

A framework for combining Business Intelligence & Agent-based Simulation

Truong Minh Thai^(a), Truong Xuan Viet^(b,c,d), Frédéric Amblard^(a), Alexis Drogoul^(b), Benoit Gaudou^(a),
Huynh Xuan Hiep^(d), Le Ngoc Minh^(c), Christophe Sibertin-Blanc^(a)

(a) UMR 5505 CNRS, Institut de Recherche en Informatique de Toulouse, Université Toulouse 1 Capitole, France

(b) UMI 209 UMMISCO-IRD/UPMC, Bondy, France

(c) Faculty of Computer Science & Engineering, HCMUT, Ho Chi Minh City, Vietnam

(d) DREAM Team/UMI 209 UMMISCO-IRD, Can Tho University, Vietnam

thai.truongminh@gmail.com, txviet@ctu.edu.vn, frederic.amblard@ut-capitole.fr, alexis.drogoul@gmail.com,
benoit.gaudou@ut-capitole.fr, hxhiep@ctu.edu.vn, minhle@cse.hcmut.edu.vn, sibertin@ut-capitole.fr

Abstract— Integrated environmental modeling approaches, especially agent-based modeling ones, are increasingly used in large-scale decision support systems. A major consequence of this trend is the manipulation and generation of a huge amount of data in simulations, which require to be efficiently managed. In this paper, we present a Combination Framework of Business intelligence solution and Multi-agent platform (CFBM) and its implementation on the GAMA platform. CFBM is a logical framework dedicated to the management of the input and output data of simulations, as well as the corresponding empirical datasets in an integrated way. To this purpose we propose to couple two major methodologies: Multi-Agent Simulation (MAS) on one hand and Business Intelligence (BI) solution on the other. An integrated model of a surveillance network and Brown Plant Hopper (BPH) invasion is presented and used throughout the paper as a case study of managing the data used and generated during the life-cycle of a simulation, from its initialization using real data to the generation of simulation outcomes. We show how the proposed combination framework can turn a model into an effective decision-support system, by helping the modeler to manage what-if scenarios.

Keywords: *BI Solution, Data Warehouse, Multi-Agent Simulation, Agent-Based Model, Brown Plant Hopper, Decision Support System.*

I. INTRODUCTION

Integrated socio-environmental modeling in general and multi-agent based simulation approach applied to socio-environmental systems in particular are increasingly used as decision-support systems in order to design, evaluate and plan public policies linked to the management of natural resources [7]. For example, in our research about invasion of Brown Plant Hopper (BPH) and the effects of BPH on rice fields of the Mekong Delta region (Vietnam), we must develop and integrate several models (e.g. BPH growth model, light trap model, BPH migration model). We must also integrate data from different data sources and analyze the integrated data at different scales. In such an integrated simulation system involving high volume of data, we are not only concerned with modeling driven approach - that is how to model and combine models from different scientific fields - but also with data driven approach - that is how to handle big data from different data sources and perform analyses on the integrated data from these sources.

The basic statement we can make is that the design and simulation of models have greatly benefited from the advances

in computer science through the popularized use of simulation platforms such as Netlogo [16], Repast [2] or GAMA [12]. But this is not yet the case for the management of data, which are still managed in an ad hoc manner, despite the advances in the management of huge datasets. Such a statement is rather pessimistic if we consider recent tendencies toward the use of data-driven approaches in simulation aiming at injecting more and more data available from the field into simulated models. These are the reasons why we propose a robust handling data solution of huge datasets for multi-agent simulations.

In the following section, we first present the state of the art linking BI and MAS. The global architecture of the CFBM and its implementation in GAMA are presented in section III. In section IV, we present the integrated model of BPH invasion as a case study for the application of CFBM. Discussion of pros and cons of our framework will conclude this article.

II. RELATED WORKS AND METHODOLOGIES

A. Integration of BI solution and simulation systems

Data warehouse (DW) and analysis tools such as BI solutions can help users to manage a large amount of simulation data and to make several data analyses that support the decision-making processes [6][5]. The combination of simulation tools and DW is increasingly used and applied in different areas. For example, although [8][11] are only two applications of OLAP technologies to a special problem. These works demonstrate that a multidimensional database is suitable to store several hundreds of thousands of simulation results. Simulation models, DW and analysis tools with OLAP technologies were also involved in decision support systems or forecast systems [15][4]. In [9], Mahboubi et al. also used data warehouse and OLAP technologies to store and analyze a huge amount of output data generated by the coupling of complex simulation models such as biological, meteorological and so on. In particular, the authors proposed data warehouse and Online Analytical Processing tool (OLAP tool) as trend for storing and analyzing simulation results.

