IDEAL, BEST, AND EMERGING PRACTICES IN CREATING ARTIFICIAL SOCIETIES

Philippe J. Giabbanelli

Computer Science Department
Furman University
3300 Poinsett Highway
Greenville, SC 29613, USA
giabba@furman.edu

Alexey A. Voinov

Faculty of Engineering & IT
University of Technology Sydney
Building 11, 81 Broadway
Ultimo NSW 2007, Australia
aavoinov@gmail.com

Brian Castellani

Department of Sociology
Durham University
32 Old Elvet
Durham, DH1 3HN, United Kingdom
brian.c.castellani@durham.ac.uk

Petter Törnberg

Amsterdam Institute for
Social Science Research
University of Amsterdam
Postbox 15508, Amsterdam, The Netherlands
petter@toernberg.com

ABSTRACT

Artificial societies used to guide and evaluate policies should be built by following "best practices". However, this goal may be challenged by the complexity of artificial societies and the interdependence of their sub-systems (e.g., built environment, social norms). We created a list of seven practices based on simulation methods, specific aspects of quantitative individual models, and data-driven modeling. By evaluating published models for public health with respect to these ideal practices, we noted significant gaps between current and ideal practices on key items such as replicability and uncertainty. We outlined opportunities to address such gaps, such as integrative models and advances in the computational machinery used to build simulations.

Keywords: Agent Based Model, Complexity, Quality Assessment, Human Behavior

1 INTRODUCTION

Artificial societies provide a 'safe' platform in which what-if questions can be tested, and their effects reversed, instead of directly impacting the population in possibly irreversible ways (Epstein 2008). While the automated identification of interventions remains a challenge, artificial societies safely assist experts in *identifying* interventions. They also provide support with *evaluation*, by highlighting which variables are related to the intervention and should be monitored (Craig et al. 2008), or finding what may have caused a failure (Hammond 2015). Therefore, the possibilities offered by artificial societies are not merely to represent data but to guide and evaluate policies, in part by highlighting where to focus the next actions and data collection efforts (Silverman et al. 2015). Given such important and multi-faceted roles, it is critical

to develop high-quality models. While modelers may agree on this abstract goal, viewpoints and practices significantly differ when we aim to precisely characterize *what* makes a high-quality model.

Many studies have been devoted to quality assessment, from Sargent's research on verification and validation (Sargent 2009) to guidance on producing quality analysis (Treasury 2015) and work specific to environmental models (Bennett et al. 2013). Together, such studies articulate many high-level objectives such as quality assurance throughout model development (e.g., after designing the model, after populating it with data, after testing it, etc.) or proportionality of response (i.e., adapting the quality assessment to the level of risks given the intended use of the model). Despite offering valuable guidance to end-users and modelers, we posit that the multiplicity of high-level objectives results in several practical shortcomings. First, the gap between current practices and the large number of tasks required for producing a high-quality model (i.e., *ideal* objectives) may lead modelers to seeing ideal practices as unattainable. Second, high-level tasks can be difficult to translate into practices given the wide variety of models as well as emerging modeling practices. Moreover, since modeling is always purposeful, model quality must be estimated in the context of its goals rather than as if it was an ideal general model. Our goal is to identify the most critical gaps between practices and requirements of high-quality models, while accounting for new solutions and challenges brought through emerging practices. This objective is addressed via the following three aims:

1. Describe a core set of ideal practices for high-quality quantitative individual models.
2. Evaluate the gap between ideal practices and current practices, using a sample of artificial societies built for public health.
3. Establish high-priority areas to bridge the gap given current practices as well as solutions and challenges emerging in the short- and medium-term.

The remainder of this paper is organized as follows. In section 2, we delineate the boundaries for the application of this paper by characterizing quantitative individual-based artificial societies. In section 3, we focus on a set of seven ideal practices for high-quality models, drawn from previous studies complemented by our collective experience as modelers. As a comprehensive review of all current practices across application areas is beyond the scope of this paper, section 4 uses a sample of artificial societies for public health to contrast current practices with our seven ideal practices. Then, section 5 focuses on emerging challenges and opportunities, from integrative models and the cloud to accounting for complexity in artificial societies. Our discussion in section 6 builds on sections 3–5 to identify high-priority areas for improvements that consider both current and emerging practices. We conclude by summarizing the position of this paper on how high-quality artificial societies can be developed using quantitative models.

2 QUANTITATIVE INDIVIDUAL MODELS USED TO BUILD ARTIFICIAL SOCIETIES

Badham categorized models as qualitative aggregated (e.g., maps delineating the structure of a problem by articulating how factors are connected (Giabbanelli 2018, Giabbanelli and Baniukiewicz 2018)), quantitative aggregated (i.e., *macro-simulation* such as System Dynamics (Verigin, Giabbanelli, and Davidsen 2016)), and quantitative individual (also called *micro-simulation*). Other typologies may include semi-quantitative models (Voinov et al. 2018), such as Fuzzy Cognitive Maps (FCMs), which can use qualitative surveys and interviews as inputs (Lavin et al. 2018) but produce numerical estimates as outputs (Lavin and Giabbanelli 2017). We recently discussed how the choice of modeling technique depends on many factors such as characteristics of the problem (e.g., a geospatial policy in a heterogeneous population would suggest a quantitative individual approach), data availability, or familiarity of team members with the technique (Voinov et al. 2018). Jun *et al.* also provided a comprehensive comparison of insight provided by different approaches, and their suitability at different stages of a project life cycle (Jun et al. 2011). Our focus is on policy problems that require individual oriented modelling techniques to track individual features, behaviors, or preferences (Badham 2010).

