Technical note on Global Yield Forecasting v 1.0

Sabo, F., Meroni, M., Kerdiles, H., Vojnovic, P., Rembold, F., Matteo-Sanchis, A., Piles, M., Munoz-Mari, J.

2024

Joint Research Centre

EU Science Hub

JRC137518

Ispra: European Commission, 2024

How to cite this report: European Commission, Joint Research Centre, Sabo, F., Meroni, M., Kerdiles, H., Vojnovic, P., Rembold, F., Matteo-Sanchis, A., Piles, M. and Munoz-Mari, J., *Technical note on Global Yield Forecasting v 1.0*, European Commission, Ispra, 2024, JRC137518.

Contents

Abstract

Agricultural monitoring, particularly early warning for food security, demands real-time information on crop growth conditions to detect potential production deficits. While the Anomaly Hotspots of Agricultural Production (ASAP) focuses on qualitative monitoring of anomalies, the Global Yield Forecasting activity aims to provide quantitative yield forecasts at the national level. This collaboration between ASAP-JRC and the University of Valencia seeks to develop and test an operational crop yield forecasting system based on Machine Learning (ML) approaches, integrating FAOSTAT yield data and agrometeorological indicators. The system, embedded within the ASAP platform, aims to enhance warning services by providing quantitative yield estimates before harvest. Covering 77 countries, the methodology involves historical yield data analysis, grouping similar countries based on crop development similarity, and utilizing remote sensing and meteorological indicators. ML models, Ridge Linear Regression and Gaussian Process Regression, are tested with various input data combinations and country-pooling strategies. Yield forecasts, accompanied by explainability information, are triggered twice during the growing season. The report outlines the methodology, guidelines for forecast usage, and file naming conventions. Forecasts, available in pdf and csv formats, include input data details, pooling strategy, forecasted yield, accuracy metrics, and feature importance plots. By providing comprehensive guidance, the report facilitates users in understanding and interpreting forecast outputs, thereby aiding decision-making processes for agricultural management and food security on a large scale.

1 Introduction

Agricultural monitoring, and particularly early warning for food security, requires near real-time information on crop growth conditions for early detection of possible production deficits. While the Anomaly Hotspots of Agricultural Production (ASAP, https://agricultural-production-hotspots.ec.europa.eu/), early warning system currently focusses on qualitative monitoring of several types of anomalies, the Global Yield Forecasting activity attempts to enhance the current system by providing quantitative yield forecasts at the national level.

With the objective of providing quantitative yield estimates in advance of harvest, a collaboration between the ASAP-JRC and University of Valencia was established for developing and testing an operational crop yield forecasting system at country level based on Machine Learning (ML) approaches to relate FAOSTAT yield data and agro-climatic data. The current version of the operational system is the first attempt to include crop yield estimation in the ASAP platform.

Yield forecasting is available for 77 countries of interest worldwide, which requires a systematic approach to address a wide variability of cropland systems with different crops, cropping seasons, climates, and crop management practices.

Yield forecast figures are provided with performance statistics and explainability information to support the user's analysis. Understanding the underlying causes of the model's judgments is key in practical applications. Thus, in addition to suppressing model forecasts when poor model performances are detected and providing statistical performances, we extract variable importance models to facilitate the analysis of the yield forecasts. In this manner, analysts can check the reliability and the consistency of the forecasts.

In this report, we describe the method used in the current version of the forecasting system and provide guidelines on how to use the forecasts. More details on the methods can be found in (Piles et al., 2021). The full methodology is currently being revised within a collaboration between JRC, FAO-GIEWS and University of Valencia.

2 Methodology summary

We forecast country-level yield for 77 countries of interest (Figure 1, list in Annex 2), using FAOSTAT yield data as response variable and ASAP agrometeorological data as predictors.

Historical crop yield data were downloaded from FAOSTAT at national level for the main crops of each country (https://www.fao.org/faostat/en/#data). We focus on the prediction of a selected number of staple crops (Table 1). A crop is considered to be a main crop in a country if it is among the most cultivated crops (mean harvested area is greater than 10% of the total harvested area). Typically, in each country we retain different main crops (Table 1).

