Toward a simulation modeling platform for studying cropping systems management: the Record project

P. Chabrier, F. Garcia, R. Martin-Clouaire, G. Quesnel and H. Raynal

INRA, UR875, Unité de Biométrie et Intelligence Artificielle, Castanet-Tolosan, France
Email: {chabrier,fgarcia,rmc,raynal}@toulouse.inra.fr

Keywords: *modelling, simulation, cropping systems, crop management*

EXTENDED ABSTRACT

The pressing need in adapting and improving farming practices has motivated research efforts to extend the classical analysis of plants and crops to consideration of more complex cropping systems up to farming systems and small regional agro-ecosystems. Sustainability concerns raise production management issues that are gaining increasing attention in research. This shift of scientific interest induces new demands of modeling and simulation capabilities. For instance, features concerning biophysical aspects (processes at different spatial and temporal scales) and production system dimensions (structures, resources, activities organization, constraints, individual preferences and goals ...) must be integrated.

Our research institute has launched the development of a simulation modeling platform, called RECORD, for studying cropping systems management. The target software is intended to foster sharing of domain concepts, componentizing and re-using of models, and to ease complex model development. This desire translates into requirements and challenges such as:

- supporting closer mapping between the structures of the natural system and those involved in the model (not only vector-based representation but all object-oriented, graph-based representations for instance);
- putting decision-making aspects on an equal foot with biophysical aspects, which is not trivial because different modeling paradigms are usually employed (rule and discrete event dynamics versus differential equations);
- coupling modules through either conceptual embedding in the framework at design time or well-engineered computational connection;
- promoting modular modeling without forcing the use of rather flat atomic components.

Numerous modeling frameworks have been developed in ecology, environment, hydrology, and agriculture. However, most of them focus on biophysical processes and give much less room to the consideration of human activities and social aspects. In particular the modeling of management behavior is often dealt with superficially. The opposite is true for the modeling frameworks developed in production management (e.g. Petri nets) but they suffer from limited capabilities to be tightly coupled with continuous processes. Although inspiring in the current design phase none of the existing platforms are fully satisfactory, hence our RECORD project.

After a survey of the modeling needs induced by cropping systems studies, this paper outlines the fundamental requirement that the modeling platform should satisfy at the technical and conceptual levels. Four frameworks reported in the literature are then reviewed: Simile, MODCOM, DIESE and DEVS-VLE. These frameworks were selected as representative of recent simulation software solutions. Their evaluation was made by examining their intrinsic properties and, for some, by applying them to the recoding of typical cropping system simulation models.

On the basis of this examination and key criteria related to systemic foundation, object-oriented and multi-formalism modeling, ability to couple models, extensibility and efficiency of the framework. We conclude with orientations chosen for the future of the RECORD project. We intend to build the RECORD platform on top the VLE environment that embeds the DEVS formalism. This choice, which satisfies the identified criteria, comes with multi-formalism modeling capabilities and a time-explicit simulation engine. The RECORD project focuses then on the development of some extensions targeted to the specific needs of cropping system studies. In particular, one of them should be dedicated to the production management aspects.

INTRODUCTION

The study and design of agricultural production systems rely increasingly on the use of computer models that represent the structure and dynamics of the underlying biophysical system together with the human activities involved in controlling the production processes. In the case of crop production, the pressing need in adapting and improving farming practices has brought up ambitious projects that go beyond the traditional studies of crop yield response to environmental factors and management options. Typically, more complex systems such as cropping systems or small regional agro-ecosystems have to be investigated, which requires modeling of multi-scale and multi-dimensional processes that might be biophysical (crop functioning), technical (implementation of cultivation operation) or cognitive (production management decision making).