The key difference with aggregate models is that individual models explicitly represent and and account for specific features of each entity whose heterogeneity in state, space, or decision making may affect the model's outcome. For instance, in a small-scale model of grazing, grazers and patches of grass are heterogeneously distributed in space and may be represented explicitly, whereas the rain supporting the growth of grass may be aggregated instead of being related to individual clouds. The positions in this paper hold for discrete models, in which a simulation progresses through discrete units of time, also known as 'ticks'. Specifically, we include three types of quantitative individual models:

1. Cellular Automata (CA) are composed of a set of cells arranged in a regular lattice (e.g., a two-dimensional grid). CA account for spatial heterogeneity and, unless modified in sophisticated versions, generally do not account for heterogeneity in contact patterns (Dabbaghian et al. 2010, Giabbanelli, Babu, and Baniukiewicz 2016).
2. Network models are composed of a set of nodes or vertices, with connections between two nodes known as edges or links. Such models account for heterogeneity in contact patterns but less often in space or decision-making (Giabbanelli et al. 2012, Shoham et al. 2015).
3. Agent Based Models (ABM) are composed of agents, which can interact with the space and with each other. Such models can thus account for spatial heterogeneity, heterogeneity in contact patterns (between agents), and heterogeneity in decision-making (Giabbanelli and Crutzen 2017, Khademi et al. 2018). While artificial societies are usually presented as agent-based models, some *may* actually be network models (Giabbanelli and Crutzen 2013) or cellular automata[1]. Being specific on the type of model allows to more precisely identify which aspects of heterogeneity are represented, and for which requirements in terms of data and computations.

3 IDEAL PRACTICES

We created a list of seven items (numbered *i-vii*) drawn from general simulation methods (i.e., time frame, sensitivity and uncertainty analyses, replicability), aspects to represent specifically in artificial societies (i.e., social interactions and heterogeneity), and the role of data for high-fidelity artificial societies (i.e., multiple datasets and individual level datasets). Since validation (i.e., ensuring that the model is sufficiently accurate for its purpose) and calibration (i.e., adjusting the model parameters within margins of certainty) are a well scrutinized condition for peer-reviewed modeling studies, we did not include them in our list of *ideal* practices. We now briefly define each of the items, how they should *ideally* be done, and why they matter.

3.1 Ideal practices for simulation in general

i. The dynamic models included here represent **time**, not abstract udpate steps. The *time frame* is the total length of simulated time, and the *time granularity* or 'temporal resolution' is the length of time simulated by one 'tick' or update in the model. Ideally, authors should state the time frame and granularity, and justify both given the model's context and/or statistical methods. In particular, policy-oriented models should have a time frame in line with the implementation and evaluation of policy performance in the real-world. For example, in British Columbia, general elections are held every 4 years; consequently, a model aiming at informing policymakers may need to provide results within 4 years rather than within decades (Verigin, Giabbanelli, and Davidsen 2016). Second, statistical methods can identify the appropriate length of time to reach a stable output (i.e., "selecting the run-length") (White Jr, Cobb, and Spratt 2000). Similarly, we can assess when a change in granularity starts to significantly affect model outputs. Without justifying the time frame and time granularity, erroneous inferences may be made regarding the potential long-term benefits of an intervention, results may incorrectly portray a transient output (useful as meta-stable state) as final result, or the model may not adequately support the decision making activities of its intended users.

---

[1]Some ABM *can* a generalization of cellular automata (adding heterogeneous decision-making and spatial mobility)and network models (by adding the environment to a social network). However, ingredients such as social networks are not a *necessary* part of an ABM, thus an ABM is not *always* a generalization of the other two types of model.