Table 1. Number of countries with specific main crops

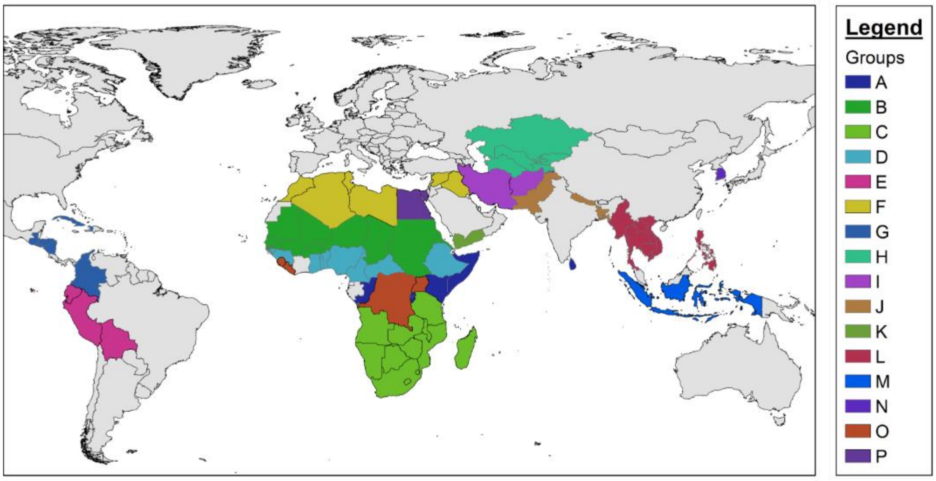
Crop	Number of countries
Maize (corn)	43
Rice	28
Wheat	18
Sorghum	18
Cassava, fresh	16
Beans, dry	16
Millet	10
Barley	10

Source: JRC analysis

The time series of crop yield used in this study spans the period 2002 – 2022([1]) (last year available in FAOSTAT database, typically two years before the current one). Although a longer archive of crop yield records is available backward in time, we are limited to 2002 because we use MODIS NDVI data, available from that year on. Therefore, the forecasting operates in a data poor environment, with about 20 yearly samples per country. To increase the sample size, we group similar countries together so that the forecasting model can be trained with a larger sample . Similarity is here defined as the temporal similarity in crop development, as assessed by the ASAP satellite-derived Land Surface Phenology (LSP). We clustered the 77 countries (Figure 1) using LSP which resulted in 16 groups, each with a different number of countries sharing similar LSP. Extraction of agro-climatic indicator was focussed on the growing season for each of the groups and this was defined by the within-group average start and end of season.

([1]) The study utilized FAOSTAT data up to the year 2022, which was the most recent available at the time.

Figure 1. Study area including 77 countries clustered in 16 groups (list of countries in Annex 2).



Source: JRC analysis

Several countries have two (or more) growing seasons per year (e.g. wet season and dry season rice in South East Asia, or short and long rainfall maize in East Africa) whereas FAOSTAT provides one yield data point per year representing the two seasons. Therefore, for such countries, the two seasons were merged into one, ranging from the start of the first to the end of the second season. Details about the clustering process and the link with the correct FAOSTAT year are given in Annex 1.

The following remote sensing and meteorological indicators were extracted from ASAP raster data (https://agricultural-production-hotspots.ec.europa.eu/download.php) over the time domain defined by per-group growing season defined above, and aggregated at national level using the ASAP cropland area fraction image (i.e. the percentage of cropland in each grid cell):

- Maximum of NDVI (MODIS dedicated ASAP processing line);
- Average air temperature (ECMWF);
- Sum of radiation (ECMWF);
- Sum of rainfall (CHIRPS and ECMWF).

This selection includes the variables used in regional yield forecasting (Meroni et al., 2021) but restricts the number of features (aggregated indicators) to one per year to avoid overfitting.

In addition to these predictors, we also fed the models with the FAOSTAT yield values for the three years preceding the year being estimated. In the following, we will refer to such features as "technological time trend" or "trend". Therefore, although the investigated time series is currently spanning the period 2002-2022, we downloaded the FAOSTAT yield data from 1999 to estimate the trend for years 2002 to 2004.

We systematically tested two machine learning (ML) models (Ridge Linear Regression, RLR) and Gaussian Process Regression, GPR) and two possible sets of input data: only trend data (TR) and trend data plus remote sensing and meteorological indicators (RSTR). The TR models based on only the yields of the 3 previous years are included in the processing to act as a baseline benchmark, i.e. **good performances of a TR model in relation to similar performances of a RSTR model indicates that the forecasted yield should be taken with care as it fairly insensitive to agrometeorological observations and mostly driven by trend.**

Remote sensing and meteorological features do not contain information about soil, management practices, crop varieties and other unobserved variables that can influence the relation between features and yield at the country. For instance, if we assume that rainfall determines yield in rainfed areas, the same rainfall quantity will not necessarily result in the same yield in two different countries. One way to convey this country-specific information to the model is by adding a categorical variable representing the country. Therefore, independently from the type of input data and ML model type, we feed the model with a country identifier as one-hot encoded variable. The influence of categorical features (country ID) can be assessed in the feature importance plots.

