



**HAL**  
open science

## **An implementation of framework of business intelligence for agent-based simulation**

Minh Thai Truong, Frédéric Amblard, Benoit Gaudou, Christophe Sibertin-Blanc, Viet Xuan Truong, Alexis Drogoul, Hiep Xuan Huynh, Minh Ngoc Le

► **To cite this version:**

Minh Thai Truong, Frédéric Amblard, Benoit Gaudou, Christophe Sibertin-Blanc, Viet Xuan Truong, et al.. An implementation of framework of business intelligence for agent-based simulation. 4th Symposium on Information and Communication Technology (SoICT 2013), Dec 2013, Da Nang, Vietnam. pp.35-44, 10.1145/2542050.2542069 . hal-04082724

**HAL Id: hal-04082724**

**<https://hal.science/hal-04082724>**

Submitted on 26 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>  
Eprints ID : 12829

Official URL: <http://dx.doi.org/10.1145/2542050.2542069>

**To cite this version** : Truong, Minh Thai and Amblard, Frédéric and Gaudou, Benoit and Sibertin-Blanc, Christophe and Truong, Viet Xuan and Drogoul, Alexis and Hyunh, Xuan Hiep and Le, Minh Ngoc *An implementation of framework of business intelligence for agent-based simulation*. (2013) In: 4th Symposium on Information and Communication Technology (SoICT 2013), 5 December 2013 - 6 December 2013 (Da Nang, Viet Nam).

Any correspondance concerning this service should be sent to the repository administrator: [staff-oatao@listes-diff.inp-toulouse.fr](mailto:staff-oatao@listes-diff.inp-toulouse.fr)

# An Implementation of Framework of Business Intelligence for Agent-based Simulation

Thai Minh Truong,  
Frédéric Amblard,  
Benoit Gaudou

Christophe Sibertin-Blanc  
UMR 5505 CNRS-IRIT/ Université  
Toulouse 1 Capitole, France

thai.truongminh@gmail.com  
frederic.amblard@ut-capitole.fr  
benoit.gaudou@ut-capitole.fr  
sibertin@ut-capitole.fr

Viet Xuan Truong,  
Alexis Drogoul  
UMI 209 UMMISCO-  
IRD/UPMC, Bondy, France

txviet@ctu.edu.vn,  
alexis.drogoul@gmail.com

Hiep Xuan Huynh  
DREAM Team/UMI 209 UMMISCO-  
IRD, Can Tho University, Vietnam

hxhiep@ctu.edu.vn

Minh Ngoc Le  
Faculty of Computer Science &  
Engineering, HCMUT, Ho Chi Minh  
City, Vietnam

minhle@cse.hcmut.edu.vn

## 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 and the analysis on integrated data, which we propose as a coupling of 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. This model is used throughout the paper as a case study of CFBM regarding to the database accesses, the integration of internal and external processing, and the model validation tasks. We show how the proposed combination framework can be an appropriate choice for all decision-support systems with a huge data, hence the modelers can be also encouraged to develop the big model in a large scale.

## Keywords

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

## 1. 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 [11]. For example, in our research about invasion of Brown Plant Hoppers (BPH) and their effects on rice fields in the Mekong Delta region of Vietnam, we must develop and integrate several models (e.g. BPH growth model, light-trap model, rice growth model, BPH migration model). We must also integrate data from different data sources and analyze the integrated data at different scales. In such integrated simulation system involving high volume of data, we are not only concerned with modeling driven approach – that is how to model and combine coupling 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 currently, 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 [24], Repast [2] or GAMA [20]. This is not yet the case for the management of data, which is still managed in an ad hoc manner, despite the advances in the management of huge datasets (data warehousing solutions for instance). 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 based simulations.

In our research, the first question, we must tackle is “What is the general architecture that can adapt several purposes: model and execute multi-agent simulation, manage input/output data of simulations, integrate data from different sources and analyze high volume of data?”. To solve this problem, we studied on several researches and solutions, related to simulation, management and analysis on big data. We argue that BI solution is a good way to handle and analyze big data. Because a BI solution contains a data warehouse, integrated data tools (Extract-Transform-Load tools) and Online Analytical Processing tools (OLAP tools), they are very adapted to manage, integrate and

analyze on huge amounts of data. Our answer for the first question is the combination framework proposed in Section 3.

The second problem that needs to be solved in our research is "How to combine the power of the two disciplines (BI and MAS) into a multi-agent based simulation system with a huge amount of data?" The solution was proposed in this article is improve agent-based platforms by adding new features, such as deep interactions with data warehouse systems.

In the following sections, we first present the state of the art of works linking BI and MAS in Section 2. The global architecture of the combination framework of BI solution and multi-agent platform and its implementation in GAMA is presented in Section 3. In Section 4, we present the integrated model of Brown Plant Hopper invasion as a case study for the application of CFBM. Discussion on advantages and disadvantages of our framework will conclude this article.