The state of the art demonstrates therefore the practical possibility and the usefulness of the combination of simulation, data warehouse and OLAP technologies. It also shows the potential of a general framework that is, to our knowledge not yet proposed in the literature.

B. Model of BPH invasion

To build a dynamics system of the BPH propagation, the migration laws take an important role in the model. Coupled map lattice approximations are introduced in [1]. This model is based on an individual-based model to simulate the migration behaviors of individuals via a cellular automaton. Two different agent-based BPH migration models have also been developed in [10]. In addition, it requires a huge amount of data. We will thus use it as case study for our framework.

III. COMBINATION FRAMEWORK OF BI SOLUTION AND MULTI-AGENT PLATFORM

A. CFBM - a logical framework to combine BI solution and Multi-agent platform

CFBM is formed by three systems and it supports four tools: model design tool, model execution tool, execution analysis tool and database tool. The architecture of the CFBM is illustrated in Figure 1.

1) Simulation system

The simulation system plays two roles: model design tool and model execution tool. It is composed of a multi-agent platform and a relational database. This system helps to implement simulation models, execute models and handle their input/output data.

Three layers with five components compose the simulation system. The **simulation interface** is a user environment that helps the modeler to design and implement his models, executes the models and visualizes results. **Multi-agent simulation models** are a set of multi-agent based models. They are used to simulate phenomena that the modeler aims at studying. The **SQL-agent** is a particular kind of agent that supports Structured Query Language (SQL) functions to manage the input/output of simulation models. The **reality database** is used to store empirical data gathered from the target system that are needed for the simulation and analysis phases. **Simulation data** is used to manage simulation models, simulation scenarios and output results of the simulation models.

2) Data warehouse system

The data warehouse system plays the role of a database tool. This system is divided into three parts. **ETL**(Extract-Transform-Load) is a set of processes with three responsibilities. First, it extracts all kind of data (empirical data and simulation data) from the simulation system. Second, ETL transfers the extracted data into an appropriate data format. Finally, it loads the transferred data into a data warehouse. **Data warehouse** is used to store historical data, which are loaded from simulation system by ETL. **Data mart** is a subset of data stored in the data warehouse and it is a data source for the concrete analysis requirement.

3) Decision support system

In CFBM, the decision support system plays the role of analysis tool. It is a software environment supporting analysis, decision-making features and visualization of results. In our design, we propose to use existing OLAP analysis tools, or a multi-agent platform with analysis features or a combination of both options. The decision support system of CFBM is built on

four parts. **Analysis interface** is a user interface used to handle analysis models and visualize results. **Multi-agent analysis models** are a set of agent-based analysis models. They are created based on analysis requirements and handled via analysis interface. **MDX-agent** is a bridge between multi-agent analysis models and data marts. This agent supports MultiDimensional eXpressions (MDX) functions to query data from a multidimensional database. **OLAP analysis tools** are analysis software packages that support OLAP operators.

