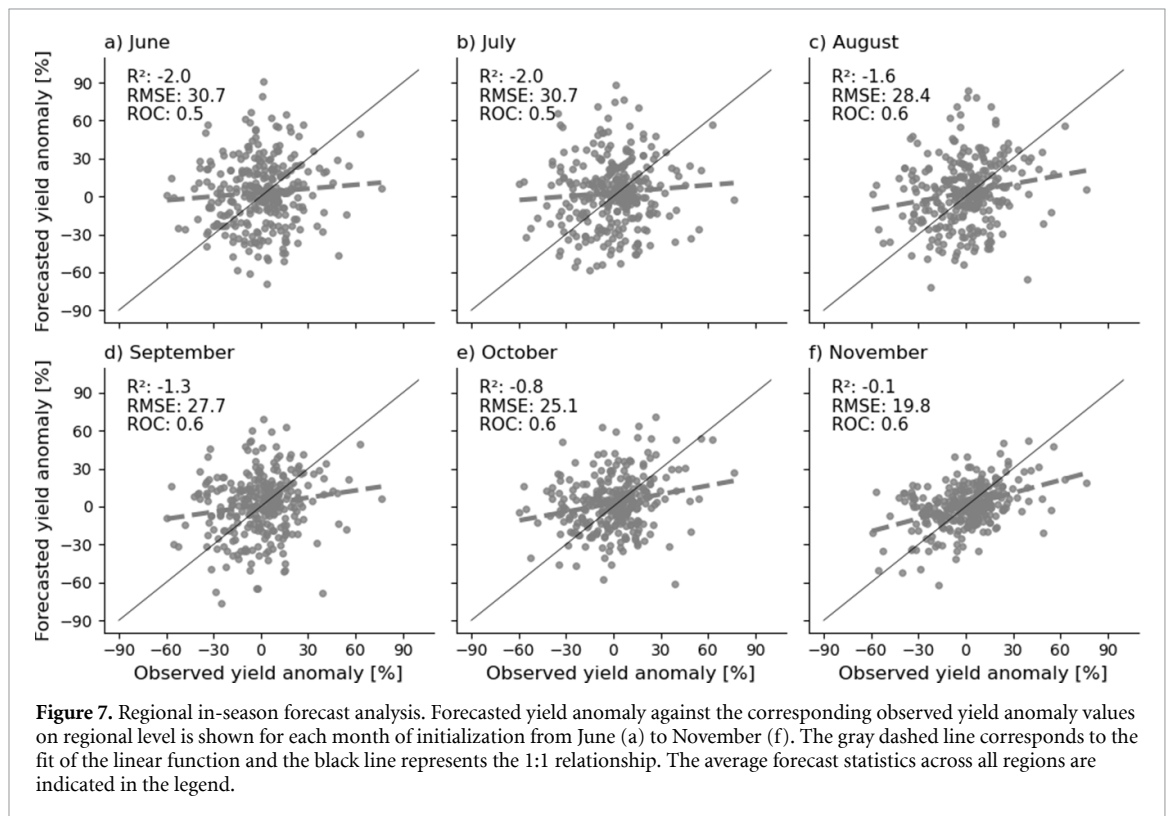


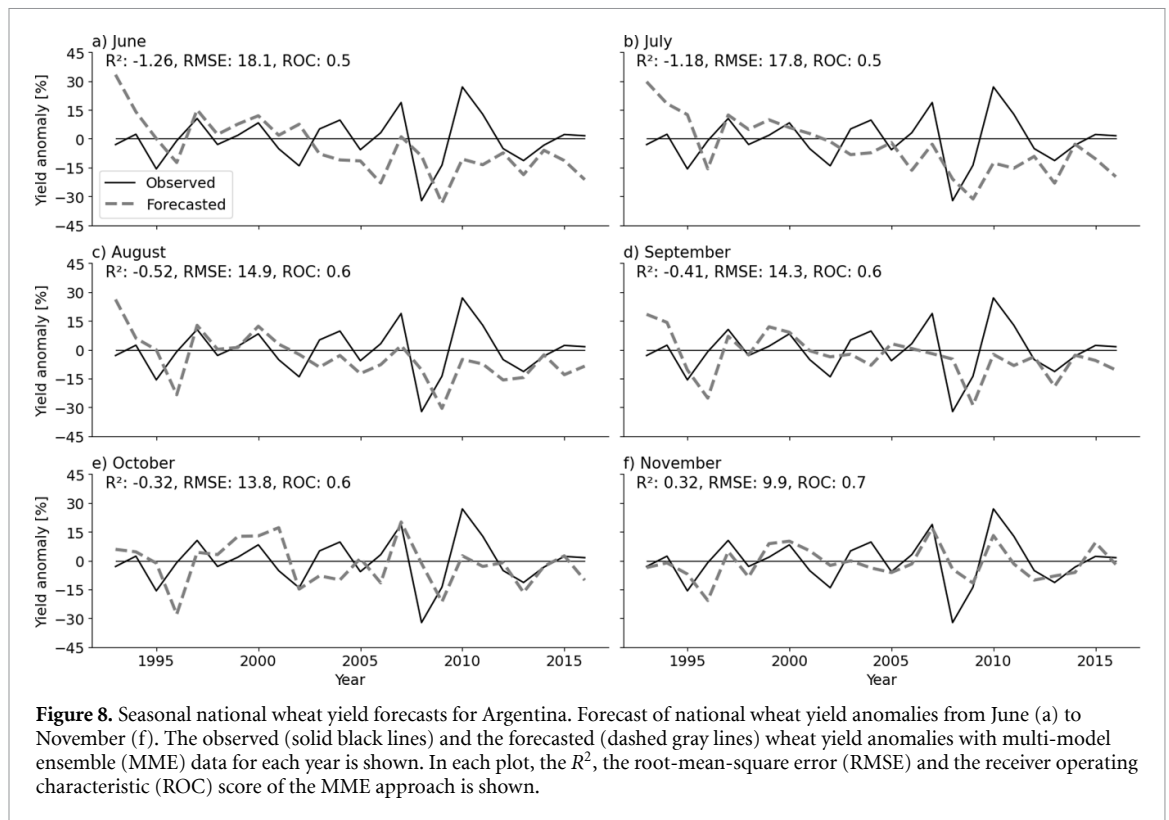
Figure 7. Regional in-season forecast analysis. Forecasted yield anomaly against the corresponding observed yield anomaly values on regional level is shown for each month of initialization from June (a) to November (f). The gray dashed line corresponds to the fit of the linear function and the black line represents the 1:1 relationship. The average forecast statistics across all regions are indicated in the legend.

4. Discussion

We proposed an approach of forecasting regional and national wheat yield variability in Argentina before harvest using monthly meteorological indices from the last four months before harvest (August–November) with a 3-member-MME consisting of ECMWF, NASA and METFR data. This is the first crop model study that compared as many as 10 SCMs to build a more accurate MME for in-season yield forecasts. On national level, the suggested approach

performed poorly from the beginning of the season in June until October, with $R^2 \leq 0$ or a ROC score ≤ 0.6 . However, at the beginning of November, one month before harvest begins, national yield forecasts were possible with R^2 of 0.32, RMSE of 9.9% and a ROC score of 0.7. On regional level, the approach performed poorly with skillful forecasts in only two regions in November and one region for October and November.

