

# FROM SIMULATED MODEL BY BIO-PEPA TO NARRATIVE LANGUAGE THROUGH SBML

Dalila Hamami<sup>1</sup> and Baghdad Atmani<sup>2</sup>

Computer science laboratory of Oran (LIO)

<sup>1</sup>Department of Computer Science, Mostaganem University, Algeria

<sup>2</sup>Oran university, Algeria

## ABSTRACT

*Many theoretical works and tools on epidemiological field reflect the emphasis on decision-making tools by both public health and the scientific community, which continues to increase.*

*Indeed, in the epidemiological field, modeling tools are proving a very important way in helping to make decision. However, the variety, the large volume of data and the nature of epidemics lead us to seek solutions to alleviate the heavy burden imposed on both experts and developers.*

*Among the important steps of modeling and simulation: model validation. It refers to the process of determining how well a model corresponds to the system that it intended to represent. So the question is: what happens if the model is invalid? Do we need to reproduce another one, or just optimize the existing one?*

*In this paper, we present a new approach consisting on the passage of an epidemic model realized in Bio-PEPA to a narrative language using the basics of SBML language. Our goal is to allow on one hand, epidemiologists to verify and validate the model, and the other hand, developers to optimize the model in order to achieve a better model of decision making. We also present some preliminary results and some suggestions to improve the simulated model.*

## KEYWORDS

*Epidemiology, simulation, modelling, Bio-PEPA, narrative language, SBML, model validation.*

## 1. INTRODUCTION

In recent years, biotechnology improved the knowledge of epidemiological pathogens, and developed effective ways to fight against these epidemics. Currently, several outbreaks are in vogue, and developing factors helped build huge data banks [1]. Therefore, the amounts of raw data are too large to be analyzed manually by both experts and computer scientists who need to understand the epidemiological domain.

Due to the variety of biological data and the nature of epidemics [2], the validation method seems to be the perfect way, to ensure the perfect model or to optimize it. A set of recommendation are presented in the literature review [3, 4, 5], among which a verifying model should [6]:

- Make biological sense,

- Mimic real life,
- Be fit for the use they are designed for,
- Carried out a sensitivity analysis to assess the influence of uncertain parameters on the model outcome.

Unfortunately, by analyzing these recommendations, it appears that, the adopted methods were insufficient and therefore inefficient, so a new approach is used: the development of an interface between expert and computer scientist who is no longer required to start from the "whole" to achieve the "perfect" model. This interface converts the model created by the developer, to understandable language by the expert, so that he can check the validity of the model by: settings, rules, constraints, etc, and finally allow the developer to review and optimize the model, so that, any implementation of prevention and control are carried out for appropriate treatment.

The rest of the paper is organized as follows. Section 2 presents a brief review of epidemiological modeling and why do we need to use the narrative language?. Section 3, presents a few existing works on translating narrative language to simulated models and why do we need to translate the simulated model to the narrative languages?. A description of our model in SBML (Bio-PEPA), and how to perform its translation into narrative language, in Section 4. Section 5 describes the details of information on testing and evaluation and some discussion are presented in Section 6. Section 7 summarizes the work done and also some suggestions to improve the model.

## 2. MODEL OPTIMIZATION

Modeling and simulation plays a critical role in estimating the potential impact of outbreaks of highly contagious diseases, either in human population, such as tuberculosis, HIV, or in animal population, as highly pathogenic avian influenza. Although in nowadays epidemiological models are refined, but the quality of the results of modeling studies, still depends on the quality, the pertinence and accuracy of the data on which they are based, and on validity, in both, of the model themselves and its conceptual specification. For this, these models should be subjected to careful and ongoing evaluation and scrutiny.

Because in epidemiological studies, major decisions are made on the basis of the results of modeling studies, it is important to know that these studies are appropriate, accurate and correct. The result is that the key step in the evaluation of epidemiological models is a model validation, which refers to the process of determining how well a model corresponds to the system that it intended to represent [7]. As well defined by Schlesinger [8], model validation is the "substantiation that a ... model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model".

Reeves et al [7], reviews a set of works about validation methods, where, several authors detailed, taxonomies of the methods used to validate model [9, 3, 4].

Taylor et al [4] specified that there no easy test of a model's validity. He described perfectly validation by four notions (methods):

- Valid models should make biological sense: The model should be examined to ensure that all the epidemiological knowledge that influence outbreaks of the disease have been included. In this phase, it is quite interesting to detail document describing the conceptual model, by a series of discussion and evaluation of the details of the model's operation. Hervey et al [10], gave a large importance to this phase, in his study. He reviewed all workshops and meeting done

about NAADSM (the North American Animal Disease Spread Model), in particular Foot-and-Mouth Disease which was carried out by experts in the field of FMD and modeling. Among the suggested improvements were inclusion of various livestock species, production systems, and a variety of mitigation strategies, as well as extension of the utility of the model to diseases other than FMD. Also, as mentioned by Taylor [4]:” Ferguson and others [11] found that estimating the spatial transmission kernel by retrospectively fitting a model to the epidemic data produced a wider kernel than that derived from the tracing data provided”.

