Annals of Tourism Research Empirical Insights 1 (2020) 100005

Contents lists available at ScienceDirect

Annals of Tourism Research Empirical Insights

journal homepage: www.elsevier.com/locate/annale

Computational modelling and simulations in tourism: A primer

Rodolfo Baggio [a,b]

[a] Master in Economics and Tourism and Dondena Center for Research on Social Dynamics and Public Policy, Bocconi University, via Röntgen, 1, 20136 Milan, Italy
[b] National Research Tomsk Polytechnic University, 30 Lenin Avenue, Tomsk 634050, Russia

ARTICLE INFO

Article history:
Received 22 September 2020
Received in revised form 26 November 2020
Accepted 30 November 2020
Available online xxxx

Keywords:
Complex systems
Modelling
Simulations
Numerical and computational methods

ABSTRACT

The aim of this contribution is to briefly sketch and discuss the main issues that concern the activities of modelling and simulating complex phenomena and systems. The focus is on numerical and computational techniques. We discuss the validity of these methods and examine the different steps to be taken for ensuring a correct, accurate and reliable implementation. The approach is essentially of general methodological nature, regardless of specific techniques or tools.

1. Introductions

For the last 400 years, roughly from the age of Galileo, Descartes, Bacon, Newton and the like, making science (or solving a problem) has meant systematically building and organizing knowledge, shaped as a set of verifiable explanations and predictions about some domain of interest. Whether *hard* (physics, biology, astronomy etc.) or *soft* (psychology, sociology, political science etc.), most disciplines have adopted, for fulfilling those goals, a series of methods that mainly rely on collecting some kind of empirical observation and analyze them with an ever increasing array of quantitative and qualitative techniques of increasing sophistication.

This approach has worked quite well until we realized that the fundamental task, the empirical observation, does not always work in a satisfactory way or at all. In many situations, in fact, observations are impossible, or very expensive in terms of time and resources and, what is more, the conditions of the *experiment* cannot be changed, thus leaving to some form of speculation the inference of what could happen if some of the variables or parameters of the trial would be altered. This is, as we all know, the case of the multitude of social, political and economic systems and phenomena that interest us, tourism included, obviously.

A classical approach to a problem (scientific or practical) unfolds over a number of well-known steps: statement of the objective(s), identification of the best and more suitable methods for the analysis, collection of the data needed, analysis (descriptive and inferential), presentation of the results. This, that many recognize as, and still call, although not completely correctly (see e.g. Andersen & Hepburn, 2016), the *scientific method* can, however, be changed and adapted when some of these steps are not feasible. The change consists of modifying the empirical aspects of the method and exploit the capabilities made available by the modern technological tools for simulating all the conditions, aspects and configurations we might need or want to explore. We resort thus to employ abstract models on which numerical simulations can be performed.

The famous aphorism attributed to the statistician George Box: "Essentially, all models are wrong, but some are useful" seems to be a quasi-standard introduction to works on modelling and simulations. Although looking like a joke, the statement holds an essence of truth: modelling is more an art than a science. And, what is more, is an art of which we are just starting to see the full potential, due to the quite recent development in the technological instruments (hardware, software) and the techniques (algorithms) and their wider availability.

In the tourism and hospitality domain that of numerical models and simulations is still at an early stage of development, but the works produced so far show interesting and insightful outcomes (see e.g. Johnson et al., 2017). As Galileo would have said (from his letter to Orso d'Elci dated 25 December 1617, (Galilei, 1832: 133–134): "…this is a complete art, although just born, based on principles and means that are new, but noble and commendable, and needs to be embraced, cultivated and promoted, so that with exercise and time it will be possible to benefit from the fruits of which it has in itself the seeds and the roots."

Objective of this contribution is to present a brief summary of the main ideas and techniques at the basis of simulation models and to provide some recommendations for ensuring rigor, reliability and validity in the activity and its outcomes. To this extent no discussion will be made here as to specific techniques or algorithms. Rather, general issues will be raised, that can

E-mail address: rodolfo.baggio@unibocconi.it..

http://dx.doi.org/10.1016/j.annale.2020.100005

be applied independently from the peculiar procedures or environments addressed by the reader.

In other words, the paper seeks to provide answers to the main questions that concern the meaning of models and simulations and the main issues to be addressed when choosing a technique and designing and implementing a simulation model.

Moreover, we focus here on computational (numerical) models and simulations. More details and examples of the most relevant and used methods in the tourism and hospitality domain can be found in some recent books such as those of Baggio and Baggio (2020) or Baggio and Klobas (2017).

The paper is organized as follows. After a brief discussion on the validity of the modelling approach for representing real world situations, a description of models and simulations is provided. The following sections contain the discussion of the main issues that concern the design and implementation of a simulation model and some general guidelines for choosing the best technique and publishing the results. The concluding remarks close the paper.

