Crop monitoring and yield forecasting at global level
The GLOBCAST project from the European Commission

Raúl López-Lozano
European Commission,
Joint Research Centre (JRC), Institute for Environment and Sustainability (IES), Monitoring Agricultural Resources Unit (MARS), Ispra (Italy)

Bettina Baruth
European Commission,
Joint Research Centre (JRC), Institute for Environment and Sustainability (IES), Monitoring Agricultural Resources Unit (MARS), Ispra (Italy)

Mohamed El Aydam
European Commission,
Directorate-General for Agriculture and Rural Development (DG AGRICULTURE), Arable Crops Unit,
Brussels (Belgium)

Eric Willems
European Commission,
Directorate-General for Agriculture and Rural Development (DG AGRICULTURE), Arable Crops Unit,
Brussels (Belgium)

Tomás García Azcárate
European Commission,
Directorate-General for Agriculture and Rural Development (DG AGRICULTURE), Observatory of Agricultural Markets, Brussels (Belgium)

Acknowledgments

The GLOBCAST project phases 1 -3 (2008 – 2015) were funded by DG AGRICULTURE.

The useful inputs from Dr. Maurits van den Berg to the present manuscript are sincerely acknowledged.

Introduction

Recent trends in global agriculture prices have brought a new scenario for agricultural policies worldwide. Increased world demand for agricultural products combined with interannual fluctuations of global production mostly caused by climate variability have been an important cause for price volatility in agricultural markets, and social unrest in many parts of the world.

In this context, crop monitoring and yield forecasting play a
major role in anticipating supply anomalies, thus allowing well-informed timely policy action and market adjustment, preventing food crises and market disruptions, reducing market speculation, and contributing to overall increased food security. This is the objective of international initiatives such as AMIS (Agricultural Market Information System) at global level established in 2011 after the request of G-20[1] or MED-Amin in the Mediterranean, launched in 2014 and coordinated by CIHEAM[2].

The European Commission (EC) DG AGRICULTURE launched and financed a three-phase project named GLOBCAST in 2011, an initiative to demonstrate the feasibility and operationalization of a project for crop monitoring and forecasting in the world's main grain producing areas. This project includes the set-up of two main systems. The first system, is dedicated to the estimation of crop areas through the integration of area sampling and remote sensing methods (Carfagna and Gallego, 2005). The second one, presented in this paper, is about forecasting yields. The global yield forecasting system represents an extension of the current European system (the MARS Crop Yield Forecasting System, or M-CYFS) to other world regions with a high relevance on agricultural market prices.

The main output of the current European system, run since 1992 by the Monitoring Agricultural Resources Unit (MARS) of the EC Joint Research Centre, is a monthly bulletin on crop monitoring and yield forecasting in Europe [3] with quantitative yield forecasts and country analysis reports on crop development.

This article presents the basic principles of this new global crop yield forecasting system: the conceptual design of the system, the regions of the world to be covered, and the specific technical solutions to be implemented.

Agricultural regions of interest

GLOBCAST is focused on seven large grain producing regions with a high relevance on agricultural market prices
(see the figure 1)

The global system to be constructed considers these geographical areas as "windows" of the system, all of them sharing the same basic principles and data structure, but with some features (crop models, static information layers, etc.) implemented differently on each window depending on the region-specific requirements and data availability:

- *The EU and its neighbourhood*, including the EU-28 member states, plus neighbouring countries in the Mediterranean and Eastern Europe with a relevant influence on agricultural markets: Ukraine, Turkey, Belarus, Morocco, Algeria, Tunisia, Egypt and Libya. The analysis and crop yield forecasting for the EU-28 is focused on the following crops: soft wheat, durum wheat, winter barley, spring barley, rye, triticale, maize, rice, sunflower, potato, sugar beet, and rapeseed. In neighbouring countries, the focus is on the main winter and summer crops: soft wheat, durum wheat, barley, maize, and sugar beet. The EU-28 and some of the neighbouring countries are covered by the European window of the M-CYFS.