In addition, we systematically tested three country-pooling strategies when training the model for main crop C in country Co belonging to group G:

1. include all countries of group G where crop C is present, no matter if C is a main crop or not (pooling strategy "All");
2. include all countries in group G where crop C is a main crop ("Main");
3. use only data from country Co ("Individual)".

A leave one year out (LOYO) cross-validation was used for model evaluation. After testing all the 12 configurations reported in Table 2, we selected the one providing the highest R_{cv}^2 for each crop-country combination. It is noted that the selected model for a given crop and country may be different from the one selected for the same crop in another country.

***Table 2**: Summary of model configuration tested for the estimation of a main crop in a given country group*

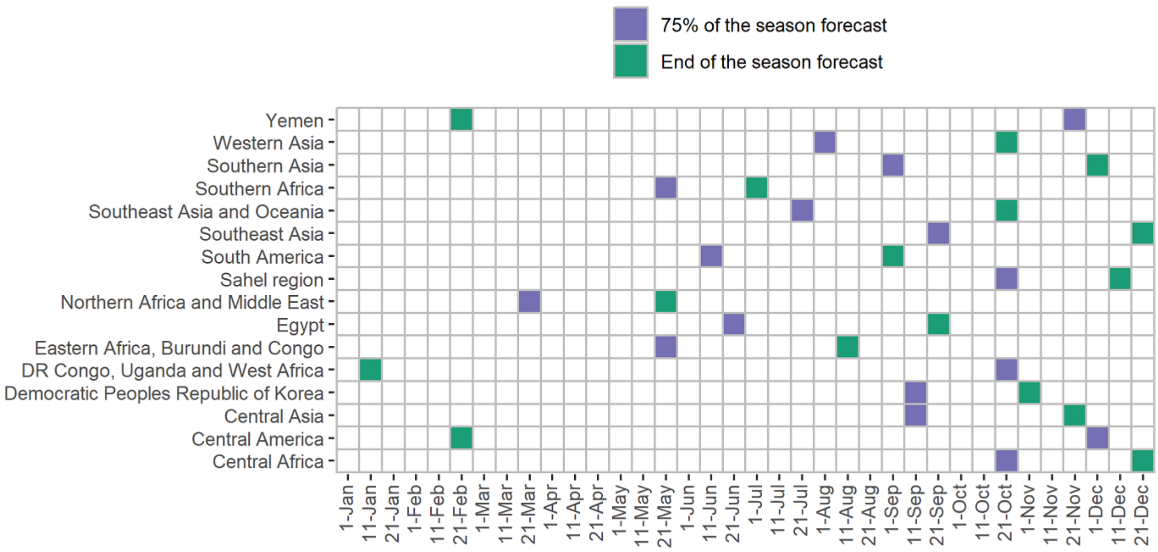
ML model		Input data		Pooling strategy	
RLR GPR	×	TR RSTR	×	All Main Individual	= **12 configurations**

Source: JRC analysis

Only models that achieved $R_{cv}^2 > 0.3$ were considered suitable for operational forecasts and retained. If no valid model is retained for a given crop-country, no forecast is available**. When the selected model is trend (TR), forecasts should be considered with caution.** When the selected model includes remote sensing and meteorological indicators (RSTR), variable importance plots are also provided.

Forecasts for a specific group of countries are triggered twice during the solar year and at specific and fixed times of the year, corresponding to the time when the group has experienced 75% and 100 % of its average crop growing period, respectively. For example, the North Africa group has an average growing season of 23 dekads (i.e. 10-day periods), starting on dekad 27 (end of September) and ending on dekad 14 (mid May) according to ASAP LSP. Forecasts for this group will be made on dekad 8 (mid March) at 75% of the growing season (23 dekads x 0.75) and on dekad 14 at the end or 100% of the average growing period. The yield forecast calendar for all the groups of countries is shown in Figure 2.

Figure 2. Calendar of yield forecasting time by country groups

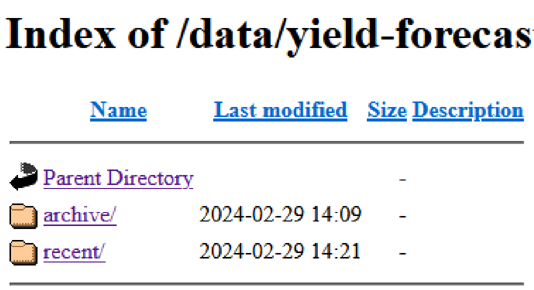


Source: JRC analysis

Yield forecasts are available on the ASAP webpage https://agricultural-production-hotspots.ec.europa.eu/data/yield-forecast/ and include the forecast for several regions (groups of countries). The new forecasts will be triggered according to the forecast calendar of Figure 2 and they are available on the website after maximum of 7 days.