- Valid models should mimic real life: the model gives the same output as the real system over a range of variables. The problem is that, all modelling groups affirm that their models could reproduce the course of an epidemic with reasonable accuracy. However, the level of proof of validity this provides is challenged by the fact that some of the models were parameterized using statistical methods in the aim to provide a matching to the real data, as done by Keeling et al [12]. Low-Beer et al [13], studied VIH disease, throughout the study, it was widely reported that the model produced similar outcomes, therefore supporting its validity. However, not all aspects of model produced conclusions about control policy that were alike in every detail. They stated that their model did not model the dynamic aspects of sexual mixing and transmission, and should be used for short-term forecasts of 3-5 years, assumptions which belongs to the reality and not still realized. Skvortsov et al [6], developing a complex agent-based model to simulate an epidemic outbreak and to validate it they using a simple mathematical model by comparing the results with a SIR model, the later is derived assuming uniform mixing among all members of the population, whereas a real population has people with a wide range of contact rates. Taylor et al [4], concluded that, the importance of studying carefully the field data from different areas, in order to better understand the relationships between control policies and disease dynamics cannot be overstressed. Indeed, models can never be substitutes for careful analysis of field data.

- Valid models should be fit for the use they are designed for: the important thing is not the model's validity/invalidity, but model's usefulness. As presented by Sanson et al [5], they achieved the model validation by comparing three scenarios over different control strategies or resource constraints. They proved that each outcome helped in making decision, even if it did not match reality.

- Sensitivity analysis should be carried out to assess the influence of uncertain parameters on the model outcome: used particularly when the values assigned to model parameters are uncertain due to lack of good quality data. Sensitivity analysis helped to detect parameters for which model is more sensitive than the case in real life and reveal that the model is invalid. As specified by Hughes et al [14]: “in the absence of data from the household survey of tuberculosis disease, the model was validated by comparing its output with TB incidence data for Zimbabwe and characteristics typical of epidemics of any infectious disease”. They noted that in future work they have to use a model population that is sufficiently large for TB disease to become endemic and include the effect of interactions of the model population with a background pool of TB infection. This leaves suggests that their used parameters was uncertain and they have to test others. Waller et al [15] proposed the use of Monte Carlo hypothesis tests, which compare a single set of outcome data from a real system to multiple model-generated outcome data sets, the drawback of this type of validation was reported by Anderson’ works [16, 17], he studied the outbreaks of FMD, where that of 2001, resulted in the infection of over 2,000 herds and that of 2007, resulted in the infection of only eight infected herds, then which data is better for validation step and consistent for study.

Summing up, this review helped us to conclude that to optimizing a model, we need verifying it, which means validating it, and to achieve this, we have to take in consideration two points; the inputs and the outputs of model. The input data are used to extract relevant parameters that will influence model outcomes. The output data (results of a system) are used to provide a basis for comparison with model's produced outcomes.

In particular cases, we have access to information pertaining to only a single outbreak of disease in a particular set of circumstances, which is inconsistent for the study. In still other instances, models are developed to explore hypothetical scenarios, where some information is generally established to inform model inputs, but there can be no data on the system outputs. Whatever the form or source of data used to inform models, their correctness and validity should be considered. These observations lead us to come back to the paramount factor of study who is expert (epidemiologist).

So, in the presence or absence of input and output data, the expert alone can devise a scheme to monitor the epidemic and the validation results. Only problem, that he is considered as an expert in epidemiology and not expert in modeling and simulation, and it is no way that he could understand the developer language, as simple as it seems.

To overcome this drawback, the revolution of using narrative language appears. Where some works offer interface to helping expert interacts with developer. The following sections review the principal existing work on this field.

### **3. FROM NARRATIVE LANGUAGE TO A MODEL**

Develop and use a good epidemiological model, remains to this day a very attractive idea, and to achieve it, many researchers are struggling between having to choose the best tools and methods, or to conduct a thorough training in the field in question, and often they find themselves stagger between them. However, others, give little importance to neither one nor the other, rather they prefer to save their energy and adopt a technique completely original that is to transform the context expressed by an expert directly in a simulated model, as it was presented by Georgoulas and Guerriero in 2012 [18], for translating the narrative language in a "Bio-PEPA" formal model. In 2007 and 2009, Guerriero et al [19, 20] studied the translation of narrative language in a "model Beta-binders" and "a bio-inspired process calculus". They proposed an approach to biochemical model specification which was based on a narrative language, where expert as biologist or epidemiologist specify their model, by providing a textual description of the system and its evolution. In their description, the experts should specify the set of list composing the model as compartments, entities, reactions describing the dynamic of system with their rate parameters. The authors have assumed that it would be better to simplify the communication between experts and developers by providing a simple interface that would allow both the expert to insert their information and the developer to manipulate only the code without worrying too much about understanding everything. This approach has been baptized the passage from narrative language to a model. Although this work is regarded as a large opening in the field of modeling, however, the question arises, what is a narrative language? What happens to existing models?.

- **Narrative language**

The narrative language is a formal language, allowing to the experts in general and biologists in particular expressing the system and its dynamics, using terms that are well known and common

in their natural language. As defined by Guerriero et al [20], a model in narrative language is composed of four sections (by respecting the copyrights of authors, we prefer to retain the important sentences definition in its entirety): “

- The description of the biological compartments in which the involved entities can be located during the evolution of the system;
- The description of the entities composing the system;
- The description of the occurring reactions;
- The narrative description of the evolution of the system, i.e. the list of the occurring events.

**A compartment** is identified by an integer number; moreover, its name, size, and number of spatial dimensions can be specified. Compartments could represent cellular or sub-cellular compartments, but also abstract locations.

**A component** is identified by its name, and it can be seen as a list of interaction sites. Each site is defined by a name and a state. Also, the initial quantity/concentration of the component should be set.

**A reaction** is identified by an integer number; its type and the reaction rate parameter should also be specified. A reliability value can be associated to each numerical value (e.g. rate parameters and initial quantities); it is a percentage value that can be used to distinguish between values that are certain because obtained from wet experiments, and others which are the result of not verified assumptions. Modellers can take this information into account during the important step of model refining by means of parameter space search and sensitivity analysis.