So far little methodology is shared among the modelers that tend to develop each their own ad-hoc modeling approach and simulation software. For a research institute or scientific community, much efficiency would be gained by promoting the application of systematic modeling principles, the use of the same tool box and the sharing of models or model parts. The purpose of the RECORD project that is presented in this paper is to build a modeling/simulation software platform tailored to the needs of a large range agro-ecosystem studies, especially on the modeling of spatial and decision-making processes. More specifically, the purpose of the project is to open the simulation arena to a much wider set of participants, foster the adoption and sharing of modeling/simulation principles and tools, and facilitate the application of computer modeling to the study of complex agricultural processes in support of science, interactive learning, and policy making. The target software is intended to encourage a compositional modeling approach that mimics as much as possible the task of constructing by assembling parts and, consequently, ease the re-using of model components. Although RECORD is a modeling/simulation platform rather than a general simulation model, a repository infrastructure may be used to exploit existing model component developed with RECORD.

Most of the available modeling/simulation platform are based on a single representation formalism (e.g. differential equation) and are not suitable to incorporate heterogeneous pieces of knowledge such as discrete event processes or hierarchical structures. RECORD relies on a multi-formalism modeling and provides software coupling facilities to connect, if necessary, black box models with components developed in the native RECORD modeling language.

In this paper, we present the main functional requirements of RECORD, and some preliminary choices concerning the platform architecture. We then review the pros and cons of various attempts reported in the literature. Finally we conclude with orientations chosen for the close future of the RECORD project.

1. MODELLING AND SIMULATION OF CROPPING SYSTEMS

1.1. Modeling objectives

The design of innovative farming systems satisfying desirable agronomic and environmental functionalities is one of the current research challenges in agriculture. Results are expected on different issues such as the design of cropping systems with low levels of inputs like water, chemical pesticides and with low impacts on environment. The design of integrated pest management strategies for a better control of insects, weeds and diseases, is also a pressing need in agriculture. In this context, prospects on a better use of natural resources and production factors, by combining convenient species or genotypes and by adapting technical practices are necessary. Modeling and simulating is an attractive option to achieve this objective. Dealing with complexity in farming systems studies requires a multi-disciplinary modeling approach where interactions between agronomical, environmental and socio-economics components can be explored. The methodology and software tools to elaborate and implement such models certainly have a great importance in the use, diffusion and generalization of these models. A modeling and simulation platform well suitable to the study of cropping systems is needed.

In the last decade, beside the tendency toward multidisciplinary approach, agricultural modeling studies have tried to give particular attention to the human role and its impact on production and on system externalities. Most of the time, the human role was simulated by decision making with fixed options or very simple rule-based procedures (Bergez *et al.,* 2001; Chatelin *et al.,* 2005). The design of innovating cropping system needs to simulate with more accuracy real management practices (Aubry *et al.,* 1998). One needs to consider some relevant features like: the farmer's material conditions, his understanding about the biophysical functioning of his system, his personal know-how, the possible interactions between technical operations, the fact that decision-making process is both reactive and goal oriented and deals primarily with the issue of dynamic work scheduling and resource allocation (Garcia *et al.,* 2005). RECORD should allow the simulation of management models, coupled with biophysical models.

Some processes like the genes or nitrogen flux, the pests, weeds or diseases diffusion, the management of water, have to be treated in a spatial two or three dimensional context. The spatial granularity is defined through elementary spatial units such as agricultural parcels in many cropping system studies. The spatial extent is rather wide; it may be from the farm field up to an area of several square kilometers. The organization of space, the exchanges between parcels or soil layers, spatial units heterogeneity need to be considered. Agronomist researchers often encounter difficulties when they have to integrate these aspects in their models. Similarly, the temporal context might be difficult to handle in complex systems with heterogeneous temporal granularities. The time horizon can be from a few months up to several years, in which case it is important to take into account the cumulative aspects of the processes. So, one of the major challenges of RECORD will be to offer modelers some support in the implementation and coupling of heterogeneous temporal and spatial dimensions.

The processes involved in cropping systems (plant growth, human decision, pest diffusion ...) are heterogeneous, and it is important to give the researchers the opportunity to choose the more suitable formalisms. In order to open widely the platform to all the approaches of dynamical models encountered in the INRA modelers community, we thus decided to make potentially no restrictions on the type of cropping system models that could be developed or plugged in within the RECORD platform.

1.2. Working with cropping system models

Another key feature of the platform concerns the different services it will offer for the simulation activity. They concern three major issues: The management of data, the statistical analysis of the simulation results and some optimization tools for the design of innovative farming systems (Wallach, 2006).

