



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Individual-Based Models for Public Health

Benjamin Roche<sup>\*,†,1</sup> and Raphaël Duboz<sup>‡,§</sup>

<sup>\*</sup>Univ Paris 06, Sorbonne Universités, IRD, UMI 209, UMMISCO, Bondy, France

<sup>†</sup>UMR 224 MIVEGEC, IRD/Université de Montpellier/CNRS, Montpellier, France

<sup>‡</sup>UMR ASTRE CIRAD-INRA, Montpellier, France

<sup>§</sup>Epidemiological and Public Health Unit, Institut Pasteur du Cambodge, Phnom Penh, Cambodia

<sup>1</sup>Corresponding author: e-mail: benjamin.roche@ird.fr

---

### ABSTRACT

Today, infectious diseases represent a threatening concern for human health. Understanding their transmission, and possibly forecasting the dynamics of these pathogens, represents both a scientific and sanitary emergency. To this goal, mathematical modeling has been a widely used tool. Nevertheless, they have important limitations to explicitly model the mechanisms involved in the infectious processes at the individual level. Thanks to the increase of computing capacity, computational models such as individual-based models (IBMs) are very relevant for understanding the complexity of mechanisms at the individual level that can be involved in disease outbreaks. Their computational formalism allows a large flexibility, while they rely on the same philosophy than current models in mathematical epidemiology that have proved their relevance. In this chapter, we review the main qualities of IBMs, what kind of new knowledge they can bring and they have already produced in epidemiological modeling. Then, we highlight their caveats and what could be developed during the future years to make IBMs a more reliable and useful approach.

**Keywords:** Individual-based model, Agent-based model, Computational biology

---

## 1 BACKGROUND

Today, infectious diseases represent a threatening concern for human health. In addition to the main pathogens affecting human populations, such as *Plasmodium falciparum* causing malaria or influenza viruses that together kill hundreds of thousands of human lives throughout the world (Murray et al., 2012), an increasing number of pathogens are emerging throughout the world, such as the MERS Coronavirus in Saudi Arabia or the Ebola virus in West

Africa to cite a few (Jones et al., 2008). Understanding their transmission, and possibly forecasting the dynamics of these pathogens, represents both a scientific and sanitary emergency.

To this goal, modeling has been a widely used tool (Hufnagel et al., 2004; Najera, 1974). First, mathematical models gave crucial insights for fundamental infectious diseases management (Anderson and May, 1991). For instance, the probability of epidemics and its expected magnitude can be easily calculated through mathematical models by deriving the basic reproductive ratio ( $R_0$ ), which quantifies the number of secondary cases that a single infectious individual will produce when introduced within a population entirely naïve from an immunological point of view (Heesterbeek, 2002). These models have also allowed understanding that imperfect vaccination could lead to an increase of the mean age of infection, which is problematic for many childhood infections because of the higher probability of severe cases when these diseases occur after childhood (Keeling and Rohani, 2008).

In addition to these fundamental estimations, mathematical models have also been used in emergency situation. For instance, the 2001 outbreak of foot-and-mouth disease in the United Kingdom has seen a team of mathematical modelers being in charge of shaping the public health interventions, and especially the number and locations of animals to kill, to mitigate the explosiveness of the epidemics (Keeling et al., 2001). Despite many criticisms regarding the huge number of slaughtered animals (Kitching et al., 2006), partly justified regarding the lack of model validation due to our poor knowledge on the pathogen biology at that time, epidemics has been successfully controlled.

Nevertheless, the mathematical models have important limitations to explicitly model the mechanisms involved in the infectious processes at the individual level, such as evolution within large strain assemblages, complex spatial configuration shaping contacts, or heterogeneous individual behaviors facing an epidemic. Considering the huge amount of computational power we have access today, we have the opportunity to explore mechanisms of diseases spread, surveillance, and control at very fine scale (individuals), and how these mechanisms are related with what is observed at the population level. This improves our understanding of fundamental processes involved in diseases spread, and may improve as well forecasting potential of these models.

Broadly, two main alternatives can be invoked. First, a forecasting goal can be satisfied with big data approach. Based on the extensive exploration of large datasets (such as requests made on Google website or message contents from digital social networks) combined with cutting-edge statistical modeling, it is possible to forecast the dynamics of an epidemics reasonably well in advance (Ginsberg et al., 2009; Salathé et al., 2013). However, such approach is similar to a “black box” because it does not rely on a mechanistic formulation of the problem. Therefore, it hardly participates to the

understanding of pathogen transmission; neither forecast the dynamics of rare events.

As we said, computational models such as individual-based models (IBMs) are very relevant for understanding the complexity of mechanisms at the individual level that can be involved in disease outbreaks. Their computational formalism allows a large flexibility, while they rely on the same philosophy than current models in mathematical epidemiology that have proved their relevance.

In this chapter, we review the main qualities of IBMs, what kind of new knowledge they can bring and they have already produced in epidemiological modeling. Then, we highlight their caveats and what could be developed during the future years to make IBMs a more reliable and useful approach.

## 2 BROAD MODEL PHILOSOPHY

By definition, an IBM focuses at the individual scale. It is worth pointing out that we will use the term “individual-based models” in a broad sense throughout this chapter, i.e., encapsulating a large gradient from basic models with few individuals interacting in simple ways like cellular automata until very complex situations where individuals can interact with other individuals directly and/or through modifications of their environment, and where individuals can exhibit complex behaviors, such as what is generally assumed for agent-based models (ABMs). It is worth to notice here that ABMs are developed in many disciplines, from distributed artificial intelligence in computer sciences, to sociology, geography, or ecology. In this chapter, we use the term IBM because in epidemiology, IBMs clearly refer to the modeling of individuals within a population.