***ii.* Sensitivity and Uncertainty Analyses** are necessary to build confidence in conclusions obtained from a model. Sensitivity analysis usually refers to *parametric sensitivity*, which assesses how sensitive (e.g. in terms of variance) the outputs are to changes in the inputs (Richiardi et al. 2006, Saltelli et al. 2008). Due to the potential for interactions between parameters, an ideal approach for parametric analysis would explore the entire parameter space, for instance using a factorial analysis to account for second-, third-, and higher-order interactions between parameters (Lavin and Giabbanelli 2017). Intuitively, it creates an 'ablation testing' as parts of the model are selectively turned off (e.g., agent behaviors) to evaluate their joint effects on outcomes. Sensitivity analysis can also be broadened to include an "algorithmic sensitivity" that checks how sensitive the results are to algorithmic choices, such as synchronous versus asynchronous updates. *Uncertainty analysis* should include a comprehensive assessment of the possible sources of uncertainty (including structural, parameter, stochastic, and data uncertainty), their impact on a model, and the means by which uncertainty was addressed in the model (Briggs et al. 2012, Saltelli et al. 2008, Caro, Briggs, Siebert, and Kuntz 2012). As models are often used as authoritative evidence, overconfidence in their results is a real and common problem (Robinson 2014). It is thus important to convey to possible users such as policymakers what affects a policy's outcome and the extent to which we can trust the simulated outcomes. While the primary purpose of uncertainty analysis is to help determine the overall level of confidence one can have in the results, it can also be used to determine the importance of collecting additional information when a decision is being made (Briggs et al. 2012) and very much helps in model calibration. Uncertainty analysis thus unlocks additional benefits for developing artificial societies, such as guiding data collection efforts. Performing sensitivity analysis has numerous benefits, including ensuring the robustness of the model (or identifying inputs that are not robust), increased understanding of the model structure, and finding potential errors in the relationship between input and output variables (Pannell 1997, Lee et al. 2015).

iii. Being able to achieve the same results as claimed in a previous study ranges from repeatability (confirming the methods via the same team and experimental setup) to replicability (confirming the results via a different team and the same experimental setup) and reproducibility (confirming the conclusions via a different team and experimental setup) (Association for Computing Machinery 2018). We focus on **replicability**, as another team of modelers should be able to build the model from the authors' descriptions, and reach the same conclusion using the same data. Ideally, a model should thus be (i) fully documented so it can be re-created (e.g., using formal definitions, pseudo-code, equations, or publicly accessible source code) and (ii) data necessary to run the model should be publicly accessible (Richiardi et al. 2006). Researchers have warned of a looming replication crisis in many different fields, with many scientists being unable to reproduce the results of other research, and often even their own (Baker 2016). In modeling and simulation, efforts should be made to achieve as much technical transparency as possible while still respecting intellectual property rights (West 2016). For artificial societies specifically, there is a 'continuum' of *access* to a model. A model's source code may be accessible on a third-party repository (e.g., www.comses.net or www.osf.io) in a commonly used language with supporting documentation. It may also be hosted on one's website, which may not be available in the future. Similarly, there is a continuum in description. We may have ad-hoc/incomplete descriptions or standard protocols, such as the ODD protocol or its extensions (Grimm et al. 2010). Finally, there is a continuum of data access. We may need to ask the authors, which offers no guarantee and may extensive administrative procedures (e.g., internal review board), or have access to a *generator* to produce synthetic data deemed sufficiently realistic, or access the dataset.

3.2 Ideal practices to represent key aspects of artificial societies

***iv.* Social interactions** are comprised of the connections between entities (i.e., the social relations) and the ways in which entities influence each other through these connections (i.e., the social function). Social influence is an integral component of an artificial society, as exemplified in artificial societies for health studies (Hill 2009, Ball and Crawford 2010). Ideally, an artificial society should include explicit interactions between pairs of entities, using a social network with sufficiently accurate structure and function, in addition to representing mediating social norms that are expressed system-wide or in specific sub-groups. There are

multiple consequences when these conditions are not met. Using a grid as a social network may lead to different simulation outcomes than when using real-world or stylized (e.g., small-world/scale-free) networks. Using the right social network structure but ignoring that social influences are mediated by individual contexts may lead to erroneous estimates of social effects (Giabbanelli and Crutzen 2013), as individual contexts may act as a barrier against social influences or amplify the effects triggered by social influences.

v. **Heterogeneity** in states and decision-making is one of the hallmarks of complexity in public health (Finegood 2011, Drasic and Giabbanelli 2015). Ideally, model parameters should account for relevant sources[2] of heterogeneity (e.g., by drawing from a distribution fitted to data rather than using a single value), but should not exaggerate this heterogeneity by ignoring dependencies between parameters (e.g., correlations between parameters may prevent a model from initializing values independently). Assumptions of homogeneity such as 'one size fits all' interventions have long been met with criticism from practitioners in several fields. For instance, they observed that success only occurs when a person happens to meet the average profile assumed for the intervention (Epstein et al. 1998, Harkaway 2000). Consequently, a model must appropriately model heterogeneity in the population, as it pertains to the objective of the model (Weinstein et al. 2003).

3.3 Ideal practices to ground artificial societies in data

vi. A model may integrate several datasets for several aspects, such as geographical data for the environment, social network data for social influences, and census data for population-level features. Ideally, the datasets that most impact a model's outcome should come from **different sources**. This is different from using one data source with multiple time points (i.e. a longitudinal dataset). While a single comprehensive dataset may meet the needs of a model, there are several reasons to use different sources. First, it is important for *consistency*, defined by the Bradford Hill criteria as the ability to observe consistent findings *with different samples*. For example, we should know whether the conclusion of a model holds for different population samples or are highly specific to (possibly unknown) features of the population used in a specific sample. Second, taking data from multiple yet similar data sources can aid in mitigating the bias from any of the individual data sources (Ioannidis 2016). Thus making use of different compatible data sources will increase the generalizability of a model as well as lessen the bias that occurs within individual data sources.

