Crop Monitoring as an
E-agricultural tool in
Developing Countries

ASSESSMENT REPORT ON THE BIOMA PLATFORM FOR RICE MONITORING

Reference: *E-AGRI_D33.1_Report on the assessment of the BioMA platform for rice monitoring and yield forecasting in Jianghuai plain*
Author(s):Valentina Pagani, Davide Fumagalli, Simone Bregaglio, Caterina Francone, Wang ZhiMing, Roberto Confalonieri
Version: 1.0
Date: 31/10/2013

DOCUMENT CONTROL

Signatures

Author(s) : Valentina Pagani
Roberto Confalonieri
Davide Fumagalli
Simone Bregaglio
Wang ZhiMing
Caterina Francone

Reviewer(s) :

Approver(s) :

Issuing authority :

Change record

Release	Date	Pages	Description	Editor(s)/Reviewer(s)
1.0	31/10/2013	27	D33.1 - Report	

TABLE OF CONTENT

LIST OF FIGURES

EXECUTIVE SUMMARY

BioMA is a platform developed for analysing and running - on explicit spatial units - modelling solutions based on biophysical models. The platform is composed of different integrated tools easily allowing the user to create specific configurations, in this case for comparing (i) the outputs simulated by different crop models, (ii) as well as their capability of forecasting yields, without the need for opening and modifying the underlying databases.

The user can especially visualize and compare the outputs and forecasting results derived from the modification of management options or from the simulation of crop limitations (i.e., water stress, biotic and abiotic damages).

In this work, BioMA was applied to simulate rice growth and development in Jiangsu province (China) in the period 1990-2010. The biophysical models CropSyst, WARM and WOFOST, previously calibrated (during this project) for the specific conditions explored, were used and compared for the spatial simulations, whereas – to assess the advanced functionalities of the platform –only WARM's outputs were post-processed to obtain the yield forecasting. The evaluation of the multi-model approach for yield forecasts will be performed in the coming weeks and presented in the E-AGRI deliverable D33.2 (month 36).

CropSyst, WARM and WOFOST differently reacted to the modification of management and to the simulation of the plant-pathogens interaction. Indeed, the values and patterns followed by the output variables strictly depend on the approach used by each simulation model to reproduce crop growth and development. The variability of official yields explained by the statistical post-processing of simulated data changed when management options were modified and increased if the diseases limited outputs were inserted as indicators in the regression analysis. In the specific case, the coefficient of determination increased for about 5% using the disease-limited indicators.

This is the first time a simulation/forecasting environment including plant-diseases interaction is presented and evaluated.

1. Introduction

The monitoring of crops in large areas during the growing season and the timely forecasting of crop yields at regional or national level sustain the agricultural policies and limit the speculations. The final values of yields are officialised and broadcasted by the government only some months after crop cycles are completed (harvest).

Various types of yield forecasting systems were developed during the last decades. The existing systems differ for the adopted technique, the spatial scale and the sources of information used. The most complex – and powerful – yield forecasting systems are based on the use of information from remote sensing and crop simulation models, individually or coupled in integrated systems.

Reliability of forecasts based on biophysical models is affected by some limitations, especially concerning the uncertainty of inputs (driving variables, parameters, management), crop distribution and varieties used and to the absence of the simulation of key processes, such as pathogens impacts on plants. Moreover, the existing systems do not allow the user to easily create and customize configurations and to obtain yield forecasts directly from the spatially simulated outputs.

BioMA (Biophysical Model Applications) is a platform that gives the opportunity to run multi-model simulations against a shared spatially-explicit database. The user can generated and compare the outputs of different modelling solutions, that are compositions of discrete entities (i.e., components) encapsulating description (ontology) and models for a specific domain within the simulated system. UNIMI.Diseases is a software component containing models to estimate the impacts of plant disease epidemics on plant growth and yield. It consists of four modules providing a generic frame to simulate disease development. A key aspect of the framework is the transparency which allows for quality evaluation of outputs in the various steps of the modelling work. In this part of the E-AGRI activities, the UNIMI.Disease component was linked to the three models used in the project (WARM, CropSyst, WOFOST) and included in the BioMA platform.

The software used for yield forecasts was the software CST, developed for the JRC's MARS (Monitoring Agriculture with Remote Sensing) project in 1994, and used to post-process simulation outputs. The software has been integrated in the simulation-forecasting environment of BioMA.

2. Materials and methods

2.1. The BioMA platform

The BioMA platform allows the user to integrate a personalized modelling solution, loading the data sources and modifying parameters values. In addition to the simulation tool, BioMA is also integrated with the "map and data visualizer", a tool for mapping the different outputs of simulations.

Moreover, the platform is also connected to the CST statistical tool, a system developed for statically processing simulated and official yields of past seasons to forecast production for the current year. The statistical tool retrieves data from a database and the simulation results can be written to that database too.

This structure easily allows comparing results obtained from the crop models WARM, CropSyst and WOFOST, implemented via the CropML library in BioMA. It is also easy to visualize the effects derived from the change of management and the differences between potential production and production limited by water stress, biotic or abiotic damages.

The following paragraphs summarize the main options which can be used and modified by the user inside the simulating, mapping and statistical tools.

2.1.1. The simulating tool

Figure 1 shows the graphical interface of BioMA. The user can load a solution earlier prepared, using the toolbar button encircled in the figure, or can construct a new configuration, inserting all the information and parameter sets needed.

The first step for the creation of a new solution is the choice of the crop simulation model, the period of simulation and the weather database, indicated in Figure 2 with a red, blue and green square, respectively. Subsequently, if the crop grows in water limited conditions, information about physical-chemical and hydraulic characteristics of soils can be added (see Figure 2, yellow square).

Bioma simulates plant growth and development limited by water stress and diseases if the corresponding components (see Figure 2, violet square) are enabled.

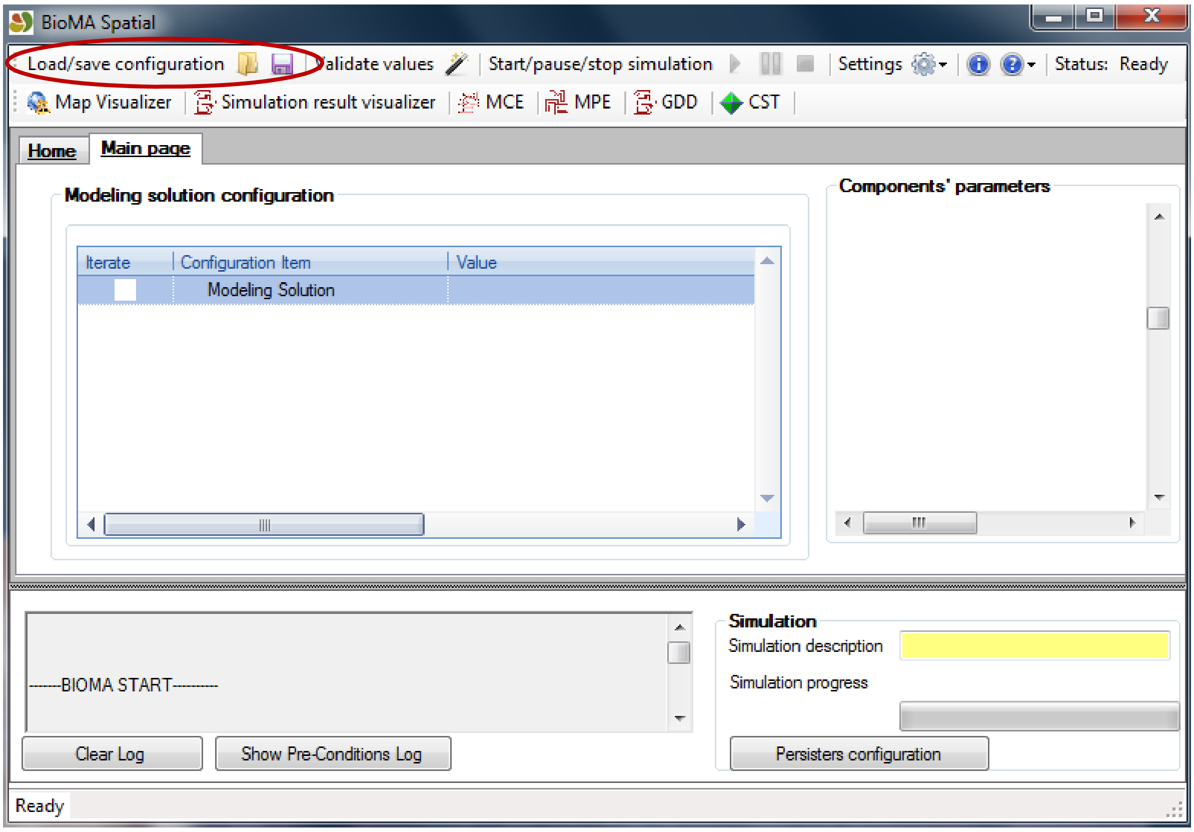

Figure 1 Graphical interface of the simulating tool of BioMA

The user can especially plan, unlike the existing forecasting systems, the agricultural management actions with the use of the application "AgroManagement Configuration Generator" (see Figure 3). The component couples rules (red square) and impacts (blue square): if a specific rule is satisfied (e.g. a specific day of the year, a value of average air temperature repeated for an established number of consecutive days,..), an agro-management event is triggered (e.g. planting, harvesting,..). For each couple rule-impact the user can set the values of parameters using the dedicated box. At the end the agro-management configuration can be saved and visualized.