Data: Model simulation needs a lot of input data. Weather, soil, variety, socio-economics data are acquired from literature surveys, maps and databases. The simulation activity would be improved if some databases and GIS were directly interfaced. Some models generating artificial data, like random weather generators (RWG) or crop generators for virtual land allocation should also be associated to the RECORD platform.

Statistical analysis: Simulations produce numerous outputs. At least one statistical tool like R (R Development Core Team, 2006), will be available from the platform to analyze them. Moreover, RECORD will provide special support on some identified statistical methods that are necessary for agronomic researchers when they construct or use models. These methods are used to evaluate the performances of the models (model evaluation), understand the links between input and output variables (sensitivity analysis), and lower the model errors (parameter estimation and data assimilation).

Optimization: The scientific purpose of the RECORD project is to define innovative farming systems in a context of climate change and sustainable development. In this context, some methods such as the optimization by simulation, multi-criteria approach or data mining seem to be full promise tools that need to be integrated to the platform (Bergez *et al*. 2006).

2. TOWARD A PLATFORM ARCHITECTURE

Due to the large number of functional requirements for the RECORD project, starting from scratch such an ambitious modeling platform would not be efficient. We thus considered the option of re-using and extending existing softwares. In order to analyze and qualify existing solutions, we made explicit some fundamental requirements, at technical or conceptual levels, for the different platform users.

2.1. Different needs for different users

When we consider the community of users the RECORD project has to serve, we can distinguish between modeling scientists or modelers, and computer software engineers. Modelers are focused on the activities of designing and analyzing models, and are helped in these tasks by software engineers, or programmers, that code parts of the model, run simulations and obtain experimental data.

Modelers require a platform where they can easily implement their models, using dedicated languages as close as possible to the mathematical and logical formalisms they use in their scientific publications. Most of them are specialist of one sub-system (the soil or the plant dynamics, the farmer's decision process, etc.) and they need to integrate big parts of other modeling projects in order to build all together simulation models for analyzing cropping systems.

Software engineers require a clear framework, where they can be efficient in their tasks of implementing model components and running simulations. Many services should be available from this framework, like simulation engines, or visualization tools for easing the development of graphical user interfaces.

At that time, the RECORD platform is a research tool and is not designed for being used as an operational decision support tool in a management context.

2.2. Different criteria to consider

We identified several main issues to consider in the definition of the platform architecture.

A systemic and object-oriented approach

A main contribution of the framework should be to encourage modelers to adopt a systemic approach when they model cropping systems. For that, we propose to provide a convenient way to design clearly hierarchically structured models, and to make available to users all the mechanisms for easily defining systems and subsystems through their interactions, their states, their inputs and outputs and their dynamics. We see this systemic approach as a way of enhancing model intelligibility and reusability, by allowing a clear identification of component boundaries, and by easing their extraction in the process of elaborating new models.

This systemic approach is clearly suitable but must be accompanied with capabilities to adapt existing model components. Some specific features might need to be added, enriched or removed. Facilities for modeling by composition are important but the possibility to create brand new components and models is fundamental as well. The object-oriented paradigm and the incorporated ideas of polymorphism, inheritance, interface-based communication are offering clean and efficient means in the modeling process. In other words, the objected-oriented representation should ideally be embedded with the systemic structuring approach.

A multi-formalism framework

We already saw that cropping system modeling could require the use of different modeling formalisms that have to be simulated all together. The architecture of the framework should then clearly separate the layer dedicated to the scientist user, based on representation languages directly related to the modeling formalisms used by modelers, from the layer that realizes integration of these different modeling formalisms. We must then consider that there are two specific levels for describing frameworks: the modeling level and the implementation level (Van Evert *et al.* 2005). The underlying requirement of this architectural design is then the need of strong theoretical basis and existing algorithms to address the required formalism heterogeneity through these two levels.

Coupling models

The framework should support the co-existence of different types of coupling between several model components. To build a new model, components might be associated in a very loose way. In that case the idea is to manage parallel simulation of different models, with some data exchange at run time between these models seen as communicating software components. When several models need to be closely connected and require frequent synchronizations, sophisticated protocols are required in order to avoid inefficiency problems. In this case, it seems preferable to have a single simulation machinery and to force all models to conform to a common implementation-level formalism. Such a formalism should be able to deal with a large range of simulation models, either continuous or of discrete event type.