vii. Separately from the matter of accessing data or whether the differences between entities are modeled (i.e., treatment of heterogeneity), we also need appropriate **individual level data** to build artificial societies. Virtual individuals (i.e., agents, cells, or nodes) have attributes that can distinguish them from each other (Nianogo and Onyebuchi 2015, Kavak et al. 2018). Ideally, entities in a model would be initialized from individual-level datasets, in which one row of the dataset corresponds to the parameter values for one simulated entity. Several problems occur when using individual-level models without individual-level data. First, a distribution would have to be assumed for each attribute, which may be erroneous, for example when assuming normally distributed values while they may not be. Second, even when the right distributions are found for each attribute in isolation, we note that attributes can be highly correlated (Giabbanelli, Jackson, and Finegood 2014). For example, an agent's food intake level is closely related to the agent's level of exercise. A study by KD Hall *et al.* suggested that a *small* average daily difference between the two would be enough to explain the observed average weight gain in the population (Hall et al. 2011). By ignoring this dependency, a model could create a population whose dynamics at baseline are already erroneous[3]. In sum, a model should be able to initialize its entities from sufficient data to match the level of diversity required in its intended use.

---

[2]Spatial granularity is related to heterogeneity. If the cells representing the environment are too big, then there may be no environmental heterogeneity to worry about, and thus less potential for heterogeneity in how individuals interact with the environment.

[3]Models *could* use multivariate distributions to capture correlated distributions (Kuhl et al. 2010). Since each distribution can have a very different form (e.g., power-law for income, normal for height), few multivariate distributions are flexible enough for this purpose (e.g., the Johnson system (DeBrota et al. 1988)) and they can be difficult to fit to the data (Flynn 2006). This difficulty may explain why models either capture correlations via individual-level data or resort to including very few (if any) correlations.

4   A SAMPLE OF BEST PRACTICES FOR ARTIFICIAL SOCIETIES IN PUBLIC HEALTH

Current studies may not accomplish all seven *ideal* tasks detailed in the previous section. To estimate the gap between current and ideal practices, we analyzed the *best* practices in a random sample of 11 journal articles in which artificial societies were developed for public health on obesity between 2013 and 2016. To limit the potential for bias, we selected studies in which none of the authors of this paper were involved.

Justifications for the **time frame** were provided in all but two studies (Mooney and El-Sayed 2016, Zhang et al. 2015b). Justifications varied between the studies. Three studies justified their time frame by reaching a steady state (Blok et al. 2015, Hammond and Ornstein 2014, Trogdon and Allaire 2014). Four studies considered that "social influences take time to converge" and settled for 3 years (Li et al. 2016), 4 years (Widener et al. 2013), the lifetime of a cohort (El-Sayed et al. 2013), or as long as three generations (Orr et al. 2014). Two studies used a time frame that matches their data, such that the baseline can be initialized from the data and predictions can be validated against real-world data (Wang et al. 2014, Zhang et al. 2015a). The **time granularity** received much less attention: there was no clear justification in most studies (Hammond and Ornstein 2014, Mooney and El-Sayed 2016, Orr et al. 2014, Trogdon and Allaire 2014, Zhang et al. 2015a, Zhang et al. 2015b, El-Sayed et al. 2013), and only an implicit link with policy objectives in other studies, such as the need to simulate days because public nutrition policies typically report servings per day.

The use of **sensitivity analysis** was almost equally divided between not doing any, for five studies (Blok et al. 2015, Widener et al. 2013, Wang et al. 2014, Zhang et al. 2015a, Zhang et al. 2015b), and performing a sensitivity analysis of only one factor at a time in the other six studies. We note that only one study went beyond single factor sensitivity analysis, by also including a factorial design that accounted for interactions between parameters (Orr et al. 2014). We found no assessment of **uncertainty** in the articles.

Replicability was particularly problematic across all studies. Model description consisted of narratives in all studies but one, which also provided a standardized description using the ODD protocol (Wang et al. 2014). No study provided algorithmic descriptions either via flowcharts or pseudocode. Studies provided neither the code of their model, nor data on which to run it. While data may often be obtained from national surveys (e.g., AddHealth, ECLS-K, NHANES, BRFSS, NLSY97), results may still not be replicable because no authors provided their scripts for data cleaning, or determined how many times their experiments had to be repeated. Studies either did not appear to run their stochastic models several times to compute average results (Hammond and Ornstein 2014, Li et al. 2016, Widener et al. 2013, Wang et al. 2014), or used ad-hoc numbers of repeated runs ranging from 20 (Orr et al. 2014) to 1000 (Mooney and El-Sayed 2016).

Social interactions between pairs of agents occurred in all but three studies (Blok et al. 2015, Widener et al. 2013, Wang et al. 2014). In several cases, interactions were highly abstract, such as being more likely to be obese simply by connecting with obese agents (El-Sayed et al. 2013). **Heterogeneity** of agent attributes was present in all but two studies (Wang et al. 2014, Mooney and El-Sayed 2016) but it was often limited to randomly assigning values from probability distributions (Blok et al. 2015, El-Sayed et al. 2013, Hammond and Ornstein 2014, Trogdon and Allaire 2014). Most notably, we found that heterogeneity was only addressed by varying the attributes of agents but not their behaviors: all agents followed the same rules.