In software engineering, we talk about *objects*, which represent individuals that are instance of a *class*, which are abstract pattern of an object (such as human, vector, or pathogen in our case), and include *methods* (i.e., functions that an object can execute). Then, each individual has a set of attributes and functions. For epidemiological systems, classic attributes are the infectious status of the individual, such as Susceptible (who can be infected), Infectious (who can transmit the pathogen), or Recovered (who cannot be infected anymore) to frame it within the SIR philosophy (Keeling and Rohani, 2008). Obviously, any kind of status could be considered and other attributes than infectious status can be included, such as time of infection, spatial location, age, etc.

The *methods* of each individual characterize the processes that could be involved at this individual scale. In epidemiological systems, transmission and recovery are obviously naturally present. Nevertheless, these processes could rely on the individual itself (such as for recovery that will only depend of host attributes) or on the population, as it is the case for transmission that

will depend on the number of individuals in contact with the focal host. Each of these methods is used to characterize the change of host's attributes at the next time step.

Indeed, IBMs are generally discrete time models. To avoid important edge effects, it is required to properly schedule the updating of the individuals states. Two main methods exist. The first one consists in updating every individuals states at time  $t$  considering the individuals states at time  $t - dt$ . This method respects the causality principle and considers that individuals instantaneously change their states at the same time  $t$ . The corresponding scheduling algorithm is as follow:

```

FOR T IN 1:TMAX
  FOR I IN 1:NB_INDIVIDUALS
    INDIVIDUAL[I].STEP()
  END
  FOR I IN 1:NB_INDIVIDUALS
    INDIVIDUAL[I].UPDATE()
  END
END

```

where functions STEP() and UPDATE() compute and update the states of the individual, respectively, as represented by object INDIVIDUAL[I].

The second method consists, for each time step, to randomly draw an individual from the list of individuals, compute its new state, and then repeat the drawing for all individuals. This method considers that individuals randomly interact between the time  $t$  and  $t + dt$ . Here, interactions are asynchronous and continuous within one time step. The corresponding scheduling algorithm is as follow:

```

FOR T IN 1:TMAX
  RANDOMLY_SORT(LIST_OF_INDIVIDUALS)
  FOR_ALL INDIVIDUAL IN LIST_OF_INDIVIDUALS
    INDIVIDUAL.STEP()
    INDIVIDUAL.UPDATE()
  END
END

```

### 3 MODELING IBMs

#### 3.1 Specification

The specification corresponds to the description of a model using a particular formalism, being mathematical, graphical, textual, or hybrid. Knowing reproducibility to be a key process in the scientific activity, the specification of IBMs appears to be fundamental. It enables to communicate the model to pairs for validation and verification. Without unambiguous specification, pairs

cannot reproduce the model, and then the simulation results. They are many different types of specification for IBMs: from pseudo-code, like in the previous section, to graphical modeling language such as UML (Unified Modeling Language, see below). Nevertheless, most of the specifications remain ambiguous and need to be complemented by a documentation.

Grimm et al. have proposed the “Overview Design Details” (ODDs) protocol to communicate IBMs (Grimm et al., 2006). Using ODD can improve significantly the replication of IBM simulations. The major drawback of ODD remains in the ambiguity of its terminology and the absence of a rigorous description of the simulator, i.e., the software that supports the implementation of the IBM (Duboz et al., 2012). In a recent paper, Donkin et al. (2017) show that using ODD does not fully support replication by implementing the same IBM on two different simulation platforms (Repast and NetLogo). They found that the results not only differ in magnitude but also in trends leading to different conclusions regarding the question the model is supposed to address. This work highlights the tremendous importance to be able to specify IBMs rigorously. Despite the fundamental importance of replication in science, few works have been done regarding IBMs (Axtell et al., 1996; Bajracharya and Duboz, 2013; Donkin et al., 2017; Railsback et al., 2006; Wilenksy and Rand, 2007). Of course, replication is time consuming and perceived as non-rewarding. It is nevertheless mandatory if we want IBM to be considered as a valid scientific tool.

### 3.2 Unified Modeling Language and Individual-Based Modeling

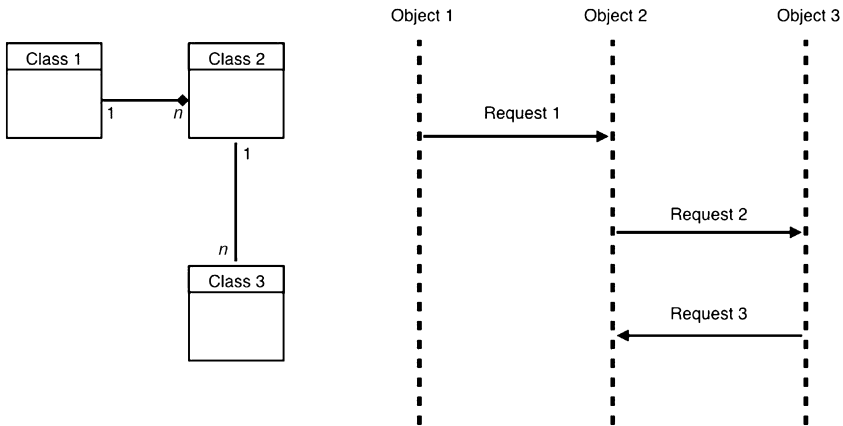
The main language used to model IBM is the UML. Developed during the 1980s to improve software design with the increasing use of object-oriented languages, this language allows a graphical representation of the interactions between the different *classes* involved in the software (class diagram) as well as the behavior of the most important *methods* (Fig. 1).

