International Congress on Environmental Modelling and Software

7th International Congress on Environmental Modelling and Software - San Diego, California, USA - June 2014

Jun 17th, 10:40 AM - 12:00 PM

Reimplementation and reuse of the Canegro model

Tommaso Stella
Universita degli Studi di Milano, tommaso.stella@unimi.it

Valentina Pagani
Universita degli Studi di Milano

Caterina Francone
Universita degli Studi di Milano

Sevim Seda Yamaç
Universita degli Studi di Milano

Giacomo Finotto
Universita degli Studi di Milano

See next page for additional authors

Stella, Tommaso; Pagani, Valentina; Francone, Caterina; Yamaç, Sevim Seda; Finotto, Giacomo; Bregaglio, Simone; Ceotto, Enrico; and Confalonieri, Roberto, "Reimplementation and reuse of the Canegro model" (2014). *International Congress on Environmental Modelling and Software*. 10.
https://scholarsarchive.byu.edu/iemssconference/2014/Stream-H/10

Presenter/Author Information

Tommaso Stella, Valentina Pagani, Caterina Francone, Sevim Seda Yamaç, Giacomo Finotto, Simone Bregaglio, Enrico Ceotto, and Roberto Confalonieri

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Reimplementation and reuse of the Canegro model

Tommaso Stella[a], Valentina Pagani[a], Caterina Francone[a], Sevim Seda Yamaç[a], Giacomo Finotto[a], Simone Bregaglio[a], Enrico Ceotto[b], Roberto Confalonieri[a]

[a]*Università degli Studi di Milano, Department of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, Cassandra lab, via Celoria 2, I-20133 Milan, Italy, tommaso.stella@unimi.it*
[b]*Consiglio per la Ricerca e la sperimentazione in Agricoltura, Via di Corticella 133, 40128 Bologna, Italy*

Abstract: Model reuse can be limited by software design, which often forces third parties to completely rewrite new versions of existing models before adapting them to new needs. This tendency removes resources from the development and improvement of models, and from the extension of their domain, leading to the proliferation of software having many different implementations of the same algorithms. A component-oriented paradigm allows these limitations to be overcome, facilitating massive model reuse and extension. This study presents the application of component-oriented principles to the reimplementation of the sugarcane (*Saccharum officinarum* L.) model Canegro (DSSAT v4.5) in a framework-independent component following the BioMA (Bio-physical Model Applications) architecture. Both opportunities of reuse and extension of the component are discussed in this work. Reuse was emphasized through the development of a model-based yield forecasting system for sugarcane in Brazil. The BioMA Spatial application was applied to perform (i) the spatially distributed simulations with the Canegro component, and (ii) the yield forecasting at the state level. The potential for extension of the Canegro component was demonstrated by its straightforward adaptation for giant reed (*Arundo donax* L.), a promising energy crop that shares several morphological and physiological features with sugarcane.

***Keywords**:* Canegro, reusability, extensibility, Arungro, yield forecast

1. INTRODUCTION

Although the advantages of model reuse and extension are well known and widely recognized within the international modelling community (Holzworth et al., 2010), the design of agro-environmental models often prevents – to a large extent – these activities, forcing third parties interested in modifying an existing model to re-implement it from the beginning (Donatelli et al., 2008). This time-consuming process removes resources that might otherwise be available for model improvement; moreover, it often leads to solutions again characterized by a low level of extensibility. In other cases, reimplementation leads to new versions of a model that are extensible but tightly coupled to specific frameworks, and this generates dependences that, in turn, limit the model reusability in other simulation environments (Donatelli et al., 2012). Resources needed for reimplementation increase with model complexity and this, coupled with the difficulties in improving or extending the way processes are formalized, often restricts the opportunity to revise and improve agro-ecological models only to their original developers.