A single study used **different sources** by assessing whether parameter estimates for the model depended on the site for data collection (Zhang et al. 2015b). Only two studies used **individual level data** to match agents' attributes with respondents' characteristics (Wang et al. 2014, Zhang et al. 2015a).

5   EMERGING PRACTICES

The results reported in the previous section show that some items are often performed (e.g., justification of the time frame) but others are relatively rare (e.g., replicability, use of data). This gap between ideal and current practices may change in the future. On the one hand, efforts are ongoing to improve several items such as replicability. On the other hand, emerging practices can widen the gap if we do not plan guidance

and assessment efforts accordingly. To envision future needs in developing artificial societies, this section focuses on two emerging opportunities: the development of integrative models and, related, advances in the computational machinery to build simulations (Gilbert et al. 2018).

5.1 Complexity and Uncertainty in Artificial Societies

Societies contain ontologically diverse systems. For instance, an infectious disease spreading between individuals in a society is a fundamentally different type of system compared to the institutions and healthcare systems making policy-decisions to respond to this spread – and this implies different approaches and affordances of knowledge-claims. Various categorizations have been suggested to capture these ontological differences. While terminologies differ, these categorizations tend to separate between *simple*, *complicated*, and *complex* systems, the latter category often separated between *restricted complex* or *general complex* (Andersson and Törnberg 2018, Morin 2007). Simple systems, like pendula, tend to be analytically solvable, and possible to predict with high precision. Complicated systems, like the machinery of a car, tend to have sophisticated components, often fulfilling separate functions, that interact in simple and structured way. Such systems respond well to top-down approaches (e.g. taking them apart and studying their components) and can be accurately predicted, generally with numerical solutions. Restricted complex systems, like traffic jam, tend to consist of many components displaying simple behaviors, but the interaction between components is strong and unstructured, often leading to unexpected macro dynamics. These systems require a bottom-up approach, i.e. individual-oriented modelling (Badham 2010), and are *uncertain*, in the sense that their future states are highly sensitive to initial conditions, which implies that knowledge-claims are reduced to predicting their dynamics and identifying their control parameters (Cvitanovic et al. 2005). Finally, *general complex* systems, like the long-term development of the automobile industry, have both sophisticated components and unstructured interactions between them, and tend to be *ontologically uncertain*: it is uncertain even what components or interactions they will contain at a future time (Lane and Maxfield 2005).

These differences in system types means that, as we have already suggested, the appropriate models also differ in important ways. First, the knowledge claims should be a function of the certainty of the systems. For instance, we should not aim for perfect prediction of future states in complex systems, as the model would then have a higher level of certainty than the system it represents. Secondly, a model should belong to the same ontological category as the system under study: a top-down, complicated model of a traffic jam is unlikely to capture its emergent dynamics. However, general complex systems are a special case, as their modeling implies the constant risk that the model will become as unmanageable as the system itself, without affording any type of prediction. This deep uncertainty leads to valuing a qualitative understanding of such systems. This may be achieved for instance by taking an integrative approach in which each model attempts to capture some aspects of the system (Andersson et al. 2014), and are tied together (e.g., through a narrative approach). All of which takes us to the issue of integrated and integral models.

5.2 Integrated and integral models

Numerous legacy models have been developed by specialists in their scientific fields and tested in many applications. It makes sense to reuse these assets as "building blocks" when designing more sophisticated, potentially transdisciplinary models. However, linking pre-existing models together, matching the variables, scales and resolutions, may be challenging. Modelers may use different paradigms, assumptions, spatial and temporal representations, which need to be carefully analyzed and adjusted to avoid creating 'integronsters' – constructs that are valid software products but useless as models (Belete, Voinov, and Laniak 2017).

Currently there are several efforts to develop the standards and software tools that provide for model coupling, such as the open modelling interface and environment (OpenMI) (Moore and Tindall 2005). However, no tool has succeeded so far in connecting artificial societies, and hardly any of them have been applied to

model social processes and dynamics. Models built for social sciences largely remain integral, that is developed and maintained by the same team as a more-or-less unified integral whole. The modeling systems that are used, even when modular by nature, are not designed to integrate legacy models as is, but rather serve as tools for future model development, offering input, output and project support tools and common library standards, into which different modules are invited to be translated. Again, as usually with integral modeling, new code is tested and reimplemented before it is included into the component libraries, and there is usually the same team of modelers involved in doing this for a particular model development. Even when the same framework is used to develop different models, there are rarely modules from previous models by other teams taken and incorporated into new models.

6 DISCUSSION

Artificial societies can play an important role to guide and evaluate policies. Several articles offer guidance in quality assessment for artificial societies, but this guidance may be disconnected from current practices. We described how seven items should be done, and why they matter. This list is not exhaustive: for example, social interactions may be shaped not only by social networks (as we examined) but also by how the (geo)space is represented. Given that standardization efforts are ongoing, our list may be amended through discussions in workshops with simulation experts (Collins et al. 2015) and/or by examining the results when these items are applied on the literature.