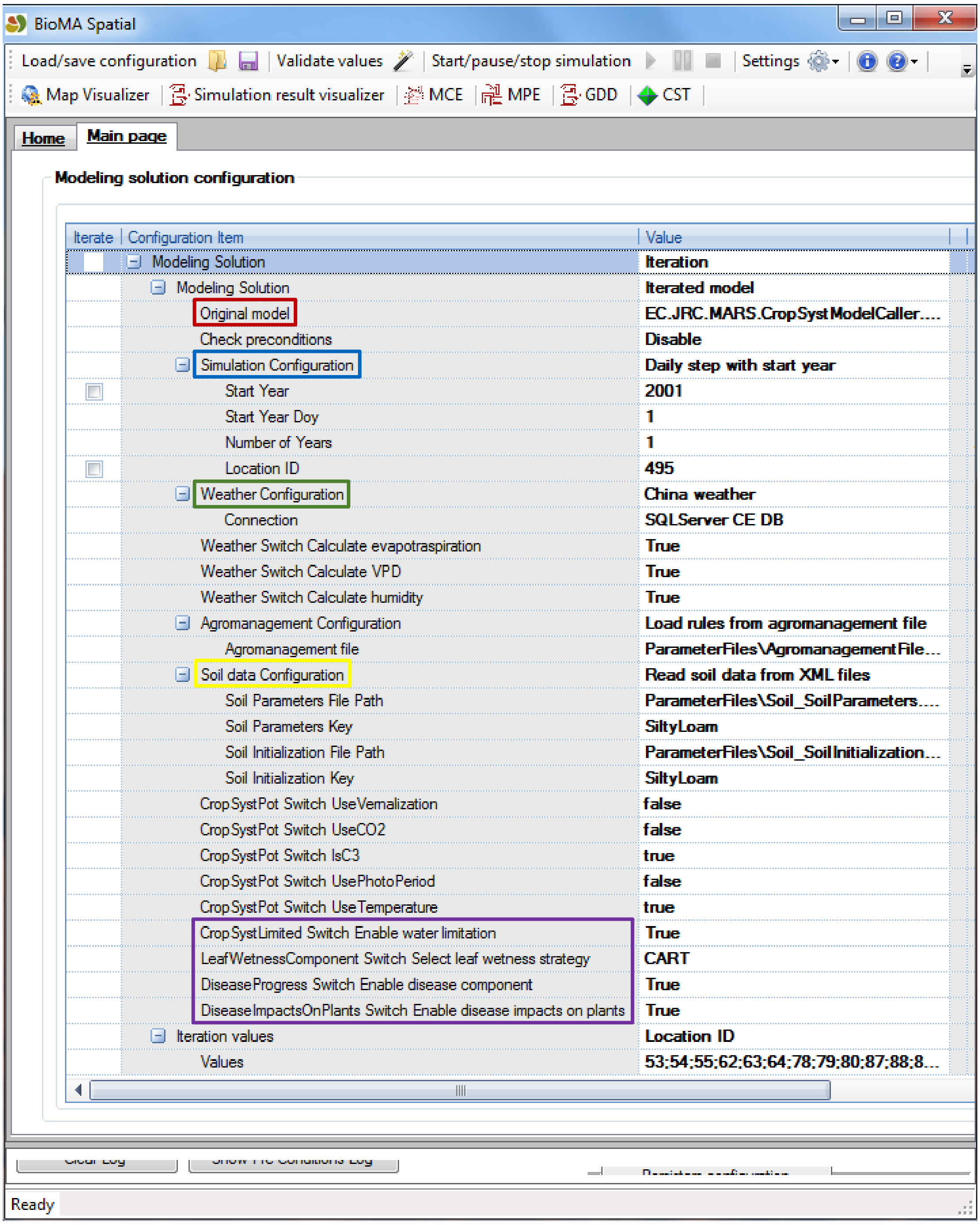

Figure 2 Components of the modeling solution configuration

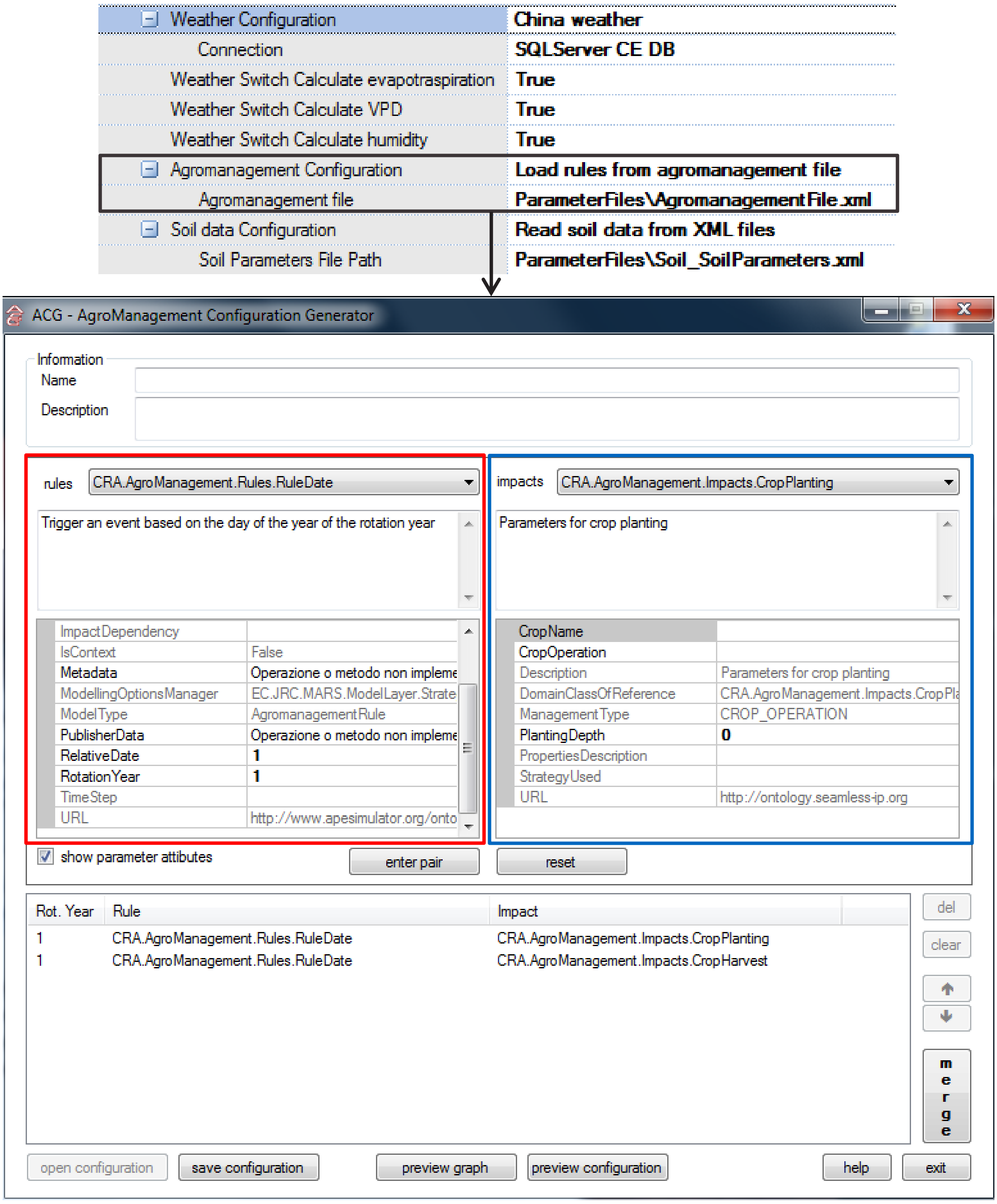

Figure 3 The component 'Agromanagement Configuration Generator'

In the right box of the Bioma interface the user can upload and modify the parameters files, referred to crop growth and development and to the soil and diseases components. The modeling solution configuration setted will be spatially applied to 25 × 25 km cells of a grid covering the studied area. The dimension of each cell derives from the resolution of the weather data, which is 1 degree latitude × 1 degreee longitude. The itaration button, which is the last in the 'modeling solution configuration' box, allows to select the cells where the simulation will be performed, through the component 'Location selector' (see Figure 4). The map and the grid were previosly uploaded and then the user can manually select the cells through the buttons enclosed with the red square in the figure.