2. A brief epistemological digression

Models are built for representing reality by abstracting what we think are the fundamental features of a real system for some specific purpose. On top of a model, a simulation is the representation of a process that allows exploring different assumptions, configurations and dynamic evolutions. For many researchers, trained to study phenomena and systems by observing, experimenting and testing a *real* situation, these abstractions might pose big questions on their capability to provide *knowledge*.

The philosophical discussion about modelling and simulations and their representative power of reality has intensified with the diffusion of the use of computerized numerical tools and extends over a range of theoretical, formal, and practical questions across diverse disciplines. The main question concerns the extent to which models and simulations represent their target systems or phenomena, whether they teach us about the nature of reality or can only represent some selected aspects of the world we delve into, and to what extent they are a valid substitute of the conventional empirical methods for understanding, describing or predicting configurations and behaviors.

The simple fact that a formal and commonly accepted definition of the terms is still to be shaped is a good indicator of the wideness of the debate. What we have is only a certain agreement on the view of models as functional entities, that allows to explore how different functional perspectives provide indications about various conceptions of their ontology, that is how concepts relate to reality (Gelfert, 2017). Moreover, when simulations are at play, philosophy of science has generated a substantial debate about their compliance with traditional philosophical descriptions of the scientific practice, especially for what concerns the role of theory in the design and justification of physical models (Nersessian & MacLeod, 2017).

The representation of reality is one of the fundamental question philosophers have tried to answer for centuries. Since Plato's well-known allegory of the cave (The Republic. Book VII) the reflection on our understanding, and our ability to imagine and describe the world around us has played a crucial role in the history of human thinking. The advent of modern science and its methods has provided further arguments concerning the validity and the reliability of the ways we derive the *laws* that are assumed to govern the multitude of phenomena we see and experience. The basic question is: do the pictures we craft convey what they are intended to represent? and to what extent they do so?

The question is important, since, even if not always explicitly recognized, the way we frame it, and answer, can be quite important in helping researchers to ponder upon the actual modelling practices. As Graebner (2018: 2.5) states, a definite epistemological framework "requires … to be very precise and explicit on how the model is expected to improve understanding of the target system."

A good epistemological framework, such as the one proposed by Graebner, is based on a series of steps (we shall discuss them in the rest of this paper) and put a good emphasis on the verification and validation (this empirical, i.e. based on some sets of observations) of both the models and simulations inputs and outputs. This echoes many other philosophers' positions, such as the one expressed by Wittgenstein his Tractatus Philosophicus (Wittgenstein, 1921: 2.223): "In order to tell whether a picture is true or false we must compare it with reality."

In designing models and simulations we also make extensive use of analogy (Turner, 1955). When a similarity can be established between different phenomena, the functions of elements in different systems or between their configurations, some structural relation can be mirrored (maybe more simply) in a known environment and a model can be designed which can be beneficially used in different settings. The effectiveness of this approach has been proved extensively (Gentner, 2002; Krieger, 2005; Wigner, 1960). What is more, from an epistemological viewpoint, it has even been claimed that theories (or models) not exhibiting a formal analogy to some other existing theory, would provide little means to understand how they could be applied to concrete problems (Nagel, 1961).

A more complete discussion on these questions is outside the scopes of this paper, but the reader is invited to go through (besides the cited paper by Graebner, 2018) the good reviews by Frigg and Hartmann (2020) and the first part of the Springer Handbook of Model-Based Science dedicated to the chief theoretical and philosophical issues of models and simulations (Magnani & Bertolotti, 2017).

In any case, the conclusion is that, once satisfied certain reliability and validity requirements (typically methodological), a model of whatever nature and a computational simulation have the same legitimacy as an empirical observation or an experiment. Thus, a properly designed and validated model and the associated simulations have all the rights to be part of, and enrich, the possible ways to make experiments in the tourism and hospitality domain (see: Viglia & Dolnicar, 2020). More or less as it is happening in mathematics, where the use of computer-generated proofs, although still debated, are being more and more accepted as 'proofs' (see e.g. the report on a panel held at the 2018 International Congress of Mathematicians in Davenport et al., 2018).

3. Models and modelling

Although having many different peculiarities that depend on the specific environments, disciplines and issues tackled, many of the fundamental methodological concepts are common to the general practice of modelling and simulation, so that we could even envisage the presence of a *discipline* per se (Silvert, 2001).

We consider a model to be a concise, workable and predictive representation of the system or phenomenon built to meet a specific goal. And a simulation as an approximate reproduction of the operation of a model that portrays its operation over time.