Finally, **the evolution of the system** is described by means of a narrative of events. This narrative is a sequence of basic events, each of which is a textual description of a reaction involving at most two components. Events can be grouped into processes.

**An event** is identified by an integer number; in addition to the textual semi-formal description of the event, the identifier of the reaction associated with the event should be specified. The description of the event is a string of the form **if condition then event\_descr**, with the conditional part being optional.

A condition is a string of the form *component is state, component\_site is state, component is in compartment*, etc. Multiple conditions can be specified by separating them with the keyword **and**. *Event\_descr* is a string of the form *component reaction* for monomolecular events, or *component reaction component* for bimolecular events. Table 1, shows an example of events expressed in narrative language.

Table 1: List of events in narrative language (an extract taken from [20])

| <b>Id</b>               | <b>Description</b>   | <b>react</b> |
|-------------------------|--|--------------|
| <b>LIF-LIFR binding</b> |  |              |
| 1                       | <b>If</b> LIFR.LIF is not bound <b>and</b> LIF is not bound <b>then</b> binds LIFR on LIF      | 1            |
| 2                       | <b>If</b> LIFR.LIF is bound <b>and</b> LIF is bound <b>then</b> LIF unbinds LIFR on LIF        | 2            |
| 3                       | <b>If</b> gp130.LIF is not bound <b>and</b> LIF is not bound <b>then</b> LIF binds gp30 on LIF | 3            |
| 4                       | <b>If</b> gp130.LIF is bound <b>and</b> LIF is bound <b>then</b> LIF unbinds gp130 on LIF      | 4            |

This section helped us to understand the narrative language's importance to modelling, why in the following work, which were largely inspired from Guerriero's work [20], we suggest, to do the reverse of him, kept the existing model and improve it, which means, translating the Bio-PEPA model to a narrative language.

#### **4. FROM MODEL TO A NARRATIVE LANGUAGE**

In order to respond to the issues raised in the previous section and based on the principle defined above, we propose an approach whose aim is to preserve existing models and also to optimize it by allowing an incremental model implementation.

An extensive literature search, which focused on methods of modeling with both, analytical and decision support tools, as well as the translation of the model in other specific formats, for approaching the narrative language, we were able to highlight Bio-PEPA [2, 21], which is formal language based on the process algebras recommended for biochemical systems and which was perfectly suited to epidemiological field. Beyond this definition, Bio-PEPA is equipped with an extension that allows translating any model in Bio-PEPA in XML format better known as SBML (Systems Biology Markup Language).

##### **4.1. Bio-PEPA (Biological Performance Evaluation of Process Algebras)**

Bio-PEPA is a tool, method and language based on process algebra. These are described by mathematical formalisms used in the analysis of concurrent systems [2, 22, 23], which consist of a set of processes running in parallel, can be independent or share common tasks. As it was defined in [21], the Bio-PEPA language is 7-tuple (V, N, K, FR, Comp, P, Event) Where:

- V is a set of locations,
- N is a set of auxiliary information,
- K is a set of parameters,
- E is a set of functional rates
- Comp is the set of species
- P is the component model.
- Event is the set of events.

#### 4.1.1. Characteristics of Bio-PEPA

The main features which are provided in Bio-PEPA are:

- provides a formal abstraction of biochemical systems and further epidemiological systems.
- Allows expressing any kind of interaction law expressed using functional rates.
- Allows expressing the evolution of species and their interaction.
- Defined syntax and structural semantics based on a formal representation.
- Provides the ability to perform different types of analysis from the model (continuous time Markov chain, the stochastic simulation algorithms, differential equations).

#### 4.1.2. Bio-PEPA Syntax

As defined by [24, 21], Bio-PEPA syntax is described by:

$S ::= ( \lambda, k ) \text{ op } S ; S ; S := S + S ; S := C$  where

$\text{op} = \ominus \oplus \odot$  And

$S ::= S \ S \ S(x)$

Where, S: describe the species (different types of individuals); P: the model describing the system and interaction between species. The term  $( \lambda, k ) \text{ op } S$ , express that the action is described by k rate and performed by the species S, “op” define the role of S.  $\text{Op} = \{ \lambda : \text{reactant}, \lambda : \text{product}, \oplus : \text{activator}, \ominus : \text{inhibitor}, \odot : \text{generic modifier} \}$ .

#### 4.2. Systems Biology Markup Language (SBML)

SBML (The Systems Biology Markup Language) is a markup language based on XML (the eXtensible Markup Language). In essence, an XML document is divided into hierarchically structured elements from a root element. Syntactically, the elements of an XML document are marked in the document itself by opening and closing pairs of tags, each element consists of a name that specifies its type, attributes, and content (elements or text).

SBML is a set of constructions' elements specific to the systems biology, defined in an XML schema. It has been adapted to the epidemiological models.

The SBML language is divided into hierarchically structured elements which are a syntax tree of language as an XML schema.

As it was defined in section 4.1, an epidemiological model is defined in Bio-PEPA by a set of compartments, species and reactions described by rates and parameters. SBML do the same by using tags and attributes [25, 26]. Figure 1 shows the general organization of SBML's TAG which are described in the following [27]:

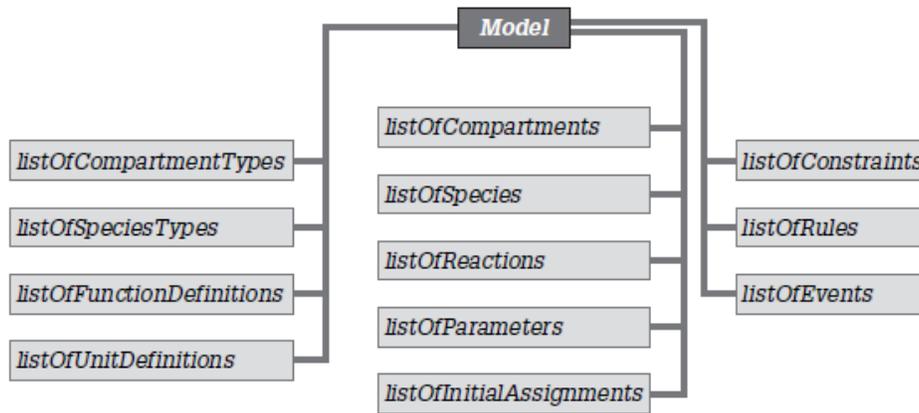


Figure 1. General organization of SBML language.