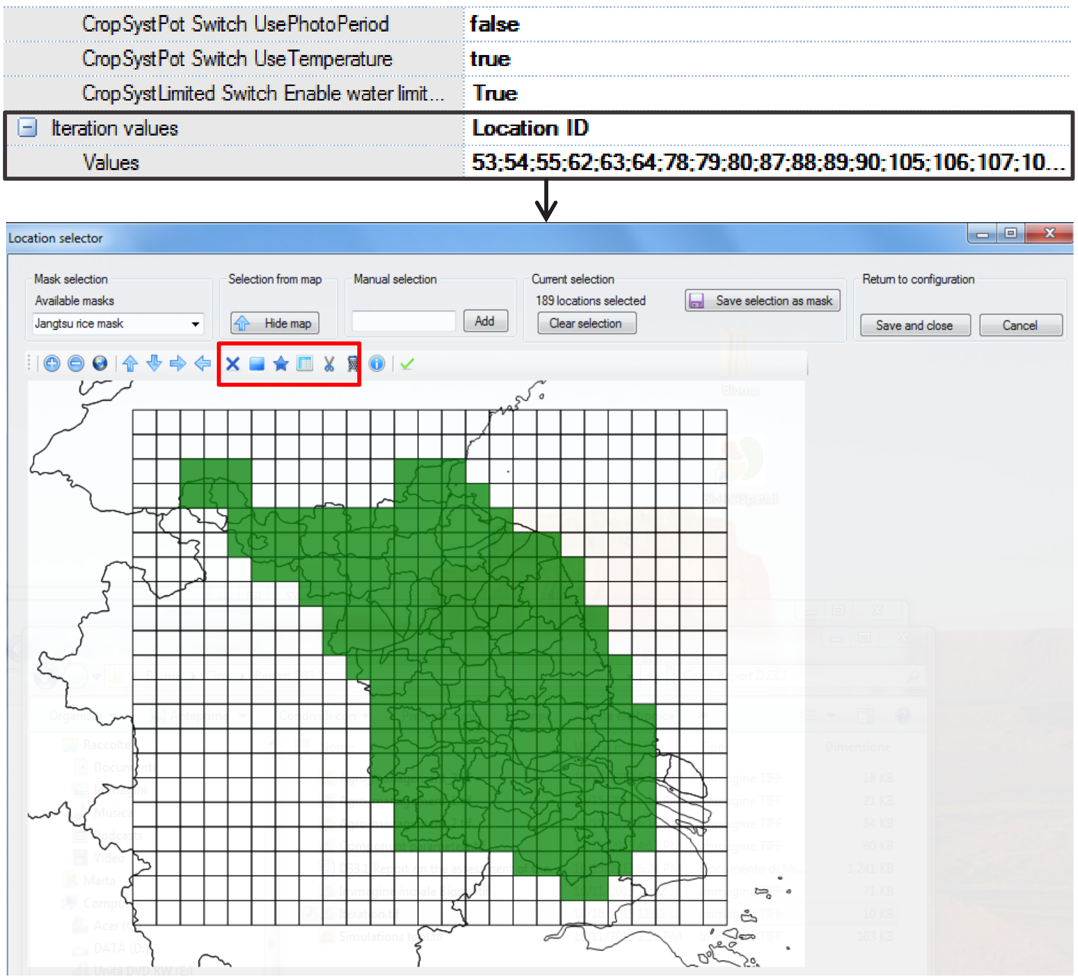

Figure 4 The component 'Location selector'

After setting the configuration the user decides where the outputs will be saved by the system, using the button 'Persisters configuration' enclosed by a red square in Figure 5. For enabling and using the other two tools integrated in BioMA, the outputs have to be stored in the mapping database and in the Access database employed by the statistical tool for yield forecasts. At the end the user validates all the inputs of configuration and starts the simulation, using the buttons enclosed by the green circle in Figure 5.

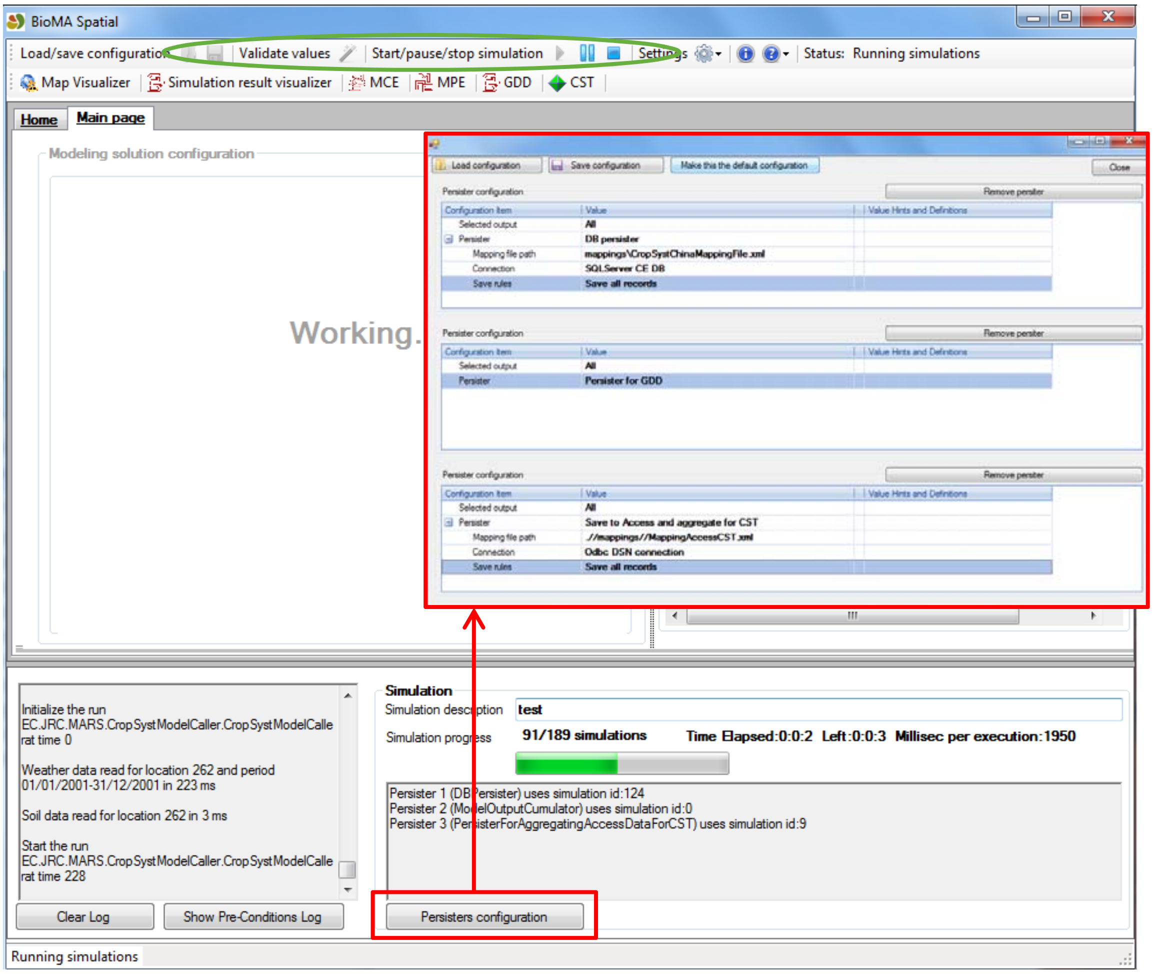

Figure 5 The function 'Persisters configuration'and the beginning/running of the simulation

2.1.2. The mapping tool

The outputs of the simulations can be spatially visualized through the use of the 'map and data visualizer' tool, shown in Figure 6. It is possible to select the simulated potential or limited variables to show in the map (i.e. aboveground biomass, yield, leaf area index or development stage). The user can chose if visualize the cumulated, average, maximum or minimum value of the selected variable during the growing season.

