International Congress on Environmental Modelling and Software

6th International Congress on Environmental Modelling and Software - Leipzig, Germany - July 2012

Jul 1st, 12:00 AM

An extensible, multi-model software library for simulating crop growth and development

Roberto Confalonieri

S. Bregaglio

T. Stella

G. Negrini

Marco Acutis

See next page for additional authors

Confalonieri, Roberto; Bregaglio, S.; Stella, T.; Negrini, G.; Acutis, Marco; and Donatelli, Marcello, "An extensible, multi-model software library for simulating crop growth and development" (2012). *International Congress on Environmental Modelling and Software*. 132.
https://scholarsarchive.byu.edu/iemssconference/2012/Stream-B/132

Presenter/Author Information

Roberto Confalonieri, S. Bregaglio, T. Stella, G. Negrini, Marco Acutis, and Marcello Donatelli

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

An extensible, multi-model software library for simulating crop growth and development

R. Confalonieri[a], S. Bregaglio[a], T. Stella[a], G. Negrini[a], M. Acutis[a], M. Donatelli[b,c]
[a] *Department of Plant Production, University of Milan, Italy*
[b] *Research Centre for Industrial Crops, Agriculture Research Council, Italy*
[c] *Joint Research Centre, European Commission*
Email: roberto.confalonieri@unimi.it

Abstract: The availability of different crop models and of a variety of techniques to evaluate their behaviour led to a change of paradigm in crop models use. Modellers are now looking beyond the idea of groups of users and developers grounded on a specific model, and international initiatives focusing on model improvement basing on intercomparison and knowledge sharing are recently catalysing the attention of the international community. Also, the analysis under environmental conditions of no adaptation by crops, such as the ones of climate change scenarios, demand for extension of crop models to account for extreme events, diseases and pests impact, difficult to implement into legacy code.

The Crop Models Library (CropML) is a framework-independent MS .NET software component where different pure (e.g., WOFOST, CropSyst, WARM), hybrid and new modelling solutions for crop growth and development are implemented following a fine level of granularity, according to a high level software architecture. CropML can be extended by third parties and is distributed at no cost with a software development kit, including documentation of code and algorithms and sample applications. CropML provides modellers with an environment favouring the hybridization of models with parts from others, the evolution of existing approaches, and the possibility of analysing and easily comparing diverse modelling solutions.

As an example, a new generation of SUCROS-type models has been developed and included in the component. Comparison of the standard and of the new version of the WOFOST model carried out using data from rice field experiments revealed an increase in accuracy and robustness with less than half of the parameters used by the standard version of the model. These results support the idea that high-level technology for models formalization can favour the development of the models themselves.

Keywords: crop models; software component; WARM; WOFOST; CropSyst.

1 INTRODUCTION

Crop models are increasingly used to analyse the interactions between plants and factors driving their growth, like weather and soil, under different climate and management scenarios. A quantitative understanding of the complex dynamics influencing the productivity and the environmental impacts of cropping systems is in fact crucial to properly define strategies for their management, both in the short and in the medium term.

In spite of the variety of crop models available, and of the differences – sometimes radical – in the approaches they use for crop growth and development, the choice of the model is almost always performed without any regard to its suitability with respect to the specific study to be carried out. The reasons are the most disparate, but can be grouped in two main categories: the “faithfulness” to a certain modelling school and the fact that a model was used in previous studies, that means the availability of knowledge on the algorithms, of parameterizations, and of routines for I/O data conversion. In any case, the fact that such nearsighted behaviours could lead to select the most suitable model for a specific study is purely accidental.

It is possible to identify, for a modelling study, a variety of objectives (e.g., research, making predictions, decision making) and conditions of application, in turns defined by the availability of data and resources, by the spatial scale considered, etc. Objectives and conditions of application should then be used to select a certain number of quantitative metrics for evaluating relevant model features [Donatelli and Confalonieri 2012]. Such metrics, focusing on different aspects of model accuracy, structure, and behaviour, should lead to define a criterion to be used to rank the available approaches, in order to allow the identification of the most suitable for the specific study.