Extensibility

The framework has to provide flexibility, compatibility, but also extensibility. It is clear that it will be impossible to solve all needs in one step, and the project development should be iterative. New needs are also expected in the future, like specific visualization tools, data access methods, statistical and optimization algorithms. Software engineers in research teams should thus have the possibility to create, implement and commit add-ons on the platform. The structure of the underlying framework should thus be well engineered, and formal specifications should be available.

Simulation efficiency

Scientists are tackling more an more complex modelling projects. Multi-simulation analysis and optimization approaches require millions of simulation runs. Simulation of cropping systems at a regional scale requires the storage of a huge amount of simulation data at run-time. The question of time and space efficiency of modeling and simulation is thus a critical issue to consider.

3. EXPLORING EXISTING SOLUTIONS

We present here some free software frameworks for simulation that we have explored for the RECORD project. There exists today more than one hundred simulation packages (Rizzoli, 2007) and thus our approach had been to evaluate a few typical integrated environments, representative of recent simulation software solutions. We only considered in this exploration simulation environments relying on rigorous formalisms for describing systems dynamics and systems composition. We thus ignored solutions only designed for linking model engines by data exchange at run time, that follow a software component based approach like in the OpenMI framework, used for integrated water management modeling (Gregersen and Blint 2004). The retained frameworks were evaluated by examining their intrinsic properties and, for some, by applying them to the recoding of typical cropping system simulation models such as MODERATO (Bergez *et al.*, 2001) for maize irrigation management, focusing on the main criteria we have just developed above.

3.1. Simile, a visual software.

Simile (Muetzelfeldt and Massheder, 2003) proposes a visual modeling environment, on a declarative basis. It is similar to other environments like Stella, Modelmaker or Vensim. All these graphical softwares are designed to help modelers in developing dynamic system models, typically represented by stock and flow diagrams. Simile has been developed primarily for ecological and environmental research, and thus also supports spatial and individual-based modeling on an object-based representation. However in these frameworks, discrete event formalisms are generally not fully supported.

Simile supports modular modeling, with plug-and-play modules and displays. Its declarative model representation makes model manipulation easy. Simile can generate C++ model code for fast execution.

With Simile-like software, model coding is almost unnecessary. Development by software engineers mainly concerns display modules. The Simile software does not appear to be designed for allowing extension for some new specific modeling domains like cropping system management.

3.2. MODCOM, a library of dedicated classes

MODCOM (Hillyer, 2003) is a modeling framework for ecosystems that essentially aims at helping modelers in designing models by coupling model components or modules. MODCOM is a procedural approach, and model components need to be programmed. The recent version of MODCOM was written for Microsoft's .NET framework. It is a software library of classes and interfaces written in C# language.

MODCOM is a framework that allows the linking of components modeling dynamical systems as set of ordinary differential equations. Each model, or component, is associated with a set of state-variables, and the task of the component is to compute the rate of state-transition for the specified time step of the simulation. MODCOM handles communication between linked components, organized as a collection of simulation objects which are managed by a simulation environment, that offers high-level simulation services, numerical integration of differential equations, data management, data analysis, etc. Communication between models relies on input and output ports, but the framework also allows the management of discrete events.

MODCOM does not target explicitly spatial or agent modeling. However, a generic AgroManagement component has been developed (Donatelli *et al.*, 2006), based on the MODCOM event mechanism. This component implements a collection of decision rules that are examined every simulated day. Finally, MODCOM does not allow hierarchical or composite modeling, where model components could be defined in terms of other model components.

In MODCOM, modeling is essentially coding. The proposed framework is clear and well structured, and as long as one considers ODE (ordinary differential equation) models, implementation is quite straight forward. For modeling projects that require other simulation mechanisms than ODE integrators, the use of a high-level language like C# makes possible to broaden the applicability of the framework, but with the possible loss of a formal model specification.