- *Russia and Kazakhstan*. In these countries the crops considered are winter wheat, winter barley, spring wheat, spring barley, rapeseed, and sugar beet.
- *China*, where the main crops of interest are: wheat, maize, and two seasons for rice (harvested in summer and harvested in autumn).

- *India*, where the monitoring and yield forecasting activities will include winter wheat, early season rice (so-called *rabi* rice), and late season (*kharif*) rice.
- *South America*. This geographical window includes Argentina and Brazil, where the main crops to be covered are wheat, barley, maize (cash crop), soybean and sugarcane.

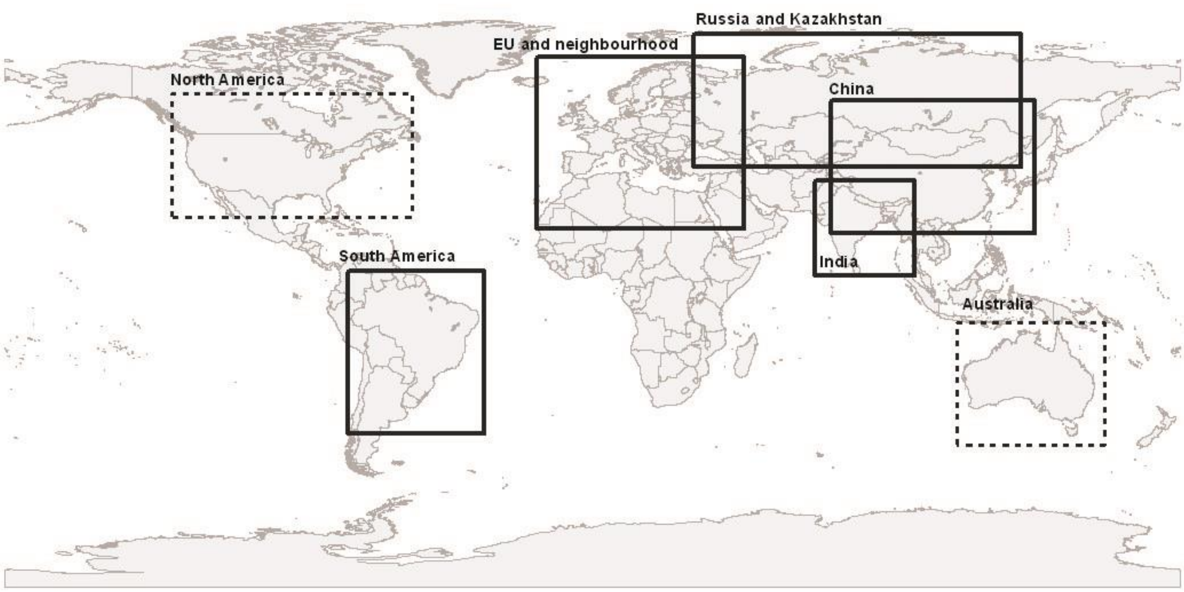
- *North America*. This window englobes the United States and Canada for wheat, barley, rapeseed (canola), maize, soybean and sugar beet.

- *Australia*, where only wheat will be analysed.

Figure 1

[1] See http://www.amis-outlook.org/
[2] See https://med-amin.ciheam.org/en/
[3] http://mars.jrc.ec.europa.eu/mars/Bulletins-Publications

"Windows considered in GLOBCAST. Solid squares identify those windows considered as priority in the system set-up"



General principles for a global crop yield forecasting system: the M-CYFS concept