These considerations assume a great importance in light of how the same models can be ranked in a complete different way according to the evaluation metrics used. Confalonieri et al. [2010], while simulating rice in Northern Italy using different approaches, pointed out that WOFOST [van Keulen and Wolf 1986] was the model with the highest modelling efficiency [Nash and Sutcliffe 1970], whereas WARM [Confalonieri et al. 2009] was that with the highest robustness. The same models achieved markedly different values for the Akaike's Information Criterion [Akaike 1974], suggesting a high degree of overparameterization for WOFOST [Confalonieri et al. 2009], which instead resulted the best in terms of plasticity [Confalonieri et al. 2012].

The overcoming of the idea that a unique model can be the most suitable for whatever modelling study strongly asks for libraries of approaches, sharing the same data structures for being interfaced with other components of the simulation environments (e.g., I/O services, advanced tools for sensitivity analysis and parameterization, results visualization). This would strongly favour their context-specific evaluation for specific studies or their parallel use in multi-model studies aimed at providing distributions of the outputs form different models (e.g., EU-FP7 Project E-AGRI).

Moreover, the availability of different approaches in the same environment, if implemented with a high level of granularity, favours (i) the reuse of algorithms (approaches for modelling specific processes and subprocesses are shared among various models), (ii) models hybridization [e.g., Bregaglio et al. 2012], and (iii) model evolution, via the substitution of parts of the code with improved algorithms.

The aims of this paper are:

- to present a multi-model library for modelling crop growth and development (Crop Models Library, CropML), where some well-known approaches to crop growth and development are implemented;
- to present a new generation of SUCROS-type models, as an example of how the conceptual and software architecture of CropML can effectively support model improvement.

2 CROPML, A LIBRARY OF CROP MODELS

2.1 Software design

The software design promotes reusability by limiting dependencies (only on the component CRA.Core.Preconditions in the specific case) and providing a semantically rich, public interface (IStrategyCropML). It allows third parties to extend the component with other modelling approaches, by adding new strategies and extending the Domain Classes (data-types, used as either input or input-output parameters in the interface of the component and implementing attributes of each variable). The design-by-contract approach [Meyer 1997] is used, requiring pre- and post-conditions to be respected. Further details about the software design are provided by Donatelli and Rizzoli [2008] and Donatelli et al. [2009].

CropML is distributed in a Software Development Kit including hypertext files documenting the code, the implemented approaches, software design and use, C# source codes of sample clients, and Visual Studio sample projects for extending the component (http://agsys.cra-cin.it/tools/cropml/help/).

CropML interfaces, domain classes and strategies, can be inspected via the external application named Model Component Explorer (http://agsys.cra-cin.it/Tools/MCE/help/).

Figure 1 shows the software design of the CropML. Crop models algorithms are coded into two different DLLs: CropML and CropML_WL (which extends CropML) respectively for potential and water limited crop growth and development. For both CropML and CropML_WL, interfaces and data structures are defined in separated DLLs (bridge pattern), i.e., CropML.Interfaces and CropML_WL.Interfaces, to keep separated the description of the biophysical domain from the modelling representations of processes characterizing the dynamics of the domain itself. Ontology of model parameters – related to the modelling representation and not of the domain – is instead encapsulated in CropML and CropML_WL. This implementation of the bridge pattern allows the substitution of modelling approaches in applications without changing the interfacing with between domain description and I/O services.