Many different types of models exist: conceptual, statistical, physical, mathematical, numerical and so on (see e.g. Baggio & Baggio, 2020). Here, as said, we focus on numerical (computerized or computational) models.

To better and more formally characterize a model we can adopt the definition by Marvin Minsky (1968: 426): "To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A". It is important to note then, as Minsky continues: "The model relation is inherently ternary. Any attempt to suppress the role of the intentions of the investigator B leads to circular definitions or to ambiguities about "essential features" and the like. It is understood that B's use of a model entails the use of encodings for input and output, both for A and for A*. If A is the world, questions for A are experiments. A* is a good mode of A, in B's view, to the extent that A*'s answers agree with those of A's, on the whole, with respect to the questions important to B."

In short, we need an observer, an objective of the observation, a certain phenomenon or system and the synthetic representation of this object of study: the model.

Methodologically, the steps to follow are not that different from what most researchers are used to: state the problem and how to measure the outcomes, scan past possible designs and implementations, build a suitable synthetic representation, verify and validate the model by using empirical data (input and/or output) if available, or use some established method for checking the consistency of the model, or use analogy with some similar setting. Finally, the model can be used as basis for a simulation in which different variables and parameters are explored and the resulting output configurations and behaviors analyzed.

Two considerations are important here. The first one concerns the well-known dichotomy between qualitative and quantitative approaches. Modelling activities definitely need both views and their separation is almost impossible. Good quantitative explorations provide a sound basis for the knowledge of the structural and dynamic features of our object of study, while without a deep qualitative expertise of the object and its history is quite difficult to correctly and meaningfully interpret the outcomes (Mariani & Baggio, 2020; Olsen, 2004). In fact, all experts consider a qualitative description (a conceptual model) the essential starting point.

Moreover, solving the dichotomy opens the way to a remarkable possibility to experiment with different techniques. As Gummesson (2007: 226) well observes: "By abolishing the unfortunate categories of qualitative/quantitative and natural sciences/social sciences that have been set against each other, and letting them join forces for a common goal – to learn about life – people open up for methodological creativity, therefore qualitative and quantitative, natural and social are not in conflict but they should be treated in symbiosis".

The second consideration is about the term *synthetic*, previously used several times. The general idea is that a model, to be useful, needs to have a certain level of complication in order to include the relevant features of our object of study but does not need to be excessively complicated resulting otherwise unmanageable and risking to produce statistical or computational artefacts that may hinder its capabilities. A modeler makes extensive use of the Ockham's razor. The implicit epistemological claim is that there are certain properties of the object (phenomenon or system) under investigation that must be incorporated for the model to be functional and helpful, but that an excessive number of features make the model too complicated conceptually and, at times, too heavy computationally, with an increased probability to generate outcomes difficult to interpret or even unreliable.

Models are *purposeful representations*. Systems and phenomena are not modelled per se, but only with respect to a specific objective. The word purposeful is essential, as, obviously, a model cannot be built if we do not know why, and models can be assessed and evaluated only with regard to their purpose (Starfield, Smith, & Bleloch, 1990). It can also be seen as a filter whose function is that of discriminating among the many possible items to be included and decide whether a certain element is of importance for the stated purposes. Thus, for example, a model designed for estimating the travel decisions of a tourist will be very different from a model dealing with the sustainability of tourists' flows in a destination, despite the fact they focus on the same object (a destination). How to attain the right level is, as for many other techniques, a matter of experience and of trial and error.

The foundation for a good computational model is a conceptual representation of the problem at stake. This is typically a qualitative description of the object of study that contains the main elements, their relationships and the general meaning of what the processes involved are.

More formally a conceptual model can be defined as (Robinson, 2008: 283): "… a non-software specific description of the computer simulation model (that will be, is or has been developed), describing the objectives, inputs, outputs, content, assumptions and simplifications of the model."

The objective is to provide a clear and rational guidance for all the subsequent activities, to clarify possible ambiguities and ensure that no different interpretations of the terms and concepts involved 'create confusion' about the functioning or the scope of the model. A conceptual model is explicitly chosen to be completely independent of specific implementation technique (Embley & Thalheim, 2011). It is the 'research question' of a modelling and simulation work.

3.1. Model design

In approaching the operational definition of a model, the best suggestion, even more so when not much experienced, is to start from a simple (even simplistic) realization and proceed to augment (complicate) it step by step always verifying the consistency of what done. This is the well-known KISS (Keep it Simple Stupid) approach (Edmonds, 2017; Sun et al., 2016).

There are several elements to be taken into account. The main concern:

- the representation of the different elements that play a role in the system analyzed along with the features they possess that are relevant for the objective set;
- the type and intensity of their interactions;
- the rules they adopt to make the decisions requested by the setting devised;
- the environment in which they operate and the relationships (influences, effects etc.) with external background;
- the timings defined by the schedules, the frequency or the duration of the actions performed, the synchronism of these actions and their priorities.