### 3.3 Toward a Formal Specification of IBMs by Using the Discrete Events Specification System

A possible answer to the specification problem in IBMs is to ground the specification into existing abstract formalisms that have been developed by simulation practitioners in the field of “automatic.” In particular, states machines and computation theory are both very inspiring. The most promising formalism for IBMs specification is probably the Discrete Events Specification System (DEVS), an operational formalism<sup>3</sup> to specify discrete events simulation models.

---

<sup>3</sup>An operational formalism is a specification having a strict equivalence with the algorithms that implement it. It is independent from the programming language used for its implementation.



**FIG. 1** Examples of UML diagrams. (*Left*) Class diagram, showing relationships between different classes. (*Right*) Sequence diagram, showing behavior of a specific method.

DEVS is derived from discrete mathematics and relies on systems theory. It involves enough generalism to formalize discrete and continuous time models and can also merge continuous and discrete time dynamics together. (A comprehensive introduction to DEVS is available in [Zeigler et al. \(2000\)](#).) From DEVS, several extensions have been proposed for specific purposes. For instance, spatially explicit systems can be formalized using Cell-DEVS, differential equations systems by using QSS (a quantized version of DEVS to solve differential equations), etc. An interesting property of DEVS extensions models is that they are all formally equivalent to DEVS models. Therefore, any models specified using a particular extension can be formally and computationally coupled with other models in a DEVS framework. This property ensures the scalability of models and promotes models reuse. Furthermore, DEVS proposes a compositional and hierarchical approach to modeling. Therefore, elementary models (called atomic models) can be coupled together to build a coupled model that can be view as an atomic model at upper hierarchical level.

Since the beginning of this century, DEVS is used for the specification of IBM ([Kim and Kim, 2001](#); [Zhang et al., 2011](#)). One specific feature of IBM is that individuals can continuously appear and disappear, change their behaviors, and change the way they interact with others. Then, the structure of such a model is dynamic, conversely to differential equations models that possess a static structure. It is now recognized that the dynamic structure of IBMs can be specified with DS-DEVS (Dynamic Structure DEVS ([Duboz et al., 2006](#); [Zeigler et al., 2000](#))), and recent works still improve IBM specification within this DEVS framework ([Won Bae and Il-Chul, 2015](#)).

## 4 WORKING WITH MEAN-FIELD AND INDIVIDUAL-BASED MODELS

As said previously, many mathematical models have been developed on numerous diseases at a population scale, and it will be a pity to not rely on these models when we add details at the individual scale. Conversely, building an aggregated model from an IBM can enlighten how properties emerge at the population level (Huet and Deffuant, 2008). Nevertheless, going up and down in scales is not that trivial. Indeed, parameters in population mathematical models (also called mean-field models, MFMs, because they are continuous states models assuming well-mixed populations and the ergodic hypothesis) are generally expressed as rates at a population scale, while parameters in IBMs have generally to be expressed as a probability at an individual scale (deterministic IBMs are not very common and will therefore not be addressed in this chapter). If we assume that time between state transitions are exponentially distributed, the individual probability can be calculated from the corresponding rate used in MFMs (Keeling and Rohani, 2008) as following:

$$P(\text{event}) = 1 - \exp.(-\text{rate} \times t)$$

But it underlines that the duration  $t$  of a time step is dramatically important. While a large time step will allow faster model execution, it will also dangerously overaggregate the event distribution and can therefore bias completely the dynamics. While a time step of 1 day was generally considered in epidemiology (Keeling and Rohani, 2008), it has been shown that a finer time resolution is required, even with very simple dynamics (Roche et al., 2011).

### 4.1 IBMs and MFMs of the Same System

By comparing an IBM with its equivalent MFM, we can learn what are the approximations made by the MFM that can lead to significant differences with the IBM simulation results. In their paper, Bonté et al. (2012) built two epidemiological models of the same system. The first one is an IBM and the second one is the corresponding MFM derived using the moments approximation technique. This work highlighted the importance for the models to capture the displacement behaviors and the contact processes at the individual level in the study of disease spread.

In their recent paper, Özmen et al. (2016) built several MFMs and IBMs of the well-documented Spanish Flu epidemic. They show clearly that the underlying modeling hypothesis in MFM and IBM can lead to significantly different simulation results for the simulation of the same phenomenon. This result is extremely important as both models can be used for decision-making.



## 4.2 The Coupling of IBM With MFM to Enable Scale Transfer Modeling (Multiscale Modeling)

Another interesting use of both approaches is to couple them to form a single model of the considered system at different space and timescales. Doing this, it is possible to simulate a scale transfer between two hierarchical levels, the individual level (IBM) and the population level (MFM) (Duboz et al., 2003). This coupling enables to study how individual behavior (foraging in the mentioned study) impacts the population dynamic. In this approach, the IBM is considered as a virtual experiment to compute the parameters of the MFM. This technique has the potential to greatly reduce computation time, which represents one of the main limitations of complex IBMs. Indeed, the MFM can deal with any size of population and the IBM is considered as a population sample that is coupled through individual statistics computed from simulation outputs.

## 5 SPECIFIC USES OF IBMs

Mathematical models have proved that they can provide crucial insights into our understanding and our ability to forecast epidemiological dynamics. Therefore, it is important to identify what kind of improvements IBMs can bring. We will here review nonexhaustively some of the most important opportunities offered by this modeling approach, which can be also combined together.