Evaluating a sample of studies with respect to our seven items revealed a gap between current and ideal practices. We found that the time granularity was scarcely justified while uncertainty was generally unaddressed and replicability is challenging across all studies. In addition, most studies relied on single sources and their findings do not easily generalize. Ongoing initiatives may address some of these shortcomings, such as the growing focus on replicability in the modelling and simulation community. Nevertheless, several challenges remain. If a society is, by definition, a highly complex system, what is the necessary level of complexity needed to accurately (or at least sufficiently) model a particular topic (Castellani et al. 2016)? Given that not all complex systems are the same (Byrne and Callaghan 2013, Andersson et al. 2014, Morin 2007), how do these differences translate to practices in building artificial societies? As key societal drivers (e.g., built environment, social norms on eating behavior) are interdependent (thus creating *nexus issues* in policy analysis), how do we integrate models of complex systems? We investigated these questions by detailing the challenge of uncertainty, complexity, and integration.

There are not only challenges, but also emerging opportunities. A particularly cutting-edge and promising opportunity is the mixed-methods work being done to integrate quantitative individual modeling (which is our focus in this paper) with other forms of computational modeling, statistical analysis, qualitative inquiry and narrative analyses. The map of the complexity sciences (http://www.art-sciencefactory.com/complexity-map_feb09.html) exemplifies this integration. In public health, examples include integrating agent-based modeling with case-based methods or digital sociological methods such as social/media and textual analysis. Other areas include linking micro-simulation with differential equation modeling, genetic algorithms, and machine learning (Byrne and Callaghan 2013, Gilbert et al. 2018, Castellani et al. 2016).

We also see promising advances in the computational machinery used to build simulations (Gilbert et al. 2018). This includes the development of the cloud, big data, web science and increasingly interconnected open-access statistical and computational environments (e.g., the connection between the statistical R project and the widely used NetLogo agent-based modeling toolkit). In turn, the emergence of mixed quantitative and computational modeling has led to an up-tick in the number of conferences devoted to modeling complexity from a mixed-methods perspective (as in the case of the ESRC methods conference) and a new openness on the part of public health journals and book editors to accept, review, publish and promote such highly novel methodological advances – albeit slowly and still with some resistance. Nonetheless, emergent changes, relative to the challenges of complexity, are happening, as we can no longer, as a global society, escape or ignore the complexities of the public health challenges we presently face.

7 CONCLUSION

Despite many articles on best practices, the development of artificial societies that can be used to guide and evaluate policies is limited by challenges of replicability, integration, complexity and uncertainty. This article has taken a step toward dealing with these challenges, by reviewing and systematizing best practices as well as accounting for emerging opportunities in mixed-methods and novel computational machinery. Such opportunities and systematization of best practices will contribute to increasing efficiency, validity, and accuracy in our modeling efforts.

REFERENCES

Andersson, C. et al. 2014. "Societal systems–complex or worse?". *Futures* vol. 63, pp. 145–157.

Andersson, C., and P. Törnberg. 2018. "Wickedness and the anatomy of complexity". *Futures* vol. 95, pp. 118–138.

Association for Computing Machinery 2018. "Artifact Review and Badging".

J Badham 2010. "A compendium of modelling techniques". http://i2s.anu.edu.au/sites/default/files/integration-insights/integration-insight_12.pdf.

Baker, M. 2016. "Is There a Reproducibility Crisis? Nature Survey Lifts the Lid on how Researchers View the 'crisis' rocking science and what they think will help". *Nature* vol. 533, pp. 452–454.

Ball, K., and D. Crawford. 2010. "The role of socio-cultural factors in the obesity epidemic". In *Obesity epidemiology: from aetiology to public health*, edited by D. Crawford, R. Jeffery, K. Ball, and J. Brug, pp. 105–118. Oxford University Press.

Belete, G. F., A. Voinov, and G. F. Laniak. 2017. "An overview of the model integration process: From pre-integration assessment to testing". *Environmental modelling & software* vol. 87, pp. 49–63.

Bennett, N. D. et al. 2013. "Characterising performance of environmental models". *Environmental Modelling & Software* vol. 40, pp. 1–20.

Blok, D. J. et al. 2015. "Reducing income inequalities in food consumption: explorations with an agent-based model". *American journal of preventive medicine* vol. 49 (4), pp. 605–613.

Briggs, A. H. et al. 2012. "Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6". *Value in Health* vol. 15 (6), pp. 835–842.

Byrne, D., and G. Callaghan. 2013. *Complexity theory and the social sciences: The state of the art*. Routledge.

Caro, J. J., A. H. Briggs, U. Siebert, and K. M. Kuntz. 2012. "Modeling good research practices – Overview a report of the ISPOR-SMDM modeling good research practices task force-1". *Medical Decision Making* vol. 32 (5), pp. 667–677.

Castellani, B. et al. 2016. "Cases, clusters, densities: Modeling the nonlinear dynamics of complex health trajectories". *Complexity* vol. 21 (S1), pp. 160–180.