## 2. RELATED WORKS

### 2.1 Multi-agent Platform

The basic features of a general computer simulation platform or specifically of a multi-agent based simulation platform are the capability to write models, run simulations of these models and handle results for analysis and/or validation [5][7]. Most simulation platforms (Netlogo, Repast, and GAMA) strongly support these features. But they do not concern on database functions or, when they do, these platforms support database functions with inflexibly in use. For instance, Netlogo has an external package (Netlogo-sql) that supports the execution of SQL statements but "There is no "deep" support for the Netlogo paradigm"<sup>1</sup>. Repast is a toolbox for agent-based modeling in Java, it has numerous libraries to connect the simulator to useful technologies (statistical analysis, GIS, SQL, etc.)<sup>2</sup>. But for using them, users need to have very good java programming skill. GAMA<sup>3</sup> pays a great deal of attention to the modeling of the environment and the use of GIS data but it only supported access GIS data on shape file, did not support features to query GIS data in database managements. That is inflexible for users when they only need to access on a subset of data.

These platforms support good tools for modeling driven approach but they still need additional features for data driven approach.

### 2.2 Integration of BI Solution into a Simulation System

Technologies or tools from different fields used within integrated systems become a common trend today. In particular, for the development of decision support systems or prediction systems based on simulation approaches, data warehouse (DW), online analytical processing (OLAP) technologies and simulation often represent a good solution. Data warehouse and analysis tools as a BI solution can help users to manage a large amount of simulation data and make several data analyses that support the decision-making processes [9][10]. The combination of simulation tools and DW is increasingly used and applied in different areas. In [12], Madeira et al. proposed a new approach dedicated to the analysis and comparison of large amounts of output data from different experiments or from similar experiments across different systems. Data warehouse and OLAP tools were recommended for

collecting and analyzing simulation results of a system [19]. Although [12][19] are only 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 [6][23]. Not only their researches solve specific problems but they also demonstrated a promising use of gathering and analyzing simulation results by using data warehouse and OLAP technologies. In [13], Mahboubi et al. also use 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, they propose a multidimensional data schema of a data warehouse 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 has, as far as we are aware, not yet proposed in the literature.

### 2.3 Model of BPH Invasion

To build a dynamics system of the BPH propagation, the migration laws play 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 reaction of individuals via a cellular automaton. Two different agent-based BPH migration models have also been developed in [15] and [17]. In addition, some investigations for the surveillance network model (SNM) are introduced in [21]. In agent-based modeling, the data can be modeled as a Gaussian process [18] by using Kriging estimation [3][14].

In mathematics, Ordinary Differential Equations (ODE) is a popular approach for growth models. The "Larvae-Pupae-Adult" (LPA) model is a typical model for insects growth model [4]. However, this approach has its own disadvantages, caused by the need to simplify the number of variables. In most cases the limit of parameter number can lead to the uncertainty of model.

In [15] and [17], authors succeeded in applying agent-based modeling approach to model growth of BPH and its invasion on the rice field with different factors. In [21] and [22], Truong et al. also succeeded in using agent-based modeling approach to simulate the Surveillance Network and optimization the light trap network system in Mekong Delta. But all of the models in these researches got some problems when users need to manage input/or output of simulation, to integrate simulation data with empirical data for validation, calibration or other analysis requirements as we present in Section 4.1.

In this paper, we aim at proposing a logical framework to combine data warehouse, OLAP technologies and multi-agent simulation platform to build integrated multi-agent based simulation system with large data. The logical framework and its implementation in GAMA are explained in the next section and the case study is demonstrated in Section 4.

## 3. COMBINATION FRAMEWORK OF BI SOLUTION AND MULTI-AGENT PLATFORM

### 3.1 Computer Simulation System

Fishwick (1997) defined that computer simulation is the discipline of designing a model of an actual or theoretical physical system,

<sup>1</sup> <https://code.google.com/p/netlogo-sql/>

<sup>2</sup> <http://repast.sourceforge.net/docs/RepastJavaGettingStarted.pdf>

<sup>3</sup> GAMA version 1.5 or earlier version.

executing the model on a digital computer, and analyzing the execution output. On the basis of Fishwick's definition, we can define a computer simulation system as below:

A Computer simulation system is a computation system with four components and the intercommunications between them:

- Model design tool: a software environment that supports a modeling language, notations and user interface for modeling an actual or theoretical physical system.
- Model execution tool: a software environment that can run models.
- Execution analysis tool: a software environment that supports statistical analysis features for analysis of output data of models.
- Database tool: a software environment that supports appropriate database and database management features for overall components in the system.

The components of a computer simulation system and their intercommunications are illustrated in Figure 1. This computer simulation system definition deals with two approaches: modeling driven approach and data driven approach.