For all these a numerical representation will be needed, whether some kind of measure exists or is inferred from some kind of qualitative assessment. Starting *simple* means starting with a very parsimonious choice of the variables and parameters involved. For example, if we are modelling the decision making of tourists' travels, we can start by assigning each tourist a utility, defined as probability to choose a destination based on some relevant characteristics such as the destination attractiveness. Once the model is built and shows to be able to render, at least coarsely, a known (empirically or theoretically) situation the model can be enriched by better formulating the individual utility function with the addition of more variables (preference, past experience, available income, distance etc.).

Obviously, as already said, an exploration of the literature may provide good suggestions when similar problems have been addressed and may provide the basis for the approach called TAPAS, i.e. "Take a Previous Model and Add Something" where a previously published implementation is used as starting point (Frenken, 2006).

A numerical model is rendered through a software program, thus adopting good and effective software development practices is of paramount importance (see e.g. Martin, 2008; McConnell, 2004, or the classical evergreens Brooks, 1995; Knuth, 2011). The most important of them are an extensive documentation of what done and the modularity of the implementation. Splitting the script into small units make the script more manageable and, more importantly, provides more valuable information in case of errors.

3.2. Two examples: Networks and agent-based models

As an example, let us consider networks and agent-based models. They are two classes of models that have emerged in the last years as convincing ways of representing complex phenomena and systems and that allow the researcher to explore the many static (structural) and dynamic features of settings that cannot be described employing traditional methods unless applying heavy simplifications that, however, risk hindering many important features and, anyway, are unable to render some of the important characteristics of a complex system such as its self-organization, its resilience or the spontaneous appearance of unplanned ordered configurations or behaviors (Baggio, 2008; Bertuglia & Vaio, 2005; Johnson, 2009).

Network models have been extensively studied in the last decade. The methods used come from a variety of disciplines with a basis in the mathematical theory of graphs. The basic idea is that when a complex system is considered, the most important elements are its components and the relationships that bind them. These are abstracted as nodes and links of a graph that can be then analyzed with a wealth of quantitative techniques (Baggio, 2018; Barabási, 2016; da Fontoura Costa, Rodrigues, Travieso, & Villas Boas, 2007). The most important details and issues related to the activities needed for deriving a good network model concern the definition of the items to be considered (the nodes), the description of their interactions and how these are formed and characterized (the links), and the collection and verification of the data and the estimate of the completeness of the network obtained (for what concerns tourism and hospitality see e.g. Baggio, 2018; Scott, Baggio, & Cooper, 2008, 2011).

Network science has been (and is) used in a wide range of situations and has helped improving our understanding of a wide array of phenomena and systems (da Fontoura Costa et al., 2011). In tourism it has been applied to the study of the structural characteristics of a destination, for detecting the relevant actors or groups in a tourism system or, also, for identifying and following relevant issues in the research literature (Baggio, 2017; Casanueva, Gallego, & García-Sánchez, 2016; Merinero-Rodríguez & Pulido-Fernández, 2016; van der Zee & Vanneste, 2015).

In addition, a network forms an ideal substrate for the investigation of a number of dynamic processes that are of great interest such as the diffusion of information and knowledge, the synchronization of opinions, or the response to internal or external shocks that may affect single elements or parts of the network (Barrat, Barthélémy, & Vespignani, 2008). Needless to say, once the characteristics of a dynamic process in a specific environment are known, it is possible to modify the basic model in order to optimize the system with respect to some defined goal thus providing possible scenarios to the interested parties, or to improve the position and the functioning of some specific element or group. For example, it is possible to build a network representing the flows of tourists in a small area and analyze it for understanding the main patterns in order to design better services or to manage these flows for avoiding overcrowded areas (Baggio & Scaglione, 2018a, 2018b).

In an agent-based model we start by defining the *agents*, that is the autonomous individual elements that are part of a system or involved in a phenomenon, and the elementary interactions between them. The idea, here, is to describe a system from the perspective of its constituent units and let a suitable software evolve the situation in order to assess their effects on the system as a whole. It has been shown, in a large number of different environments, that this technique is able to provide a natural and effective description of systems and processes and capture emergent phenomena. This with a high flexibility due to the possibility of varying the number and the type of agents involved and their characterizing features (Bonabeau, 2002; Gilbert & Terna, 2000; Macal & North, 2010; Wilensky & Rand, 2015). Not much used in the tourism domain, agent-based modelling has however provided a handful of interesting applications (Amelung et al., 2016; Nicholls, Amelung, & Student, 2017).