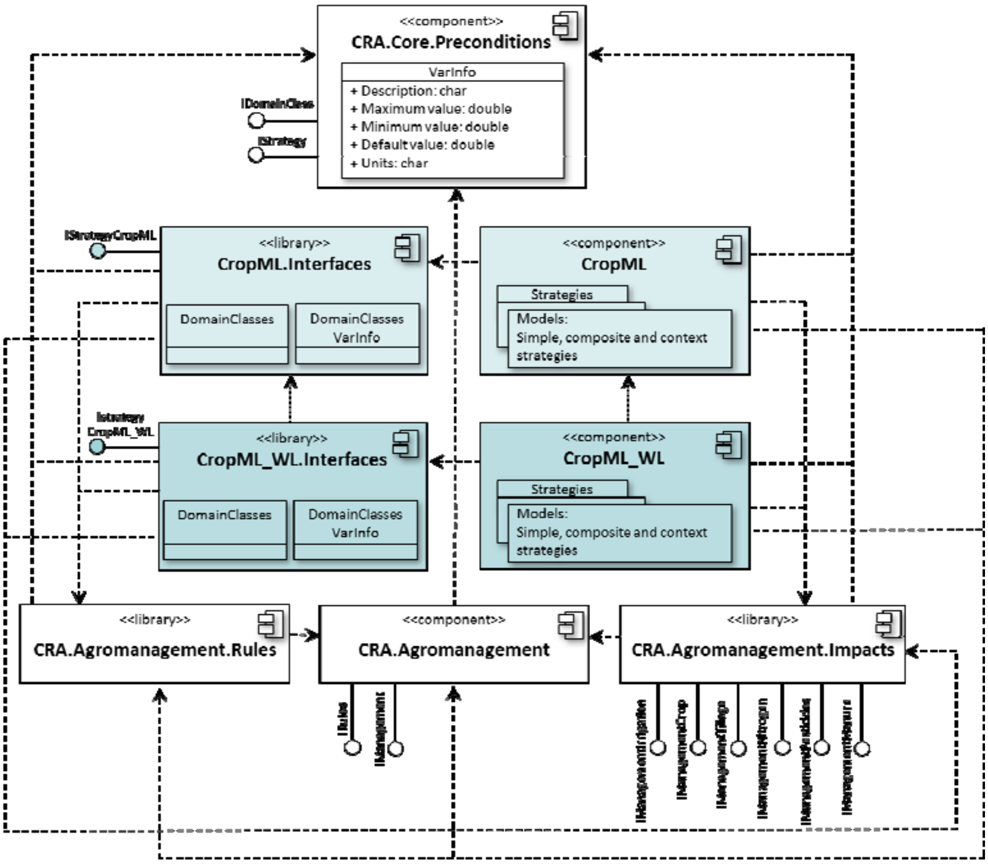

Figure 1 Software design of the CropML component.

2.2 Standard models implemented

The standard models currently implemented in CropML are WOFOST (version 7.1), CropSyst (version 3.02.23, last release of version 3) and WARM (version 2.0.1). STICS [Brisson et al. 2003] and CANEGROW [Inman-Bamber 1991] are under implementation.

WOFOST [van Keulen and Wolf 1986] simulates crop development based on thermal time accumulation, optionally corrected for a photoperiodic factor. Crop growth is reproduced using the concept of gross photosynthesis, with gross CO_2 assimilation, maintenance and growth respirations explicitly simulated. The former is derived on a daily basis with Gaussian integrations on the instantaneous CO_2 assimilation rates computed at three moments of the day and for three canopy depths. Maintenance respiration is based on the assumptions that different organs have different respiration to weight ratios, and that it is proportional to the dry weight of plant organs. The different chemical composition of leaves, stems, storage organs and roots reflects on how their growth respiration is modelled. Total dry matter production is partitioned among the different plant organs according to development-dependent coefficients. During early growth stages, leaf area is considered growing exponentially as a function of temperature. After canopy closure, leaf area index (LAI) is derived from leaf biomass using a development-dependent specific leaf area (SLA). Leaf senescence is driven both by leaves aging and by self-shading.

CropSyst [Stöckle et al. 2003] is a generic crop simulator, successfully used worldwide for many crops under a variety of agro-climatic conditions [Donatelli et al. 1997]. Crop development is simulated as a function of thermal time accumulated between base and maximum temperature, optionally corrected to account for photoperiod, vernalization and water stress. Net photosynthesis is assumed as the minimum between biomass accumulation rates calculated according to (i) a radiation use efficiency (RUE) approach and (ii) an approach based on vapour pressure deficit-corrected transpiration use efficiency. Thermal limitation is explicitly accounted for only in the RUE-based approach. Leaf area development is calculated from daily accumulated biomass using a constant SLA and an empiric parameter. Crop yield is derived by multiplying the total biomass at harvest by a harvest index, with the latter varying according to the specific sensitivity to water stress of the simulated crop or variety. Root depth is simulated as a function of leaf area development, and was set to reach its maximum at flowering in this study. Leaf senescence is calculated by subtracting the dead leaf area index to the total one, with each daily emitted green leaf unit dying once a threshold amount of degree days is accumulated.