### 5.1 Spatially Explicit Models

Classic compartmental mathematical models are able to consider a spatial dimension, through a collection of populations connected between them (Xia et al., 2004) or through a diffusion equation involved within a partial differential equations systems (Tran and Raffy, 2006). However, such models have generally a simplistic representation of space at fine scale.

Therefore, the most involved improvement of IBM has been clearly to consider a spatially explicit dimension at this fine scale. Use of IBM in this context has started very early with the development of cellular automata (Wolfram, 2002) and remains one of the main reasons to develop IBM in epidemiology.

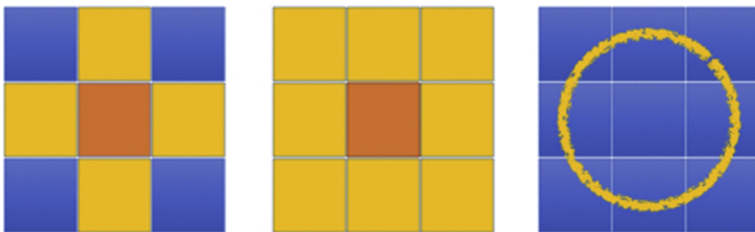
Considering an explicit spatial dimension yields to the fact that the different elements of an epidemiological system (humans, animal reservoir, and/or vectors among others) have less or more contact than with others. As a result, this explicit spatial dimension can be summarized as a network of interactions, which can fluctuate through time.

When individual movements are not considered explicitly, epidemiological studies have generally considered spatial dimension through a network of interactions among individuals. These networks could be derived from

complex socioeconomic data, with a clear aim of forecasting possible future disease outbreaks (Eubank et al., 2004). These networks could be also a result of well-known algorithms, which can generate a large gradient of network properties such as random networks with the Erdos–Rényi algorithm (Erdos and Rényi, 1959) until realistic network patterns that can mimic accurately documented network with the preferential attachment that generates scale-free networks (Barabasi et al., 1999). Generally, epidemiological studies that have relied on such networks have investigated the role of connectivity between these epidemiological elements on the different metric of disease transmission (Keeling, 1999; Roche et al., 2013) or on pathogen evolution (Read and Keeling, 2003).

These interactions networks can be the result of an explicit consideration of individual movement from one place to another. Generally, when the spatial dimension is discrete, we consider that individuals can interact with their neighborhood, i.e., with the individuals that are located closely. Thanks to an inheritance from the cellular automata era, we can consider a “Von Neumann” neighborhood (assuming that an individual can interact only with the adjacent individuals) or a “Moore” neighborhood (assuming that individuals can interact with all the eight individuals surrounding the considered one). Today, even when the spatial dimension can be still discrete, it is also common to consider a spatial radius where focal individual can interact with other elements of the model (Fig. 2). In epidemiology, these models that not relying on field data have been mostly used to identify fundamental properties of pathogen propagation with physical constraints or different kind of habitats (Roche et al., 2008).

Based on this explicit consideration of the spatial dimension, many sophistications have been made possible thanks to software design and computing resources. The recent developments of IBM platforms that allowed the integration of very complex heterogeneous data with multiple layers of spatial dimensions from Geographical Information Systems (GISs). These models are allowing to explore the spatial propagation of epidemics in a realistic context (Brown et al., 2005; Muller et al., 2004), despite their validation on epidemiological data remains extremely challenging.



**FIG. 2** Different considerations of the spatial neighborhood. (*Left*) Von Neumann neighborhood, (*middle*) Moore neighborhood, and (*right*) radius neighborhood.

## 5.2 Complex Behaviors

The second main reason to use IBMs is the potential to consider complex behaviors of the different epidemiological elements. For instance, the decision of an individual to adopt a preventive measure depends on her/his perception of risks associated with her/his actions. This perception depends on individual experiences and personal beliefs and can evolve in time. The set of possible choices is discrete and finite, and the rules selecting actions often refer to qualitative data.

Differential equations hardly represent discrete choices and qualitative information, whereas computer programming is extremely flexible and allows considering behaviors relying on many different things, such as the surrounding individuals, the environment status, etc. Therefore, integrating complex behavior has always been a key element of IBMs, mainly in ecological studies (Grimm et al., 2005).

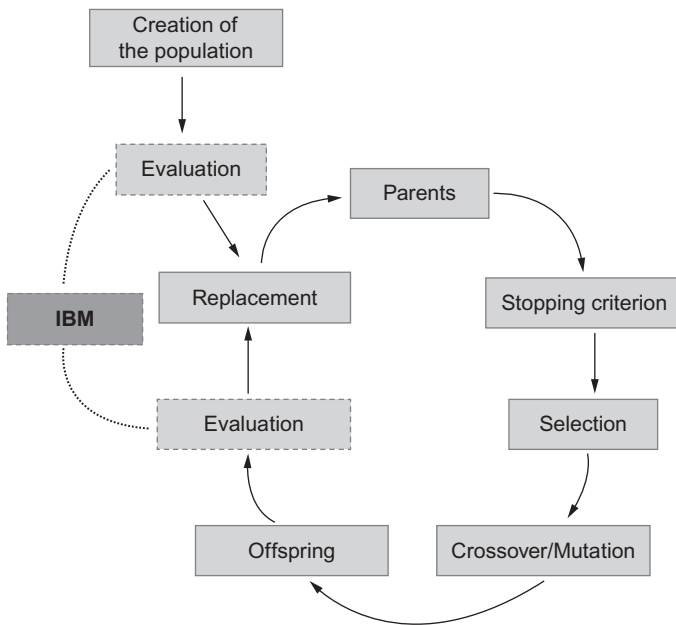
This aspect is critical for IBMs. Indeed, human behavior is increasingly recognized as a key element in pathogen transmission (Funk et al., 2010). Nevertheless, integration of behaviors in IBMs is still burgeoning in epidemiology despite the numerous models that have been developed in the field of artificial intelligence, and more specifically for ABMs. The main reason is the quasi-absence of data regarding human behavior being explicitly collected for a modeling purpose in epidemiology. It appears necessary here to collaborate with sociologists or behavioral economists for instance, to benefit from their methodologies. More generally, participatory modeling methods are needed to integrate the different disciplines and stakeholders concern with public health issues (Binot et al., 2015). Moreover, the possibility to track simple parts of human behavior in simple mathematical model relying on the classic SIR framework has been very insightful (Andrews and Bauch, 2016), but has also delayed the consideration of more complex behaviors.