Although some ready-to-use software applications exist (mainly for network analysis) the need for precise specifications of elements and interactions or processes definitely requires some sort of programming capabilities for the researcher. For more details, worked examples and a selection of useful software development environments the interested reader can check the books by Barabási (2016), Wilensky and Rand (2015) and Baggio and Baggio (2020).

4. Simulations

Determining the structural (and so static) characteristics of a system is an important matter. However, most of the systems and phenomena we deal with are of dynamical nature, so a static representation is not that useful. And even in a static case, we might want to explore different possible configurations, for example obtained by changing some of the initial conditions or of the main parameters governing the system. In an ideal case we should write a series of relations, in the form of analytic expressions and solve them. Unfortunately, the cases in which this is possible are very few. An analytical relationship has usually the form of a differential equation or of a system of differential equations and the solutions, even if we approximate and simplify the equations, is one of the less common events in mathematics (Hale & LaSalle, 1963). This is especially true when, as happens in tourism, we consider systems characterized by complex interdependencies that are, in most cases, non-linear.

In other words, by using a numerical simulation, we can change any parameter and test the behavior of the system, or how its structure changes as a response to these variations, or how structural changes influence the main parameters of interest. This may allow predicting future behaviors and evolutions, when this is theoretically and practically possible, or, better, to produce a series of scenarios that can then be profitably employed for improving knowledge, inform decisions or assess effects of some interventions. Here the use of analogy can be of great help in providing interesting and useful solutions. For example, using the similarity with the diffusion of a disease, that is well known and studied in epidemiology, we can simulate the transfer of knowledge in a destination, test the starting conditions and figure out the speed and the extent of the diffusion. It is then possible to modify the system's structure or some other relevant parameter and assess the effects of this variation (see e.g. (Baggio, 2015; Baggio & Cooper, 2011; Del Chiappa & Baggio, 2015).

For their nature no 'ready-made' software exists for addressing a numerical simulation. The researcher has to write own code using one of the many computer languages existing. Among the most used are Matlab (commercial, see e.g. Hanselman & Littlefield, 2004) and Python (free, see e.g. Lutz & Ascher, 2013) that provide good development interfaces and are quite efficient and effective. Moreover, for their characteristics, these languages are relatively easy to grasp. Agent-based models, instead, require some specially built environments, able to ease customization and conditionally control the agents and their behaviors and perform the different iterations. In this area the most known is NetLogo (free, see e.g. Wilensky & Rand, 2015). All these tools have wide communities of users that publish their codes and discuss issues and problems, and are precious sources for TAPAS approaches.

5. Calibration, verification and validation

Once a model has had an initial definition and configuration needs to be thoroughly verified and adapted to the specific situation through some tuning of the quantities used to define its components. Typically, we shall have a number of parameters, quantities used to describe objects statically that will (normally) not change during the various possible changes. Examples are the quantity of agents, their initial conditions (values of characterizing features) the number of agents considered for the basic interactions, and so on. In addition, a number of variables are defined that represent the state of the model and will change depending on the evolution of the interactions between the agents forming the outputs of the implementation.

5.1. Calibration

The calibration phase consist of determining the best values for the model's parameters, where by best we mean the values that more closely correspond to a real situation, or provide outcomes as close as possible to a real setting. This calibration is usually obtained by exploring the parameter space (i.e. changing the possible values of the parameters) and checking the results to verify, for some known setting, the agreement with the system or phenomenon under study, by comparing them to other known realizations, or by establishing their plausibility with some knowledgeable informant. In some cases, the calibration may also involve modifications to the algorithms used or to optimization mechanisms if present. The initial values are to be derived from some empirical data, if available, or from expert knowledge or past literature. As an example, Balbi, Giupponi, Perez, and Alberti (2013) calibrated a model on the effects of climate change on an alpine region tourism based on empirical data and assessing the results of their model with local stakeholders.

Together with a calibration of the different parameters it is important to perform a sensitivity analysis. This is done by 'exercising' the model with different parameters' combinations in order to establish how they, or their variations, influence the outcomes and identify the most relevant and critical, or those who have no practical influence and can then be eliminated, simplifying the model. Moreover, a sensitivity analysis can be a useful aid in the validation of the code written (Iooss & Lemaître, 2015; Pianosi et al., 2016).

5.2. Verification and validation

Verification and validation, although looking synonymous, have distinct meanings, mainly rooted in computer science and software engineering. V*erification* is used for expressing the activities of evaluating the computational implementation of a model in terms of the objectives stated, and the correspondence and the agreement with the conceptual model. V*alidation* refers to the assessment of the credibility and reliability of model or the simulation in representing the subject modelled.

In other words, verification focuses more on the assessment of concepts and inferences in the process of programming, observing and interpreting a computational implementation, validation concerns the evaluation of the inferences and concepts as representations of the phenomenon or system under consideration.