Collins, A. et al. 2015. "A call to arms: Standards for agent-based modeling and simulation". *JASSS* vol. 18 (3), pp. 12.

Craig, P. et al. 2008. "Developing and evaluating complex interventions: the new Medical Research Council guidance". *BMJ* vol. 337, pp. a1655.

Cvitanovic, P. et al. 2005. "Chaos: classical and quantum". *ChaosBook.org (Niels Bohr Institute, Copenhagen 2005)* vol. 69.

Dabbaghian, V. et al. 2010. "A cellular automata model on residential migration in response to neighborhood social dynamics". *Mathematical and Computer Modelling* vol. 52 (9-10), pp. 1752–1762.

DeBrota, David J. and others 1988. "Input Modeling with the Johnson System of Distributions".

Drasic, L., and P. Giabbanelli. 2015. "Exploring the interactions between physical well-being and obesity". *Can J Diabetes* vol. 39, pp. S12–S13.

El-Sayed, A. M. et al. 2013. "Are network-based interventions a useful antiobesity strategy? An application of simulation models for causal inference in epidemiology". *American journal of epidemiology*, pp. kws455.

Epstein, J. 2008. "Why Model?". *JASSS* vol. 11 (4).

Epstein, L. et al. 1998. "Treatment of pediatric obesity". *Pediatrics* vol. 101 (3), pp. 554–570.

DT Finegood 2011. "The complex systems science of obesity".

Flynn, M. 2006. "Fitting human exposure data with the Johnson S(B) distribution". *J Expo Sci Environ Epidemiol* vol. 16 (1), pp. 56–62.

Giabbanelli, P. et al. 2012. "Modeling the influence of social networks and environment on energy balance and obesity". *Journal of Computational Science* vol. 3, pp. 17–27.

Giabbanelli, P., and R. Crutzen. 2013. "An agent-based social network model of binge drinking among Dutch adults". *JASSS* vol. 16 (2), pp. 10.

PJ Giabbanelli and P Jackson and DT Finegood 2014. "Modelling the joint effect of social determinants and peers on obesity among Canadian adults".

Giabbanelli, P. J. 2018. *Analyzing the Complexity of Behavioural Factors Influencing Weight in Adults*, pp. 163–181. Springer.

Giabbanelli, Philippe J. and Babu, Guru Jagadeesh and Baniukiewicz, Magda 2016. "A Novel Visualization Environment to Support Modelers in Analyzing Data Generated by Cellular Automata".

Giabbanelli, P. J., and M. Baniukiewicz. 2018. *Navigating Complex Systems for Policymaking Using Simple Software Tools*, pp. 21–40. Springer.

Giabbanelli, P. J., and R. Crutzen. 2017. "Using Agent-Based Models to Develop Public Policy about Food Behaviours: Future Directions and Recommendations". *Computational and mathematical methods in medicine* vol. 2017.

Gilbert, G. et al. 2018. "Computational modelling of public policy: Reflections on practice". *JASSS* vol. 21 (1), pp. 1–14.

Grimm, V. et al. 2010. "The ODD protocol: a review and first update". *Ecological Modelling* vol. 221, pp. 2760–2768.

Hall, K. et al. 2011. "Quantification of the effect of energy imbalance on bodyweight". *Lancet* vol. 378 (9793).

RA Hammond 2015. "Appendix A: Considerations and Best Practices in Agent-Based Modeling to Inform Policy".

Hammond, R. A., and J. T. Ornstein. 2014. "A model of social influence on body mass index". *Annals of the New York Academy of Sciences* vol. 1331, pp. 34–42.

Harkaway, J. 2000. "Obesity and systems research: the complexity of studying complexities". *Familities, Systems & Health* vol. 18 (1), pp. 55–59.

Hill, A. 2009. "Social and psychological factors in obesity". In *Obesity: Science to Practice*, edited by G. Williams and G. Fruhbeck, pp. 347–366. Wiley-Blackwell.

Ioannidis, J. P. 2016. "Anticipating consequences of sharing raw data and code and of awarding badges for sharing". *Journal of clinical epidemiology* vol. 70, pp. 258–260.

Jun, G. T. et al. 2011. "Development of modelling method selection tool for health services management: From problem structuring methods to modelling and simulation methods". *BMC health services research* vol. 11 (1), pp. 108.

Kavak, Hamdi and others 2018. "Big data, agents, and machine learning: towards a data-driven agent-based modeling approach".

Khademi, A. et al. 2018. *An Agent-Based Model of Healthy Eating with Applications to Hypertension*, pp. 43–58. Springer.

Kuhl, M. E. et al. 2010. "Univariate input models for stochastic simulation". *Journal of Simulation* vol. 4 (2), pp. 81–97.

Lane, D. A., and R. R. Maxfield. 2005. "Ontological uncertainty and innovation". *Journal of evolutionary economics* vol. 15 (1), pp. 3–50.

Lavin, Eric A and others 2018. "Should we simulate mental models to assess whether they agree?".

Lavin, Eric A and Giabbanelli, Philippe J 2017. "Analyzing and simplifying model uncertainty in fuzzy cognitive maps".