WARM [Confalonieri et al. 2009] calculates development rate as a function of thermal time using the curvilinear response proposed by Yin et al. [1995], optionally corrected for photoperiod and water stress. Daily biomass accumulation is simulated as a function of intercepted radiation, modulating RUE according to temperature (Yin function), senescence, saturation of the enzymatic chains, atmospheric CO_2 concentration, and diseases. Above-ground biomass (AGB) is partitioned to the different plant organs according to a set of development-dependent beta and parabolic functions, driven by a single parameter (partitioning to leaves at emergence). LAI is derived by multiplying leaves biomass by a SLA that decreases till mid-tillering using a quadratic function and is assumed as constant from mid-tillering to physiological maturity. Leaves senescence is simulated according to leaves aging. A micro-meteorological module, implemented in a separated component, is used to account for floodwater effect on vertical thermal profile, in turns allowing to provide temperature at the meristematic apex for development and spikelet sterility, and mid-canopy temperature for thermal limitation to photosynthesis.

2.3 Guidelines for models implementation in CropML

One of the main guidelines for the implementation of models in CropML is that approaches for representing biophysical processes are not unique. This leads to the concept of alternate modelling solution, assumed as valid for all the different levels of organization of processes. A strategy can be defined as the smallest part of an algorithm for which an alternate approach exists – or it is reasonable to imagine that it will exist in the future. In practice, a strategy is the result of a process aimed at increasing progressively the level of granularity of the implementation that ends just before the pieces of algorithms lose their internal coherence. This coherent modelling units (i.e., strategies) can be composed (composite strategies) at different hierarchical levels, thus coherently modelling processes of increasing complexity. Changes in the algorithms due to, e.g., the fact that a specific phenological stage has been reached, are managed by dynamically switching between different modelling units at run-time (strategy pattern). The façade pattern allows accessing simple and complex (composite) strategies using always the same, simple interface.

As an example, Figure 2 shows the strategy diagram of the WOFOST implementation in CropML, underlying the level of granularity and the way modelling units are composed at different levels of organization.