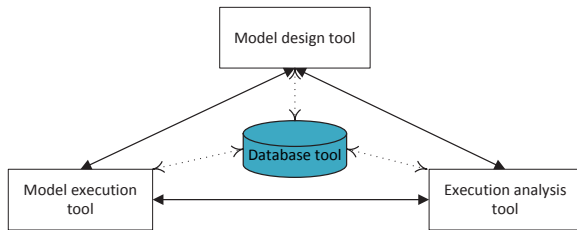


Figure 1: Computer simulation system

The Combination framework of Business Intelligence Solution and Multi-agent platform (CFBM) was designed based on our computer simulation system definition. It means that we have designed CFBM with four major components: The model design tool, the model execution tool, the execution analysis tool and the database tool and the interconnections between. The architecture of the CFBM is summarized in Figure 2. In this framework, we use a BI solution as a database tool, a multi-agent platform as model design tool and model execution tool. For the execution analysis tool, we can either use OLAP analysis tool or use analysis features of the platform (implemented as an external plug-in for the platform, for instance R scripts).

### 3.2 CFBM - a Logical Framework to Combine BI Solution and Multi-agent Platform

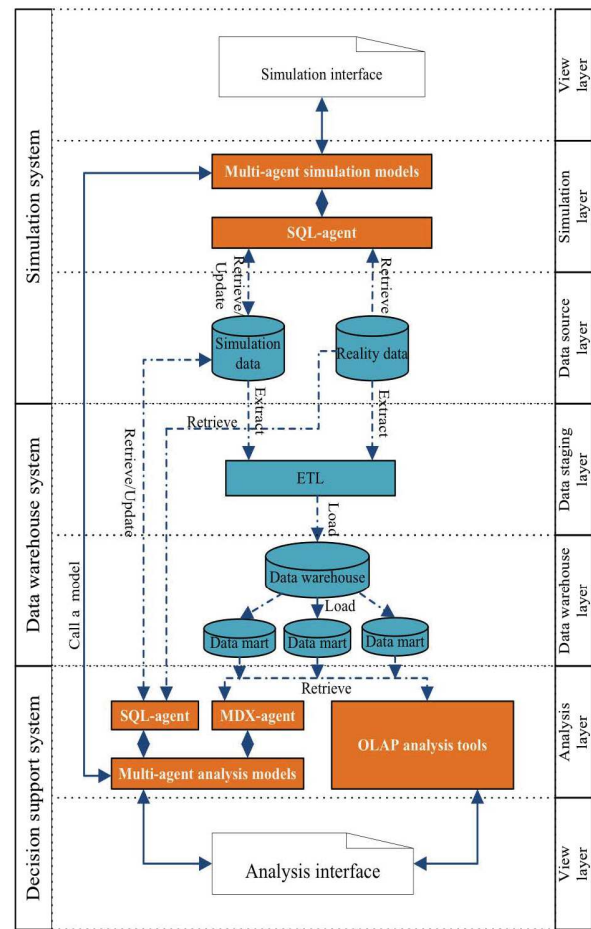
The CFBM is based on three major systems and divided into seven layers. The function of each part is detailed in following sections.

#### 3.2.1 Simulation System

The simulation system is composed of a multi-agent platform with a relational database. This system helps to implement models, execute simulations and handle their input/output data. This part plays the role of an Online Transaction Processing (OLTP) or of an operational source system. It is an outside part of the data warehouse [10].

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, execute

simulations and visualize results. **Multi-agent simulation models** are used to simulate phenomena that the modeler aims at studying. The **SQL-agent** plays the role of the database tool and can access to the relational database. It is a particular kind of agent that supports Structured Query Language (SQL) functions to retrieve simulation inputs from simulation data or reality data, to store output simulation data into simulation data databases and to transform data (in particular the data type) from simulation model to relational database, and conversely. Reality data and Simulation data are relational databases. **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 the simulation models, simulation scenarios and output results of the simulation models. These two data sources will be used to feed the second part of the framework, namely the Data warehouse system.



Note: - - - -> Data flow  
 ———> Intercommunication

Figure 2. Combination framework of BI solution and multi-agent platform (CFBM) architecture.

#### 3.2.2 Data Warehouse System

The data warehouse system is understood as a part of the BI solution. It is very important part to integrate data from different sources (simulation data, reality data and others external data) and it is used as data store to feed data for decision support systems. The data warehouse 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 transforms the extracted data into an appropriate data format. Finally, it loads the transferred data into the data warehouse. **Data warehouse** is a relational database used to store historical data loaded from simulation systems and from other sources by the ETL. **Data mart** is a subset of data stored in the data warehouse and it is a data source for the concrete analysis requirement. We can create several data marts depending on our analysis requirements. Data mart is a multidimensional database, which is designed based on multidimensional approach. It uses star join, fact table and dimension table to present the structure data mart. With multidimensional structure, data mart is particularly useful to help users to improve the performance of analytic processes.

### 3.2.3 Decision Support System

In CFBM, the decision support system component is a software environment supporting analysis, decision-making features and visualization of results. In our design, we propose to use existing OLAP analysis tools, a multi-agent platform with analysis features or a combination of both options. The decision support system of CFBM is built with 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 the analysis interface. One of the key points of our framework is the fact that multi-agent analysis models and multi-agent simulation models are implemented with the same modeling language in the same platform hence they can communicate easily with each others. MDX-agent is a special kind of agent, which supports MultiDimensional eXpressions (MDX) functions to query data from a multidimensional database. **MDX-agent** is a bridge between multi-agent analysis models and data marts. It is used to retrieve data from the data marts in the data warehouse system. Multi-agent analysis models can use SQL agents (same SQL-agents as in the simulation system) or MDX agents to access data appropriately. OLAP analysis tools are analysis softwares that support OLAP operators.