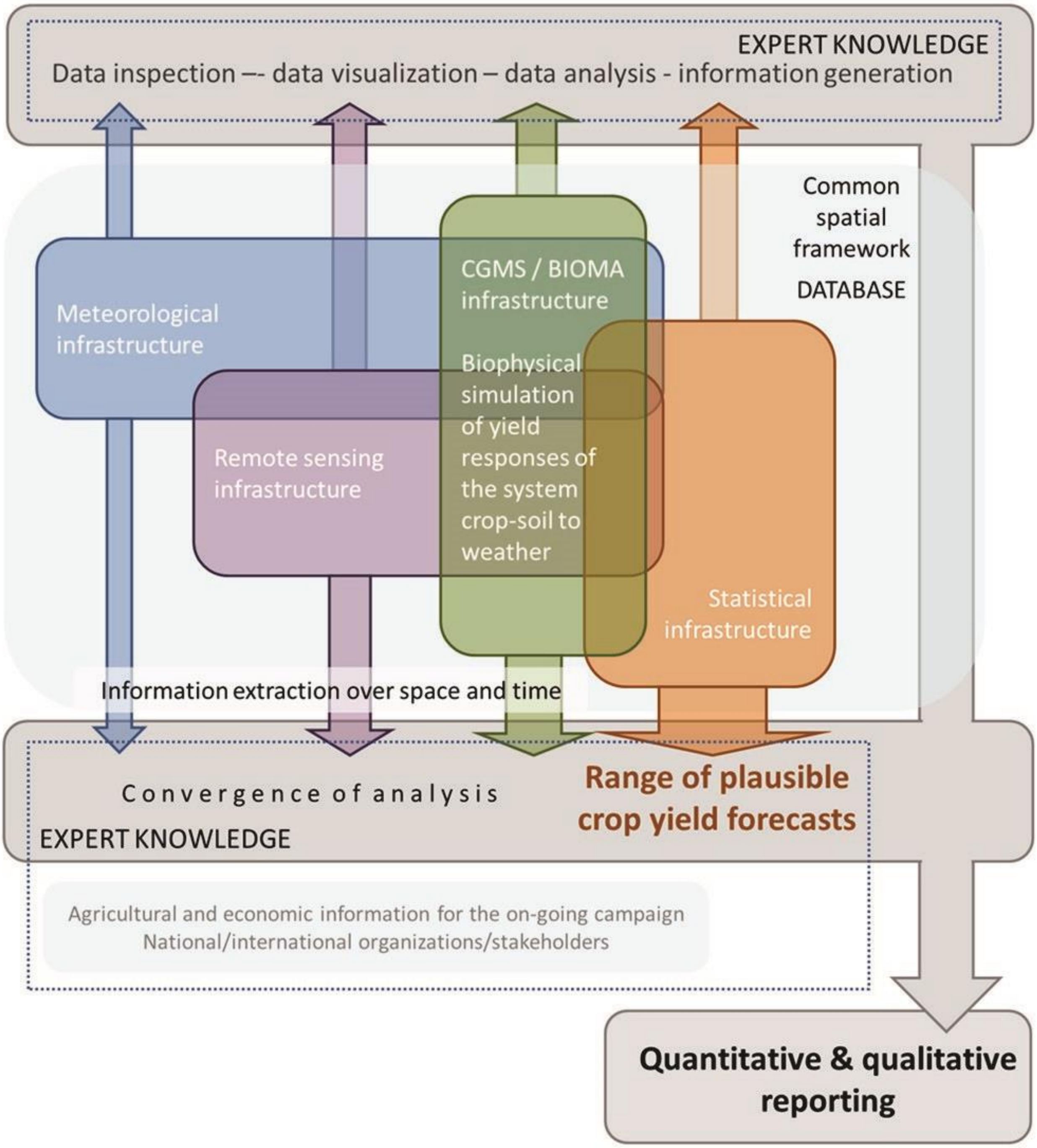
The crop yield forecasting system that is envisaged to be implemented is based on the principles of the current M-CYFS running at EU-28 level[4].