The component-oriented paradigm allows these limitations to be overcome favouring model reuse by limiting dependencies, specifying interfaces, and encapsulating the algorithms in discrete units (Szyperski, 2002). As a consequence, models implemented in framework-independent libraries can be effectively used, composed and improved by third parties (Argent, 2005) by implementing the design pattern Adapter (Gamma et al., 1994), i.e., by developing adapters to specific frameworks. An example of the application of these principles to agro-environmental models is given by the BioMA (Biophysical Model Applications; http://bioma.jrc.ec.europa.eu/index.htm; Donatelli et al., 2012) platform of the European Commission, where the focus is shifted from the framework to the development of independent software, intended as libraries of models for the simulation of different sub-domains (e.g., crop growth and development, soil hydrology, plant-pathogen interactions).

This paper presents a component-based reimplementation of the sugarcane model Canegro (Inman-Bamber, 1991); in particular, the version of the model we started from is the one implemented in the DSSAT suite (Jones et al., 2008). UNIMI.Canegro was re-built as a framework-independent .NET 3.5 software library, characterized by a fine level of granularity. Each simulated process (e.g., photosynthesis) is made up of independent basic units (e.g., light interception, carbon fixation, maintenance and growth respirations) which can be easily substituted by alternative approaches. The main advantages of this kind of re-implementation lie (i) in the possibility of sharing knowledge via ready-to-use software units, because of the fine granularity and of the absence of dependencies, and (ii) in the ease of reuse and extension of the component algorithms. Both the possibilities of reuse and extension of the component are explored in this paper. The former, via the reuse of UNIMI.Canegro in a model-based yield forecasting system, considering that crop forecasting is one of the greatest potential applications of crop modelling (Bannayan and Crout, 1999). The system was developed to predict sugarcane yield in the state of São Paulo, responsible for 60% of sugarcane production in Brazil. The latter, with the extension of the component to the simulation of giant reed (*Arundo donax* L.), a promising energy crop (Pilu et al., 2012) which shares several morphological and physiological traits with sugarcane.

2. MATERIALS AND METHODS

2.1 Software Architecture

The software design promotes component reusability by limiting dependencies and providing a semantically rich, public interface (i.e., IStrategyCanegro). According to the design pattern façade, this interface is implemented by all the simple and composite model units included in the UNIMI.Canegro component (i.e., a strategy). More specifically, a simple strategy is an indivisible unit of algorithm coherently representing a sub-process, i.e., the smallest piece of an algorithm for which alternate approaches exist or could exist in the future. Simple strategies are composed into objects of increasing complexity, that, according to the composite pattern, are in turn subject to composition, leading to a hierarchical structure culminating with a composite strategy that represents the whole model (Figure 1). This architecture allows the extension of the component by simply adding new strategies implementing original or existing modelling approaches to reproduce the process of interest. The implementation of the bridge pattern implies the separation of model algorithms from data-types structures (i.e., domain classes, that can be extended independently) in two different components (Donatelli and Rizzoli, 2008). This pattern allows the substitution of modeling approaches, that are non-unique by definition, without changing the interfacing between I/O services and domain description, that does not vary according to the modelling approach used. The domain classes describe the biophysical domain by including inputs and outputs of the model with their attributes, whereas the ontology of the parameters, related to the specific modeling representation and not to the domain, is made available via the related strategies. The coherence of input, output and parameter values with their ontology can be verified through the test of pre- and post-conditions, according to the design-by-contract approach (Meyer, 1997). UNIMI.Canegro is distributed in a Software Development Kit including hypertext files documenting the code, the implemented approaches, the software design and the code (Visual Studio 2010) of a sample application illustrating how to use it. Strategies, domain classes and interfaces can be inspected via an external application named Model Component Explorer (http://agsys.cra-cin.it/Tools/MCE/help/). UNIMI.Canegro can be coupled to other available .NET framework-independent components for the simulation of, e.g., soil water balance (http://agsys.cra-cin.it/tools/soilw/help/) or plant-pathogen interactions (http://agsys.cra-cin.it/tools/diseases/help/).