Rather obviously, the verifiability of a model or a simulation depends on the process and the tools used to develop that object. The use of widely employed high-level simulation packages, libraries or languages can simplify this phase, since, often, common model building blocks are available, and these are typically already well verified (Balci, 1998; David, Fachada, & Rosa, 2017; Sargent, 2013).

Moreover, as said, these environments benefit from wide and collaborative user communities that can be of great help in solving computational (and non) issues. The most famous example is the huge Stack Exchange network (https://stackexchange.com/), an ensemble of question-and-answer websites on topics in diverse fields, many of which dedicated to professional and enthusiast programmers such as those grouped in the Stack Overflow network (https://stackoverflow.com/). This is an important point since a developer might be able to correct many errors, also with the help of the debugging facilities of the development environment used but might be unable to fully seize 'logical' mistakes or computational artefacts that may produce unexpectedly wrong outcomes.

The validation phase passes through the joint assessment of the modeler and field experts that have the task of judging the truthfulness of the implementation. Typically, the elements on which a validation is based are (following Robinson, 1999):

- data: determining whether all the data used are of sufficient accuracy;
- white-box: determining the extent to which the different parts of the implementation accurately represent real world elements;
- black-box: testing whether the overall model suitably and accurately represent the object of study;
- experimentation: assessing the validity of the 'experimental' procedures (initialization, run length, replications, sensitivity etc.);
- solution: gauging the extent to which the solution is a 'reasonable' and meaningful outcome.

Obviously, as also Robinson states (1999: 68): "It is not possible to prove that a model is absolutely correct. Therefore, model verification and validation is concerned with creating enough confidence in a model for the results to be accepted. This is done by trying to prove that the model is incorrect. The more tests that are performed in which it cannot be proved that the model is incorrect, the more confidence in the model is increased. For verification and validation the general rule is: the more the better."

6. Choosing the modelling and simulation approach

A difficult issue in simulation and modelling concerns the choice of the specific technique to be used. There are probably many possible approaches and tools that can be used in answering a defined question and the question is: which one is the best?

Since a model is a synthetic representation of some real world entity, we adopt some simplification based on assumptions that do not have a general validity but are bound to the peculiarities of what examined. This means that a model which validly explains a certain situation might be completely unsuitable in another. This intuitive statement is also formally demonstrated in a series of theorems known as no free lunch (NFL) theorems (Wolpert, 2001; Wolpert & Macready, 1997). The essence is that given the biases inevitably introduced by the assumptions made, the cost of finding a solution, averaged over all the possible similar problems, is the same for any solution method. Therefore, no solution offers a 'short cut', or, in other words: how 'good' a model is depends on its alignment with the actual problem, and there is no universal better model.

Said that, we may focus here on two elements that are important in choosing the approach to follow: the type of problem or research question and the data. To these we should add the resources (hardware and software) available or known to the researcher (for a more extensive discussion see e.g. Baggio & Baggio, 2020; Robinson, 2011).

The first element, via a well-designed conceptual model, determines at least a class of possible choices. So, for example, if we are interested in comparisons, classifications or causal relationships we will go for some of the well-known statistical techniques or, more recently to some machine learning algorithm. If the structure of a system is to be investigated a network analysis is able to provide good outcomes and can be used, without big difficulties, for simulating several dynamic processes. If we have to deal with heterogeneous interacting agents, an agent-based model is the natural choice. Some of the possible modelling techniques (statistical, machine learning, or network analysis, for example) may have specific requirements on the data in terms of types, quantity and quality (see e.g. Baggio & Klobas, 2017), while an agent-based model can be built with very little restrictions or even in absence of empirical data, provided some robust theoretical basis or good expert knowledge exist. Lastly, the hardware and software facilities available may restrict the possible choice. Some of the modelling or simulation techniques are quite *hungry* in terms of computational efforts and time and may require the use of specific platforms or languages often not easily 'reachable' or known by the researcher.

It must be noted, finally, that these elements (problem, data, resources) are quite often strongly interconnected and play iteratively so that the final decision is, inevitably, a compromise between the different requirements and constraints. Here too, one of the most important elements is the experience that the modeler has acquired.

7. Publishing

As discussed previously, the validity of a computational model or simulation is a delicate matter. Here we want to draw the attention of the reader to the necessity and importance of making available all the elements of the model or simulation designed and executed. There is no common or standard way to publish these items. A good description of the algorithms and a pseudocode of the scripts used is the minimum reasonable level of publicity to provide. Together with these, the data and the parameters used, and the actual code should be uploaded to one of the many platforms existing such as GitHub (https://github.com/). In the case of agent-based models the Overview, Design concepts and Details (ODD) protocol has shown to be quite effective in documenting what done, and a dedicated repository exist: CoMSES (https://www.comses.net/).