The M-CYFS is a system that makes use of several models and tools producing a large volume of information in near real time (NRT) on meteorological and crop indicators. Nevertheless, the system is driven by expert knowledge, in which a team of analysts play the central role in evaluating and interpreting numerous indicators, and selecting those that are determining crop yields to produce a reliable yield forecast. A conceptual design of the system is given in figure 2

[4] http://marswiki.jrc.ec.europa.eu/agri4castwiki/index.php/Main_Page

Figure 2

Conceptual design of the MARS Crop Yield Forecasting System (M-CYFS) with the five system components: meteorological data infrastructure, remote sensing data infrastructure, crop modelling infrastructure, statistical analysis infrastructure and the common spatial framework



Data production of the M-CYFS relies on five system components:

- *Meteorological data infrastructure.* This component accounts for the acquisition and processing of meteorological data in NRT (temperature, precipitation, incoming radiation, etc.) relevant for crop monitoring.

- *Crop modelling infrastructure.* This system component uses NRT meteorological data, soil information, and agronomic parameters (collected from field experiments, scientific literature, etc.) for the crops monitored to run crop models. The modelling platform BioMA[5], developed at MARS, constitutes the engine of this component. The outputs are a set of simulated crop indicators (total above-ground biomass, leaf area index, relative soil moisture, biomass in storage organs, crop development stage) describing –according to the crop model assumptions– the effects of weather conditions on crop growth and yield.

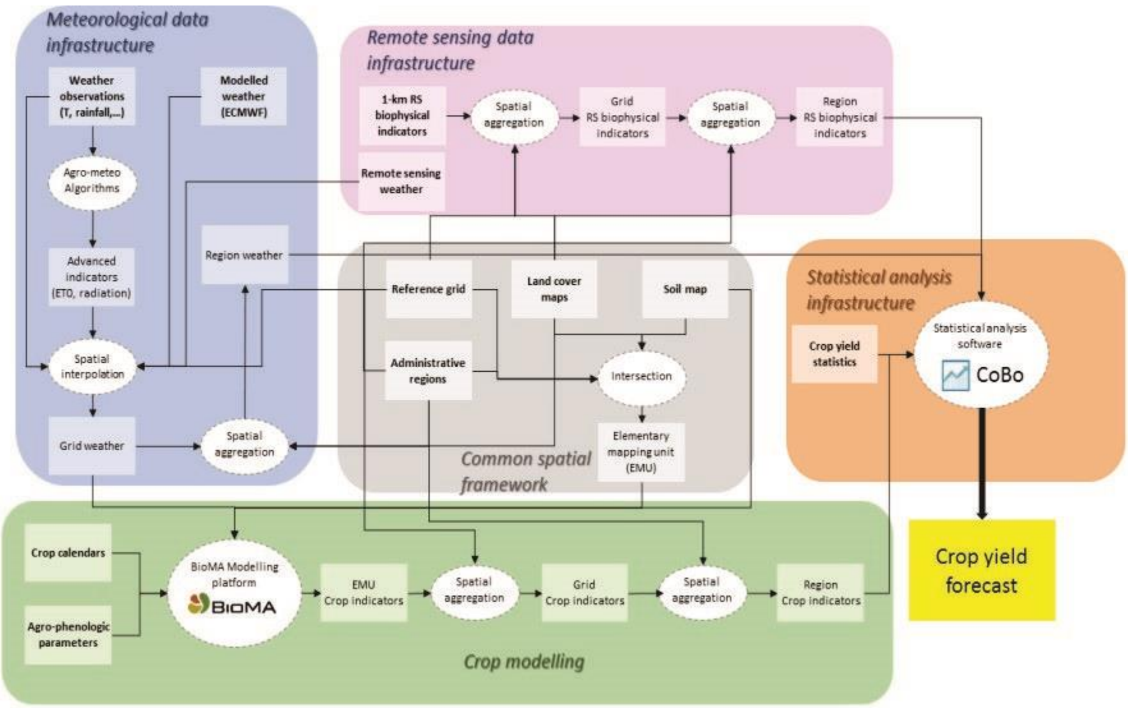
- *Remote sensing infrastructure.* This includes the acquisition and processing of satellite imagery at medium-low spatial resolution (about 1-km) to produce biophysical vegetation indicators, such as the normalized difference vegetation index (NDVI), the fraction of absorbed photosynthetically active radiation (fAPAR) or the leaf area index (LAI). These indicators are used to assess the actual green biomass formation and canopy vegetative status over arable land areas.
- *Statistical infrastructure.* This component encompasses the methodological framework to use the outputs from the previously mentioned system components to produce a yield forecast. The following statistical procedures are used: (i) analysis of yield trends, (ii) simple/multiple linear regression between indicators and official yield statistics, and (iii) similarity analysis to identify historical years that are similar to the forecast season in terms of sets of selected indicators (e.g. simulated leaf area and soil moisture behaviour over the season). A specific software application CoBo (acronym for Control Board) was developed inhouse to accommodate these statistical tools, and to manage and store all the statistical processes executed by the analysts to produce their forecasts. This ensures a full traceability of the process and it permits to assess the overall quality of the system outputs.