The key points of the CFBM architecture are that it contains and adapts the four features of a computer simulation system (model design, model execution, execution analysis and database management). All these functions are integrated into one multi-agent platform. The data warehouse manages the related data. The analysis models and simulation models can interact with each other. Using the CFBM architecture, we can build a simulation system not only suitable for modeling driven approach but also for data driven approach. Furthermore, CFBM brings special benefits for building simulation system with complex requirements such as integrating and analyzing on high volume of data.

## 3.3 Implementation of CFBM in GAMA

In the previous section, we have described two kinds of multi-agent based models (multi-agent simulation model and multi-agent analysis model), which can be built in a multi-agent platform and can retrieve or update many kinds of data via special agents (SQL agents and MDX agents). We have chosen to implement CFBM into the GAMA platform following the software architecture illustrated in Figure 3.

### 3.3.1 Software Architecture of CFBM in GAMA

**Presentation tier** plays the role of view layers in the CFBM architecture. In our implementation, the GAMA user interface plays this role. It can be used to write models or to analyze models

by using agent based approach, execute models and visualize results of models in different modes of views (text, chart, GIS or 3D).

**Logic tier** coordinates the application process commands, as it plays the role of two layers in the CFBM architecture (simulation layer and analysis layer). In our implementation, it contains four components. The Analysis-agent is designed to supply statistical analysis functions. The statistical functions may be built-in functions or called functions from an external application (e.g. R via RScript). The Analysis-agent is responsible of the analysis of data that are generated by models or data retrieved from a database. An SQL-agent is responsible of the retrieval as well as the update on relational database. A MDX-agent is binding on retrieval data from data marts. SQL agents and MDX agents were implemented as plug-in of GAMA. A modeler can easily use them via primitives of the Gama Modelling Language (GAML), the language used to write models in GAMA. The GAMA-database API is responsible for translating database requirements from GAML to SQL query or MDX query appropriately, transforming retrieved data type to GAMA data type and vice versa. The API has been developed as an application library of the CFBM. In addition, we can add several database functions in the same way we did with the SQL-agent and MDX-agent based on this API.

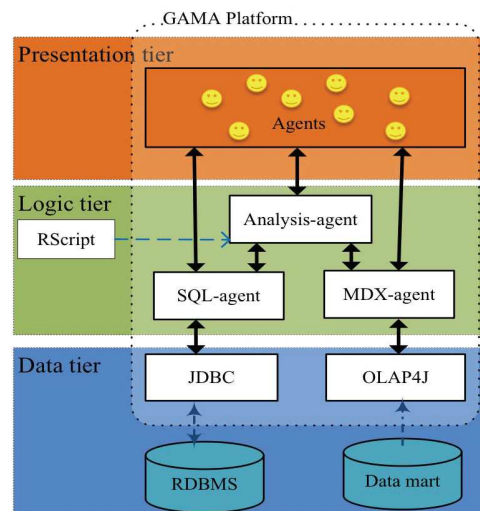


Figure 3. Software architecture of CFBM in GAMA.

**Data tier** plays two roles in the framework: 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. JDBC is a Java API allowing connecting to and querying data from relational databases. OLAP4J is also a Java API to connect to and query data from a multidimensional database. RDBMS plays different roles (Real data, Simulation data or Data warehouse). Data mart plays the same role as described in Section 3.1 of this section.

### 3.3.2 Database Access Features in GAMA

Thanks to added database features of GAMA, we can create agents and define the environment of the simulation by using data selected from database or store simulation results into relational databases. It helps us to gain flexibility in the management of simulation models and the analysis of simulation results. In this part, we only demonstrate some basic SQL related functions, which have been implemented in GAMA.

For using SQL functions, we must define at least an agent that is endorsed by the SQLSKILL skill<sup>4</sup> or inherits from SQL agent. For example:

```
entities {
  species toto skills: [SQLSKILL]
  {
    //insert your descriptions here
  }
}
```

Then all access database requirements of simulation model will be done via the defined agent using the SQLSKILL skill to query data. Depending on the activity (reading from database or writing into database), the defined agent will auto transform data from data type of database to data type of GAMA or on the contrary from GAMA type to database types.