3.3. DIESE

DIESE (Martin-Clouaire and Rellier 2003) is a modelling and simulation framework designed to support the study of dynamical systems. DIESE relies on object-oriented design principles and the discrete event simulation paradigm. It enables to model continuous, discontinuous aspects of the functioning of the system that is the subject of the simulation study. A system is represented through its structural, functional and dynamic aspects that, at the modelling level, translate respectively into as set of entities, processes and events. A general simulation mechanism iteratively processes the next event scheduled in the agenda, which fires asynchronously the processes embedded in it and causes state changes of the system. Information about the state of the system is virtually fully accessible to any process at any time. Entities, process and events are generic constructs that can be used directly to model biophysical system with, for instance, plant, leaf and fruit as entities and plant growth and development as processes, and events initializing or ending processes.

In order to ease the modelling of production system, the above constructs have been specialized to represent the production management aspects. Special entities, such as operations, activities and resources have been introduced to model the logic and temporal articulations involved in an intended management strategy. These extensions enable to represent flexible plans. Specific processes have been developed to realize a dynamic and context-dependent interpretation of a plan and, more generally, simulate decision-making and action execution procedures.

At implementation level, DIESE is a C++ package that comes with two additional tools aiming at the model developer and the simulation analyst. The first one facilitates the specification and/or specialisation of the classes of entities, processes and events needed in the system domain of interest. It can then put together these pieces and generate an executable program from the so-specified system model. The second one enables to specify an instance of the system of concern by instantiating the classes written by the model developer. This tool provides also some capabilities to setup batches of simulation trials and display output dynamically during simulation.

Currently, no ready-to-use facility is provided for modeling spatial processes (e.g. pest propagation) and multi-agent modeling (e.g. interacting farms) is not supported. DIESE is presently used in several production system studies (see, for instance, Chardon *et al.*, 2007).

3.4. DEVS and VLE

The Discrete Event System Specification (DEVS) is a well known approach to discrete event system modelling and simulation (Zeigler *et al.*, 2000). It is a general formalism that allows a modular and hierarchical representation of system models. Atomic DEVS models are specified by their internal state, by a set of input and ouput ports through which external events are received, processed and sent, depending on transition and output functions. The time a model can stay in its state is also explicitly represented. DEVS coupled models are defined by coupling DEVS models, the output events of one model being converted into input events of other models. Atomic or coupled DEVS models can be seen as modular components comprising input and output interfaces in the form of their input and output ports that are simulated by DEVS abstract simulators.

One key advantage of the DEVS approach is that it allows multi-modeling, that is, system modeling with different mathematical formalisms. Two main approaches are possible. First, one can specify all models in DEVS and use a common DEVS simulator. Different DEVS extensions have been developped specially for modelling and simulating ODE, cellular automata, agents. A second possibility consists in wrapping some non-DEVS models within a DEVS interface, and then coupling the different simulators themselves through the concept of DEVS-Bus.

There are several modeling environments implementing DEVS abstract simulators such as DEVSJAVA, CD++ or VLE (Virtual Laboratory Environment). VLE (Quesnel *et al.*, 2007) is a modelling and simulation platform for complex environmental systems studies. It can integrate heterogeous formalisms by specifying DEVS models developped in various programming languages, through the use of a complete and portable API written in C++. The VLE framework has been designed to be efficient, easily modifiable and includes several DEVS extensions.

4. DISCUSSION AND CONCLUSION

The RECORD project is now at a point where we can start adapting and extending some existing modeling platforms, in order to study cropping systems management by simulation of biophysical and decisional processes. The different solutions we considered were evaluated with regards to some key criteria related to systemic, object-oriented and multi-formalism modeling, ability to couple models, extensibility and efficiency of the framework.

Visual modeling environments like Simile present highly interesting advantages when one is concerned by simulating small and simple system dynamics models, but they do not fit most of the functional requirements of the RECORD project. Environments like MODCOM, DIESE, or VLE appear to fit better with the development of larger and more diverse models as expected with RECORD. However, MODCOM is mainly dedicated to dynamical models represented by flat sets of ODE or difference equations, and does not really support multi-formalism modeling, although it can manage discrete events. DIESE shares with MODCOM this same last limitation, but proposes richer operational concepts for representing structural features of the systems. Another originality of DIESE stems from its capability to deal with complex flexible management strategies.