2.2 Component reuse: the sugarcane yield forecasting system

The UNIMI.Canegro component was implemented in the software platform BioMA, which is developed for analysing and running –on explicit spatial units– modelling solutions based on biophysical models. The component-based structure allows developing a personalized configuration, choosing the crop simulation model and the processes influencing plant growth (e.g. water stress, effects biotic and abiotic factors) and modifying the related parameters.

In this work, the Canegro component was used to simulate sugarcane growth and development in potential conditions in the State of São Paulo. The set of parameters describing the morpho-physiological characteristics of the crop was that obtained by Marin et al. (2012), based on a set of experimental data from the sugarcane cultivar RB867515, extensively grown in Brazil. The resprouting date was fixed at the beginning of January and the simulation keeps on until the end of the succeeding year, considering that RB867515 is a medium-late maturing cultivar generally harvested in August-September. The modelling solution configuration setup was 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 European Center for Medium-Range Weather Forecasts (ECMWF) Era-Interim meteorological data used to perform the simulations.

The outputs simulated in each grid cell were stored in the database every ten days and were then aggregated at state level on the basis of the crop percentage present in each elementary unit. The latter information was derived from the georeferenced database made available online by the Brazilian Enterprise for Agricultural Reasearch (EMBRAPA). The model indicators together with some agro-meteorological indexes aggregated at the decade when the forecast is needed were post-processed together with a 20 years (i.e. 1991-2011) series of statistical sugar yields. Assuming that the simulated variables and the official yields are, respectively, the independent and dependent variables of a linear multiple regression, a leave-one-out cross-validation procedure was applied to identify the regression model able to explain the highest percentage of the inter-annual variability in sugarcane yields.

2.3 Component extension: from Canegro to Arungro

The interest in giant reed as an energy crop has increased only recently and this is probably the reason why crop models for simulating this species are not available. Moreover, some peculiar features of giant reed (e.g., the presence of a rhizome and its role in determining the rate of stalk emission) make generic crop simulators (e.g., CropSyst, WOFOST, STICS) unsuitable and attempts to target the calibration of their parameters would likely lead to poor performance and to inconsistent calibrations. The Canegro model, in its component version, was used as a starting point for the definition of a new, explanatory giant reed model (Arungro, hereafter). This choice is suggested by the affinity of giant reed to sugarcane, which share many morphological and physiological traits. Canegro is detailed in the description of sugarcane growth and development processes, that can be modulated via the calibration of model parameters to properly reproduce giant reed behavior. For the processes where the differences between the two species required the formalization of new algorithms, the extensibility of UNIMI.Canegro simplified the modification of specific modelling approaches via the addition of alternative strategies. An example for this refers to the way leaf area index (LAI) and its photosynthetically active fraction (GLAI) are simulated: observations, indeed, revealed differences between the two species, mainly related with the different dynamics of leaf development and senescence. For this aspects, the peculiar giant reed features were thus considered by developing two new strategies for substituting two of those composed in the UNIMI.Canegro strategy LeafAreaIndexC (see Figure 1), thus keeping most of the Canegro strategies unchanged. Moreover, the software architecture favored the revision of some of the Canegro strategies involved in biomass partitioning and maintenance respiration, allowing also the inclusion of the effect of rhizome biomass on tiller population at spring restart. Accounting for all the modified strategies, the extended version of the component (i.e., UNIMI.Arungro) shares with UNIMI.Canegro about 70% of the strategies. The Arungro model was calibrated and evaluated using data collected in two experimental sites located in the alluvial plain of the Po Valley (Northern Italy) characterized by mild continental climate and different management conditions.