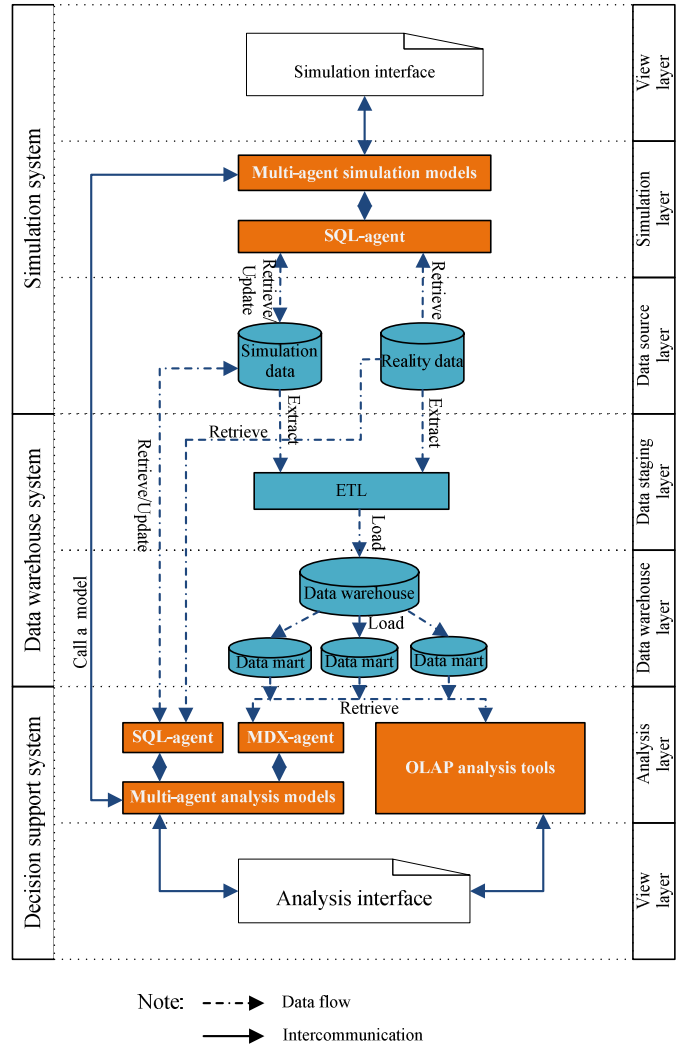


Figure 1. Combination framework of BI solution and multi-agent platform architecture

B. Implementation of CFBM in GAMA

We have chosen to implement CFBM into the GAMA platform following the software architecture illustrated in Figure 2.

Presentation tier plays the role of view layers in CFBM architecture. In our implementation, the GAMA user interface plays this role. It is used to write models or analyze models, execute the models and visualize results of models in different modes of views (text, chart, GIS or 3D).

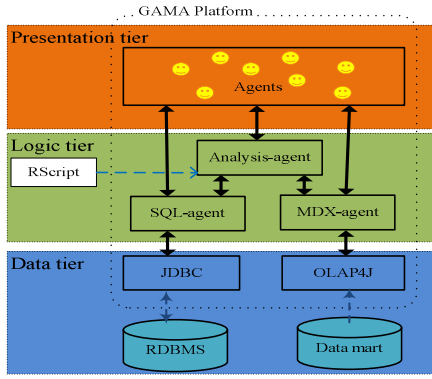


Figure 2. Software architecture of CFBM in GAMA

Logic tier coordinates the application process commands, as it plays the role of two layers in CFBM architecture (simulation player and analysis layer). In our implementation, it contains four components. **Analysis-agent** is designed to supply statistical analysis functions. **SQL-agent** is responsible for retrieval as well as update on relational database. **MDX-agent** is binding on retrieval data from data marts. **RScript** works as an external service.

Data tier plays two roles: data source layer and data warehouse layer. The main functions of this tier are to store and retrieve data from a database or a file system. OLAP4J is a Java API to connect to and query data from a multidimensional database.

IV. APPLYING THE CFBM TO BROWN PLANT HOPPER PREDICTION MODEL

A. Integrated Model of BPH Prediction Model and Surveillance Network Model

The integrated model is composed of two sub-models, both implemented in the GAMA platform: (1) BPH Prediction Model, a co-model of growth & migration models and (2) Surveillance Network Model (SNM). Inputs and outputs of the integrated model are handled via the CFBM in GAMA.

1) BPH prediction model

To predict the BPH density, we propose a BPH prediction model. This model contains two sub-models: BPH Growth Model and BPH Migration Model.

The studied region is discretized by a grid of square cells, which is appropriate for implementing a cellular automaton in agent-based modeling. This cellular automaton is considered as a stochastic model where the observed variable is modeled as a random process. Each cell in the stochastic model represents a square area of the real system. The prediction model is based on interactions between two sub-models:

▪ BPH migration model

The migration process of BPHs in the studied region is modeled by a dynamical moving process in cellular automata.

Denoting $x(t)$ the number of adult BPHs at time t . The migration model essentially determines the outcome $x_{out}(t)$ at next time $t + 1$ from a specific source cell and the rates of $x_{out}(t)$ moving to all destinations at time $(t + 1)$. Destination cells are determined by the semi-circle under the wind, while the radius

of the circle is determined by the wind velocity and the migration time in a day. The local constraints are also considered by two combinational indices: attractiveness index and obstruction index [14].