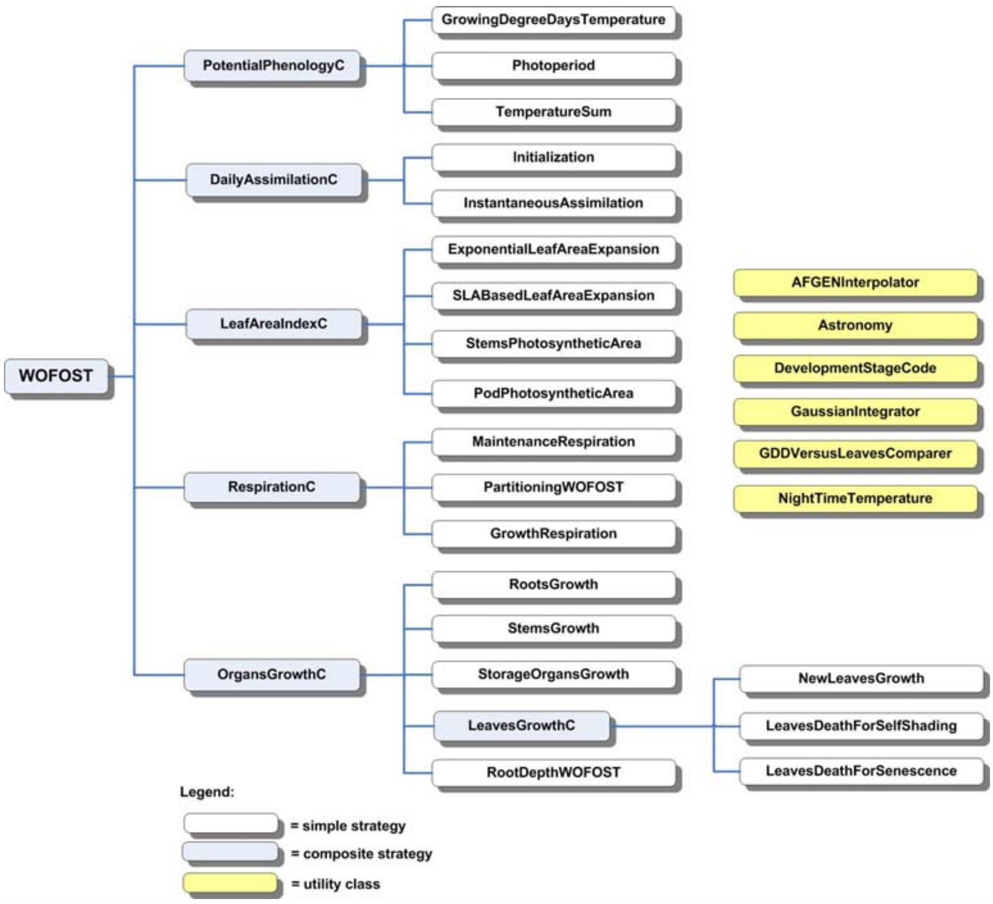

Figure 2 Strategy diagram of the WOFOST implementation in CropML.

2.4 Benefits deriving from the CropML design

The design described above ensures a rational and effective development of models. The high level of granularity allows to implement new models by reusing strategies already available (models often shares parts of the algorithms). As an example, half (8 out of 16) of the simple strategies of our CropSyst implementation were already developed from WARM. And similar levels of reuse could be reached in case other models based on net photosynthesis (e.g., those belonging to the CERES, STICS) will be implemented. The level of reuse is of course maximized in case models from the same family will be implemented, like in the case implementing models like SUCROS or ORYZA starting from the WOFOST implementation presented in Figure 2.

The benefits of such a design are also related to the extension of the features of existing models. As an example, the development of a version of WARM based on hourly response to temperature for both development and photosynthesis – crucial for properly exploring conditions the crop is not adapted to – made a large reuse of strategies from the daily version (15 out of 19 strategies reused). Another advantage of this pronounced level of modularity (each strategies is the smallest coherent and "independent" modelling unit) is the easiness of testing new approaches for specific sub-process in an environment that allows parallel runs with the original approach (the one modified) and with other, different models.

3 A NEW GENERATION OF SUCROS-TYPE CROP MODELS

3.1 Criticalities in the standard version of SUCROS-type models

WOFOST, as well as the other models belonging to the same family (e.g., SUCROS, ORYZA), is very detailed in the representation of the processes involved with crop growth and development. In light of the idea that the most suitable approach – among those available – should be selected for a specific modelling study, WOFOST could surely represent the best solution in cases where a high adherence to the modelled entities is required [Confalonieri et al. 2012].

However, its over-parameterization exposes users to non-negligible risks because of the large number of freedom degrees during the parameterization, that could increase the risk of including the effect of site- or season-specific factors in parameters values, that should instead describe only plant morphological and physiological features. This could strongly decrease the robustness of the model [Confalonieri et al. 2010]. Part of the reasons explaining the large number of parameters of this family of models is related to the high level of detail used to represent biophysical processes, and should be considered as a positive, intrinsic feature of SUCROS-type models. But the main reason responsible for the over-parameterization is the presence of tables (AFGEN) for describing changes in the values of some parameters as a function of temperature or development stage, with the "points" of the AFGEN tables linearly interpolated. Apart from the potential negative effects on parameterizations robustness, the presence of AFGEN tables could create problems to the coupling of the model to tools for sensitivity analysis and automatic calibration. The reason is that the exploration of the parameters hyperspace performed by sampling algorithms used in sensitivity analysis or by optimization tools could lead to situations without any biological meaning, where, e.g., partitioning of photosynthates to storage organs in cereals decreases during the reproductive phase. Some authors [Confalonieri 2010, Ceglar et al. 2011] succeeded in performing Monte Carlo based sensitivity analysis on this kind of models only at cost of strongly reducing the number of points in the AFGEN tables, in order to limit the overlaps among the distributions of parameters values for different values of the variables driving the changes in the parameters themselves.

Another critical aspect of this family of models is related to the division of the canopy in a number of layers, for which instantaneous gross assimilation rates are calculated. This number appears to be arbitrary, e.g., it is three for WOFOST and five for SUCROS, with no justification in both cases and it is constant from the post-emergence phases to maturity, regardless from differences in the canopy structure during the crop cycle. Moreover, the division of the canopy in different layers is explicitly used for some processes (e.g., gross photosynthesis) and not for others. As an example, leaves death does not take into account the position of leaves within the canopy, neither for aging nor for self-shading, leading to a situation where last emitted LAI units (the youngest, with lower SLA compared to the those emitted during early stages) die exactly like the oldest one, or with dead leaves shading green ones.