In Argentina, grains per unit area are mostly defined during the critical period from October to



November, through the establishment of the number of spikes per unit area and the definition of the number of fertile florets per spike at anthesis. As the meteorological indices for November are more accurately forecasted from the MME towards the end of the season, the crop yield model makes better forecasts with an increased R^2 and a lower RMSE. The end-of-season model that uses ERA features instead of forecasted features from SCMs achieved a R^2 of 0.33, a RMSE of 9.8% and a ROC score of 0.83 on national level and a R^2 of 0.24, a RMSE of 17.7% and ROC score of 0.65 on regional level. Our approach suggests that large MMEs are not needed to provide more accurate national in-season wheat yield forecasts and is in alignment with the results from climate studies [18]. One possible explanation on why the best MME found in our study was a three-member-MME with data from ECMWF, NASA and METFR is that their forecast errors have lower model-to-model correlations than other SCMs (supplementary figure S6). In addition, only using weather conditions from the last four months before harvest as predictors were sufficient to discriminate between positive and negative yield anomalies one month before harvest. As outlined above, this could be because the most important numerical yield component, the number of grains per unit area, is defined during the final months before harvest. It is important to clarify that another numerical yield component, grain weight, is also defined during the last part of the critical period, however, the ability to explain the variations in yield of this component is much more limited

than that of the number of grains per unit area [36, 59]. Wheat yield estimation and forecast performance increased when aggregated from regional to national level. This has been observed in similar studies [10] with a possible explanation being that yield on regional level is overestimated in some regions and underestimated in others and through the aggregation the effects were reduced and averaged out. For national November forecasts, we analyzed model uncertainty, which varies among seasons, but is in general relatively small based on the three MME members used in the study (figure S7(a)). However, little uncertainty, i.e. small differences between the forecasts of the two models, does not necessarily indicate high precision, i.e. a small error in yield forecasting (figure S7(b)). Based on hindcasts, our results of 9.8% RMSE for national end-of-season-estimates, 9.9% RMSE for national in-season November forecasts and 19.8% RMSE for regional in-season November forecasts are comparable to another study by Franch *et al* for Argentinian wheat yield with 11%, 12.4% and 22.5% RMSE, respectively [28]. The performance difference may be explained by the fact that Franch *et al* [28] employed a statistical method to forecast future vegetation indices for in-season yield forecasts, whereas our approach is based on forecasted weather predictors from dynamical SCMs. Furthermore, their subnational approach uses administrative units, while we use the regions from the 'Map of Argentinian wheat sub-regions and other winter cereals' [35] that are based on climate and soil conditions.

While the results are promising and encourage an application to other crops and regions, there are some limitations and obstacles that require further attention. To have a consistent dataset, we decided to use reanalysis data (ERA) which is available in the same gridded format like the SCM data [13, 21]. It has been shown that reanalysis data can differ from *in situ* observations, especially in mountainous regions [60, 61]. Although our study region is defined by flat terrain across the Argentinian Pampas, differences between ERA data and actual ground conditions may have affected the model performance. The MAE was calculated over the entire 24 years of the study period. In the validation setting, however, when forecasting for a specific year, the MAE would only be known for all other 23 years except the current year. For example, in September, the error between forecasted and ERA precipitation for November is not known because the ERA data is not available yet and hence, SCMs cannot be selected on the basis of this year. We chose this approach to show the potential in case the true accuracy of all SCMs would be known for the target use case. In an operational setting, we recommend to calculate the MAE and to identify the best SCMs using all available past years, where a true performance ranking is likely to converge to the same SCMs. Furthermore, the SCMs from the NMME were only available at monthly resolution, which restricted the introduction of additional day-based indices, such as killing degree days [62]. Future studies may attempt to include even more SCMs that are currently certified by the World Meteorological Organization [63] than the ones studied here. Future work may use our findings to propose a probabilistic yield forecast model [64] instead of the deterministic regression approach chosen here. This would facilitate a quantification of reliability, which is crucial for transparent communication and for the interpretation of the outputs [65]. One possible approach to achieve this would be the calculation of yield forecasts with each ensemble member from the individual SCMs to obtain a yield forecast range [64]. We used the ensembles mean of the SCMs to obtain one output per SCMs which may have caused a loss of signal [18]. In addition, since each ensemble member is a physically plausible forecast, averaging these forecasts leads to an output that is not based on the underlying physical processes anymore [18]. Better yield forecasts may also be possible if non-linear machine learning models are used, such as Random Forest [66]. Lastly, vegetation indices can be used as additional features to weather indices and might lead to improved forecasts [67].