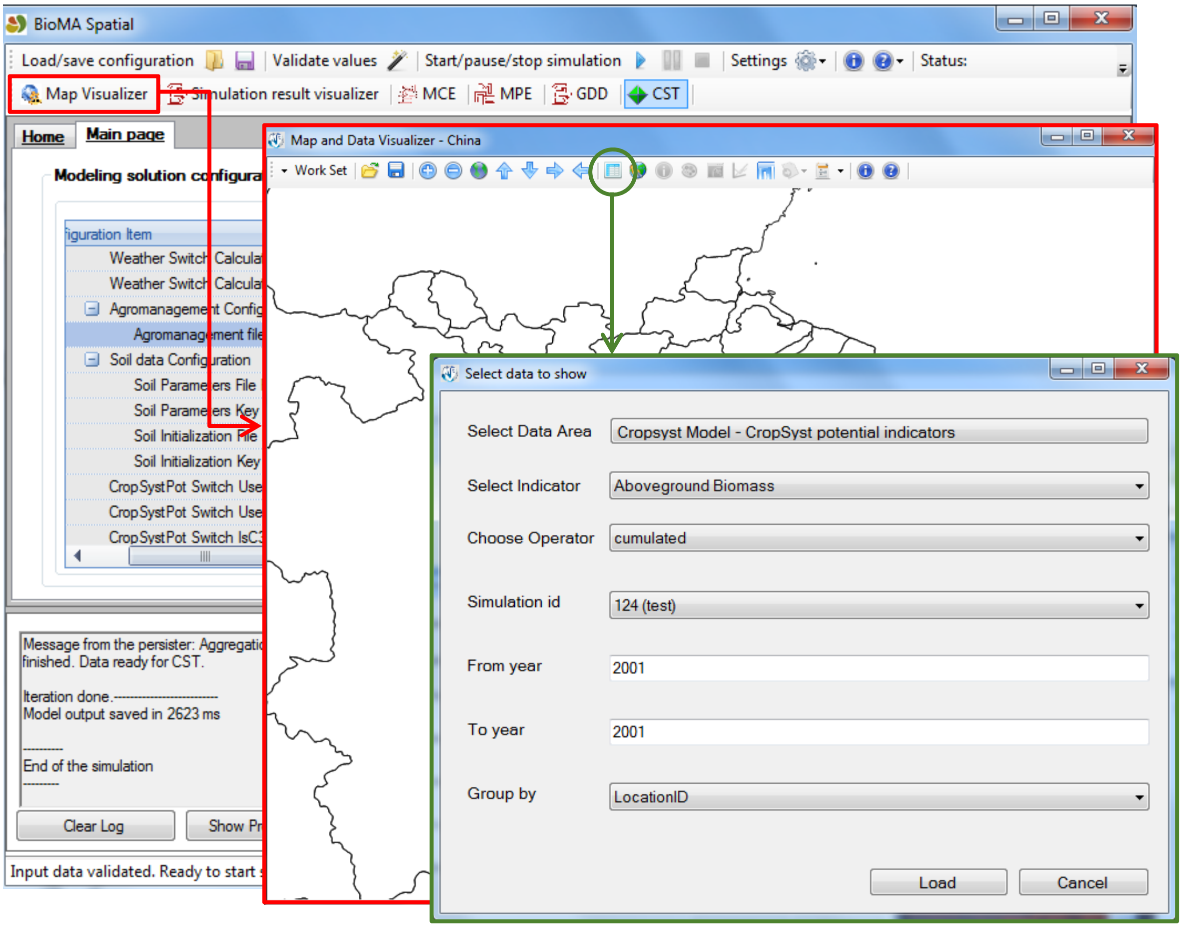

Figure 6 'Map and data visualizer' tool

An example of the map obtained is shown in Figure 7. The user can set the legend and the map layout and export the final map. The most relevant function is the possibility to create the graph of the simulated trend of model outputs during the growing season, selecting the cell and the variable of interest. The graph and the relative CSV can be exported.

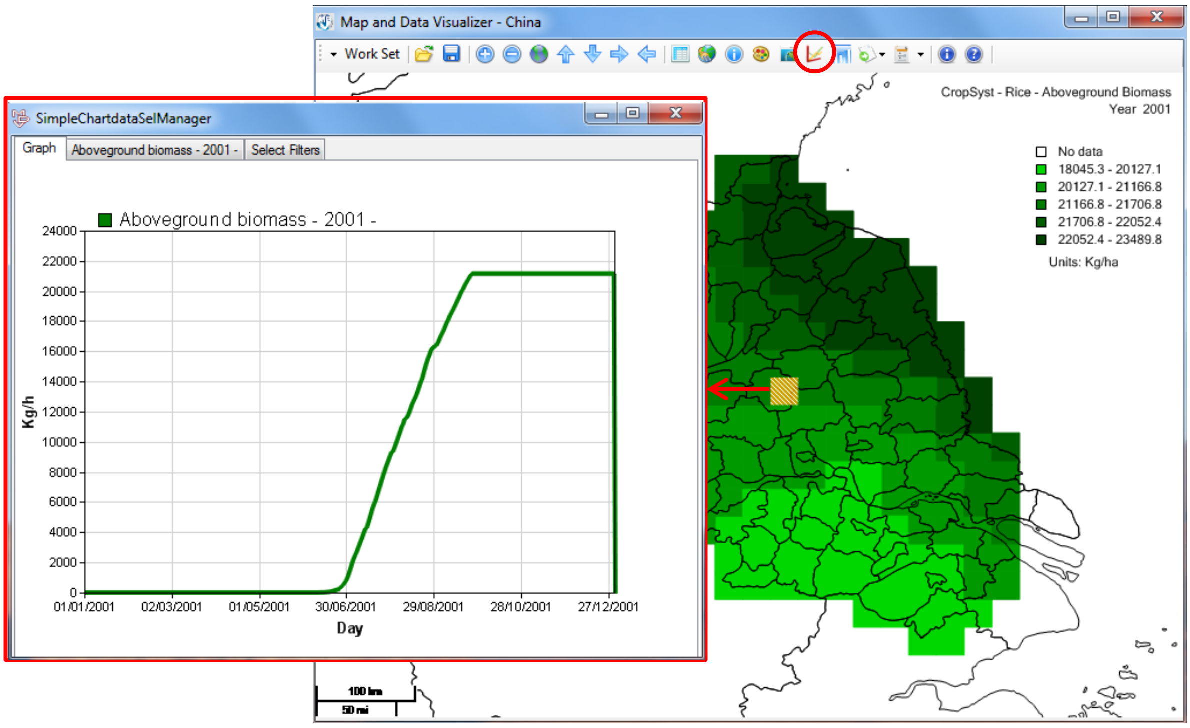

Figure 7 Example of a map shown the pattern of one model output and a variable trend of one cell obtained with the 'map and data visualizer' tool

2.1.3. The statistical tool

The outputs simulated in each grid cell are stored in the database every ten days. In the same database, the aggregated variables are saved. The latters are derived from the data simulated at the elementary simulation units and form the percentage of crop present in each cell. The user can modify the database, adding the values of cultivated area of each cell and the values of official yields for the available years. The button enclosed by a red circle in Figure 8 directly connects the simulation tool with the CST statistical tool. The first step to trigger the statistical analysis is the selection of the crop, the studied area and the decade of the year to obtain the yield forecasting (see Figure 8, green box). Then the user can select the years used by the software to find the regression model able to better explain the variability of official yields (see Figure 8, yellow box). Part of this variability is explained by the introduction of technological improvements. The trend of this phenomenon (e.g. linear, quadratic,..) can be automatically tested by the software or defined by the user (see Figure 8, blue box).