For example, we can select data from a database and use selected data to create agent location as following steps:

- Step 1: Define parameters of connection and the selection statement:

```
map PARAMS <-
  [dbtype::'sqlite', 'database': './includes/bph.sqlite'];
string LOCATIONS <- 'SELECT ID_4, Name_4, geometry
  FROM bph
  WHERE id_2=38253 or id_2=38254;';
```

- Step 2: Use a defined agent with SQLSKILL skill (e.g. agent *toto*) to select data and create *location* agents:

```
ask toto {
  create location from: list(self select
    [params:: PARAMS, select:: LOCATIONS])
  with: [
    id:: 'id_4',
    name:: 'name_4',
    shape:: 'geometry'
  ];
}
```

We can also write simulation results into a database by using *insert* function of agent *toto*:

```
ask toto {
  do action: insert {
    arg params value: PARAMS;
    arg into value: 'SIM_RESULTS';
    arg columns value: [
      'model_id',
      'scenario_id',
      'replication_no',
      'step_no',
      'measure_id',
      'measure'
    ];
    arg values value: [
      model_id,
      scenario_id,
      replicate,
      time,
      measure_id,
      measure
    ];
  }
}
```

<sup>4</sup> Any kind of agent in GAMA can be endorsed by one or several skills. A skill provides new built-in attributes or built-in actions the agent can perform. A skill thus adds new capabilities to an agent.

```
];
}
}
```

The database access features in GAMA support most common query statements (CREATE, UPDATE, INSERT, DROP) and do not request a deep knowledge in SQL query language and database management from the user. The main benefit of our implementation in GAMA is the fact that the database access is directly integrated into the GAML modeling language, SQLSKILL can access with almost GIS database such as SQLite, PostgreSQL, MySQL and SQL Server and MDXSKILL can interact with multidimensional databases. More details can be found in "Database access user guide"<sup>5</sup> of GAMA version 1.5.1.

## 4. APPLYING THE CFBM TO AN ENVIRONMENTAL SURVEILLANCE NETWORK MODEL FOR INSECT ECOSYSTEM

### 4.1 Motivation

This section introduces a study case of CFBM for an environment surveillance network model applied in an insect ecosystem. An integrated model of the surveillance network and the BPH prediction one is presented in Section 4.2. Before applying CFBM, this integrated model worked as a stand-alone model and required many single operations for almost all works: data preparation, data analysis, data aggregation, etc. With CFBM, we have more support to couple these operations into a uniform framework and the application is really more flexible and portable. Some main conveniences according to which CFBM can be a good choice for our model, and also for other big models:

- Data preparation: An integrated model of an environment surveillance network on an insect ecosystem is built for an application at the regional scale. The input data of this model are divided into four main groups: Insect sampling data (daily sampling data of about 10 insect species at more than 300 light-traps in the Mekong Delta region of Vietnam), Administrative regions (province, district and small town scales), Land uses (seasonal crops of multiple tree species of the region), Natural phenomena (wind, temperature, humidity, etc). In the case of a development of the model without using CFBM, for an example with only one insect species in three provinces (as in our scenarios in this paper), we need to manually filter the data by some tools and the selected data can be stored in some individual files. These files will be used as the input of the model. With CFBM, these works are just replaced by some queries applied on database as presented in Section 3.3.2. This flexibility allows the modelers to easily change their temporal or spatial scale of their scenarios. In other words, CFBM supports to select and update the appropriate data at every micro scale of the model for every simulation step. Furthermore, the input and output data are also distributed and portable.
- Data analysis & processing: Beside built-in operators of GAMA, a model can use many libraries of operators from both the RDBMS (with the storage procedures) and the RScript service (with data mining operators). For example, in a same simulation process

<sup>5</sup> [https://code.google.com/p/gama-platform/wiki/Ext\\_DB\\_SQL151](https://code.google.com/p/gama-platform/wiki/Ext_DB_SQL151)

we can apply the Kriging estimation by adding the Gaussian noises with support of RScript, but apply the smooth processing in GAMA.

- **Data aggregation:**  
Almost all agent-based models are stochastic. Therefore, a research question is often answered by an aggregation of multiple simulation results. With CFBM, the status of each simulation can be archived in the database, and then easily aggregated after executing a batch of simulations.

## 4.2 Integrated Model of BPH Prediction Model and Surveillance Network Model

This section aims at introducing the main process for implementing a specific agent-based model with support of CFBM. This integrated model is composed of two sub-models and both are implemented on the GAMA platform (Figure 4): (1) BPH Prediction Model, a co-model of Growth and Migration models and (2) Surveillance Network Model (SNM). Input and output of the integrated model are handled via the CFBM in GAMA. **REALITY DATA** (Empirical data) such as administrative boundary (region, river, sea region, land used), light-trap coordinates, daily trap-densities, rice cultivated regions, general weather data (wind data), station weather data (temperature, humidity, etc.), river and sea regions are used as input of simulation model and they are also used as validation data for the model. The **SIMULATION DATA** essentially contains the trap-densities collected by the surveillance network model.

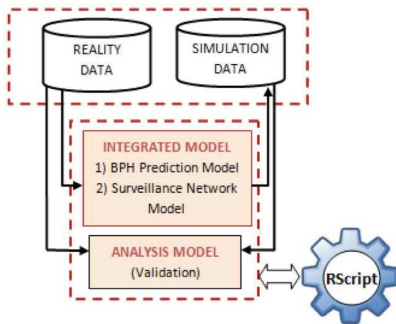


Figure 4. Architecture of the integrated model.

This integrated model reads the BPH trap-density data from real context and returns the simulated data for the future time. The input data also contain different related parameters, which directly affect the dynamics of BPH, such as the rice-cultivated regions, different kinds of weather and river/sea regions. The BPH trap-density is the main parameter of this prediction model. The surveillance network model reflects the plant protection activity in Vietnam, with a light-trap network to monitor the densities of multiple adult insects.