At this juncture, we defend a solution that would retain the object-oriented modeling approach of DIESE, VLE and MODCOM and that would rely on a strong mathematical formalism for describing systems dynamics and systems composition. We intend to build the RECORD platform on top the VLE environment that embeds the DEVS formalism. This choice, which satisfies the criteria of section 3.2, comes with multi-formalism modeling capabilities and a time-explicit simulation engine. The RECORD project focuses then on the development of some DEVS extensions targeted to the specific needs of cropping system studies. In particular, one of them concerns the production management aspect and should greatly benefit from the DIESE experience. For more about RECORD, see http://record.toulouse.inra.fr.

5. REFERENCES

Aubry C., Papy F., Capillon A. 1998. Modeling of decision making processes for annual crop management. *Agricultural Systems* 56: 45-65.

Bergez J.-E., Debaeke P., Deumier J.-M., Lacroix B., Leenhardt D., Leroy P., Wallach D. 2001. MODERATO, an object-oriented tool for designing maize irrigation schedule. *Ecological Modelling* 137, 43-60.

Bergez J.-E., Garcia F., Wallach D. 2006. Representing and optimizing management decisions with crop models. In Working with Dynamic Crop Models: Evaluating, Analyzing, Parameterization and Applications. Wallach D., Makowski D, Jones J.W. (eds.), Elsevier, Amsterdam.

Chardon X., Rigolot C. *et al.* 2007. MELODIE : a whole-farm model to study the dynamics of nutrients in integrated dairy and pig farms. Proc. of MODSIM 2007

Chatelin M.-H., Aubry C., Poussin J.-C., Meynard J.-M., Massé J., Verjux N., Gate P., Le Bris X. 2005. Déciblé, a software package for wheat crop management. *Agricultural Systems* 83(1), 77-99.

Donatelli M., van Evert F., Di Guardo A., Adam M., Kansou K. 2006. A component to simulate agricultural management. *Proc. iEMSs congress*, Vermont, USA, July.

Garcia F., Guerrin F., Martin-Clouaire R., Rellier J.-P. 2005. The Human Side of Agricultural Production Management – the Missing Focus in Simulation Approaches. *Proc. of the MODSIM 2005 Int. Cong. on Modelling and Simulation*, Melbourne.

Gregersen J.B., Blint M. 2004. OpenMI: The Essentials Concepts and their Implications for Legacy Software. Proc. of the IEMSS Conf. Complexity and Integrated Resources Management, Osnabrück, Germany.

Hillyer C., Bolte J., van Evert F., Lamaker A. 2003. The ModCom modular simulation system. *European Journal of Agronomy* 18, 333-343.

Martin-Clouaire R., Rellier J.-P. 2003. A conceptualization of farm management strategies. Proc. of EFITA-03 conference, July 5-9, Debrecen, Hungary, 719-726.

Muetzelfeldt R., Massheder J. 2003. The Simile visual modelling environment. *European Journal of Agronomy* 18, 345-358.

Quesnel G., Duboz R., Ramat E., Traoré M.K. 2007. VLE - A Multimodeling and Simulation Environment. *Proc. of SCS'07 conference*, San Diego, USA.

R Development Core Team. 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rizzoli A.E. 2007. A collection of modeling and simulation resources on the internet. www.idsia.cd/~andrea/sim/simtools.html.

Van Evert F., Holzworth D., Muetzelfeldt R., Rizzoli A. and Villa F. 2005. Convergence in integrated modeling frameworks. *Proc. of the MODSIM 2005 Int. Cong. on Modelling and Simulation*, Melbourne.

Wallach D. 2006. Evaluating crop models. In Working with Dynamic Crop Models: Evaluating, Analyzing, Parameterization and Applications. Wallach D., Makowski D., Jones J.W. (eds.), Elsevier, Amsterdam.

Zeigler B.P., Kim D., Praehofer H. 2000. *Theory of Modeling and Simulation*, Second Edition. Academic Press.