5. Conclusion

In-season wheat yield forecasts can help various stakeholders to optimize their strategies and stabilize food supply [7, 8]. To forecast wheat yield variability in Argentina before harvest, we have shown how

a 3-member-MME can be applied. As medium and long range weather forecasts are expected to further improve in the near future [68, 69], in-season crop yield forecast models that employ weather forecasts can play a crucial role for food security.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://zenodo.org/doi/10.5281/zenodo.12513934>.

Acknowledgments

We acknowledge the agencies that support the NMME-Phase II system, and we thank the climate modeling groups (Environment Canada, NASA, NCAR, NOAA/GFDL, NOAA/NCEP, and University of Miami) for producing and making available their model output. NOAA/NCEP, NOAA/CTB, and NOAA/CPO jointly provided coordinating support and led development of the NMME-Phase II system. The results of our study contain modified Copernicus Climate Change Service information 2023. Neither the European Commission nor CMCC, DWD, ECCC, ECMWF, METFR or UKMO is responsible for any use that may be made of the Copernicus information or data it contains.

ORCID iD

Maximilian Zachow  <https://orcid.org/0000-0002-0525-7900>

References

- [1] Iizumi T and Ramankutty N 2016 Changes in yield variability of major crops for 1981–2010 explained by climate change *Environ. Res. Lett.* **11** 034003
- [2] Liu W, Ye T, Jägermeyr J, Müller C, Chen S, Liu X and Shi P 2021 Future climate change significantly alters interannual wheat yield variability over half of harvested areas *Environ. Res. Lett.* **16** 094045
- [3] Brás T A, Seixas J, Carvalhais N and Jägermeyr J 2021 Severity of drought and heatwave crop losses tripled over the last five decades in Europe *Environ. Res. Lett.* **16** 065012
- [4] Lesk C, Rowhani P and Ramankutty N 2016 Influence of extreme weather disasters on global crop production *Nature* **529** 84–87
- [5] Ben-Ari T, Boé J, Ciaïa P, Lecerc R, Van der Velde M and Makowski D 2018 Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France *Nat. Commun.* **9** 1627
- [6] Júnior R D S N et al 2021 Extreme lows of wheat production in Brazil *Environ. Res. Lett.* **16** 104025
- [7] Basso B and Liu L 2019 Chapter four—seasonal crop yield forecast: methods, applications, and accuracies *Adv. Agron.* **154** 201–55
- [8] Funk C et al 2019 Recognizing the famine early warning systems network: over 30 years of drought early warning science advances and partnerships promoting global food security *Bull. Am. Meteorol. Soc.* **100** 1011–27
- [9] FAS. USDA Foreign Agricultural Service 2018 Grain and feed annual (available at: www.fas.usda.gov/data/brazil-grain-and-feed-annual-3) (Accessed 1 February 2022)