Two directories are present (Figure 3): “recent” (https://agricultural-production-hotspots.ec.europa.eu/data/yield-forecast/recent/) and “archive” (https://agricultural-production-hotspots.ec.europa.eu/data/yield-forecast/archive/). In the “recent” folder the user can find the latest forecasts for each country group. When new forecasts are issued for a country group they are placed in the “recent” folder while the previous ones are moved to the “archive” folder. Note: be aware that forecasts in “recent” refer to the latest available, so when the current growing season in one group is approaching the 75% of the season forecast time the “recent” folder may still contain the end of season forecasts of the previous. Once the 75% forecast time is reached, these will be replaced by the current year forecasts.

Figure 3. Yield forecasts parent directory overview



Source: JRC analysis

The file naming convention is:

RegionName_HarvestYear_ForecastTiming.extension

where

RegionName = Group of countries belonging to a specific region and have similar crop phenology;

HarvestYear = Calendar Year of the harvest (according to FAOSTAT);

ForecastTiming = "EOS" (end-of-season) or "75p" (in-season, 75%);

Extension = pdf or csv

An example of naming for South America with harvest time in 2023 and EOS forecast is ***South America_2023_EOS.pdf*** and ***South America_2023_EOS.csv.***

Note that in the case of an in-season forecast (75p, 75% of season) made in the calendar year preceding the harvest one, the year reported is the year of harvest.

For example in Central America where 75% of the season was reached at the end of 2022 and EOS is in 2023, the naming is: ***Central America_2023_75p.pdf***, ***Central America_2023_75p.csv.***

The 75% of the season forecasts are kept in the folder "recent" until they are replaced with EOS forecasts and then they are moved to folder "archive". The EOS forecasts are kept in the folder "recent" for 3 months. Figure 4 shows an example of forecast available in the "recent" folder.

Figure 4. Available forecasts for 2023 (accessed on 19 March 2024).

Index of /data/yield-forecast/recent

Name	Last modified	Size	Description
Parent Directory		-	
Central Africa_2023_EOS.csv	2024-03-12 11:04	2.0K	
Central Africa_2023_EOS.pdf	2024-03-12 11:04	416K	
Central America_2023_EOS.csv	2024-03-12 11:04	1.3K	
Central America_2023_EOS.pdf	2024-03-12 11:04	240K	
Central Asia_2023_EOS.csv	2024-03-12 11:04	568	
Central Asia_2023_EOS.pdf	2024-03-12 11:04	128K	
DR Congo, Uganda and West Africa_2023_EOS.csv	2024-03-12 11:04	1.0K	
DR Congo, Uganda and West Africa_2023_EOS.pdf	2024-03-12 11:04	131K	
Eastern Africa, Burundi and Congo_2023_EOS.csv	2024-03-12 11:04	1.1K	
Eastern Africa, Burundi and Congo_2023_EOS.pdf	2024-03-12 11:04	267K	
Egypt_2023_EOS.csv	2024-03-12 11:04	209	
Egypt_2023_EOS.pdf	2024-03-12 11:04	3.0K	
Northern Africa and Middle East_2023_EOS.csv	2024-03-12 11:04	710	
Northern Africa and Middle East_2023_EOS.pdf	2024-03-12 11:04	158K	
Sahel region_2023_EOS.csv	2024-03-12 11:04	1.3K	
Sahel region_2023_EOS.pdf	2024-03-12 11:04	239K	
South America_2023_EOS.csv	2024-03-12 11:04	588	
South America_2023_EOS.pdf	2024-03-12 11:04	91K	
Southeast Asia_2023_EOS.csv	2024-03-12 11:04	611	
Southeast Asia_2023_EOS.pdf	2024-03-12 11:04	154K	
Southeast Asia and Oceania_2023_EOS.csv	2024-03-12 11:04	448	
Southeast Asia and Oceania_2023_EOS.pdf	2024-03-12 11:04	89K	
Southern Africa_2023_EOS.csv	2024-03-12 11:04	619	
Southern Africa_2023_EOS.pdf	2024-03-12 11:04	277K	
Southern Asia_2023_EOS.csv	2024-03-12 11:04	471	
Southern Asia_2023_EOS.pdf	2024-03-12 11:04	134K	
Western Asia_2023_EOS.csv	2024-03-12 11:04	288	
Western Asia_2023_EOS.pdf	2024-03-12 11:04	32K	
Yemen_2023_EOS.csv	2024-03-12 11:04	233	
Yemen_2023_EOS.pdf	2024-03-12 11:04	21K	