3.2 Solutions to the criticalities: an improved version of WOFOST

The CropML design favoured the implementation and testing of solutions aimed at improving the WOFOST for the aspects outlined above.

The AFGEN tables were designed to provide the SUCROS-type models with the highest flexibility for reproducing the behaviour of crops with different partitioning patterns (like those characterizing, e.g., grain and tuber crops), SLA dynamics, etc. However, the same level of flexibility could be reached by clustering crops according to their behaviour and by developing simple (with few parameters) functions, specific for each crop typology, for substituting AFGEN tables. As an example, for the way crops allocate resources to the different plant organs, we used here a partitioning approach derived from that implemented in WARM (see section 2.2), which can be considered valid not only for rice but for a variety of cereals like wheat, barley, etc. For temperature effects on thermal time accumulation rate and on photosynthesis, we used the Yin function [Yin et al. 1995], like in WARM. This led to strongly decrease model complexity without penalizing its adherence to the real system.

In order to allow the model to explicitly consider the vertical dimension of the canopy, we

- included in WOFOST a model for plant height,
- developed a function for deriving the number of canopy layers as a function of development stage (from one after emergence to 20 at anthesis),
- related each canopy layer to a specific canopy depth,
- attributed dead leaves starting to the bottom of the canopy, with layers saturating (of dead leaves) with a bottom-up dynamic.

In practice, we gave an explicit vertical dimension to the canopy (Figure 2), with relevant micrometeorological implications, e.g., the microclimate in the bottom layers is warmer since dead leaves do not transpire, like in real canopies, etc.

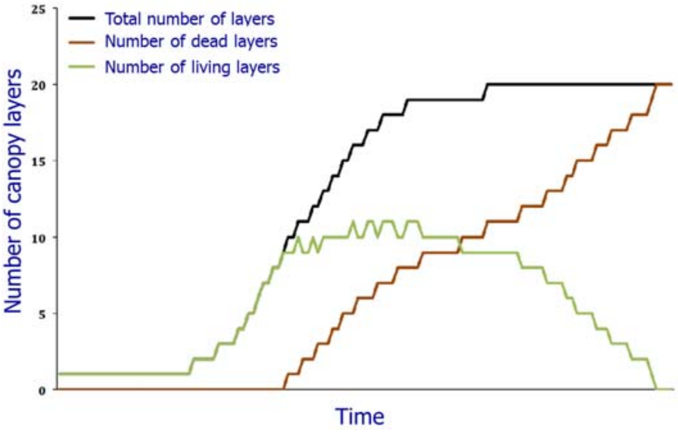

Figure 3 Dead and living canopy layers during crop development.

3.3 Evaluation of the performances of the standard and of the improved versions of WOFOST

The performances of the new version of WOFOST (WOFOST_GT2) for rice simulations was compared with the standard one by using data collected during various field experiments carried out in Northern Italy between 1989 and 2004, described in detail by Confalonieri et al. [2009]. For the standard version of the model, the evaluation metrics published in the same study were used.

Evaluation results (averaged for calibration and validation) are shown in Table 1. Under the conditions explored, the new version of the model achieved the best evaluation metrics for accuracy (RRMSE) with only 36 parameters out of the 106 present in the standard version. This allowed the new version of the model to be much more efficient (AIC consider both accuracy and complexity). The decrease in the number of freedom degrees during the calibration led also to increase model robustness (the optimum value for the indicator is 0).

Table 1 Evaluation metrics resulting from the comparison of the standard version of WOFOST and of the improved one, in terms of accuracy, robustness, and complexity.

Evaluation metric	WOFOST	WOFOST_GT2
RRMSE (relative room mean square error, %)	25.9	19.8

EF (modelling efficiency, Nash and Sutcliffe 1970)	0.931	0.939
I_R (Robustness indicator, Confalonieri et al. 2010)	0.387	0.342
AIC (Akaike's Information Criterion, Akaike, 1974)	280.9	112.9
Total no. of parameters for potential conditions	106	36

4 CONCLUSIONS

The availability of libraries of approaches for modelling crop growth and development is crucial to allow users to select the best approach for specific studies. In this study, we presented a framework-independent, extensible component with a high-level conceptual and software architecture, with some well-known approaches available and other under implementation. Apart from making available different models within the same environment, the component design demonstrated to favour model improvement, with a new generation of SUCROS-type models derived from the standard WOFOST version with the aim of increasing the efficiency and robustness of this family of models.