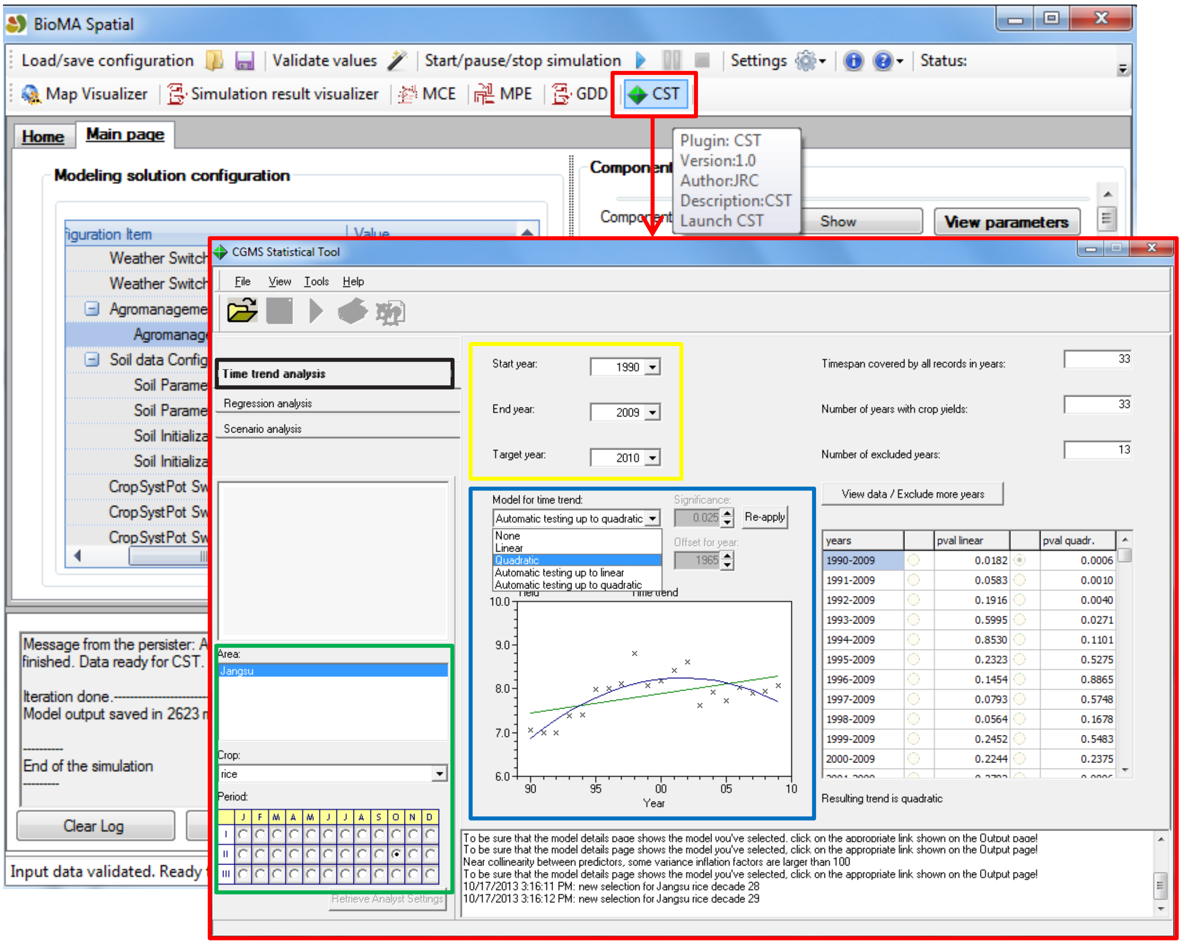

Figure 8 CGMS statistical tool: selection of the crop, area, simulated period and model of the time technological trend

In the statistical analysis section (see Figure 9) the user can decide which simulated variables include as independent variables in the regression model (red box). The options of the regression analysis can also be chosen (blue box). In particular the tool can find the regression models relating official yields with each single simulated indicator or it can obtain a multiple regression combining different simulated variables with official yields.

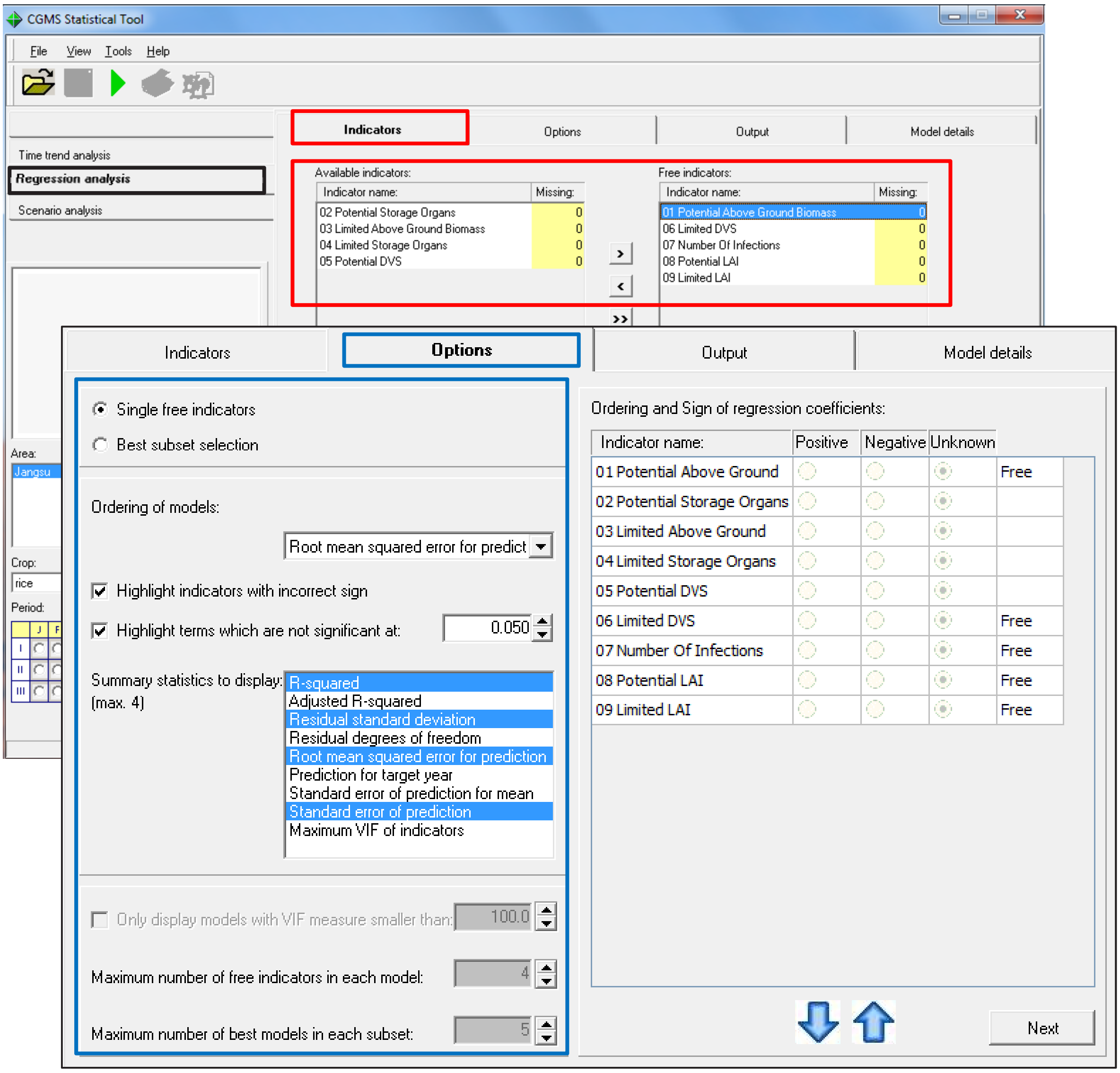

Figure 9 CGMS statistical tool: selection of simulated indicators and options of statistical analysis

At the end of the running, the user can visualize all the regression models calculated by the tool and the related statistical metrics (see Figure 10, red box). The details of each regression model can be shown (blue box): the summary statistics, the regression coefficients and the tabular and graphical comparison between official and processed yields.

Indicators | Options | **Output** | Model details

Model					t-values					
consists of quadr. trend (forced) and free:	R-squared	Residual standard deviation	Root mean squared error for prediction	Standard error of prediction	01	06	07	08	09	lir te
none	64.61	0.32	0.34	0.39	-	-	-	-	-	
+ 08	77.94	0.26	0.31	0.33	-	-	-	-3.110	-	
+ 06	72.29	0.29	0.32	0.36	-	2.106	-	-	-	
+ 09	72.01	0.29	0.32	0.36	-	-	-	-	-2.056	
+ 01	68.31	0.31	0.35	0.39	1.366	-	-	-	-	
+ 07	66.52	0.32	0.36	0.41	-	-	-0.955	-	-	
+ 07 + 08	80.33	0.25	0.29	0.33	-	-	-1.350	-3.246	-	
+ 08 + 09	80.00	0.25	0.31	0.33	-	-	-	-2.448	1.241	
+ 01 + 08	80.49	0.25	0.31	0.33	1.399	-	-	-3.060	-	
+ 06 + 08	78.02	0.27	0.32	0.35	-	0.231	-	-1.977	-	
+ 06 + 07	73.70	0.29	0.33	0.38	-	2.024	-0.896	-	-	
+ 07 + 08 + 09	83.44	0.24	0.29	0.32	-	-	-1.706	-2.879	1.620	
+ 01 + 06 + 08	82.69	0.24	0.29	0.32	1.945	-1.336	-	-2.889	-	
+ 01 + 08 + 09	84.35	0.23	0.31	0.31	1.974	-	-	-3.046	1.859	
+ 06 + 07 + 08	80.33	0.26	0.31	0.35	-	0.048	-1.283	-2.173	-	

Copy to clipboard | Legend: wrong sign | not significant | both not good | Save

Figure 10 CGMS statistical tool: regression models calculated by the tool and corresponding statistical indices

2.2. Spatially distributed simulation tests

In the present report the BioMA platform was applied to rice growing in Jiangsu province (China), using the models CropSyst, WARM and WOFOST previously calibrated for the specific climatic and environmental conditions (see E-AGRI report D32.3). The simulations were performed using ECMWF meteorological data at 25 × 25 km spatial resolution. For this first application the simulations were performed on twenty years (i.e. 1991-2010).