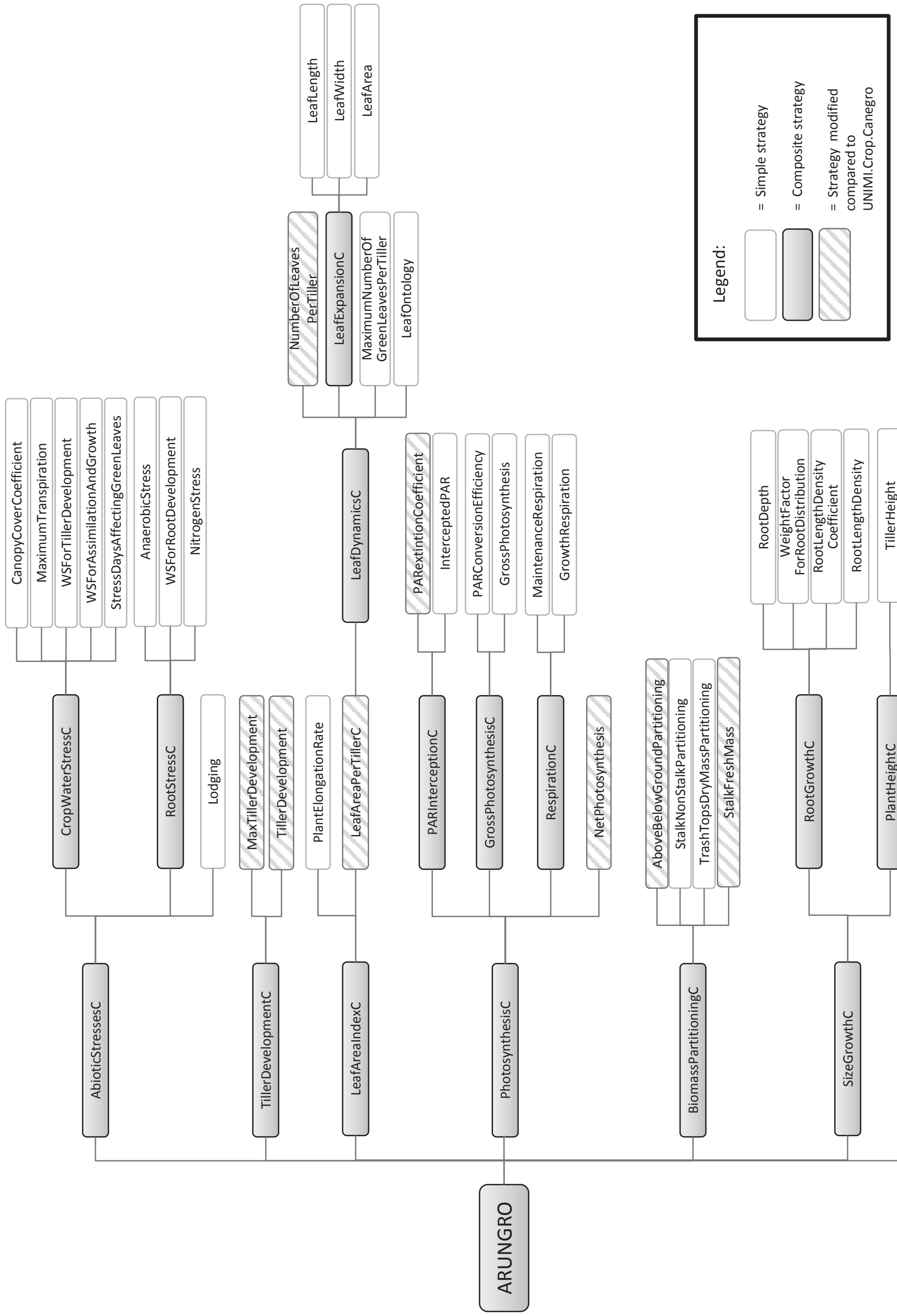

Figure 1. Strategy diagram of Arungro. Differences with the Canegro model are highlighted.