## 5.3 Multistrains Pathogens

Another possibility is also to expand the possibility of classic mathematical models. To this extent, considering large assemblages of pathogen strains represent a crucial example. Many pathogens show an antigenic variability that constrains vaccine development and therefore disease control, such as in the case of influenza or Dengue viruses. In populational mathematical models, the strain space has to be fixed, which could also generate some edge effects around the first or the last strains (Gog and Grenfell, 2002). This constraint is removed thanks to the computational formulation because this strain space can be extended dynamically, while it can nevertheless rely on the same compartmental formulation (Roche et al., 2011).

## 6 IBM CALIBRATION

IBM calibration or optimization is an important issue. For IBM, as for any kind of models, it may happen that some of the parameters cannot be estimated experimentally. In such case, the modeler could fit model parameters to reproduce collected data and then calibrate the model. The main difficulty is the stochastic nature of IBM. We have therefore to use Monte Carlo techniques to fit parameters, which are difficult to implement and for which the convergence toward an optimum cannot be assert. Moreover, this problem is particularly important with IBMs when one single simulation may require minutes or hours to be achieved. Heuristic techniques, such as evolutionary algorithms, can reduce the computation time for parameter exploration (Fig. 3) (Duboz et al., 2010).



**FIG. 3** General outline of an evolutionary algorithm. *Creation of a population* using a uniform random drawing of parameter values within the search space. *Evaluation*: Computation of the phenotype and the fitness of each individual. It consists in using the IBM instead of a mathematical evaluation function. *Selection*: Choice of some individuals that will become parents based on the fitness. *Crossover/mutation*: Use of parents to create new individuals, the offspring. Several parents' genotypes can be mixed (crossover) and/or modified (mutation). *Evaluation*: Computation of the phenotype and the fitness of the offspring. *Replacement*: Choice of the individuals (among parents and offspring) which will constitute the next generation (based on the fitness). From Duboz, R., Versmisse, D., Travers, M., Ramat, E., Shin, Y.J., 2010. Application of an evolutionary algorithm to the inverse parameter estimation of an individual-based model. *Ecol. Model.* 221, 840–849.

Nevertheless, convergence is still an issue and IBMs calibration remains difficult when the number of parameters is large. Some methodological developments in Approximate Bayesian Computation (ABC) have started addressing this problem. ABC has the potential to optimize an IBM's structure and parameters within an established statistical framework. In their paper, [Van der Vaart et al. \(2015\)](#) showed that ABC can potentially represent uncertainty in IBM parameters, structure, and predictions. However, work remains to be done in developing ABC for IBM with a large number of parameters.

## 6.1 Sensitivity Analysis

Sensitivity analysis focuses on studying uncertainties in model outputs because of uncertainty in model inputs. In other words, the aim of sensitivity analysis is to assess the influence of model's parameter values and model structure on its output. Sensitivity analysis is divided into two main approaches: local and global. The local approach looks at the variation in the output of the model when parameters vary around their reference values. The global approach considers the whole values intervals for each parameter.

When using sensitivity analysis with IBM, we consider model-free methods, for which the explicit mathematical formulation of the model is not necessary ([Saltelli et al., 2008](#)). With sensitivity analysis, we can identify the parameters that we have to know precisely (conversely roughly) to increase the confidence we have in simulation results. The method is particularly useful in the design phase of the model, when the modeler wants to identify what are the important processes to consider when trying fitting the simulation output to data. Then, the modeler increases his knowledge on the mechanisms responsible for generating the model output.

One of the most interesting methods for IBM is the variance decomposition method ([Ginot et al., 2006](#)). It consists in using the analysis of variance method (ANOVA) to study the part of variance in the output originating in the variance of the input parameters values. It is possible to study linear and nonlinear direct effects and the interaction effects between parameters. Ginot et al. also showed how local sensitivity analysis can be done for all the parameters, showing how their influence can vary over time. In epidemiology, [Nsoesie et al. \(2015\)](#) developed an IBM for influenza spread. They considered public health-relevant outcomes of the model: time to epidemics peak, proportion of infected people at the peak, and total attack rate. Their results suggested that small changes in some of the model parameters could significantly influence the characteristics of an epidemic.

## 7 BIOLOGICAL KNOWLEDGE GAINED THANKS TO IBMs

The first objective of IBMs was clearly to consider an explicit spatial dimension at a low scale combined with the need to represent individual heterogeneity.

Therefore, many different scenarios have been tested to simulate the propagation of numerous emerging pathogens, such as Ebola disease for instance (Merler et al., 2015). For other pathogens like influenza viruses potentially pandemic, the simulation of complex spatiotemporal dynamics (Ferguson et al., 2005) may have also been used to quantify how different kind of interventions may impact this spatiotemporal dynamics (Longini et al., 2005). Such approach has been even used in decision-making, in the case of the foot-and-mouth epidemics in the United Kingdom (Keeling et al., 2001, 2003).