- **Model** : An SBML model definition consists of lists of SBML components located inside the tags : `<model id="My_Model" > ...</model>`.
- **listOfFunctionDefinitions**: The mathematical functions that can be used in the other part of the model are defined in this section.
- **listOfUnitDefinitions** : these units are used to explicitly specify: constants, initial conditions, the symbols in the formulas and the results of formulas.
- **listOfCompartments** : Is an enclosed space in which the species (species) are located.
- **listOfSpecies** : To specify the different entities in the model regardless of their nature, where one type of species "listOfSpeciesTypes" can be specified.
- **listOfReactions** : Any process whereby the transfer of a species from one compartment to another.

We specify that Representation and semantics of mathematical expressions are defined in the SBML using MathML.

### 4.3. Relation of Bio-PEPA to SBML

The principal notions which rely Bio-PEPA to SBML are summarized in table 2, (this table was directly extracted from [2]).

Table 2: Summary of mapping from SBML to Bio-PEPA (taken from [2])

| <b>SBML Element</b>  | <b>Corresponding Bio-PEPA component</b>   |
|----------------------|---|
| List of Compartments | Bio-PEPA compartments   |
| List of Species      | Species definitions (Name, initial concentration and compartment). Step size and level default to 1. Also used in species sequential component definitions. |
| List of Parameters   | Bio-PEPA parameter list. Local parameters renamed to include re- action name.   |
| List of Reactions    | Species component definitions and model component definition.   |
| Kinetic-Laws         | Bio-PEPA Functional rates   |

Table 2, describes the mapping from SBML to Bio-PEPA, where each element in SBML, matches perfectly to element in Bio-PEPA.

As seen in figure 1, SBML schema contains:

- A ListOfCompartments section. Each compartment in this list is directly matched to a compartment in Bio-PEPA, which defines compartment by compartment name matched to an identifier 'ID' attribute in SBML. The compartments in Bio-PEPA, is also defined by its size. This is also mapped to the attribute "compartment size" in SBML.
- A ListOfSpecies section. Each species in this list is directly mapped to a species in Bio-PEPA, which defines species by name, the initial concentration, the enclosing compartment name and the 'unit' for the species concentration. All these information are matched respectively to SBML species section which include attributes for ID, initialConcentration, compartment and unit.
- A ListOfParameters. Each parameter in this list is directly matched to a parameter in Bio-PEPA, which defines parameters by name and value, those are mapped to the "name" and "value" attribute in SBML.
- A Kinetic-Laws included in ListOfReactions. Each Kinetic-Laws in this list is directly matched to a KineticLaw (Functional rate) in Bio-PEPA, which expresses it by a name of functional rate matched to an integer number in SBML and an expanded formula in terms of species and parameters matched to a list of species and parameters used in MathML link to express what kind of relation is between each element of formula.
- A ListOfReactions. Each reaction in this list is directly matched to reaction in Bio-PEPA, which defines a dynamic of species using a set of operation as: reactant ( ), product ( ), and generic modifier ( $\odot$ ), and functional rates defined above. These information are matched in SBML to a ListOfProducts, a ListOfReactants and ListOfModifiers (respectively).

## 5. IMPLEMENTATION

For implementing our approach, we resumed work that we had already started in [21] which was to reproduce the spread and vaccination protocol of chickenpox in Bio-PEPA, as shown in Figure 2.

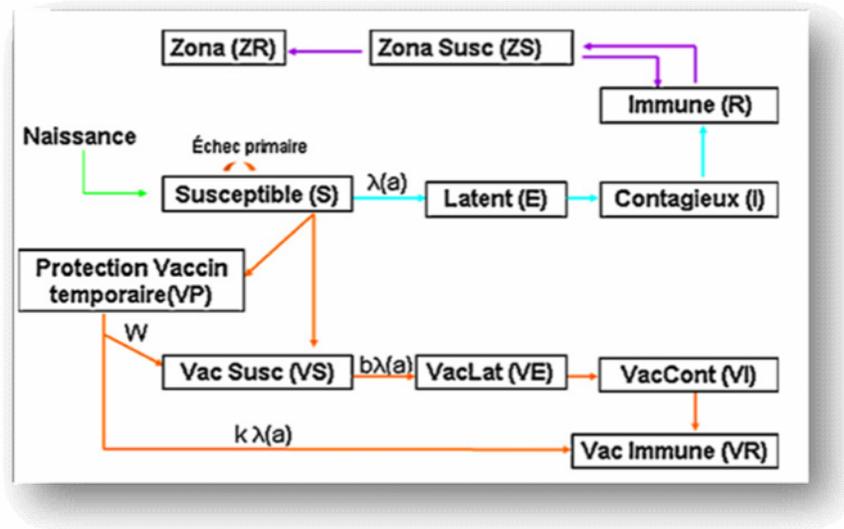


Figure 2. Model structure (taken from [28]).