In the first part of the work, the options implemented in BioMA for spatially monitoring crop growth, described in Paragraph 2.1.1 and 2.1.2, were modified and applied. The results obtained for year 2003 are shown as example. Specifically, the following tests were applied:

a) Application of the simulation tool to the three models in potential conditions. The rule-impact management selected was that based on a specific date for planting and harvesting. The sowing date was set to 1th June and the harvesting at the end of the year. The simulations were performed in 189 grid cells (25 ×25 km) where rice is cultivated, according to the rice mask supplied by the Jiangsu Academy of Agricultural Science (see E-AGRI report D32.4). The sets of physiological parameters which regulate crop growth and development inside the models were derived from the calibration activity (see E-AGRI report D32.3).
b) Change of management options. As example the date of planting was moved up from 1th June to 1th May to analyse the effects of possible shift in sowing dates on final production.
c) Simulation of diseases and study of the effects on rice growth. In this preliminary study the default values for *Pyricularia Oryzae* (agent of blast disease) were used for parameterizing the diseases component.

2.3. Yield forecasting tests

In the second part of the work, the whole workflow needed to reach the forecasted yield was carried out using the simulation model WARM. Rice official yields were supplied by the Jiangsu Academy of Agricultural Science. Official yields and simulated indicators from 1990 to 2009 were used by the CGMS statistical tool to find the best regression models, which were applied to forecast rice yield of 2010. Two tests were performed to demonstrate that the modifications on the simulation tool largely affect the results of yield forecasting:

a) The management options were changed moving up the sowing day from 1th June to 1th May, as done in the spatially distributed simulations test b). The results obtained using the two different planting dates and considering only potential indicators are shown and compared.
b) Spatially distributed simulations influenced by the effects of pathogens were performed. The effects of disease-limited indicators on rice yield forecast are shown and discussed.

3. Results and discussion

3.1. Spatially distributed simulations tests

The outputs obtained from the three tests are shown in Figure 11 (aboveground biomass) and Figure 12 (leaf area index). The maps were created using the mapping tool integrated in BioMA (see Figure 7).

a) Final biomass and maximum leaf area index simulated by CropSyst, WARM and WOFOST are shown in Figure 11a and Figure 12a, respectively. This first test shows that the values of the two variables and the patterns simulated by the three models were different. Each biophysical model is effectively characterized by a specific approach in reproducing processes related to crop growth and development, thus differently reacting to various climatic and environmental conditions. The topic is specifically discussed in the AGRI report D32.4.

b) Figure 11b and Figure 12b show the effects of the change of management on the accumulation of biomass and LAI, respectively. In the specific case the sowing date was moved one month forward. The final biomass simulated by the three models maintained almost the same pattern. However it can be noted a general increase of the biomass accumulated. This phenomenon was probably due to the anticipation of sowing date which exposed the crop to lower temperatures, causing a delay of development and an extension of the cycle.

   The anticipation of planting date caused in each simulation model a homogenization of LAI values, which did not follow a trend. It can be also noted a marked increase of LAI values simulated by WOFOST.

   The user can also decide to change the type of rule, not using a fixed sowing and harvesting date, but triggering the management event if a particular condition happens (e.g. fixed numbers of rainy days, air temperature higher than a fixed value for a set number of days, etc.).

c) The results of the third test are shown in Figure 11c and Figure 12c. The sowing date was the same applied in test a), thus isolating and studying the effects of diseases on plant growth. The theoretical effects on the plant are the same for the three models. However, the different realizations in terms of approaches used to formalize each step of the disease progress led to markedly different responses in CropSyst, WARM and WOFOST. The highest impacts were shown by CropSyst with the level of production almost halved (i.e. from 15 - >19 t ha^{-1} to <7 – 11 t ha^{-1}). WARM and WOFOST kept the same spatial trend of potential production reducing the accumulation of biomass especially in the grid cells where a greater number of infections was simulated (see *Figure 13*).

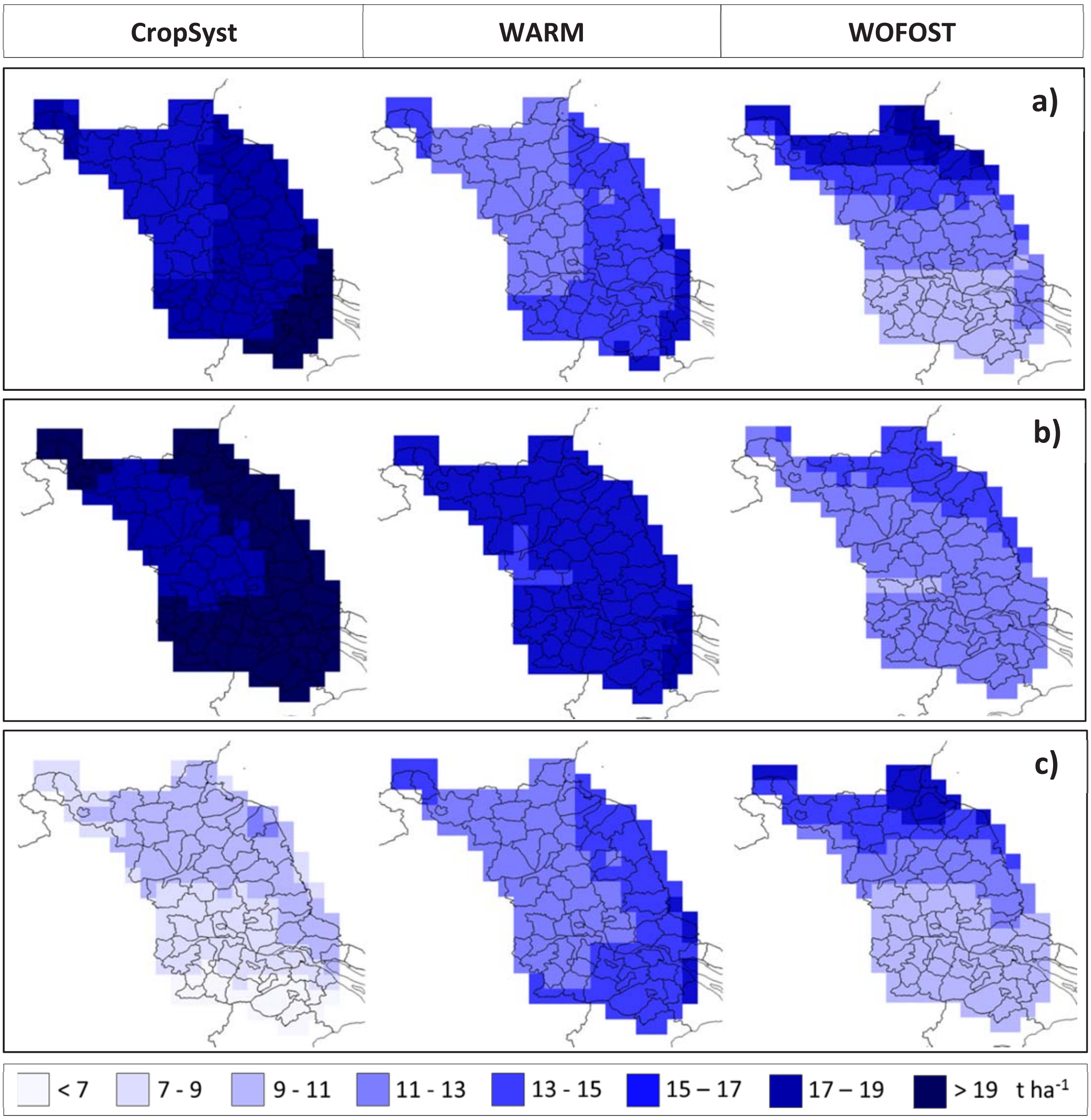

Figure 11 Final aboveground biomass simulated by CropSyst, WARM and WOFOST a) in potential conditions; b) after the change of management options; c) accounting the effects of diseases.