3. RESULTS AND DISCUSSION

3.1 Analysis of forecasted yields

The regression model performing best in the period 1991-2011 was able to explain the 66.4% of the variability of sugarcane official yields, using four indicators as regressors. The selected regressors were both agro-meteorological indexes like, e.g., the cumulated reference evapotranspiration in the growing season, and aggregated outputs of the crop model. The result is particularly satisfactory in light of the absence of a technological-driven trend influencing sugarcane yields in the state of São Paulo. The steep increase in sugar production outlined by historical data is in fact due to the expansion of sugarcane cropped area, rather than to an improved efficiency of the cropping system. Similar yield forecasting systems can explain 80-90% of the inter-annual yield variability, however the trend alone explains even more than half of it (Pagani et al., 2013). The performances of the BioMa-based system are therefore comparable to other widely used forecasting systems. Moreover, the system reproduces quite accurately the fluctuations of the official yields time series, except for the period 1993-1996 (Figure 2). In particular, it shows the best performances in the window 1998-2001, reproducing precisely the marked decreasing trend from 1999 to 2000. The regression model was then applied to the target year (i.e., 2012): the increasing trend –in respect to 2011– was well reproduced by the forecasting system, although a marked overestimation of the yield was observed. A similar overestimation can be noticed also in the previous two years (2010 and 2011). Predictions are expected to improve once the adherence of forecasting system to the real system will be improved. To reach this goal, when the BioMA based system will be applied at national level, it will be extended with a component for the simulation of the soil water balance (automatic rules for irrigation will be implemented where needed). Moreover, different parameterization of the model will be defined in order to characterize the most important cropped varieties. According to the different managements (i.e., planting and harvest dates), different instances of the crop model will be run simultaneously, aggregating model outputs coherently with the relative importance of each management strategy.

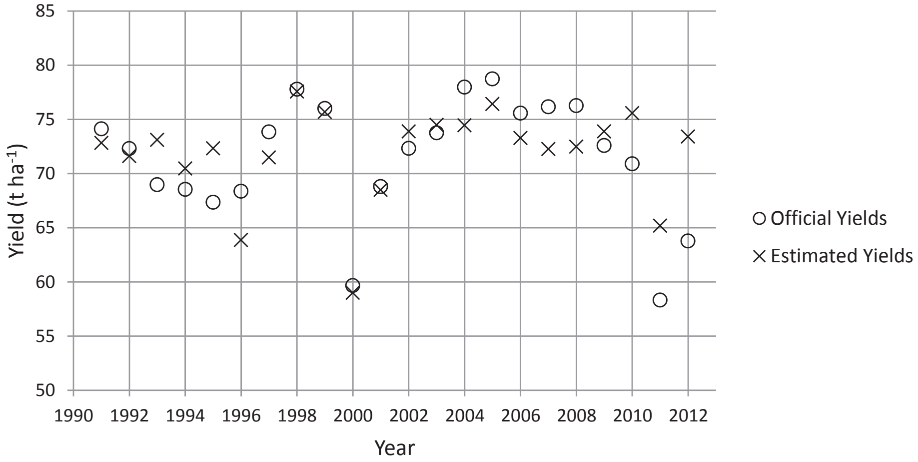

Figure 2. Comparison between official yield statistics and estimated yields in the period 1991-2012.

3.2 Arungro performances

Although the literature about the crucial variables needed as input by the model was poor, parameters describing development progress and leaf characteristics are in the range of those proposed elsewhere. Simulation outputs highlighted a satisfactory agreement with measured data, both with calibration and evaluation datasets. Relative root mean square error metrics (RRMSE i.e., the root of the mean square difference between predicted and observed values divided by the average of observed values) ranged –in calibration– from 20% to 59% and 21% to 25% for organ biomasses and stem height, respectively. The variability of the observed plant features, stressed by error bar magnitudes (Figure 3), underlined the high degree of heterogeneity characterizing the crop. Moreover, the decidedly different planting density used in the two experimental sites tends to strengthen the variability of the samples, due to the clear edge effect characterizing some of the plots. The model performed better in estimating organs growth in evaluation datasets, with aboveground mass (AGB) RRMSE ranging from 18% to 41%. Evaluation also highlighted the necessity of improving the simulation of stem height, which reached with this dataset RRMSE values above 32%. However, this variable has no effect in the estimation of LAI or photosynthesis, and therefore its wrong estimation would not produce errors in the estimation of giant reed productivity. The more marked deviation from observations shown by LAI, which highlights a slight overestimation (Figure 3b and 3c), can be partly explained by the adoption of a gap fraction-based measurement method in one of the two experimental sites. The maximum measurable LAI values by camera devices are generally lower than the direct ones, since a gap fraction saturation effect is typical for close canopy conditions (Francone et al., 2014). Conversely, the key variable in view of bioenergy application (i.e., AGB) was accurately simulated by Arungro (Figure 3a and 3b), as demonstrated by the performance indices very close to the ones obtained by other well-known crop models (e.g., Confalonieri et. al, 2009; Brisson et al., 2002).