The overall scheme of our approach is defined by three main steps:

- Formulation of the epidemic model in Bio-PEPA: definition of compartment, species and reactions.
- Exporting SBML file.
- Representation in narrative language: analysis of SBML file, displaying a detailed report, validation by the expert.

### 5.1. Description of model structure

Our approach, as it was structured, allows us to share our work in two main stages, the first is to develop a model with Bio-PEPA (Formulation from epidemic model in Bio-PEPA), a work that has already been done [21] and demonstrated the importance of using such a tool.

The second part (Exporting SBML file from Bio-PEPA, Representation of SBML text in narrative language) is developing a module that would translate the Bio-PEPA code in a language understood by the expert, who may easily check whether the contents of the model is adequate to the example and thus validate it.

#### 5.1.1. Chickenpox model in Bio-PEPA

To better understand the modeling process, we have taken and explained in this section the most important part of the code Bio-PEPA model of chickenpox [28]. (For clarity of the document, we have listed a few parts of model).

**1. Location:** To express the seven age groups of the model, we have represented it as compartments.

```
location Age1 in world : size = sizeAge, type = compartment;
.....
location Age7 in world : size = sizeAge, type = compartment;
```

**2. Functional rate:** Describes the interaction law between compartments.

**Exposition** =  $\lambda \cdot S \cdot I$ ; describes the contact between susceptible (S) and infected (I) with  $\lambda$  rate.

.....  
**LostVaccin** =  $W \cdot VP$ ; defines the rate of immune lost (W) of those protected by vaccination (VP).

**3. The species :** are the system entities expressed by operations describing their evolution.

**S** = [(**Exposition,1**) **S** + (**Vaccination\_1,1**) **S** + (**Vaccination\_2,1**) **S**]; explains what happens if executes Exposition function, Vaccination\_1 or Vaccination\_2.

Some lines of Bio-PEPA code are shown in Figure 3. We could note that even if the Bio-PEPA language is simple and easy for developer, however, remains an ambiguous part face which is set epidemiologist who cannot verify the validity of information represented by the developer, as the epidemiologist cannot understand the Bio-PEPA code.

We can extract from this figure, two important points, firstly the representation of chickenpox model in Bio-PEPA (right of the figure), and another hand the results of a simulation graph summarizing the status of various species (left of the figure).

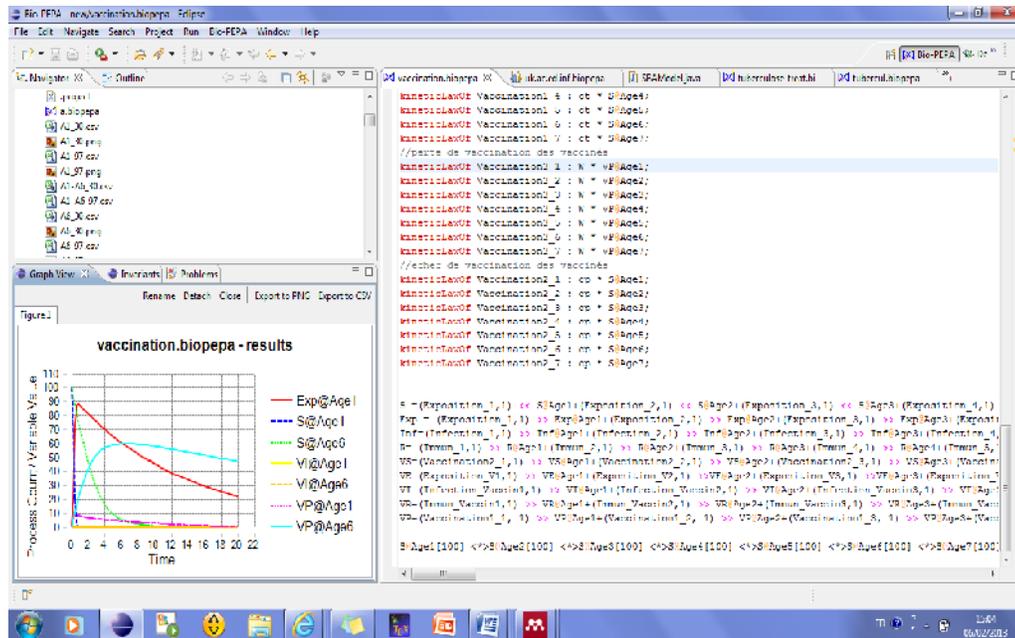


Figure 3. Global view of Bio-PEPA model.

### 5.1.2. Exporting SBML file from Bio-PEPA

Bio-PEPA provides the ability to export the model as an SBML file. As shown in Figure 4: The resulting text describes all the tags and attributes as they were presented in Section 4.2, corresponding to our model of chickenpox. It should be remembered that to study an epidemic, we must take into consideration: the environment "space", "time", and various other functions.

SBML can express perfectly each section describing the elements defined in Bio-PEPA.

### 5.1.3. The Chickenpox model in narrative language

To work with SBML, we need to perform a literature research, about tools analyzing and interpreting this type of descriptor, the latter revealed the JDOM tool (Java Document Object Model) [29], is an open source library for manipulating XML files in Java..

The main features of DOM are:

- The DOM model (unlike on this point to another famous API: SAX) is a specification that has its origins in the w3C consortium.
- The DOM model is not only a multi-platform specification, but also multi-languages: as Java, JavaScript, etc.
- DOM presents documents as a hierarchy of objects, from which, more specialized interfaces are themselves implemented: Document, Element, Attribute, Text, etc. With this model, we can treat all DOM components either by their generic type, "Node", or by their specific type "element, attribute", many methods of navigation allow navigation in the tree without having to worry about the specific type of component Treated.