Lee, J.-S. et al. 2015. "The complexities of agent-based modeling output analysis". *JASSS* vol. 18 (4), pp. 4.

Li, Y. et al. 2016. "Social norms and the consumption of fruits and vegetables across New York City neighborhoods". *Journal of Urban Health* vol. 93 (2), pp. 244–255.

Mooney, S. J., and A. M. El-Sayed. 2016. "Stigma and the etiology of depression among the obese: an agent-based exploration". *Social science & medicine* vol. 148, pp. 1–7.

Moore, R. V., and C. I. Tindall. 2005. "An overview of the open modelling interface and environment (the OpenMI)". *Environmental Science & Policy* vol. 8 (3), pp. 279–286.

Morin, E. 2007. "Restricted complexity, general complexity". In *Worldviews, science and us: Philosophy and complexity*, pp. 5–29. World Scientific.

Nianogo, R., and A. Onyebuchi. 2015. "Agent-based modeling of noncommunicable diseases: a systematic review". *American Journal of Public Health* vol. 105 (3), pp. e20–e31.

Orr, M. G. et al. 2014. "Reducing racial disparities in obesity: simulating the effects of improved education and social network influence on diet behavior". *Annals of epidemiology* vol. 24 (8), pp. 563–569.

Pannell, D. J. 1997. "Sensitivity analysis of normative economic models: theoretical framework and practical strategies". *Agricultural economics* vol. 16 (2), pp. 139–152.

Richiardi, M. G. et al. 2006. "A common protocol for agent-based social simulation". *JASSS* vol. 9.

Robinson, S. 2014. *Simulation: the practice of model development and use*. Palgrave Macmillan.

Saltelli, A. et al. 2008. *Global sensitivity analysis: the primer*. John Wiley & Sons.

R. G. Sargent 2009. "Verification and validation of simulation models".

Shoham, D. et al. 2015. "Modeling social norms and social influence in obesity". *Current epidemiological reports* vol. 2, pp. 71–79.

Silverman, B. et al. 2015. "A systems approach to healthcare: agent-based modeling community mental health, and population well-being". *Artificial Intelligence in Medicine* vol. 63, pp. 61–71.

Treasury, HM 2015. "The Aqua Book: guidance on producing quality analysis for government".

Trogdon, J. G., and B. T. Allaire. 2014. "The effect of friend selection on social influences in obesity". *Economics & Human Biology* vol. 15, pp. 153–164.

Verigin, Tanner and Giabbanelli, Philippe J. and Davidsen, Pål I. 2016. "Supporting a Systems Approach to Healthy Weight Interventions in British Columbia by Modeling Weight and Well-being".

Voinov, A. et al. 2018. "Tools and methods in participatory modeling: Selecting the right tool for the job". *Environmental Modelling & Software* vol. 109, pp. 232–255.

Wang, Y. et al. 2014. "Examining social norm impacts on obesity and eating behaviors among US school children based on agent-based model". *BMC public health* vol. 14 (1), pp. 1.

Weinstein, M. C. et al. 2003. "Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices Modeling Studies". *Value in health* vol. 6 (1), pp. 9–17.

West, R. 2016. "Data and statistical commands should be routinely disclosed in order to promote greater transparency and accountability in clinical and behavioral research". *Journal of clinical epidemiology* vol. 70, pp. 254.

White Jr, K Preston and Cobb, Michael J and Spratt, Stephen C 2000. "A comparison of five steady-state truncation heuristics for simulation".

Widener, M. J. et al. 2013. "Agent-based modeling of policies to improve urban food access for low-income populations". *Applied Geography* vol. 40, pp. 1–10.

Zhang, J. et al. 2015a. "Leveraging social influence to address overweight and obesity using agent-based models: the role of adolescent social networks". *Social Science & Medicine* vol. 125, pp. 203–213.

Zhang, J. et al. 2015b. "Network interventions on physical activity in an afterschool program: An agent-based social network study". *American journal of public health* vol. 105 (S2), pp. S236–S243.

AUTHOR BIOGRAPHIES

PHILIPPE J. GIABBANELLI is an Assistant Professor in the Computer Science Department at Furman University (USA). He received his Ph.D. from Simon Fraser University (Canada). His research interests include machine learning and simulation models. His email address is giabba@furman.edu.

ALEXEY A. VOINOV is a Distinguished Professor of Spatio-Temporal Systems Modeling for Sustainability Science at University of Technology Sydney (Australia). He received his Ph.D. from Moscow State University (Russia). His research interests include integrative models. His email address is aavoinov@gmail.com.

BRIAN CASTELLANI is a full Professor of Sociology at Durham University (United Kingdom). He received his Ph.D. from Kent State University (USA). His research interests include computational modeling and complexity science. His email address is brian.c.castellani@durham.ac.uk.

PETTER TÖRNBERG is a researcher in sociology at the University of Amsterdam (Netherlands). He received his Ph.D. from Chalmers University of Technology (Sweden). His research interests include complex systems and interpretative computational methods. His email address is petter@toernberg.com.