- [10] von Bloh M *et al* 2023 Machine learning for soybean yield forecasting in Brazil *Agric. For. Meteorol.* **341** 109670
- [11] Laudien R, Schauburger B, Makowski D and Gornott C 2020 Robustly forecasting maize yields in Tanzania based on climatic predictors *Sci. Rep.* **10** 19650
- [12] Vijverberg S, Hamed R and Coumou D 2023 Skillful U.S. soy yield forecasts at presowing lead times *Artif. Intell. Earth Syst.* **2** e210009
- [13] Bento V A, Russo A, Dutra E, Ribeiro A F S, Gouveia C M and Trigo R M 2022 Persistence versus dynamical seasonal forecasts of cereal crop yields *Sci. Rep.* **12** 7422
- [14] Peng B, Guan K, Pan M and Li Y 2018 Benefits of seasonal climate prediction and satellite data for forecasting U.S. maize yield *Geophys. Res. Lett.* **45** 9662–71
- [15] Zachow M, Júnior R D S N and Asseng S 2023 Seasonal climate models for national wheat yield forecasts in Brazil *Agric. For. Meteorol.* **342** 109753
- [16] Baker L, Charlton-Perez A and Mattu K L 2023 Skilful sub-seasonal forecasts of aggregated temperature over Europe *Meteorol. Appl.* **30** e2169
- [17] Hagedorn R, Doblas-Reyes F J and Palmer T N 2005 The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept *Tellus A* **57** 219–33
- [18] Knutti R, Furrer R, Tebaldi C, Cermak J and Meehl G A 2010 Challenges in combining projections from multiple climate models *J. Clim.* **23** 2739–58
- [19] Chun J A, Kim S, Lee W S, Oh S M and Lee H 2016 Assessment of multimodel ensemble seasonal hindcasts for satellite-based rice yield prediction *J. Agric. Meteorol.* **72** 107–15
- [20] Iizumi T, Shin Y, Kim W, Kim M and Choi J 2018 Global crop yield forecasting using seasonal climate information from a multi-model ensemble *Clim. Serv.* **11** 13–23
- [21] Iizumi T, Takaya Y, Kim W, Nakaegawa T and Maeda S 2021 Global within-season yield anomaly prediction for major crops derived using seasonal forecasts of large-scale climate indices and regional temperature and precipitation *Weather Forecast.* **36** 285–99
- [22] Duan W and Wei C 2013 The ‘spring predictability barrier’ for ENSO predictions and its possible mechanism: results from a fully coupled model *Int. J. Climatol.* **33** 1280–92
- [23] Gubler S *et al* 2020 Assessment of ECMWF SEAS5 seasonal forecast performance over South America *Weather Forecast.* **35** 561–84
- [24] Ferreira G W S, Reboita M S and Drumond A 2022 Evaluation of ECMWF-SEAS5 seasonal temperature and precipitation predictions over South America *Climate* **10** 128
- [25] Zhang J, Guan K, Fu R, Peng B, Zhao S and Zhuang Y 2023 Evaluating seasonal climate forecasts from dynamical models over South America *J. Hydrometeorol.* **24** 801–14
- [26] CONAB A cultura do trigo 2017 (available at: www.conab.gov.br/institucional/publicacoes/outras-publicacoes) (Accessed 23 January 2022)
- [27] FAOSTAT 2023 Faostat (available at: www.fao.org/faostat/en/#data/QCL) (Accessed 27 April 2023)
- [28] Franch B, Vermote E, Skakun S, Santamaria-Artigas A, Kalecinski N, Roger J C, Becker-Reshef I, Barker B, Justice C and Sobrino J A 2021 The ARYA crop yield forecasting algorithm: application to the main wheat exporting countries *Int. J. Appl. Earth Observ. Geoinf.* **104** 102552
- [29] Penalba O C, Bettolli M L and Vargas W M 2007 The impact of climate variability on soybean yields in Argentina. Multivariate regression *Meteorol. Appl.* **14** 3–14
- [30] Thomasz E O, Corfield K, Vilker A S and Osman M 2023 Forecasting soybean production to enhance climate services for Agriculture in Argentina *Clim. Serv.* **30** 100341
- [31] García G A, Miralles D J, Serrago R A, Alzueta I, Huth N and Dreccer M F 2018 Warm nights in the Argentine Pampas: modelling its impact on wheat and barley shows yield reductions *Agric. Syst.* **162** 259–68