### 4.2.1 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.

In stochastic model, the studied region is modeled as a cellular automaton. Each cell in the stochastic model represents a square in the real context. A cell  $k$  in the model keeps its own local parameters:  $Z_k(s, t)$  and two additional parameters (attractiveness index and obstruction index). In fact, the prediction model composes the interactions between two sub-models:

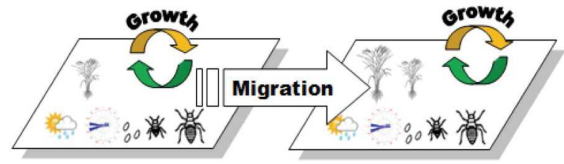


Figure 5. BPH prediction model.

#### (1) BPH migration model:

The migration process of BPHs in a study region is modeled by a dynamic moving process in a cellular automaton. In agent-based model, each cell becomes an agent which contains all important parameters of the environment, such as the trap-density vector and two combinational indices: obstruction and attractiveness indices.

Denoting that  $x(t)$  is the number of adult BPHs at the time point  $t$ . The migration model is essentially concerned with the outcome  $x_{out}(t)$  of the next time point  $(t + 1)$  from a specific source cell and the rate  $r$  of  $x_{out}(t)$  on the existing number of BPHs, which moves to all destinations at time point  $(t + 1)$ . The destination cells are determined by the semi-circle under the wind, the radius of circle is the multiplication of wind velocity and the migration time in a day.

Assuming that there are  $n$  destination cells found at the time point  $t$  for a source cell  $i$ . The ratio of BPHs from  $i$  to the destination  $j$  is decided by the attractiveness index (a combinational and standardized index calculated from different environmental factors, such as the rice cultivated regions, weather, etc). If the attractiveness index is high, then the outcome ratio of BPHs will be high, and to the contrary. The detailed explanation of attractiveness index is presented in[22].

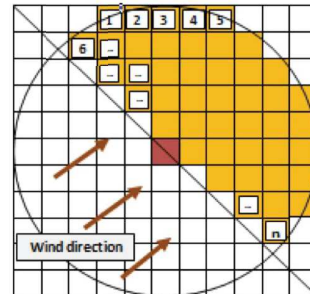


Figure 6. Brown plant hopper migration model.

#### (2) BPH growth model:

In the growth model, we apply a deterministic model of  $T$  variables where  $T$  is the life cycle of the insect. For example, BPH has a life cycle of 32 days in maximum, then we have  $T = 32$ . To simplify the implementation process, these variables will be stored in an array variable  $V$  of length  $T$  where an element  $V[t]$  marks the number of insects at age  $t$  (i.e.  $t^{\text{th}}$  day of BPH life cycle).

For each simulation step, all elements of  $V$  will be updated by the following equation:



$$V[t] = \begin{cases} \sum_{i \in T_1} V[i] * r_b * m, & t = 1 \\ V[t-1] * r_{en} * m, & t \in T_2 \\ V[t-1] * r_{na} * m, & t \in T_3 \\ V[t-1] * m, & \text{otherwise} \end{cases}$$

Where

- $V[i]$  denotes the number of insects at age  $i$ ,
- $r_{en}$  denotes the ratio of egg number able to become the nymph,
- $r_{na}$  denotes the ratio of nymph number able to become the adult,
- $r_b$  denotes the ratio of eggs can be produced by an adult,
- $m$  denotes the ratio of natural mortality,
- $T_1$  denotes the egg giving time span,
- $T_2$  denotes the egg and time span.
- $T_3$  denotes nymph time span.

#### 4.2.2 Surveillance Network Model (SNM)

This model is presented in several publications [21][22]. SNM is based on the Unit Disk Graph technique and autocorrelation between different surveillance devices. In [21], the surveillance network is modeled as a Correlation & Disk graph-based Surveillance Network (CDSN) where the edges of the graph are defined by two conditions: spatial distance and autocorrelation coefficient. We only treat the inputs and outputs of this network model in this paper. Each surveillance device contains the trap-density at a specific geographic location.

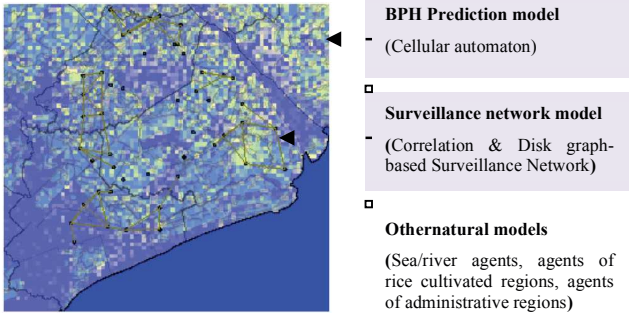


Figure 7. Integrated model.

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.