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml version="3" level="2"
xmlns="http://www.sbml.org/sbml/level2/version3">
  <model id="vaccination biopepa">
    <listOfCompartmentTypes> <compartmentType
id="Compartment"/>
    <compartmentTypeid="Membrane"/>
  </listOfCompartmentTypes>
  <listOfCompartments>
    <compartmentid="Age7" outside="world" size="100000.0"
compartmentType="Compartment"/>
    .....
    <compartment id="Age5" outside="world" size="100000.0"
compartmentType="Compartment"/>
  </listOfCompartments>
  <listOfSpecies> <species id="Exp_Age1"
hasOnlySubstanceUnits="true" substanceUnits="item"
```

Figure 4. Chickenpox model in SBML.

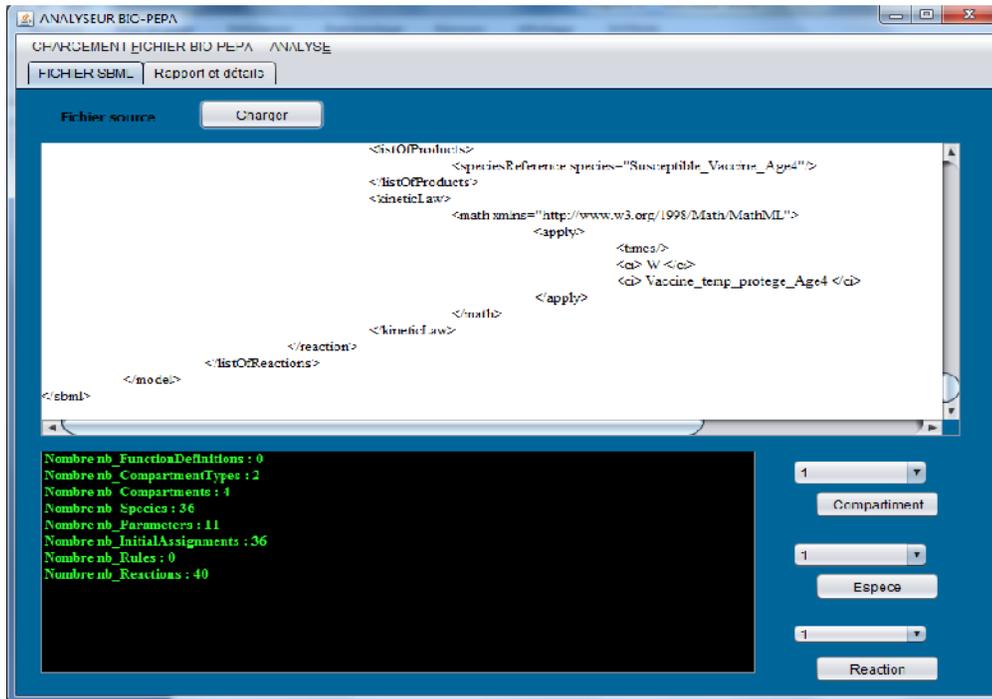


Figure 5. (a) From SBML code to narrative language.

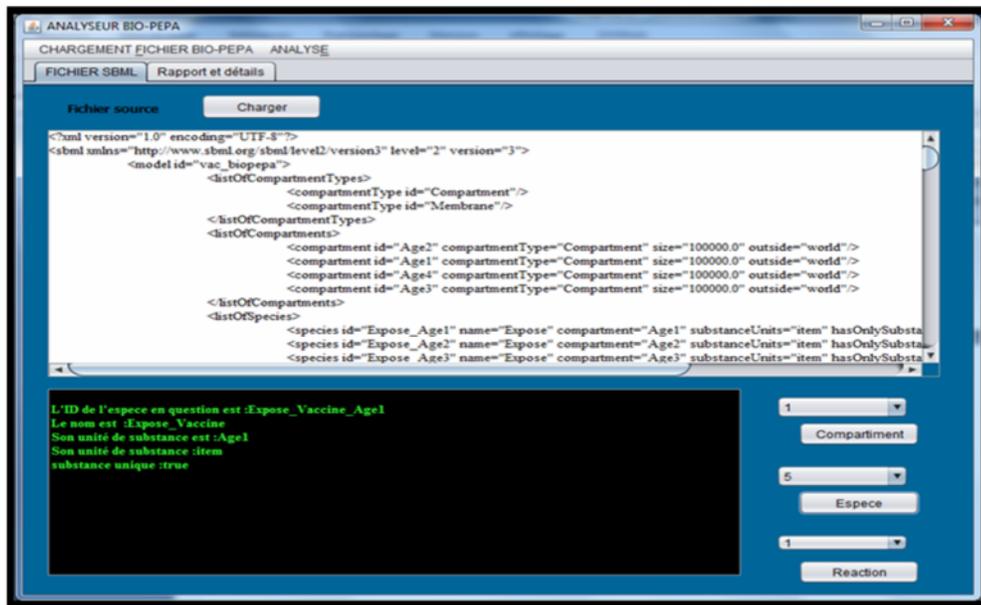


Figure 5. (b) From SBML code to narrative language.

Figures 5 (a, b) shows the interface of our application based on JDOM model and thus gathering the steps defined above. White space viewed in the figure corresponds to the loading of SBML

file, when the black area corresponds to the translation and analysis of SBML in narrative language understandable by the expert, the expert in this way has no difficulty in verifying the validity of the model. The user-friendly interface allows him to navigate the various components of the model (species, function of interaction, locations ... etc.)

To validate our application, we made a change in the initial code (Bio-PEPA) where we intentionally caused an error in our model, the generation of the latter, as shown clearly in Figure 6 that the expert could detect the species and reactions which are missing, and therefore it can easily report them to us. (The red frame line specifies the error caused by the number of species missed).

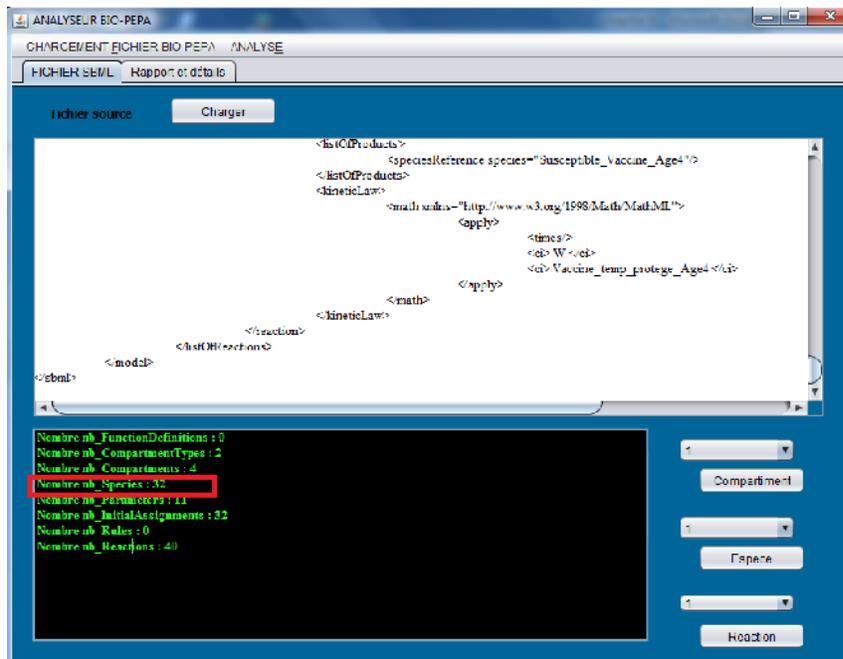


Figure 6. Detection error after translation the model into narrative language.

## 6. DISCUSSION

This paper surveyed an importance of validation in optimizing simulated model, in particular epidemiological model. It emphasized statistical a several works using techniques that yield reproducible, objective, quantitative data about the quality of simulation models. However, our analyze suggests that all these techniques remain insufficient, and that epidemiologist is the expert of the final validation. Hence the proposal to translate the model into a narrative language.

Our translation module has been implemented as an extension of the already developed Bio-PEPA plugin for the Eclipse IDE [30]. Specifically, we have added a new menu item that prompts the user to select the output type, which could performs a simulation of Bio-PEPA model, or translation of Bio-PEPA model to narrative language. However, we advise the user, to start with simulation step, to help expert to validate in first time that the model is something adequate to the reality, and if he needs to verify it thoroughly, he may perform the second step: translation.

To validate our model we choose to focus on the characteristics of the output. Specifically, we wanted to see how easy the output model is to read and how easy the expert to validate a model, and how close it is to natural language. To prove the importance of our work, we review the principal remark done by those who translated the narrative language to Bio-PEPA model. In [18, 19, 20] works, authors also focused on the same characteristics of us, they specified that obviously, as they was presented by Bio-PEPA model, these metrics are very subjective and difficult (if not impossible) to measure. However, this task seems for us, to be very easy. Therefore, the expert has just to read the output, and compared it, with its own knowledge. the alone thing is that, if there is any mistakes noted by expert in the resulting narrative model, the expert has to come back to developer, because he is the only one able to correct it.