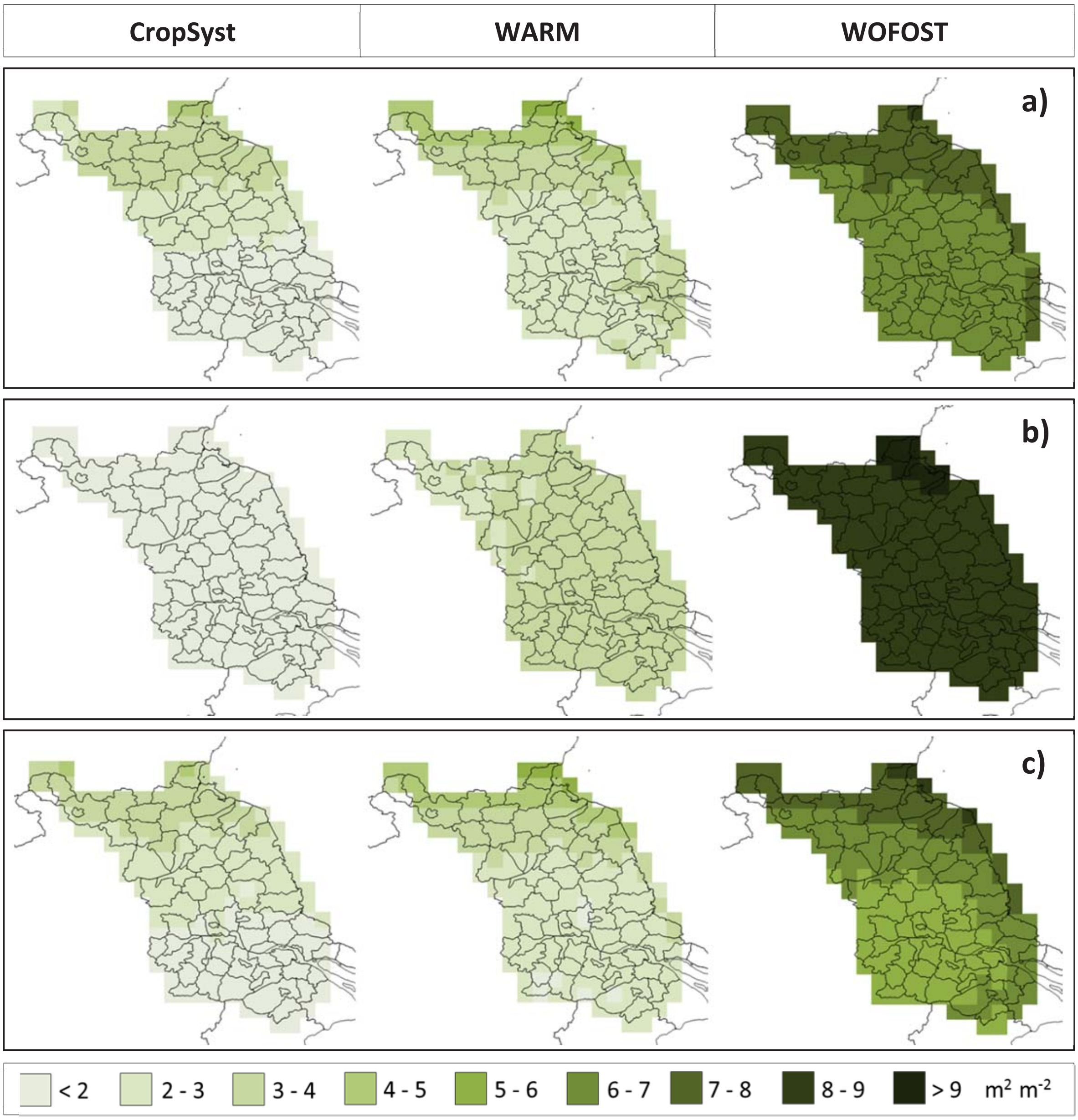

Figure 12 Maximum leaf area index simulated by CropSyst, WARM and WOFOST a) in potential conditions; b) after the change of management options; c) accounting the effects of diseases.

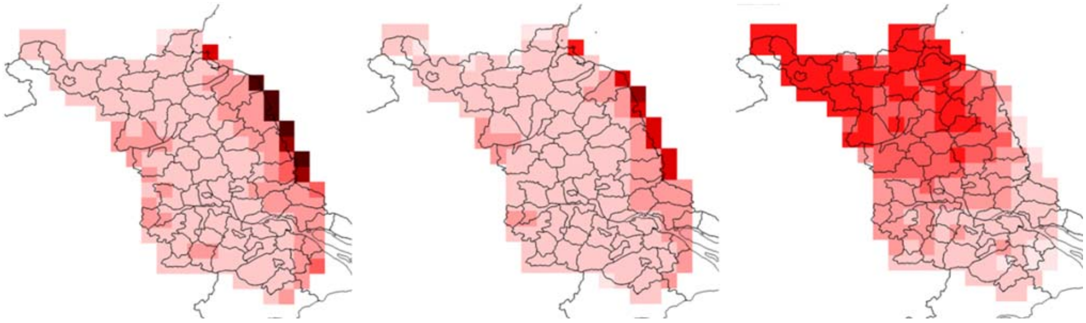

Figure 13 Number of infections influencing the accumulation of biomass simulated by CropSyst, WARM and WOFOST

3.2. Yield forecasts tests

The first step is to establish the type of time trend followed by rice official yields from 1990 to 2009. Figure 14 shows that the trend seems quadratic. This was confirmed by the automatic test applied by the statistical tool.

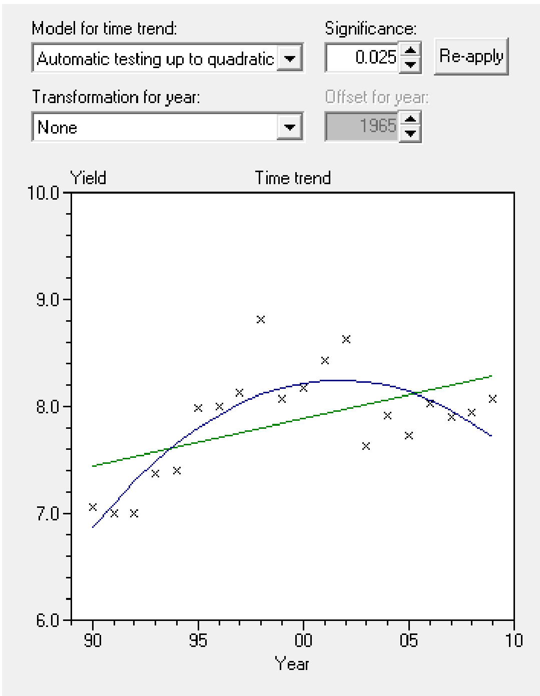

Figure 14 Time trend followed by rice official yields from 1990 to 2009

The results of the two tests are shown below:

a) It was decided to obtain the forecast at maturity which, according to the simulated development, in almost all the grid cells occurred the first decade of October and September for the sowing date set at 1th June and 1th May, respectively.

The simulated indicators used to find the best regression models able to explain the variability of official yields are:

- Potential aboveground biomass (01)
- Potential storage organs (02)
- Potential development stage code (05)
- Potential leaf are index (08)

The outputs show that the technological time trend explained 64% of the varibility of official yields. Figure 15 shows the variability explained by the simple and multiple linear regression models. The latters derived from the combination of the different simulated indicators up to a maximum of four indicators per model.

It is clear that the best result in both the tests was obtained by the multiple regression model which combined all the potential simulated indicators. However the anticipation of the sowing date caused a reduction of almost 10 % of the explained variability.

SOWING DATE: 1th June

Model consists of quadr. trend (forced) and free:	R-squared
none	64.61
+02	77.67
+08	73.32
+05	69.08
+01	68.53
+08 +02	78.93
+05 +02	78.16
+02 +01	79.74
+08 +01	75.44
+08 +05	74.19
+08 +05 +02	86.67
+08 +05 +01	84.75
+05 +02 +01	80.48
+08 +02 +01	79.95
+08 +05 +02 +01	87.39

SOWING DATE: 1th Mav

Model consists of quadr. trend (forced) and free:	R-squared
none	64.61
+02	69.32
+08	64.98
+05	64.71
+01	65.35
+05 +02	73.32
+08 +02	71.24
+02 +01	71.99
+08 +05	65.29
+08 +01	65.41
+08 +02 +01	76.33
+08 +05 +02	73.43
+05 +02 +01	75.43
+08 +05 +01	67.00
+08 +05 +02 +01	76.33

Figure 15 Regression models and percentage of variabily explained, setting up the sowing date of spatial simulations at 1th June and 1th May.

Figure 16 shows the comparison between official and estimated yields derived from the application of the best regression model. Red circles in Figure 16b underline that after the change of management the regression model markedly underestimated the observed yield of 1998 and 2009. This is the main reason causing the reduction of the coefficient of determination value.