The publication of data and code is important for several reasons. First of all, it allows the fundamental scientific activity of replication, important also for enhancing the capabilities of validating the implementations. Secondly, it allows others to build upon a realization and improve it, maybe addressing issues not considered originally or augmenting the possible 'coverage'. Finally, it contributes to the visibility and reputation of the researcher. It must be noted here that many journals (and among them the most famous and reputable) consider the availability of data and code as a prerequisite for publication. This journal, as stated in the guide for authors (https://www.elsevier.com/journals/annals-of-tourism-research-empirical-insights/2666-9579/guide-for-authors) has adopted this position.

8. Concluding remarks

Computational models and simulations have proved to be powerful, effective and efficient tools for dealing with complex systems and phenomena. They provide a way to better describe, examine, and (sometimes) predict the structural and dynamic characteristics of the object of study, and may allow creating much more truthful and reliable scenarios that can be of great help in answering research questions or solving 'practical' problems. The interest for both the academic and 'industrial' communities is very high and the reliability and validity of these techniques, when soundly based on a correct and rigorous approach, is high as we have briefly sketched in this contribution.

The tasks might seem difficult and complex, and indeed they are, quite often, and some of them may require a good deal of preparation before attempting some initial experiment. As Euclid replied to Ptolemy I "There is no royal road to geometry". But the outcomes are undoubtedly remarkable and fascinating.

A final consideration is in order. The enterprise, as seen, is not one of the easiest and requires not only a good deal of efforts and experience, but also skills and expertise at times quite technically specific, so that it is difficult they are possessed by a single individual. This poses the need to assemble a team of researchers with the different know-hows required. That is to say that the 'art' of modelling and simulation is a truly multi-disciplinary art.

Statement of contribution

Modelling and simulations, especially numerical, are one of the few possible approaches to explain and predict the structure and the behaviors of complex systems such as the social, political and economic entities of interest for the tourism and hospitlity cresearch community.

They allow to "experiment" different settings and configurations where these cannot be easily obtained in a living condition because of time, costs, ethical concerns etc.

In this paper we present the very basic methodological requirements and provide recommendations, in order to ensure the design and the implementation of rigorous, valid and reliable models and simulations.

With these attentions, the practice is able to provide useful and interesting insights into many phenomena and systems of interest and allow making better informed decisions.

Although a certain growing number of works already use these techniques, the practice is relatively unknown in the tourism domain and the paper intends to be of general guidance for those interested in employing this approach.

Declaration

I declare no conflicts of interest.

Acknowledgements

R.B. acknowledges the financial support of the Ministry of Education and Science of the Russian Federation in the framework of the Competitiveness Enhancement Program of the Tomsk Polytechnic University.

References

Amelung, B., Student, J., Nicholls, S., Lamers, M., Baggio, R., Boavida-Portugal, I., ... Balbi, S. (2016). The value of agent-based modelling for assessing tourism-environment interactions in the Anthropocene. *Current Opinion in Environmental Sustainability*, *23*, 46–53.

Andersen, H., & Hepburn, B. (2016). Scientific method. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2016 Edition)* Retrieved July, 2020, from http://plato.stanford.edu/archives/sum2016/entries/scientific-method/.

Baggio, J. A., & Baggio, R. (2020). *Modelling and simulations for tourism and hospitality*. Bristol, UK: Channel View.

Baggio, R. (2008). Symptoms of complexity in a tourism system. *Tourism Analysis*, *13*(1), 1–20.

Baggio, R. (2015). Knowledge management and diffusion: The network paradigm. In M. T. McLeod, & R. Vaughan (Eds.), *Knowledge networks and tourism* (pp. 108–125). New York: Routledge.

Baggio, R. (2017). Network science and tourism – The state of the art. *Tourism Review*, *72*(1), 120–131.

Baggio, R. (2018). Network analysis: Quantitative methods in tourism. In C. Cooper, B. Gartner, N. Scott, & S. Volo (Eds.), *The Sage handbook of tourism management* (pp. 150–170). Thousand Oaks, CA: SAGE.

Baggio, R., & Cooper, C. (2011). Knowledge transfer in a tourism destination: The effects of a network structure. In N. Scott, & E. Laws (Eds.), *Advances in service network analysis* (pp. 177–192). London: Routledge.

Baggio, R., & Klobas, J. (2017). *Quantitative methods in tourism: A handbook* (2nd ed.). Bristol, UK: Channel View.

Baggio, R., & Scaglione, M. (2018a). Destination attractions system and strategic visitor flows: An exploratory study. In F. Sánchez, C. Pautasso, & K. Systä (Eds.), *Current trends in web engineering, ICWE2018 workshops* (pp. 227–237). Cham (CH): Springer.