As said previously, IBMs can be used to offer a more flexible formalism than the mathematical one in the case where mathematical model could yield to combinatory explosion. It was the case for the genetic diversity of avian influenza viruses, where the environmental transmission route (i.e., infected ducks are excreting viral particles in their feces that can persist during several weeks in the environment) was suspected to play an important role. Nevertheless, the current multistrains mathematical models were involving mathematical assumptions in order to reduce the complexity of this problem that consequently underestimate transmission. While these assumptions are not problematic with human influenza viruses where direct transmission is short and intense, relying on these assumptions for environmental transmission, which operates on a longer term, would yield to a disappearance of this transmission route. Through the development of an IBM validated to reproduce the same dynamics than an equivalent mathematical one (Roche et al., 2011), its investigation has been able to show the importance of this transmission in avian influenza genetic diversity (Roche et al., 2014).

The Pattern-Oriented Modeling (POM) method introduced by Grimm et al. (2005) is a pragmatic method to develop IBM. It is well known in ecology. It consists in selecting observational data characterizing the processes under study, the patterns (for instance the evolution of total biomass in time if we are interested in population tree growth), and measuring the distance between observations (selecting several independent patterns) and the model output. The idea is to increment the complexity of the model until the fit with the patterns is considered acceptable (the distance between the patterns and simulation output is small). An important contribution to this method is the use of an information criterion similar to AIC to select the best model (Piou et al., 2009). As we said previously regarding IBM calibration, ABC can be used to optimize the model (Van der Vaart et al., 2015) and could be combined with an information criterion to select the best model.

When using the POM, the modeler selects the minimum set of processes and their formulation necessary to reproduce the field data. The use of a parsimony criterion increases the confidence we have on the ability of the model to effectively represent the processes that occur in the real system, therefore increasing the knowledge we have on it.

## 8 CAVEATS OF IBMs

Despite their flexibility and the increase of computing resources, there are some relevant reasons that mathematical models are still more used than IBMs. The first one is a problem of reproducibility that is inherent to the complexity of these models. For mathematical models, everything is encapsulated within a set of equations or table of rates associated with the different events. IBMs rely on complex computational code, and that is rarely made available to the whole scientific community. As a consequence, most of the results produced by these IBMs are not reproducible, which casts doubts on their validity because epidemiological outcomes could be the result of coding errors or even hidden process not explained in model description. This is even more detrimental because some specification norms (DEVS, see previously) already exist, but are not well disseminated in the epidemiological literature. This concern leads to a lack of confidence in this approach, which is extremely detrimental for the use of IBMs in many fields and not only for epidemiological studies. The effort of computational coding, which has obviously to be rewarded through scientific valorization, should nevertheless not constrain its reproducibility.

The second caveat to this approach is the rare confrontation of models simulations to real data. This is also inherent to the complexity of this model, because such comparison is challenging to conduct. It is worth pointing out that the initial goal of these models was clearly to evaluate, or at least to visualize, different scenarios, mostly of spatial propagation of some diseases. Nevertheless, the lack of confrontation with real data makes these models equivalent as some kind of thought experiments, or toy models that are impossible to validate, casting doubts on the predictions produced by them. Because the computing resources are now available to confront these simulations to real data, it would extremely relevant to develop relevant statistical framework to ensure a clear and deep evaluation of their accuracy.

The last important caveat to these IBMs relies on the difficulty to explore the whole range of parameters included in these models. Because they generally involve a high complexity, doing a sensitivity analysis could be extremely challenging and time consuming. Nevertheless, algorithms such as the Latin Hypercube Sampling (Tang, 1993) can allow an efficient exploration of this multidimensional parameter space and be combined with a decomposition of variance, in order to quantify the fraction of the variance that can be attributed to each input parameters. Today, other approaches using Monte Carlo filtering exist with the goal to identify parameter space where models outputs are significantly different from other parameter space. Because the risk of overparameterization is high with IBMs and that the parsimony principle of the modeling approach is not always satisfied, sensitivity analyses should become a standard for these models.

## 9 SOFTWARE PLATFORMS

These IBMs can be developed with any kind of computing language, especially object-oriented ones such as C++ or Java. Nevertheless, in order to focus only on the concepts involved in the model and then facilitate its coding (especially when complex interactions exist between the different objects), numerous platforms exist. We review briefly here three of them that are representative to what could exist.

*Swarm* is one of the first individual-based platforms that has been made available. Historically rooted in the field of artificial life, this platform has been originally developed at the Santa Fe Institute. As a consequence, it has been developed to offer a wide range of possibilities in order to be able to simulate a large range of systems, such as social, ecological, or epidemiological ones. Coded in Java today, this platform does not have a specific interface to develop these models, but rather offers numerous libraries to facilitate task scheduling, simulation visualization, or data interface among many other possibilities.

*Netlogo* is a platform initiated by the Northern University's Center for Connected Learning and Computer-Based Modeling. At the opposite of *Swarm*, its use is extremely simple through a friendly integrated interface. While it is consequently less flexible than low-scale platforms such as *Swarm*, *NetLogo* is extremely used for educational purposes because it could illustrate rapidly the possibilities offered by IBMs.

*GAMA* is one example of a sophisticated individual-based modeling platform, involving hybrid models (models combining simultaneously numerical resolution of integrated mathematical models and IBMs), GIS integration, and parallel simulations allowing simulations of huge ecosystems (up to millions of agents) among others. Developed by the French National Research Institute for Sustainable Development (IRD), this platform is an example of intermediate platforms between the gradient offered going *NetLogo* to *Swarm*. Models are developed through a high-level language (GAML) through a dedicated interface. Therefore, it offers more flexibility than *NetLogo* despite more complicated to use for educational purposes, but is nevertheless more constrained than *Swarm* and its Java libraries.