To overcome this drawback, the idea to couple our model with [18, 19, 20] model is very encouraging, in the aim to minimize all interaction between expert/developer.

## 7. CONCLUSIONS

Modeling and simulation are very useful to understand and predict the dynamics of various biological phenomena. The Bio-PEPA approach seems to be an interesting and powerful approach to address such problems. Through its various features it allows an easy development of the computer model and a transparent way for biologists between the real system and the built which helps a faithful representation of the phenomenon studied. Nevertheless, in case of occurrence of a new event, which has been badly treated by the developer and therefore omitted, correction model is considered a tedious task for both. This is the reason for which, we have introduced a new module (interface), where the expert can easily detect this omission and thus back to the developer, who may discern error and quickly position it, on the Bio-PEPA model. By doing so, we concluded that to validate/optimize simulated model is better to come back to the expert's knowledge than to try a series of validation methods without knowing where exactly, we have to optimize.

As perspective to strengthen this work, why not attached it to the one that was mentioned in Section 2, and thus drift toward a cyclical pattern, which would not require even the presence of developer, however, after reflection, what will become the expert, front of its multitude of information? After a brief literature review the idea of integrating it with the world of data mining would be a much better idea to fruition.

## REFERENCES

- [1] Mansoul, A., Atmani, B. (2009). Fouille de données biologiques: vers une représentation booléenne des règles d'association, In Proceedings of CHA.
- [2] Ciocchetta, F., Ellavarason, K. (2008). An Automatic Mapping from the Systems Biology Markup Language to the Bio-PEPA Process Algebra.
- [3] Sargent, R., (2009). Verification and validation of simulation models. In Rossetti, M.D., Hill, R.R., Johansson, B., Dunkin, A., Ingalls, R.G. (Eds.), Proceedings of the 2009 Winter Simulation Conference, December 13-16 2009, Austin, Texas, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, USA.
- [4] Taylor, N., (2003). Review of the use of models in informing disease control policy development and adjustment. A Report for the Department for Environmental, Food, and Rural Affairs, UK. Web page <http://www.defra.gov.uk/science/documents/publications/2003/UseofModelsInDisease-ControlPolicy.pdf>.
- [5] Sanson, RL., Harvey, N., Garner, MG., Stevenson, MA., Davies, TM., Hazelton, ML., O'Connor, J., Dube, C., Forde-Folle, KN. and Owen, K ,(2011). Foot and mouth disease model verification

- and'relative validation'through a formal model comparison.Revue scientifique et technique, International Office of Epizootics, 30.2,pp.527.
- [6] Skvortsov, ATRB., Connell, RB., Dawson, P., Gailis, R, (2007). Epidemic modelling: Validation of agent-based simulation by using simple mathematical models, MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 657-662
- [7] Reeves, A., (2012). Construction and evaluation of epidemiologic simulation models for the within-and among-unit spread and control of infectious diseases of livestock and poultry. Dissertation from Colorado State University.
- [8] Schlesinger, S., (1979). Terminology for model credibility. *Simulation* 32, 103–104.
- [9] Law, A.M., Kelton, W.D., (2000). Simulation modeling and analysis. Analysis, 3rd ed. McGraw-Hill, Boston, Massachusetts, USA.