- *Common spatial framework.* The M-CYFS is a spatialized system. The spatial framework permits an integrated analysis of all the information. It includes the datasets containing the system spatial units (reference grid, different levels of administrative units), their intersection, and the methods to aggregate spatially the indicators from their original resolution.

  The data workflow in the system is shown in figure 3. All the datasets produced by these five components are stored in a relational database (the M-CYFS DB), and can be accessed through viewing and mapping applications developed for that purpose.

[5] http://bioma.jrc.ec.europa.eu/

Figure 3
Data workflow among the M-CYFS system components to produce a final crop yield forecast Datasets are identified with boxes, whereas data processing is identified with ellipses



Implementation of the M-CYFS in different regions of the world: technical solutions to set-up the system components

The set-up of the M-CYFS in the areas of the world mentioned in Section 0 is envisaged as an incremental process of improvement from a basic system to a technically advanced and more performing one. The adoption of a technical solution for a system component, e.g. the meteorological infrastructure depends on two main criteria:

- Cost-efficiency, defined as the expected contribution of that solution to the overall performance of the system compared with the costs of implementing it
;
- The availability of required data and methods necessary to implement such a solution.

After carrying out a viability assessment of the system composition and implementation in the regions of the world covered in GLOBCAST, three technical levels (basic, intermediate, and advanced), were identified as successive stages of implementation and improvement for the different system components. These solutions are summarized in table 1 *(see the Annex).*

Basic system components

Basic components are those constituting a system able to achieve an accuracy of the crop yield forecasts comparable to the average inter-annual variability at national level; valid predictions can be produced also at sub-national level, in those specific regions where strong variations happen.

The basic option for the meteorological component is the use of weather data produced by meteorological models. ECMWF (the European centre for medium-range weather forecasts) provides the products ERA (archive reanalysis) and HRES (deterministic 10-day forecasts), based on the assimilation of observed atmospheric data from different sources (meteorological satellites, radiosondes, weather stations, etc.) into a global circulation model to produce weather indicators [6] . Although the reliability of modelled weather in the different regions of the world depends strongly on the availability of observed data to assimilate into the model, it may

[6] http://www.ecmwf.int/en/forecasts/datasets

constitute a valid option where NRT data from automatic weather stations is not available or accessible.

The crop modelling component relies on the BioMA platform with different crop models implemented: WOFOST (Van Diepen et al., 1989) and CropSyst (Stöckle et al., 2003) to simulate crop growth of most field crops; WARM (Confalonieri et al., 2009), a specific model for rice; and CANEGRO (Singels et al., 2008) for sugarcane. For all these models a basic parameter setting exists based on general agronomic studies and values reported in scientific literature. In the basic system configuration, this basic setting is used, which may present some limitations for specific crops and regions of the world.
The remote sensing infrastructure relies on 1km resolution vegetation indices such as the NDVI (Normalized Difference Vegetation Index) for these platforms: SPOT-VEGETATION, METOP-AVHRR and NOAA-AVHRR. A disadvantage of this approach is that these products are not directly comparable, which makes it difficult to maintain long-term records when the lifecycle of these platforms finish and are replaced by others. Finally, the statistical analysis is applied with the basic statistic tools implemented in the CoBo software: trend analysis, regression, similarity analysis performed at country level to produce yield forecasts.