▪ BPH growth model

In this paper, we apply a deterministic model with T variables where T is the life cycle of the insect. To simplify the implementation process, these variables will be managed by a vector V of T elements, where the element $V[i]$ denotes the number of insects at age i . The growth model will update the values of vector V for each simulation step.

2) Surveillance Network Model (SNM)

The model is presented in several publications[13]. It is based on the Unit Disk Graph technique and autocorrelation between different surveillance devices. In[13], the surveillance network is modeled as a Correlation & Disk graph-based Surveillance Network (CDSN). Each surveillance device contains the trap-density at a specific geographic location.

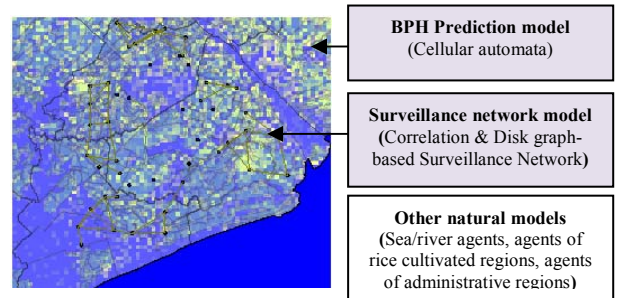


Figure 3. Integrated model of SNM and prediction co-model.

The surveillance network model is used to measure the trap-density in the prediction model. Each surveillance device agent monitors the trap-density based on the location of its cell agent. We define the trap-density as equivalent to the number of adult BPHs located in this cell agent.

B. Experiments

1) Three scenarios of the environment

This section presents three different results of prediction validation relevant to three specific scenarios for BPH cycle (BPH growth model) and wind (BPH migration model). The prediction process by simulation is applied in a short period of one month. In all three scenarios, the wind direction is set to northeast with a velocity of 9.4 ± 3.5 km/h.

TABLE I. PARAMETERS FOR LIFE CYCLE.

Parameters	Scenario 1 (Growth model)	Scenario 2 (Co-model)	Scenario 3 (Co-model)
Egg time span	7	7	7
Nymph time span	13	13	13
Egg giving time span	6	6	12
Life time cycle of BPH	32	32	32
Ratio of egg number can become the nymph	0.42	0.42	0.40
Ratio of nymph number can become the adult	0.42	0.42	0.40
Ratio of eggs can be produced by an adult	250.0	250.0	350.0
Ratio of natural mortality	0.245	0.245	0.35

2) Validation with RMSE

▪ Empirical data

We used the data from 48 light-traps of three typical provinces in the Mekong Delta region: Soc Trang, Hau Giang and Bac Lieu from January 1, 2010. We use the data of the first 32 days ($T = 32$) to apply T Kriging estimations by adding the Gaussian noises[3].

▪ **Simulation data**

Figure 5 shows the simulation data corresponding to the real data of Figure 4. The first 32 days are related to the estimation of trapped densities at the cells located by light-traps and the prediction data is considered from the 33th day.

Figure 5a demonstrates the simulation data of scenario 1 with a sinusoidal form. However, the form of empirical data is very varied, as it looks like a peak of insect amount. In Figure 5b, the form of BPH trapped density is more similar to real data. This phenomenon is explained by migration process.

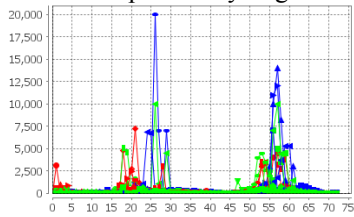


Figure 4. Empirical data from Jan 01, 2010 in three provinces: Soc Trang, Hau Giang and Bac Lieu.

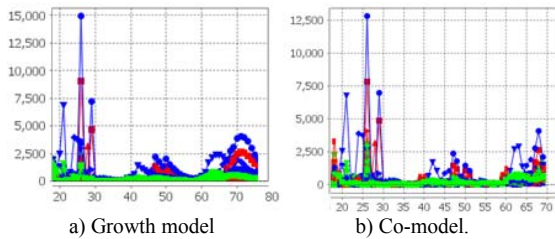


Figure 5. Prediction data in two different scenarios.

Table 2 shows the RMSE of three scenarios listed in Table 1. The RMSE values are independently calculated for two stages of simulation data:

- RMSE (Estimation): Between the empirical data and Kriging estimation (first 32 days of simulation data).
- RMSE (Prediction): Between the empirical data and prediction (from the 33th day of simulation data).