- [32] Global Yield Gap Atlas 2024 Description of cropping systems, climate, and soils (available at: www.yieldgap.org/argentina) (Accessed 17 January 2024)
- [33] IPAD. Argentina Wheat Area 2024 Yield and production Accessed 17 January 2024 (available at: <https://ipad.fas.usda.gov/countrysummary/default.aspx?id=AR&crop=Wheat>)
- [34] Ministerio de Agricultura, Ganadería y Pesca 2023 Estimaciones agrícolas (available at: <https://datosestimaciones.magyp.gob.ar/reportes.php?reporte=Estimaciones>) (Accessed 15 December 2023)
- [35] Abbate P E, Miralles D J and Ballesteros A 2021 Mapa de subregiones trigueras argentinas y de otros cereales invernales (available at: www.argentina.gob.ar/sites/default/files/subregiones_trigueras_2021-06-18_final.pdf) (Accessed 1 April 2024)
- [36] Slafer G A, García G A, Serrago R A and Miralles D J 2022 Physiological drivers of responses of grains per m² to environmental and genetic factors in wheat *Field Crops Res.* **285** 108593
- [37] Asseng S, Cammarano D, Basso B, Chung U, Alderman P D, Sonder K, Reynolds M and Lobell D B 2017 Hot spots of wheat yield decline with rising temperatures *Glob. Change Biol.* **23** 2464–72
- [38] Wall G W, Kimball B A, White J W and Ottman M J 2011 Gas exchange and water relations of spring wheat under full-season infrared warming *Glob. Change Biol.* **17** 2113–33
- [39] de la Casa A, Ovando G and Díaz G 2021 Linking data of ENSO, NDVI-MODIS and crops yield as a base of an early warning system for agriculture in Córdoba, Argentina *Remote Sens. Appl.* **22** 100480
- [40] Wang A X, Tran C, Desai N, Lobell D and Ermon S 2018 Deep transfer learning for crop yield prediction with remote sensing data *Proc. 1st ACM SIGCAS Conf. on Computing and Sustainable Societies (Association for Computing Machinery)* (Accessed 11 January 2023) pp 1–5 COMPASS’18
- [41] Copernicus Climate Change Service 2019 ERA5-Land monthly averaged data from 2001 to present ECMWF (available at: <https://cds.climate.copernicus.eu/doi/10.24381/cds.68d2bb30>) (Accessed 17 January 2024)
- [42] Copernicus Climate Change Service 2018 Seasonal forecast monthly statistics on single levels ECMWF (available at: <https://cds.climate.copernicus.eu/doi/10.24381/cds.68dd14c3>) (Accessed 18 January 2024)
- [43] IRI 2024 Dataset: models NMME (available at: <https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/>) (Accessed 18 January 2024)
- [44] APCC 2024 APCC MME individual models CLIK—Climate Information Toolkit (available at: <https://cliks.apcc21.org/dataset/model>) (Accessed 18 January 2024)
- [45] Deser C *et al* 2020 Insights from Earth system model initial-condition large ensembles and future prospects *Nat. Clim. Change* **10** 277–86
- [46] Gualdi S, Borrelli A, Cantelli A, Davoli G, Del Mar Chaves Montero M and Masina S 2020 The new CMCC operational seasonal prediction system (SPS3.5 CMCC) (available at: www.cmcc.it/publications/tn0288-the-new-cmcc-operational-seasonal-prediction-system-sps3-5) (Accessed 18 January 2024)
- [47] Fröhlich K, Dobrynin M, Isensee K, Gessner C, Paxian A, Pohlmann H, Haak H, Brune S, Früh B and Baehr J 2021 The german climate forecast system: GCFS *J. Adv. Model. Earth Syst.* **13** e2020MS002101
- [48] Lin H *et al* 2020 The canadian seasonal to interannual prediction system version 2 (CanSIPsv2) *Weather Forecast.* **35** 1317–43
- [49] Stockdale T, Johnson S, Ferranti L, Balmaseda M and Briceag S 2018 ECMWF’s new long-range forecasting system SEAS5 (available at: www.ecmwf.int/node/18202) (Accessed 18 January 2024)
- [50] Penabad E and Dorel L Description of System8-v20210101 C3S contribution 2023 (available at: <https://confluence>).