Source: JRC analysis

Forecast model outputs are provided as pdf and csv files. The csv file can be used to source the forecast values while the pdf files provides additional graphics for a quick inspection. For each country-crop combination covered, both type of files report the following information:

— the **input data** used (either the trend only, TR, or trend and remote sensing and meteorological indicators, RSTR);

— the identified best pooling strategy i.e. **Best approach** ("Main", "All", or "Individual");

— the average of the yield in the last five years ("**5 yrs avg (kg/ha)**")

— the forecasted yield for the current season ("**Forecasted yield (kg/ha)**");

— the percentage difference between the forecasted and the 5 years average yield ("**2023/5yrs (%)**");

— the coefficient of determination in prediction (LOYO) in hindcasting as a metric of accuracy ("**R2_pred**");

— the root mean square error (RMSE) in prediction (LOYO) in hindcasting as a metric of accuracy ("**RMSE_pred (kg/ha)**").

When the forecast of a given crop is missing (denoted by the "**/**" sign), it means that the minimum R^2 was not achieved.

Pdf files report the above information in Section 1, *Summary table*. An example of such table for the North Africa group is reported in Table 3.

Table 3: Example of a summary table reported in a pdf file

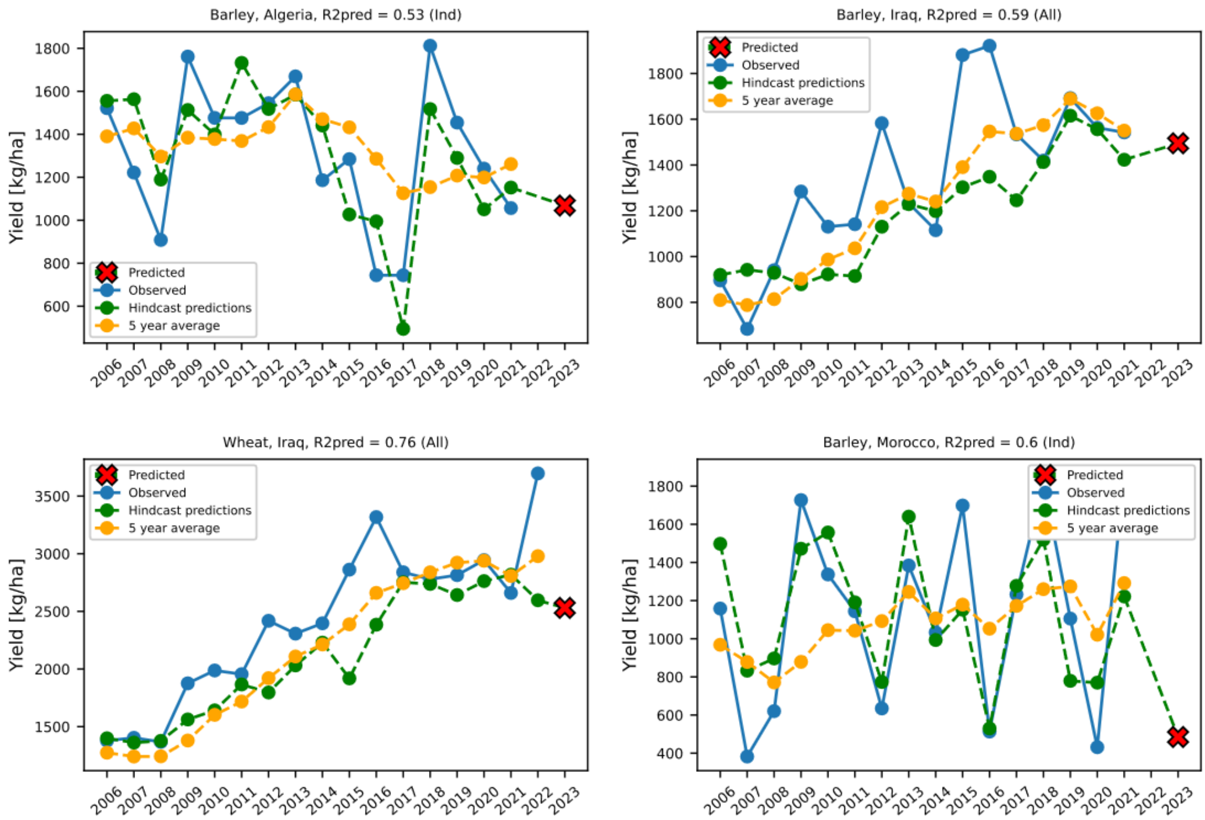
Country	Crop	Input data	Best approach	5 yrs avg (kg/ha)	Forecasted yield (kg/ha)	2023/5yrs (%)	R2_pred	RMSE_pred (kg/ha)
Algeria	Wheat	/	/	/	/	/	/	/
Algeria	Barley	RSTR	Individual	1261.44	1068.27	-15.31	0.53	207.52
Iraq	Wheat	TR	All	2978.9	2530.18	-15.06	0.76	370.12
Iraq	Barley	TR	All	1550.18	1494.77	-3.57	0.59	244.08
Libya	Wheat	/	/	/	/	/	/	/
Libya	Barley	/	/	/	/	/	/	/
Morocco	Wheat	RSTR	Individual	1724.88	1081.22	-37.32	0.63	343.08
Morocco	Barley	RSTR	Individual	1291.74	484.76	-62.47	0.6	289.31
Syrian Arab Republic	Wheat	RSTR	All	1548.24	1570.18	1.42	0.52	393.92
Syrian Arab Republic	Barley	RSTR	All	982.64	1181.48	20.24	0.56	297.1
Tunisia	Wheat	/	/	/	/	/	/	/
Tunisia	Barley	/	/	/	/	/	/	/