REFERENCES

Bregaglio, S., M. Donatelli, R. Confalonieri, and M. Acutis, Comparing modelling solutions at submodel level: a case on soil temperature simulation, paper presented at International Environmental Modelling and Software Society (iEMSs) 6th International Congress, Leipzig, Germany, July 1-5, 2012.

Brisson, N., F. Ruget, P. Gate, J. Lorgeou, B. Nicoullaud, X. Tayot, D. Plenet, M.H. Jeuffroy, A. Bouthier, D. Ripoche, B. Mary, and E. Justes, STICS: a generic model for the simulation of crops and their water and nitrogen balances. II. Model validation for wheat and corn, *Agronomie, 22*(1), 69-93, 2002.

Ceglar, A., Z. Črepinsek, L. Kajfez-Bogataj, and T. Pogacar, The simulation of phenological development in dynamic crop model: the Bayesian comparison of different methods, *Agricultural and Forest Meteorology, 15*1(1), 101–115, 2011.

Confalonieri, R., Monte Carlo based sensitivity analysis of two crop simulators and considerations on model balance. *European Journal of Agronomy, 33*(2), 89-93, 2010.

Confalonieri, R., M. Acutis, G. Bellocchi, M. Donatelli, Multi-metric evaluation of the models WARM, CropSyst, and WOFOST for rice, *Ecological Modelling, 220*(11), 1395-1410, 2009.

Confalonieri, R., S. Bregaglio, M. Acutis, A proposal of an indicator for quantifying model robustness based on the relationship between variability of errors and of explored conditions, *Ecological Modelling, 221*, 960-964, 2010.

Confalonieri, R., S. Bregaglio, M. Acutis, Quantifying plasticity in simulation models, *Ecological Modelling, 225*, 159-166, 2012.

Donatelli, M., A.E. Rizzoli, A design for framework-independent model components of biophysical systems. Proceedings of the International Congress on Environmental Modelling and Software (IEMS '08), M. Sanchez-Marre, J. Bejar, J. Comas, A. E. Rizzoli, and G. Guariso, Eds., vol. 2, pp. 727-734, Barcelona, Spain, 2008.

Donatelli, M., C.O. Stöckle, E. Ceotto, and M. Rinaldi, Evaluation of CropSyst for cropping systems at two locations of northern and southern Italy, *European Journal of Agronomy, 6*(1-2), 35–45, 1997.

Donatelli, M., G. Bellocchi, E. Habyarimana, R. Confalonieri, and F. Micale, An extensible model library for generating wind speed data, *Computers and Electronics in Agriculture, 69*(2), 165-170, 2009.

Inman-Bamber, N.G., A growth model for sugarcane based on a simple carbon balance and the CERES-Maize water balance, *South African Journal of Plant and Soil, 8*(2), 93-99, 1991.

Meyer, B., *Object-oriented Software Construction, 2nd edition*, Prentice Hall, Upper Saddle River, NJ, USA, 1254 pp, 1997.

Stöckle, C.O., M. Donatelli, and R. Nelson, CropSyst, a cropping systems simulation model, *European Journal of Agronomy, 18*(3-4), 289-307, 2003.

Van Keulen, H., J. Wolf, *Modelling of agricultural production: weather soils and crops*, Simulation Monographs, Pudoc,Wageningen, 1986.

Wang, Z.-M., B. Zhang, X.-Y. Li, K.-S. Song, D.-W. Liu, S.-Q. Zhang, Using CropSyst to simulate spring wheat growth in black soil zone of northeast China, *Pedosphere*, 16(3), 354–361, 2006.

Yin, X., M.J. Kropff, G. McLaren, and R.M. Visperas, A nonlinear model for crop development as a function of temperature, *Agricultural and Forest Meteorology, 77*(1-2), 1-6, 1995.