#### 4.2.3 Validation of Prediction Data

All related operations for validation in Analysis Model (Figure 4) are shortly introduced in this section. As mentioned, the prediction data (collected by the surveillance network model) and empirical data of all surveillance devices will be compared by several validation techniques to assess the precision of the prediction model. Some possible techniques used in this module include:

- Analysis of the correlation level between  $z_0$  and  $\hat{z}_0$ .
- Error analysis between  $z_0$  and  $\hat{z}_0$  (e.g. by RMSE).

## 4.3 Experiments

### 4.3.1 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 the north-east with a velocity of  $9.4 \pm 3.5$  km/h.

Table 1. List of agents of model

Parameters	Scenario 1 (Growth model)	Scenario 2 (Co-model)	Scenario 3 (Co-model)
$T_2$	7	7	7
$T_3$	13	13	13
$T_1$	6	6	12
$T$	32	32	32
$r_{en}$	0.42	0.42	0.40
$r_{na}$	0.42	0.42	0.40
$r_b$	250.0	250.0	350.0
$m$	0.245	0.245	0.35

### 4.3.2 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. Figure 8 shows the real trap-density data for January 1, 2010. We used the data of the first 32 days ( $T = 32$ ) to apply  $T$  Kriging estimations by adding the Gaussian noises. These estimations simultaneously assign the values for vector  $V[1..T]$  of each cell. The prediction data are taken from the  $(T+1)^{th}$  simulation step.

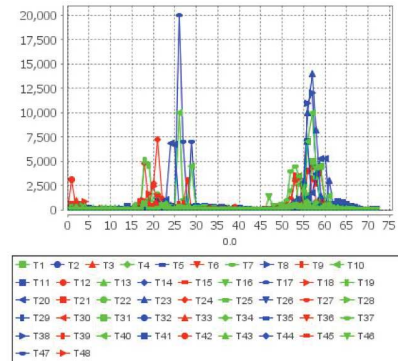


Figure 8. Empirical data from Jan 01, 2010 in three provinces: Soc Trang, Hau Giang and Bac Lieu.

In Figure 8,  $T_i$  denotes the  $i^{th}$  of these light-traps. These data are stored in the REALITY DATA database.

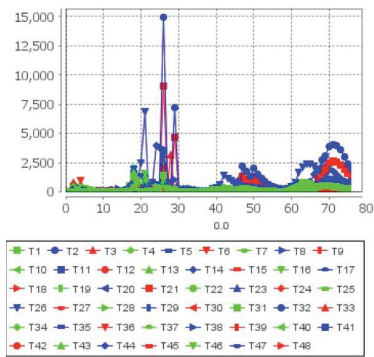
As aforementioned, the R language can be called from GAMA with support of CFBM. We need the R language to estimate the BPH density by using Kriging technique. R code files can be applied by using the `R_compute` (String RFile) or `R_compute_param` (String RFile, List

vectorParam) operators of GAMA. We develop an R code file to read the coordinates of all light-traps and their trap-densities at a specific time point from GAMA, then returns the estimation data to this platform. The estimation data also contain some Gaussian noises. The *krige* function of *gstat* library is used for this work [16]. This process is repeated  $T$  times to assign the values of vector  $V[1..T]$  of all cells in the cellular automaton.

▪ **Simulation data**

Figures 9 & 10 show the simulation data corresponding to the real data of Figure 8. The first 32 days are related to the estimation of trap-densities at the cells located by light-traps and the prediction data is considered from the 33th day.

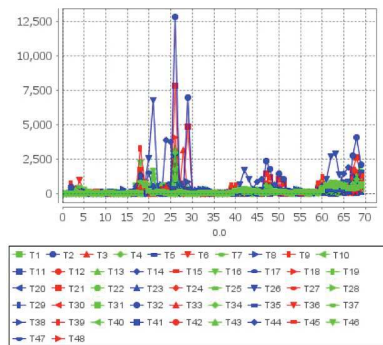
Figure 9 demonstrates the simulation data of scenario 1. The wave of prediction data in this figure is significantly smooth (like a sinusoidal form). The same sinusoidal wave form is found in [8]. However, the form of empirical data is too shocked, as it looks like a pick of insects in a short time.



**Figure 9. Simulation data from Jan 01, 2010 in three provinces (Growth model is applied).**

The BPH prediction co-model is implemented for scenarios 2 & 3. In Figure 10, the form of BPH trap-density is more similar to real data. This phenomenon is caused by migration process implemented in BPH migration model.

All simulation data are stored in the **SIMULATION DATA** database. They will be recalled in the validation process.



**Figure 10. Simulation data from Jan 01, 2010 in three provinces (Co-model is applied).**

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

## 5. CONCLUSION & FUTURE WORK

In this paper, we proposed a conceptual framework, which is adapted to multi-agent models with high volume of data. CFBM solves two problems listed in the introduction section. CFBM supports experts not only to model a phenomenon and to execute the models via a multi-agent platform, but also to manage a set of models, input and output of models, to aggregate and analyze output data of models via data warehouse and OLAP analysis tools.

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 important point of CFBM is the integration power of data warehouse, OLAP analysis tools and a multi-agent based platform that is useful to develop complex simulation systems with a large amount of input/output data. These systems can be a what-if simulation system, a prediction/forecast system or a decision support system.