Source: JRC analysis.

As already mentioned**, predictions made by trend models (TR) should be interpreted with caution** because the model forecasts are not considering the agrometeorological data from the current season at all. A selection of the TR as the best model means that the trend was the best model on historical records and that agrometeorological drivers were not found to improve the predictions. This implies, for example, **that the model will not be able to represent exceptional drought conditions in the current forecast**. Forecasts obtained with the trend model are highlighted in pink.

Section 2, *Observed and predicted yield values for main crops*, shows the hindcasting LOYO predictions (in green) compared to the actual observed yield value (in blue). Hindcasting predictions were made in the leave one year out setup where the model predicts the hold-out unseen year. Additionally, the 5-year average predictions are shown as a null model with orange color (the average yield of the previous 5 years is the prediction of the next year yield). An example of such graphics is shown in Figure 5.

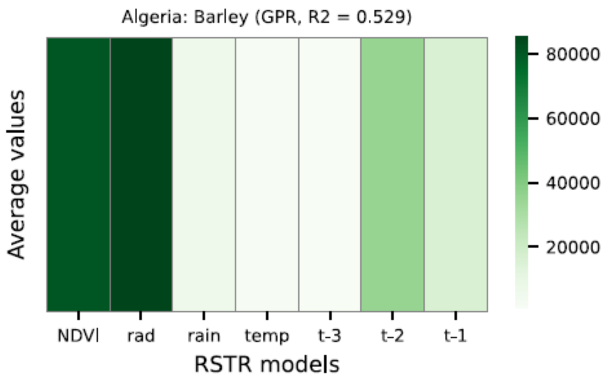
Figure 5. Observed and predicted yield values for main crops.



Source: JRC analysis

When a RSTR model is selected it means that the agrometeorological variables do improve predictions with respect to trend only. The importance of each agrometeorological variable is quantified with their feature importance plots which are also shown in the last section of the pdf file, Section 3, *Feature importance plots for RSTR models* (figure 6). The feature importance values are averaged for all the years in the hindcasting LOYO part. In the example below, for barley in Algeria, we can observe that the main features driving the forecasts are NDVI and radiation with almost no effect of temperature and rainfall indicators while previous observed yield values (the trend) have a slight impact on the final predicted of yield.

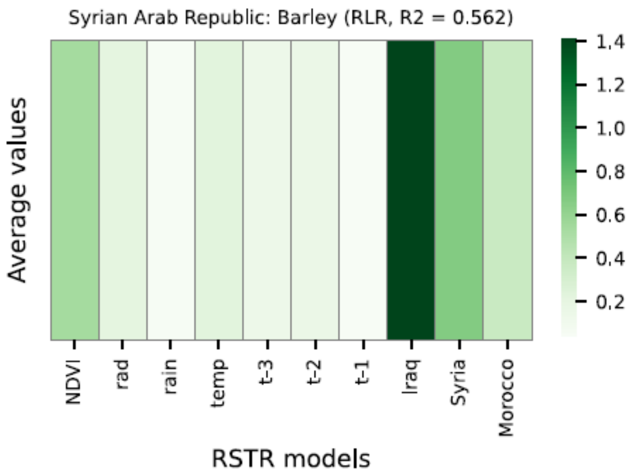
Figure 6. Example of a feature importance plot and "Individual" approach



Source : JRC analysis

In the case of “Main” or “All” approach, the feature importance plots contain also the importance of the country ID (reported as country name in the graphic) as shown in Figure 7. In this case the categorical variable ID for Iraq and Syriaand NDVI are mostly contributing to the forecasts while the trend data is not important. High importance associated with a country ID can mean that the model is using country ID information to adjust the forecasts (e.g. the yield of such specific country is largely different from the other countries of the group). Feature importance for country ids are reported for the three countries with the largest importance.