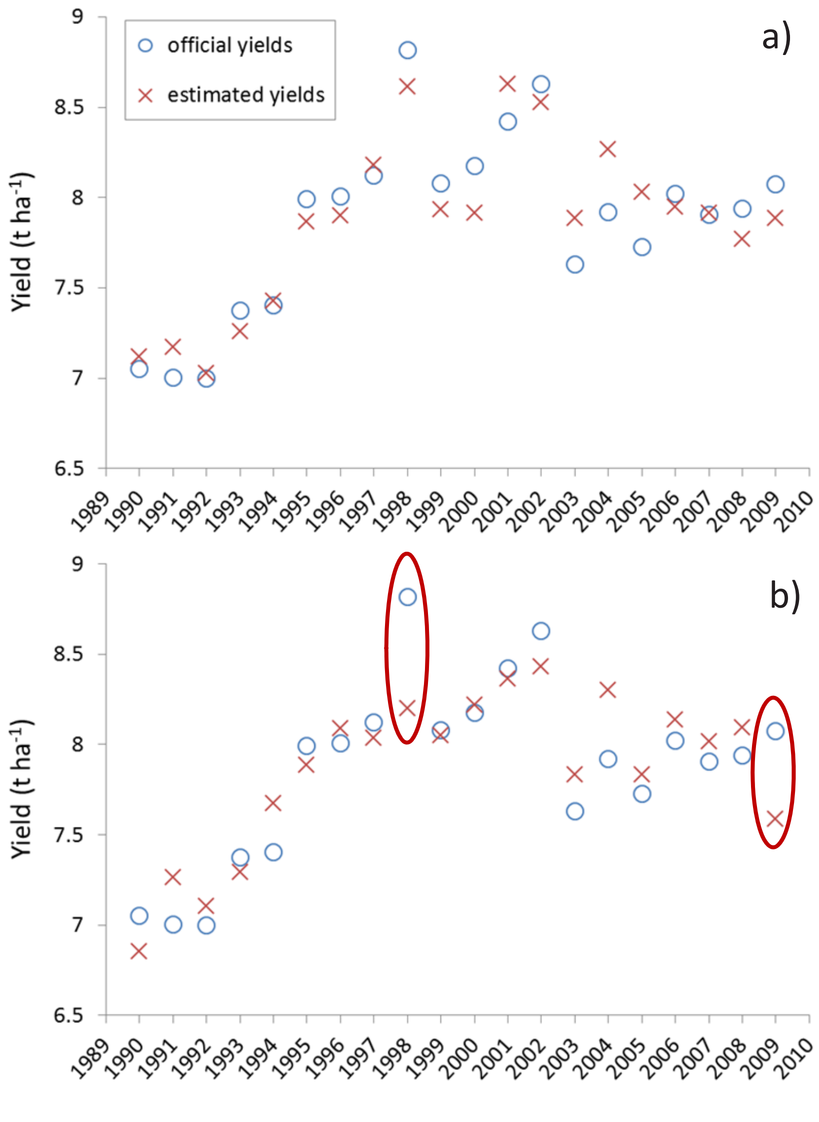

Figure 16 Comparison between observed and estimated yield values derived from the best regression model, setting up the sowing date of spatial simulations at a) 1th June 2003; b) 1th May 2003.

b) The simulated indicators used to find the best regression model able to explain the variability of official yields are:
- Potential aboveground biomass (01)
- Potential storage organs (02)

- Potential development stage code (05)
- Potential leaf are index (08)
- Limited aboveground biomass (03)
- Limited storage organs (04)
- Limited leaf area index (09)
- Number of infections (07)

Figure 17 shows the variability explained by the simple and multiple linear regression models.Considering the indicators derived from the simulation of diseases effects, the coefficient of determination of the best regression model increases from 87.39 % to 91.69 %. In the latter case the best model derived from the combination of two potential indicators (i.e. storage organs and development stage code) and two limited indicators (i.e. storage organs and leaf area index).

Model consists of quadr. trend (forced) and free:	R-squared
none	64.61
+04	80.30
+02	77.67
+08	73.32
+09	69.76
+03	69.74
+02 +04	86.12
+04 +07	80.30
+04 +05	80.99
+04 +08	81.20
+01 +04	83.42
+04 +05 +08	88.93
+02 +04 +05	86.15
+04 +08 +09	87.39
+02 +04 +08	88.14
+02 +08 +09	87.13
+03 +04 +05 +08	89.65
+04 +05 +07 +08	88.97
+02 +04 +05 +08	91.26
+01 +04 +05 +08	89.15
+02 +04 +05 +09	91.69

Figure 17 Regression models and percentage of variably explained considering both the potential and limited indicators in the statistical analysis

Figure 18 shows the comparison between official and estimated yields derived from the application of the best regression model. The inclusion of limited indicators increases the explanation of yields variability of almost 5%. This is due to a slight improvement of the yield estimation in most of the years of the time series. The most marked improvement can be visualized in the period 1998-2002, underlined by a green circle in Figure 18.

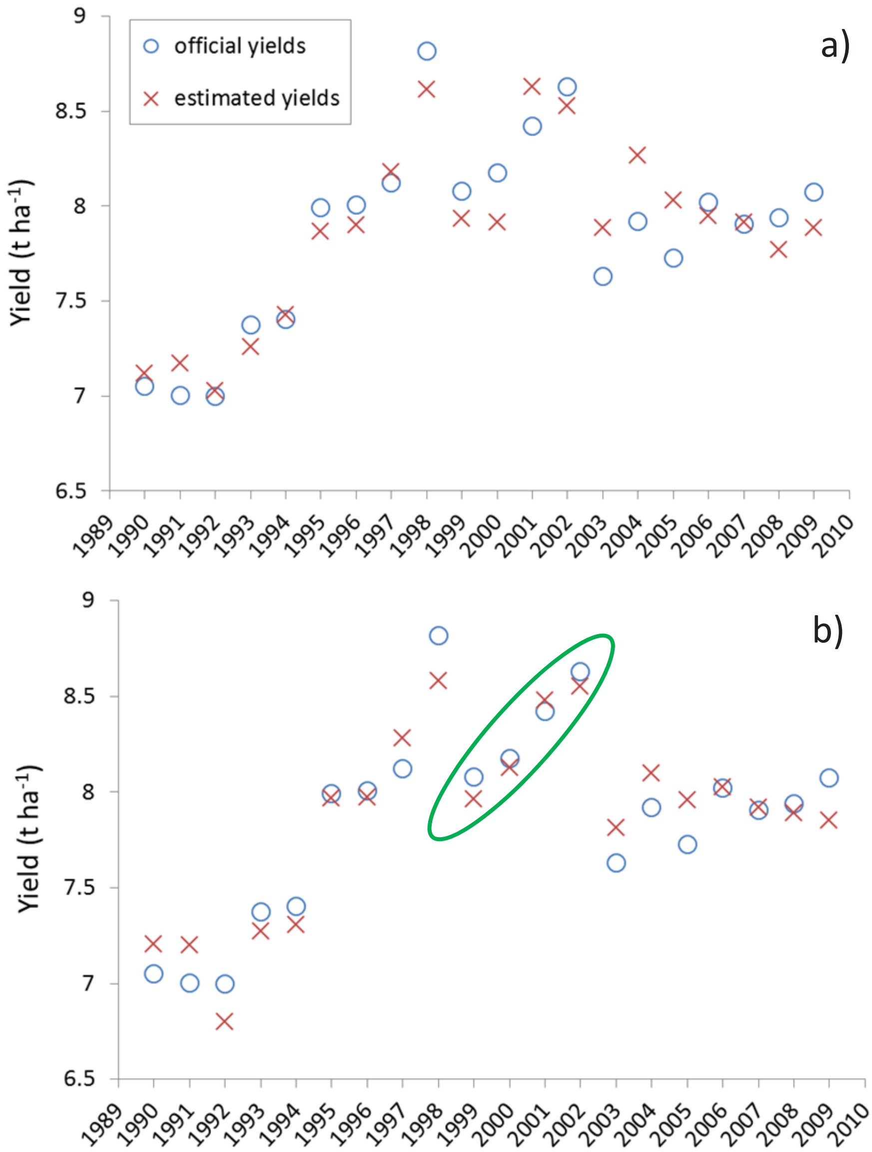

Figure 18 Comparison between observed and estimated yield values of the best regression model, considering a) potential indicators; b) potential and limited indicators.

4. Conclusions

The BioMA platform allows – via a user friendly simulation/forecasting environment – the user to:

- run multi-model simulations;
- create specific modeling solutions (e.g. modifying some management options and visualizing the different effects on the final production);
- consider the effects of biotic and abiotic damages on crop growth;
- spatially visualize and compare the outputs of simulations through the 'map visualizer tool';
- forecast yields through the CST statistical tool, considering the effects of the management change and of the simulation of biotic damages.

The user can easily run the simulations, perform mapping and apply the statistical software integrated in BioMA, without the need of opening and modifying directly any database.

In this work, the platform was specifically applied to simulate rice growth and development in Jiangsu province in the period 1990-2010.

The spatially distributed simulations tests revealed that CropSyst, WARM and WOFOST differently reacted to the modification of management options and to the simulation of pathogens impacts on the crop. The three simulation models effectively used specific approaches to reproduce crop growth and development. This phenomenon led CropSyst, WARM and WOFOST to differently react to specific climatic, management and environmental conditions. The outputs simulated by CropSyst showed the higher reduction caused by the impacts of pathogens on the crop.

The yield forecasts tests revealed that the choices achieved during the construction of the modeling solution, affected the results of the prevision. Especially, the simulation of the diseases progression and the effects on plant allowed the user to insert the related variables as indicators for the statistical analysis. The multiple regression models accounting for both potential and diseases-limited variables explained more variability of official yields than regression models considering only potential indicators. In the specific case the simulation of the effects of *Pyricularia Oryzae* caused an increase of almost 5% of the coefficient of determination of the best regression model which used only the simulated potential outputs.

This is the first time a simulation/forecasting environment including plant-diseases interaction is presented and evaluated.