In addition, CFBM has some advantages listed below:

- CFBM is an open and modular architecture. As for the implementation of the CFBM, we can use any BI solution and multi-agent platform depending on which technology is the most adapted. We can choose open source software or a commercial one. For instance, GAMA has succeeded on interaction with Pentaho Analysis Service (Mondrian) and SQL Server Analysis Service (SSAS).
- CFBM can be used to deploy in distributed environment. Assuming that there are many modelers and analyzers who work together on the same project but they are located in different locations, we can setup a simulation system and an analysis system in each location and all the data from each location can be integrated and shared via a centralized data warehouse.
- In practice, we recognize that CFBM allows handling a complex simulation system. In particular, we can build several simulation models to simulate the same phenomenon, conduct a lot of simulations for each of them and compare these simulation results (e.g. to determine which one is better for which parameters value domain). In this case, it is very difficult for modelers to manage, analyze or compare the output data of simulations if modelers do not have an appropriate tool. With help of SQL agents and relational data in the simulation system of CFBM, modelers can create a database to manage and store simulation models, scenarios and outputs of simulation models easily. In our case study, we have fully handled on input/output simulation data of models. Selecting and comparing output data between different replications or between simulation data and empirical data can be done in easier and shorter way.

Although CFBM has many advantages, it still has some drawbacks such as:

- It is very hard to implement CFBM. Because CFBM is a mix system of different applications, multi-agent platform, BI solution and analysis tools, the integration of all these software components and expertise is a complex work. However, a similar complexity is actually presented when working on integrated modeling.
- CFBM is not suitable for building a simple simulation system such as one or two models working with a small amount of data because it may take more time and workforce than other approaches.

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 CFBM in GAMA, it has succeeded in interacting with several database management systems (SQLite, PostgreSQL, MySQL and Microsoft SQL Server) and its extension is not too complex. In addition, GAMA with database access features has also special operations due to spatial databases.

In our case study, an integrated model of Surveillance network model and a BPH prediction co-model (BPH Growth & Migration Model) is 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.

As future works, on the one hand we will continue develop and improve features for special agents (SQL-agent, OLAP-agent and Analysis-agent) in GAMA described in Section 3. On the other hand, we will also apply CFBM on calibration and evaluation simulation model, multi-scale in multi-agent simulation or building what-if system, prediction system and decision support system. In addition, we will study to apply our framework to manage different simulation models, their input/output, validation and evaluation models of DREAM project<sup>6</sup>. The project aims at building integrated models, dealing thus with several models and huge quantities of data.

## 6. ACKNOWLEDGMENT

This work was partially supported by the French ANR project EmoTES.

## 7. 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] Dennis, B. et al. 2001. Estimating chaos and complex dynamics in an insect population. *Ecological Monographs*. 71, 2 (2001), 277–303.
- [5] Drogoul, A. et al. 2003. Multi-agent based simulation: Where are the agents? *Multi-agent-based simulation II*. Springer. 1–15.
- [6] 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.
- [7] Fishwick, P.A. 1997. Computer simulation: growth through extension. *Transactions of the Society for Computer Simulation*. 14, (1997), 13–24.
- [8] Hung, N.N. 2008. *Principles and Applications in Mathematical model for Biological*. Studies, Agriculture and Environment. Agriculture Publishing House, Vietnam.
- [9] Inmon, W.H. 2005. *Building the Data Warehouse*. Wiley Publishing Inc.
- [10] Kimball, R. and Ross, M. 2002. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. John Wiley & Sons, Inc.
- [11] Laniak, G.F. et al. 2013. Thematic issue on the future of integrated modeling science and technology. *Environmental Modelling & Software*. 39, (2013), 13–23.
- [12] 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.
- [13] Mahboubi, H. et al. 2010. A Multidimensional Model for Data Warehouses of Simulation Results. *International Journal of Agricultural and Environmental Information Systems (IJAIEIS)*. 1, (2010), 1–19.
- [14] Matheron, G. 1963. Principles of geostatistics. *Economic geology*. 58, 8 (1963), 1246–1266.
- [15] 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.
- [16] Pebesma, E. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*. 30, 7 (Aug. 2004), 683–691.
- [17] Phan, C.H. et al. 2010. An agent-based approach to the simulation of Brown Plant Hopper (BPH) invasions in the Mekong Delta. *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2010 IEEE RIVF International Conference* (2010), 1–6.
- [18] Rasmussen, C.E. 2006. *Gaussian processes for machine learning*. MIT Press.
- [19] Sosnowski, J. et al. 2007. Developing Data Warehouse for Simulation Experiments. *Lecture Notes in Computer Science* (2007), 543–552.
- [20] Taillandier, P. et al. 2012. GAMA: a simulation platform that integrates geographical information data, agent-

<sup>6</sup> <http://www.ctu.edu.vn/dream/>

- based modeling and multi-scale control. *Principles and Practice of Multi-Agent Systems* (2012), 242–258.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Wilensky, U. 1999. Netlogo, Technical report. *Netlogo, Technical report* (1999).