Figure 7. Example of a feature importance plot and “All” approach.



Source: JRC analysis

3 An example of yield forecast data inspection

The pdf files with the latest forecasts can be downloaded from:

https://agricultural-production-hotspots.ec.europa.eu/data/yield-forecast/recent/

As an example the table below shows Section 1 of the pdf file issued in May 2023, ***Norhern_Africa_and_Middle_East_2023_EOS.pdf***.

1. Summary table

Country	Crop	Input data	Best approach	5 yrs avg (kg/ha)	Forecasted yield (kg/ha)	2023/5yrs (%)	R2_pred	RMSE_pred (kg/ha)
Algeria	Wheat	/	/	/	/	/	/	/
Algeria	Barley	RSTR	Individual	1261.44	1068.27	-15.31	0.53	207.52
Iraq	Wheat	TR	All	2978.9	2530.18	-15.06	0.76	370.12
Iraq	Barley	TR	All	1550.18	1494.77	-3.57	0.59	244.08
Libya	Wheat	/	/	/	/	/	/	/
Libya	Barley	/	/	/	/	/	/	/
Morocco	Wheat	RSTR	Individual	1724.88	1081.22	-37.32	0.63	343.08
Morocco	Barley	RSTR	Individual	1291.74	484.76	-62.47	0.6	289.31
Syrian Arab Republic	Wheat	RSTR	All	1548.24	1570.18	1.42	0.52	393.92
Syrian Arab Republic	Barley	RSTR	All	982.64	1181.48	20.24	0.56	297.1
Tunisia	Wheat	/	/	/	/	/	/	/
Tunisia	Barley	/	/	/	/	/	/	/

Of the two main crops in Algeria, a suitable model was found for Barley only, i.e., the wheat model did not pass minimum accuracy threshold. Barley yield forecast was 1.068 t/ha, 15% below the 5 years average, well in line with the MARS yield forecasts which gave a forecast of 1.06 t/ha (May 2023 Bulletin[2]).

While no suitable models were identified for Libya and Tunisia, models providing relatively accurate forecasts for Algeria, Morocco and Syrian Arab Republic (R2_pred larger than 0.5) were found. For Morocco, despite an R2_pred of 0.6, large differences are found between our forecast (1.08 t/ha for wheat and 0.48 t/ha for barley in 2023) and the forecasts of the MARS bulletin (1.48 t/ha for wheat and 0.99 t/ha for barley).

(2) https://publications.jrc.ec.europa.eu/repository/handle/JRC133197

4 Conclusions

The collaborative effort between ASAP-JRC and the University of Valencia has resulted in the development of an operational crop yield forecasting system leveraging ML approaches. This system, integrated within the ASAP platform, offers a significant advancement in agricultural monitoring, particularly in providing real-time quantitative yield forecasts at the national level. By incorporating FAOSTAT yield data and agrometeorological indicators, the system enhances early warning services, allowing stakeholders to anticipate potential production deficits and take proactive measures.

The methodology adopted in this project involved thorough historical yield data analysis, grouping similar countries based on crop development similarities, and utilizing ML models such as Ridge Linear Regression and Gaussian Process Regression. These models were tested with various input data combinations and country-pooling strategies to ensure robustness and accuracy in yield forecasting across 77 countries.

Forecasts generated by the system are accompanied with detailed information, including input data details, pooling strategy, forecasted yield, accuracy metrics, and feature importance plots. The provision of comprehensive guidance and standardized file naming conventions facilitates users in understanding and interpreting forecast outputs, thus empowering decision-making processes for agricultural management and food security on a large scale.

The successful implementation of this operational crop yield forecasting system marks a significant milestone in agricultural monitoring and early warning systems. Moving forward, continual refinement and improvement of the system based on feedback and emerging technological advancements will be crucial in ensuring its effectiveness and relevance in addressing global food security challenges. Overall, this collaborative endeavor represents a valuable contribution towards harnessing the potential of data-driven approaches in safeguarding agricultural productivity and enhancing food security worldwide. The complete methodology is currently undergoing revision through a collaborative partnership among JRC, FAO-GIEWS, and the University of Valencia.