## 10 CONCLUSIONS

IBMs represent an intriguing opportunity in epidemiological modeling. They allow addressing epidemiological systems that are not easily addressable with classic mathematical models, such as complex spatial landscape, complex behaviors, or large pathogen assemblages among others. Nevertheless, these models have not passed the threshold of "being an opportunity" to become a reliable and widely used formalism. This is mostly because of the lack of validation of these models, which is deeply rooted in their complexity but that



could be overcome with the increasing computing resources and sophisticated statistical framework.

It is therefore crucial to highlight that IBMs and mathematical models are not competing with each other, but rather complementary. On one hand, mathematical modeling offers a compact and synthetic way to a given problem that should be favored for simple problem. On the other hand, IBMs can go further when mathematical formalism is not adequate anymore. The emergence of hybrid models (Bobashev et al., 2007), involving numerical resolution of ordinary differential equation, and stochastic IBMs confirms the complementarity of these different approaches. Nevertheless, it also highlights the relevance of the paradigm “one model for one question.”

## REFERENCES

- Anderson, R.M., May, R.M., 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford Science Publications, Oxford, UK.
- Andrews, M.A., Bauch, C.T., 2016. The impacts of simultaneous disease intervention decisions on epidemic outcomes. *J. Theor. Biol.* 395, 1–10.
- Axtell, R., Axelrod, R.M., Epstein, J.M., Cohen, M.D., 1996. Aligning simulation models: a case study and results. *Comput. Math. Organ. Theory* 1, 123–141.
- Bajracharya, K., Duboz, R., 2013. Comparison of three agent-based platforms on the basis of a simple epidemiological model (WIP). In: *Proceedings of the Symposium on Theory of Modeling & Simulation—DEVS Integrative M&S Symposium*, 6.
- Barabasi, A., Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509–512.
- Binot, A., Duboz, R., Promburom, P., Phimpraphai, W., Cappelle, J., Lajaunie, C., Goutard, F., et al., 2015. A framework to promote collective action within the one health community of practice: using participatory modelling to enable interdisciplinary, cross-sectoral and multi-level integration. *One Health* 1, 44–48.
- Bobashev, G.V., Goedecke, D.M., Yu, F., Epstein, J.M., 2007. A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches, 2007 Winter Simulation Conference.
- Bonté, B., Mathias, J.D., Duboz, R., 2012. Moment approximation of infection dynamics in a population of moving hosts. *PLoS One* 7e51760.
- Brown, D.G., Riolo, R., Robinson, D.T., North, M., Rand, W., 2005. Spatial process and data models: toward integration of agent-based models and GIS. *J. Geogr. Syst.* 7, 25–47.
- Donkin, E., Dennis, P., Ustalakov, A., Warren, J., Clare, A., 2017. Replicating complex agent based models, a formidable task. *Environ. Model. Softw.* 92, 142–151.
- Duboz, R., Ramat, E., Preux, P., 2003. Scale transfer modelling: using emergent computation for coupling an ordinary differential equation system with a reactive agent model. *Syst. Anal. Model. Simul.* 43 (6), 793–814.
- Duboz, R., Versmisse, D., Quesnel, G., Muzzy, A., Ramat, E., 2006. Specification of dynamic structure discrete event multiagent systems, *Agent-Directed Simulation (ADS 2006)*.
- Duboz, R., Versmisse, D., Travers, M., Ramat, E., Shin, Y.J., 2010. Application of an evolutionary algorithm to the inverse parameter estimation of an individual-based model. *Ecol. Model.* 221, 840–849.
- Duboz, R., Bonté, B., Quesnel, G., 2012. Vers une spécification des modèles de simulation de systèmes complexes. *Stud. Inform. Univ.* 10, 7–37.

- Erdos, P., Rényi, A., 1959. On random graphs. I. *Publ. Math.* 6, 290–297.
- Eubank, S., Guclu, H., Kumar, V.S.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 180–184.
- Ferguson, N.M., Cummings, D.A.T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., et al., 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437, 209–214.
- Funk, S., Salathé, M., Jansen, V.A.A., 2010. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* 7, 1247–1256.
- Ginot, V., Gaba, S., Beaudouin, R., Aries, F., Monod, H., 2006. Combined use of local and ANOVA-based global sensitivity analyses for the investigation of a stochastic dynamic model: application to the case study of an individual-based model of a fish population. *Ecol. Model.* 193, 479–491.
- Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. *Nature* 457, 1012–1014.
- Gog, J.R., Grenfell, B.T., 2002. Dynamics and selection of many-strain pathogens. *Proc. Natl. Acad. Sci. U. S. A.* 99, 17209–17214.
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., et al., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* 310, 987–991.
- Grimm, V., Berger, U., Bastiansen, F., Eliasenn, S., Ginot, V., Giske, J., Goss-Custard, J., et al., 2006. A standard protocol for describing individual-based and agent-based models. *Ecol. Model.* 198, 115–126.
- Heesterbeek, H., 2002. A brief history of  $R_0$  and a recipe for its calculation. *Acta Biotheor.* 50, 189–204.
- Huet, S., Deffuant, G., 2008. Differential equation models derived from an individual-based model can help to understand emergent effects. *J. Artif. Soc. Soc. Simul.* 11, 10.
- Hufnagel, L., Brockmann, D., Geisel, T., 2004. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15124–15129.
- Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. *Nature* 451, 990–994.
- Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. *Proc. Biol. Sci.* 266, 859–867.
- Keeling, M.J., Rohani, P., 2008. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, NJ.
- Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., et al., 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 294, 813–817.
- Keeling, M.J., Woolhouse, M.E.J., May, R.M., Davies, G., Grenfell, B.T., 2003. Modelling vaccination strategies against foot-and-mouth disease. *Nature* 421, 136–142.
- Kim, J.-H., Kim, T.G., 2001. DEVS-based framework for modeling/simulation of mobile agent systems. *Simulation* 76, 345–357.
- Kitching, R.P., Thrusfield, M.V., Taylor, N.M., 2006. Use and abuse of mathematical models: an illustration from the 2001 foot and mouth disease epidemic in the United Kingdom. *Rev. Sci. Tech.* 25, 293–311.
- Longini, I.M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D.A.T., Halloran, M.E., 2005. Containing pandemic influenza at the source. *Science* 309, 1083–1087.
- Merler, S., Ajelli, M., Fumanelli, L., Gomes, M., Piontti, A., Rossi, L., Chao, D., et al., 2015. Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the

- effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *Lancet Infect. Dis.* 15, 204–211.
- Muller, G., Grébaud, P., Gouteux, J.P., 2004. An agent-based model of sleeping sickness: simulation trials of a forest focus in southern Cameroon. *C. R. Biol.* 327, 1–11.
- Murray, C.J.L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A.D., Michaud, C., Ezzati, M., et al., 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010. *Lancet* 380, 2197–2223.
- Najera, J., 1974. A critical review of the field application of a mathematical model of malaria eradication. *Bull. World Health Organ.* 50, 449–457.
- Nsoesie, E.O., Ricketts, R.P., Brown, H.E., Fish, D., Durham, D.P., Ndeffo Mbah, M.L., Christian, T., et al., 2015. Spatial and temporal clustering of chikungunya virus transmission in Dominica. *PLoS Negl. Trop. Dis.* 9e0003977.
- Özmen, O., Nutaro, J.J., Pullum, L.L., Ramanathan, A., 2016. Analyzing the impact of modeling choices and assumptions in compartmental epidemiological models. *Simulation* 92, 459–471.
- Piou, C., Berger, U., Grimm, V., 2009. Proposing an information criterion for individual-based models developed in a pattern-oriented modelling framework. *Ecol. Model.* 220, 1957–1967.
- Railsback, S.F., Lytinen, S.L., Jackson, S.K., 2006. Agent-based simulation platforms: review and development recommendations. *Simulation* 82, 609–623.
- Read, J.M., Keeling, M.J., 2003. Disease evolution on networks: the role of contact structure. *Proc. Biol. Sci.* 270, 699–708.
- Roche, B., Guegan, J.-F., Bousquet, F., Daszak, P., Cunningham, A., Hyatt, A., Anderson, R., et al., 2008. Multi-agent systems in epidemiology: a first step for computational biology in the study of vector-borne disease transmission. *BMC Bioinf.* 9, 435.
- Roche, B., Drake, J.M., Rohani, P., 2011. An agent-based model to study the epidemiological and evolutionary dynamics of influenza viruses. *BMC Bioinf.* 12, 87.
- Roche, B., Benbow, M.E., Merritt, R., Kimbirauskas, R., McIntosh, M., Small, P.L.C., Williamson, H., et al., 2013. Identifying the Achilles' heel of multi-host pathogens: the concept of keystone "host" species illustrated by mycobacterium ulcerans transmission. *Environ. Res. Lett.* 8, 45009.
- Roche, B., Drake, J.M., Brown, J., Stallknecht, D.E., Bedford, T., Rohani, P., 2014. Adaptive evolution and environmental durability jointly structure phylodynamic patterns in avian Influenza viruses. *PLoS Biol.* 12.
- Salathé, M., Freifeld, C., Mekaru, S., Tomasulo, A., Brownstein, J., 2013. Influenza A (H7N9) and the importance of digital epidemiology. *N. Engl. J. Med.* 369, 401–404.
- Saltelli, A., Chan, K., Scott, E.M., 2008. Sensitivity Analysis. Wiley 494 pp.
- Tang, B., 1993. Orthogonal array-based Latin hypercubes. *J. Am. Stat. Assoc.* 88, 1392–1397.
- Tran, A., Raffy, M., 2006. On the dynamics of dengue epidemics from large-scale information. *Theor. Popul. Biol.* 69, 3–12.
- Van der Vaart, E., Beaumont, M.A., Johnston, A.S., Sibly, R.M., 2015. Calibration and evaluation of individual-based models using approximate Bayesian computation. *Ecol. Model.* 312, 182–190.
- Wilensky, U., Rand, W., 2007. Making models match: replicating an agent-based model. *J. Artif. Soc. Soc. Simul.* 10, 1–24.
- Wolfram, S., 2002. *A New Kind of Science*. Wolfram Media, Cahmpaign, IL.
- Won Bae, J., Il-Chul, M., 2015. LDEF formalism for agent-based model development. *IEEE Trans. Syst. Man Cybern. Syst. Hum.* 46, 1.

- Xia, Y., Björnstad, O.N., Grenfell, B.T., 2004. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Nat.* 164, 267–281.
- Zeigler, B.P., Praehofer, H., Kim, T.G., 2000. *Theory of Modeling and Simulation*, second ed. Academic Press, San Diego, CA.
- Zhang, B., Chan, W.K., Ukkusuri, S.V., 2011. Agent-based discrete-event hybrid space modeling approach for transportation evacuation simulation, Winter Simulation Conference (WSC), Phoenix, AZ, pp. 199–209.