- [ecmwf.int/display/CKB/Description+of+System8-v20210101+C3S+contribution](https://www.ecmwf.int/display/CKB/Description+of+System8-v20210101+C3S+contribution))
- [51] Met Office 2024 Met Office seasonal prediction system: gloSea6 (available at: www.metoffice.gov.uk/research/climate/seasonal-to-decadal/gpc-outlooks/user-guide/global-seasonal-forecasting-system-glosea6) (Accessed 18 January 2024)
- [52] Kirtman B P et al 2014 The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction *Bull. Am. Meteorol. Soc.* **95** 585–601
- [53] Switanek M B, Troch P A, Castro C L, Leuprecht A, Chang H I, Mukherjee R and Demaria E M C 2017 Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes *Hydrol. Earth Syst. Sci.* **21** 2649–66
- [54] Lenderink G, Buishand A and van Deursen W 2007 Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach *Hydrol. Earth Syst. Sci.* **11** 1145–59
- [55] Gruber M H J 1998 *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators* (Marcel Dekker) p 632 (Statistics, textbooks and monographs)
- [56] Wasserman L 2004 *All of Statistics: A Concise Course in Statistical Inference* (Springer Texts in Statistics) (Accessed 24 May 2024) (<https://doi.org/10.1007/978-0-387-21736-9>)
- [57] Hajian-Tilaki K 2013 Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation *Caspian J. Intern. Med.* **4** 627–35
- [58] Hudson D 2017 Ensemble Verification Metrics (available at: www.ecmwf.int/sites/default/files/elibrary/2017/17626-ensemble-verification-metrics.pdf) (Accessed 2 May 2023)
- [59] Alvarez Prado S, Giménez V D, Ciancio N, Alzueta I, Serrago R A and Miralles D J 2023 Grain growth and development in wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.): coordination between water content and source/sink ratio *Field Crops Res.* **302** 109100
- [60] Hofer M, Marzeion B and Mölg T 2012 Comparing the skill of different reanalyses and their ensembles as predictors for daily air temperature on a glaciated mountain (Peru) *Clim. Dyn.* **39** 1969–80
- [61] Imfeld N, Barreto Schuler C, Correa Marrou K M, Jacques-Coper M, Sedlmeier K, Gubler S, Huerta A and Brönnimann S 2019 Summertime precipitation deficits in the southern Peruvian highlands since 1964 *Int. J. Climatol.* **39** 4497–513
- [62] Pede T, Mountrakis G and Shaw S B 2019 Improving corn yield prediction across the US Corn Belt by replacing air temperature with daily MODIS land surface temperature *Agric. For. Meteorol.* **276–277** 107615
- [63] WMO. Global Producing Centres for Long-Range Forecasts 2024 World meteorological organization (available at: <https://community.wmo.int/en/global-producing-centres-long-range-forecasts>) (Accessed 31 January 2024)
- [64] Potgieter A B, Schepen A, Bridler J and Hammer G L 2022 Lead time and skill of Australian wheat yield forecasts based on ENSO-analogue or GCM-derived seasonal climate forecasts—A comparative analysis *Agric. For. Meteorol.* **324** 109116
- [65] Silver N 2012 *The Signal and the Noise: Why so Many Predictions Fail—but Some Don't* (Penguin Press) p 534
- [66] Everingham Y, Sexton J, Skocaj D and Inman-Bamber G 2016 Accurate prediction of sugarcane yield using a random forest algorithm *Agron. Sustain. Dev.* **36** 27
- [67] Cai Y et al 2019 Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches *Agric. For. Meteorol.* **274** 144–59
- [68] Buizza R, Leutbecher M and Janousek M 2017 *The Forecast Skill Horizon* (ECMWF Annual Seminar)
- [69] Google DeepMind 2023 GraphCast: AI model for faster and more accurate global weather forecasting (available at: <https://deepmind.google/discover/blog/graphcast-ai-model-for-faster-and-more-accurate-global-weather-forecasting/>) (Accessed 30 January 2024)