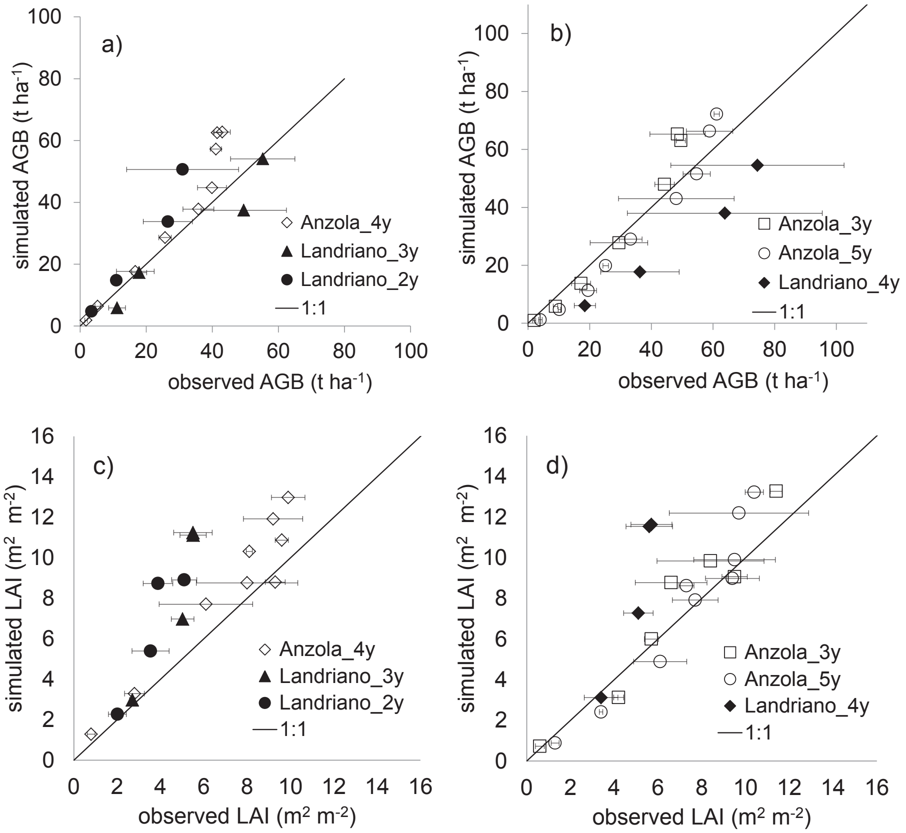

Figure 3. Comparison between observed and simulated values of AGB and LAI. a and c: calibration datasets; b and d: evaluation datasets.

4. COUNCLUSIONS

The international community recognizes the usefulness of software design favoring model reuse. However, most of the agro-ecological models are still implemented in monolithic software units using outdated software designs and technologies. This is far from limiting the problem to a programming issue, since the use of unsuitable technology for developing complex, integrated system models is likely one of the major factors limiting the formalization of new knowledge in mathematical constructs. The result is a gap between scientific knowledge and its formalization into simulation models. The reimplementation of Canegro presented in this paper is aimed at providing third parties with a version of the model explicitly designed for being easily used, composed and extended, regardless of the simulation environment. Both the features of reusability and extensibility of the component were here demonstrated via the development of a model-based yield forecasting system for sugarcane in Brazil and the extension of the model leading to the first explanatory giant reed crop model.