Intermediate system components

It is the goal with the intermediate and advanced system components to obtain crop yield forecasts with an accuracy significantly better than the average interannual change at national level and with accuracy comparable to the average inter-annual change at subnational level for the main production regions.

The intermediate level represents an incremental enhancement compared to the baseline level. In the meteorological component global datasets of weather stations are added to the system. The M-CYFS interpolates those observations from stations to produce spatially-continuous layers of meteorological indicators. The daily records of the WMO surface network stations constitute the most valid option, although in some of the countries the availability of stations is low. These data are freely accessible via the GHCN (Global Historical Climatological Network Dataset)[7] maintained by the US NOAA.

The crop modelling component is enhanced by improving the crop model calibration based on regional specific agro-climatic conditions, in partnership with local agricultural and research institutions. This implies the compilation of local agronomic data (phenological observations, biomass measurements, etc.), which permits a fine-tuning of the crop model outputs to the actual conditions in the field.

In the remote sensing component, the land vegetation biophysical products from the EC Copernicus Programme [8] represent an advancement against the basic system. Although the spatial resolution is still 1 km, the products have been conceived to be platformindependent, which permits to generate compatible long-term historical archives from different platforms: NOAA-AVHRR, SPOT-VEGETATION, PROBA-V and the future Sentinel mission. Long-term archives are crucial to understand the relationship between biomass dynamics observed by satellites and crop yields in large areas.

The statistical analysis will be performed at sub-national level, which allows the analyst to select specific methods to produce yield forecasts for regions within countries based on local agro-climatic conditions. This improvement depends, however, on the availability of reliable crop statistics at sub-national level for the different countries.

Advanced system components

In an advanced system, the meteorological component is further enhanced by including data from national networks of weather stations. The reliability of the meteorological indicators produced from weather stations is strongly determined by the spatial density of the latter, and determines as well the accuracy of crop model indicators, fed with NRT weather data. Global repositories of weather station data (like GHCN) contain only sparse data on some countries, and therefore alternative sources have to be considered to reach the required density.

The advanced crop modelling component includes the development of specific modules to simulate country or region specific abiotic stress that are locally determining yields, and that are not included in the general purpose crop models. This includes, for example, a frost-kill module to simulate effects of extreme low temperatures in Russia and

[7] http://www.ncdc.noaa.gov/climate-monitoring/

[8] http://land.copernicus.eu/global/?q=products

Kazakhstan; or modules to simulate pests and diseases, etc. Moreover, additional agromanagement information (irrigation, fertilizing, etc.) must be collected with local partners to further calibrate crop models on local conditions.

The remote sensing advanced component would incorporate high-resolution products. In many countries, the agricultural landscape is highly fragmented and 1-km images are not sufficient to isolate the response of the different crops. The forthcoming European Space Agency's (ESA) Sentinel-2 Mission would permit to exploit synergistically high and low spatial resolution data to retrieved continuous crop-specific time-series of biophysical products. In that scenario, advanced procedures assimilating remote sensing data into crop models (Dorigo et al., 2007, Guerif and Duke, 2000) could be envisaged to upgrade as well the statistical analysis component.

Outlook

In the EU the system is active, and has reached an important degree of maturity after more than 20 years, as well as in Turkey and Ukraine. The meteorological and crop monitoring components have reached an advanced technical level: the system processes every day weather data from more than 2200 weather stations, and the crop modelling component has been improved during different calibration exercises. Remote sensing and statistical analysis are already at an intermediate level, based on low resolution products (1-km SPOTVEGETATION and PROBA-V products). The forthcoming ESA Sentinel-2 mission [9] will permit to improve the system component with high resolution data.