TABLE 2. RMSE OF THREE SCENARIOS

Validation result	Scenario 1	Scenario 2	Scenario 3
RMSE (Estimation)	614.3536	657.2320	323.9902
RMSE (Prediction)	535.5026	524.7963	443.3628

V. CONCLUSION

The key features of CFBM are that it supplies four components: (1) model design, (2) model execution, (3) execution analysis and (4) database management. These components are coupled and combined in a uniformed simulation system. The most import point of CFBM is the integration power of BI solution and multi-agent based platform that is useful to develop complex simulation systems with high volume of data.

From a conceptual framework, CFBM is realized by a concrete implementation of the framework in GAMA and an

application to Brown Plant Hopper Prediction model. By mean of the implementation achieved in GAMA, it proves that CFBM is an open and modular architecture. For instance, GAMA has succeeded in interacting with several database management systems (SQLite, PostgreSQL, MySQL and Microsoft SQL Server), analysis services (Pentaho Analysis Service and SQL Server Analysis Service).

In our case study, an integrated model of Surveillance network model and a BPH prediction co-model (BPH Growth & Migration Model) are implemented as a typical application of CFBM. The validation process with three different scenarios is also executed. Applying the CFBM for this integrated model allows for a more flexible and portable mechanism in querying, updating and analyzing the simulation data for numerous research questions.

REFERENCES

- [1] Brännström, Å. and Sumpter, D.J. 2005. Coupled map lattice approximations for spatially explicit individual-based models of ecology. *Bulletin of mathematical biology* (2005), 663–682.
- [2] Collier, N. 2003. Repast: An extensible framework for agent simulation. The University of Chicago’s Social Science Research. 36, (2003).
- [3] Cressie, N.A. 1993. *Statistics for Spatial Data*, revised edition. Wiley, New York.
- [4] Ehmke, J.F. et al. 2011. Interactive analysis of discrete-event logistics systems with support of a data warehouse. *Computers in Industry*. 62, (2011), 578–586.
- [5] Inmon, W.H. 2005. *Building the Data Warehouse*. Wiley Publishing Inc.
- [6] Kimball, R. and Ross, M. 2002. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. John Wiley & Sons, Inc.
- [7] Laniak, G.F. et al. 2013. Thematic issue on the future of integrated modeling science and technology. *Environmental Modelling & Software*. 39, (2013), 13–23.
- [8] Madeira, H. et al. 2003. The OLAP and data warehousing approaches for analysis and sharing of results from dependability evaluation experiments. *IEEE/IFIP International Conference on Dependable Systems and Networks, Dependable Computing and Communications, DSN-DCC*. (2003), 22–25.
- [9] Mahboubi, H. et al. 2010. A Multidimensional Model for Data Warehouses of Simulation Results. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*. 1, (2010), 1–19.
- [10] Nguyen, V.G.N. et al. 2011. On weather affecting to brown plant hopper invasion using an agent-based model. *Proceedings of the International Conference on Management of Emergent Digital EcoSystems* (2011), 150–157.
- [11] Sosnowski, J. et al. 2007. *Developing Data Warehouse for Simulation Experiments*. *Lecture Notes in Computer Science* (2007), 543–552.
- [12] Taillandier, P. et al. 2012. GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In *Proc. of PRIMA 2012*, 242–258.
- [13] Truong, V.X. et al. 2012. *Modeling a Surveillance Network Based on Unit Disk Graph Technique—Application for Monitoring the Invasion of Insects in Mekong Delta Region*. *PRIMA 2012: Principles and Practice of Multi-Agent Systems*. Springer. 228–242.
- [14] Truong, V.X. et al. 2013. Optimizing an Environmental Surveillance Network with Gaussian Process – An optimization approach by agent-based simulation. *The Sixth International KES Conference on Agents and Multi-agent Systems – Technologies and Applications (KES AMSTA 2013)* (2013), 102–111.
- [15] Vasilakis, C. et al. 2008. A decision support system for measuring and modelling the multi-phase nature of patient flow in hospitals. *Intelligent Techniques and Tools for Novel System Architectures* (2008), 201–217.
- [16] Wilensky, U. 1999. *Netlogo*, Technical report. Netlogo, Technical report (1999).