Disclaimer

The system is in beta testing phase. Contents of pdf and csv files may be modified and/or moved to a dedicated section on the ASAP website.

References

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

Piles, M., Mateo–Sanchis, A., Munoz–Marí, J., Camps-Valls, G., Waldner, F., Rembold, F., Meroni, M., 2021. Global Cropland Yield Monitoring with Gaussian Processes, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Presented at the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 180–183. https://doi.org/10.1109/IGARSS47720.2021.9554315

List of abbreviations and definitions

Abbreviations	Definitions
ASAP	Anomaly Hotspots of Agricultural Production
ML	Machine learning
NDVI	Normalized difference vegetation index
LSP	Land surface phenology
MODIS	Moderate-Resolution Imaging Spectroradiometer
ECMWF	European Centre for Medium-Range Weather Forecasts
RLR	Ridge linear regression
GPR	Gaussian process regression
TR	Trend
RSTR	Remote sensing and trend
EOS	End of season
LOYO	Leave one year out
RMSE	Root mean square error
MARS	Monitoring agricultural resources
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
GIEWS	Global Information and Early Warning System on Food and Agriculture

List of figures

List of tables

Annexes

Annex 1. Country clustering workflow

- Determine, per country, the fraction of crop area that is active at each dekad of the year (ranging from 0 to 100% for each of the 36 dekads) according to ASAP. Note: a pixel is considered "active" in a given dekad if that dekad falls in the multi-annual average growing season period, as determined by the satellite-derived Land Surface Phenology.

- Determine the period (or periods) of the year over which at least 15% of the crop area has growing crops using satellite-derived phenology. Note that this period may be due to multiple and overlapping growing season within one country and may cross the calendar year.

- Adjust the start and end of this period for those countries where the period covers the whole year without any interruption and some other complex situation where satellite phenology is less interpretable (e.g. Viet Nam and neighbouring countries) using FAO crop calendars and ASAP analysts' knowledge.

- In case of two growing season per year, merge them to be coherent with the single value per year provided by FAOSTAT.

- Take care of identifying the correct FAOSTAT year to which the above-determined period has to be associated. According to FAO definition (that we empirically checked in a few countries for which we have more detailed data), the yield data point must be attributed to the year where harvesting starts (according to FAO definition, i.e. their calendars). Unfortunately, this start of harvest is not precisely defined by FAO and must be interpreted case by case. We went through this process to identify it for all countries, using FAO reports and analysts' knowledge.

After that, we clustered the countries to a reasonable number of groups using Ward hierarchical clustering based on the yearly profile of active area. For each of these groups of countries with similar phenology (i.e. similar time domains), we defined a common time domain (average of start and end within the group).

Annex 2 Countries and group names

Country name	Group name	Group letter
Sri Lanka*	Sri Lanka	/
Congo	Eastern Africa, Burundi and Congo	A
Burundi		
Somalia		
Kenya		
Rwanda		
Sudan	Sahel region	B
Mali		
Niger		
Senegal		
Mauritania		
Eritrea		
Gambia		
Chad		
Burkina Faso		
Mozambique	Southern Africa	C
Madagascar		
United Republic of Tanzania		
Zambia		
Malawi		
Angola		
Lesotho		
South Africa**		
Botswana		
Namibia		
Zimbabwe		
South Sudan	Central Africa	D

Cameroon		
Nigeria		
Ethiopia		
Guinea		
Guinea-Bissau		
Central African Republic		
Ghana		
Togo		
Benin		
Ecuador		
Bolivia (Plurinational State of)	South America	E
Peru		
Iraq		
Libya		
Syrian Arab Republic		
Algeria	Northern Africa and Middle East	F
Morocco		
Tunisia		
Nicaragua		
Honduras		
Guatemala		
Colombia	Central America	G
El Salvador		
Cuba		
Haiti		
Kyrgyzstan		
Tajikistan		
Uzbekistan	Central Asia	H
Turkmenistan		
Kazakhstan		
Afghanistan	Western Asia	I

Iran (Islamic Republic of)		
Bangladesh Pakistan Nepal	Southern Asia	J
Yemen	Yemen	K
Lao Peoples Democratic Republic Myanmar Philippines Cambodia Viet Nam Thailand	Southeast Asia	L
Timor-Leste Indonesia	Southeast Asia and Oceania	M
Democratic People's Republic of Korea	Democratic People's Republic of Korea	N
Cote d'Ivoire Democratic Republic of the Congo Liberia Uganda	DR Congo, Uganda and West Africa	O
Egypt	Egypt	P

*Currently being regrouped

**Forecasts available only for maize

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official