5. REFERENCES

Argent, R.M., 2005. A case study of environmental modelling and simulation using transplantable components. Environ. Model. Soft. 20, 1514-1523.

Bannayan, M., Crout, N.M.J., 1999. A stochastic modelling approach for real-time forecasting of winter wheat yield. Field Crops Research. 62, 85-95.

Brisson, N., Ruget, F., Gate, P., Lorgeou, J., Nicoullaud, B., Tayot, X., Plenet, D., Jeuffroy, M.H., Bouthier, A., Ripoche, D., Mary, B., Justes, E., 2002. STICS: a generic model for the simulation of crops and their water and nitrogen balances. II. Model validation for wheat and corn. Agronomie, 22:69-93.

Confalonieri, R., Rosemund, A.S., Baruth, B., 2009. An improved model to simulate rice yield. Agron. Sustain Dev. 29:463-474.

Donatelli, M., Rizzoli, A.E., 2008. A design for framework-independent model components of biophysical systems. In: Sànchez-Marrè, M., Béjar, J., Comas, J., Rizzoli, A.E., Guariso, G. (Eds.), iEMSs 2008 International Congress on Environmental Modelling and Software, vol. 2. International Environmental Modelling and Software Society, Barcelona, Spain, 7-10 July, pp. 727-734.

Donatelli, M., Cerrani, I., Fanchini, D., Fumagalli, D., Rizzoli, A.E., 2012. Enhancing model reuse via component-centered modeling frameworks: the vision and example realizations. In: Seppelt, R., Voinov, A.A., Lange, S., Bankamp, D. (Eds.) iEMSs 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting, Leipzig, Germany, 1-5 July, pp. 1185-1192.

Francone, C., Pagani, V., Foi, M., Cappeli, G., Confalonieri, R., 2014. Comparison of leaf area index estimates by ceptometer and PocketLAI smart app in canopies with different structures. Field Crops Res.,155, 38-41.

Gamma, E., Helm, R., Johnson, R., Vlissides, J., 1995. Design Patterns: Elements of Reusable Object-Oriented Software. Addison Wesley.

Holzworth, D.P., Huth, N.I., de Voil, P.G., 2010. Simplifying environmental model reuse. Environ. Modell. Softw. 25(2), 269-275.

Inman-Bamber, N.G. 1991. A growth model for sugar-cane based on a simple carbon balance and the CERES-Maize water balance. S. Afr. J. Plant Soil 8(2):93-99.

Jones, M.R., Singels, A., van den Berg, M., 2008. DSSAT v4.5-Canegro Sugarcane Plant Module Scientific Documentation. International Consortium for Sugarcane Modelling, South African Sugarcane Research Institute, Mount Edgecombe, South Africa, pp. 34. http://sasri.sasa.org.za/misc/DSSAT%20Canegro%20SCIENTIFIC%20documentation_20081215.pdf (last accessed 10.03.14.).

Marin, F.R., Jones, J.W., Singels, A., Royce, F., Assad, E.D., Pellegrino, G.Q., Justino, F., 2012. Climate change impacts on sugarcane attainable yield in southern Brazil. Climatic Change, 117, 227–239

Meyer, B., 1997. Object-oriented Software Construction, 2nd edition. Prentice Hall, Upper Saddle River.

Pagani, V., Fumagalli, D., Bregaglio, S., Francone, C., ZhiMing Wang, Confalonieri, R., 2013. Report on the assessment of the BioMA platform for rice monitoring and yield forecasting in Jianghuai plain. (http://www.e-agri.info/pdf/documents/D33.1_Assessment_BioMa_for_Rice_Monitoring.pdf)

Pilu, R., Bucci, A., Badone, F.C., Landoni, M., 2012. Giant reed (Arundo donax L.): A weed plant or a promising energy crop?. Afr. J. Biotech. 11(38), 9163-9174.

Szyperski, C. 2002. Component software: beyond object-oriented programming. ACM Press and Addison-Wesley, New York, USA, second edition.