- [10] Harvey, N., Reeves, A., Schoenbaum, M.A., Zagmutt-Vergara, F.J., Dubé, C., Hill, A.E., Corso, B.A., McNab, W.B., Cartwright, C.I., Salman, M.D., (2007). The North American Animal Disease Spread Model: A simulation model to assist decision making in evaluating animal disease incursions. *Prev. Vet. Med.* 82, 176–197.
- [11] Ferguson, N.M., Donnelly, C.A., Anderson, R.M., (2001). The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* 292, 1155–1160.
- [12] Keeling, M.J., (2005). Models of foot-and-mouth disease. *P. Roy. Soc. B* 272, 1195–1202.
- [13] Low-Beer, D., and Stoneburner, R.L., (1997). An age-and sex-structured HIV epidemiological model: features and applications. *Bulletin of the World Health Organization*, 75-3, pp213.
- [14] Hughes, Georgina R and Currie, Christine SM and Corbett, Elizabeth L, (2006), Modeling tuberculosis in areas of high HIV prevalence. Simulation Conference, 2006. WSC 06. Proceedings of the Winter, pp. 459-465, IEEE.
- [15] Waller, L.A., Smith, D., Childs, J.E., Real, L.A., (2003). Monte Carlo assessments of goodness-of-fit for ecological simulation models. *Ecol. Model.* 164, 49–63.
- [16] Anderson, I., 2002. Foot and mouth disease (2001). lessons to be learned inquiry report. Web page [http://webarchive.nationalarchives.gov.uk/20100807034701/archive.cabinetoffice.gov.uk/fmd/fmd\\_report/documents/index.htm](http://webarchive.nationalarchives.gov.uk/20100807034701/archive.cabinetoffice.gov.uk/fmd/fmd_report/documents/index.htm).
- [17] Anderson, I., (2008). Foot and mouth disease 2007: a review and lessons learned. Web page <http://webarchive.nationalarchives.gov.uk/20100807034701/archive.cabinetoffice.gov.uk/fmdreview/>. Last accessed June 26, 2011.
- [18] Georgoulas, A., & Guerriero, M. L. (2012). A software interface between the Narrative Language and Bio-PEPA, 1–9.
- [19] Guerriero, M. L., A. Dudka, N. Underhill-Day, J. K. Heath and C. Priami (2009). Narrative-based computational modelling of the Gp130/JAK/STAT signalling pathway, *BMC Systems Biology* 3, p. 40.
- [20] Guerriero, M. L., J. K. Heath and C. Priami, (2007). An Automated Translation from a Narrative Language for Biological Modelling into Process Algebra, in: *Proceedings of Computational Methods in Systems Biology (CMSB'07)*, LNCS 4695, pp. 136–151. URL <http://www.springerlink.com/content/vt23126776012835/>
- [21] Hamami,D, Atmani.B. (2012). Modeling the effect of vaccination on varicella using Bio-PEPA. *Proceeding Iasted* , 783-077, doi:978-0-88986-926-4
- [22] Milner.R. (1999). *Communicating and Mobile Systems: the  $\pi$ -calculus*. Cambridge University Press.
- [23] Baeten J.C.M.(2005). A Brief History of Process Algebras. *Theoretical Computer Science*, Volume 335, Issue 2-3, Pages 131-146.
- [24] Ciocchetta, F. and M. Guerriero (2009). Modelling Biological Compartments in Bio-PEPA, *ENTCS* 227, pp. 77–95.
- [25] Hucka.M, Finney.A, S. Hoops, S. Keating and N. L. Novere (2007). *Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions*. Systems Biology Markup Language, Release 2.
- [26] Hucka.M, Finney.A, Bornstein.B.J, Keating.S.M, B.E. Shapiro, J. Matthews, B.L. Kovitz, M.J. Schilstra, A. Funahashi, J.C. Doyle and H. Kitano (2004). Evolving a Lingua Franca and Associated Software Infrastructure for Computational Systems Biology: The Systems Biology Markup Language (SBML) Project, *Systems Biology*, Volume 1, Pages 41-53.

- [27] Beurton-aimar.M (2007). Langage de modélisation des réseaux biochimiques, 1–16, ECRIN-Biologie syst, Chap. 07, Page 7.
- [28] Bonmarin.I, Santa-Olalla.P, Lévy-Bruhl.D(2008). Modélisation de l'impact de la vaccination sur l'épidémiologie de la varicelle et du zona », Revue d'Epidémiologie et de Santé Publique 56 323–331.
- [29] Hunter, J. (2002). JDOM Makes XML Easy.Sun's 2002 Worlwide Java Developer Conference.
- [30] Bio-PEPA, <http://www.biopepa.org/>.