The system is active as well for Maghreb countries (Morocco, Algeria and Tunisia), Belarus, Russia and Kazakhstan, but with a lower technological level compared to the European one. Weather data for these regions rely heavily on modelled ECMWF data, with sparse observations from weather stations. The crop model calibration needs some improvements with specific datasets to implement realistic crop calendars and crop phenology parameters in many regions. The remote sensing and statistical analysis components are at the same as in Europe.

The implementation of a basic system for Argentina, Brazil, China, India, Egypt and Libya will occur during 2015 in the frame of the current GLOBCAST project. The system will rely mostly on global datasets: ECMWF weather data, a basic crop model calibration with basic data published in scientific literature, global soil and land cover maps, etc. The MARS Unit is currently working to identify local partners to exchange agronomic and meteorological information and experiences, including technology transfer and capacity building activities to these countries.

[9] https://earth.esa.int/web/guest/missions/esa-futuremissions/sentinel-2

Bibliography / more information

- Carfagna, E., Gallego, F.J., 2005. *Using Remote Sensing for Agricultural Statistics*. Int. Stat. Rev. 73, 389–404. doi:10.1111/j.17515823.2005.tb00155.x
- Confalonieri, R., Rosenmund, A.S., Baruth, B., 2009. *An improved model to simulate rice yield*. Agron. Sustain. Dev. 29, 463–474.
- Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. *A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling*. Int. J. Appl. Earth Obs. Geoinformation 9, 165–193
- Guerif, M., Duke, C.L., 2000. *Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation*. Agric. Ecosyst. Environ. 81, 57–69.
- Singels, A., Jones, M., van den Berg, M., 2008. DSSAT v4.5 - Canegro Sugarcane Plant Module.
- Stöckle, C.O., Donatelli, M., Nelson, R., 2003. *CropSyst, a cropping systems simulation model*. Eur. J. Agron. 18, 289– 307.
- Van Diepen, C.A., Wolf, J., Van Keulen, H., Rappoldt, C., 1989. *WOFOST: a simulation model of crop production*. Soil Use Manag. 5
-
  - , 1

Table 1
Summary of the technical solutions (basic, intermediate, and advanced) foreseen to construct the GLOBCAST system component

	Basic solution		**Intermediate solution**		**Advanced solution**	
	description	**datasets/methods**	**description**	**datasets/methods**	**description**	**datasets/methods**
Meteorological data infrastructure	Modelled weather data	*ECMWF daily global datasets: ERA historical archive, HRES reanalysis, OPE deterministic forecast*	Observed weather from global repositories	*WMO – GHCND daily data*	Observed weather from local networks	*Acquisition and processing on NRT of weather stations from each country network*
Crop modelling	Crop models with a baseline calibration	*WOFOST, CropSyst, WARM, CANEGRO models implemented in BioMA* *Calibrated with data from scientific literature*	Crop models locally calibrated	*WOFOST, CropSyst, WARM and CANEGRO models implemented in BioMA* *Calibrated specifically for each region using local agronomic studies*	Region-specific crop model solutions with agromanagement	*WOFOST, WARM, CropSyst and CANEGRO plus additional biotic and abiotic stress modules implemented in BioMA* *Calibrated specifically for each region using local agronomic studies and local agromanagement data*
Remote sensing data infrastructre	Low resolution datasets	*NOAA-AVHRR NDVI 1km data* *METOP-AVHRR NDVI 1km data*	Low resolution long-term biophysical products	*Copernicus global land vegetation monitoring 1-km products:* *NDVI, LAI, fAPAR, fCOVER*	Synergy low resolution and high resolution biophysical products	*Copernicus high and 1-km biophysical products and 30-m imagery (future Sentinel-2 Mission) exploited to produce crop-specific information.*
Statistical analysis infrastructure	Countrylevel statistical forecasts	*Analysis performed at country level (GAUL0) with basic statistical procedures*	Sub-national statistical forecasts	*Analysis performed at sub-national level (GAUL1) with basic statistical procedures*	Data assimilation	*Remote sensing data assimilation techniques into crop models to exploit components synergy*