Baggio, R., & Scaglione, M. (2018b). Strategic visitor flows and destination management organization. *Information Technology and Tourism*, *18*(1–4), 29–42.

Balbi, S., Giupponi, C., Perez, P., & Alberti, M. (2013). A spatial agent-based model for assessing strategies of adaptation to climate and tourism demand changes in an alpine tourism destination. *Environmental Modelling & Software*, *45*, 29–51.

Balci, O. (1998). Verification, validation, and testing. In J. Banks (Ed.), *Handbook of simulation: Principles, methodology, advances, applications, and practice* (pp. 335–393). Hoboken, NJ: John Wiley & Sons.

Barabási, A. L. (2016). *Network science*. Cambridge, UK: Cambridge University Press.

Barrat, A., Barthélémy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge: Cambridge University Press.

Bertuglia, C. S., & Vaio, F. (2005). *Nonlinearity, chaos, and complexity: The dynamics of natural and social systems*. Oxford: Oxford University Press.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of the Sciences of the USA*. *99*. (pp. 7280–7287).

Brooks, F. P. (1995). *The mythical man-month: Essays on software engineering* (anniversary ed.). Reading, MA: Addison-Wesley.

Casanueva, C., Gallego, Á., & García-Sánchez, M. R. (2016). Social network analysis in tourism. *Current Issues in Tourism*, *19*(12), 1190–1209.

Davenport, J., Poonen, B., Maynard, J., Helfgott, H., Tiep, P. H., & Cruz-Filipe, L. (2018). Machine-assisted proofs (ICM 2018 Panel) (arXiv:1809.08062). Retrieved May 2020, from http://arxiv.org/abs/1809.08062.

David, N., Fachada, N., & Rosa, A. C. (2017). Verifying and validating simulations. In B. Edmonds, & R. Meyer (Eds.), *Simulating social complexity* (pp. 173–204). Cham, CH: Springer.

Del Chiappa, G., & Baggio, R. (2015). Knowledge transfer in smart tourism destinations: Analysing the effects of a network structure. *Journal of Destination Marketing and Management*, *4*(3), 145–150.

Edmonds, B. (2017). Different modelling purposes. In B. Edmonds (Ed.), *Simulating social complexity* (pp. 39–58). Cham, CH: Springer.

Embley, D. W., & Thalheim, B. (Eds.). (2011). *Handbook of conceptual modeling*. Cham, CH: Springer.

da Fontoura Costa, L., Oliveira, O. N., Travieso, G., Rodrigues, F. A., Villas Boas, P. R., Antiqueira, L., ... Correa Rocha, L. E. (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, *60*(3), 329–412.

da Fontoura Costa, L., Rodrigues, A., Travieso, G., & Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, *56*(1), 167–242.

Frenken, K. (2006). Technological innovation and complexity theory. *Economics of Innovation and New Technology*, *15*, 137–155.

Frigg, R., & Hartmann, S. (2020). Models in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2020 Edition)* Retrieved July, 2020, from https://plato.stanford.edu/archives/spr2020/entries/models-science/.

Galilei, G. (1832). *Opere (Vol. II)*. Milano: Nicolò Bettoni.

Gelfert, A. (2017). The ontology of models. In L. Magnani, & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 5–24). Cham, CH: Springer.

Gentner, D. (2002). Analogy in scientific discovery: The case of Johannes Kepler. In L. Magnani, & N. J. Nersessian (Eds.), *Model-based reasoning: Science, technology, values* (pp. 21–39). New York: Kluwer Academic / Plenum Publishers.

Gilbert, N., & Terna, P. (2000). How to build and use agent-based models in social science. *Mind & Society*, *1*(1), 52–72.

Graebner, C. (2018). How to relate models to reality? An epistemological framework for the validation and verification of computational models. *Journal of Artificial Societies and Social Simulation*, *21*(3) (art. 8).

Gummesson, E. (2007). Case study research and network theory: Birds of a feather. *Qualitative Research in Organizations and Management*, *2*(3), 226–248.

Hale, J. K., & LaSalle, J. P. (1963). Differential equations: Linearity vs. nonlinearity. *SIAM Review*, *5*(3), 249–272.

Hanselman, D. C., & Littlefield, B. L. (2004). *Mastering Matlab 7*. Upper Saddle River, NJ: Prentice-Hall.

Iooss, B., & Lemaître, P. (2015). A review on global sensitivity analysis methods. In G. Dellino, & C. Meloni (Eds.), *Uncertainty management in simulation-optimization of complex systems* (pp. 101–122). Cham, CH: Springer.

Johnson, N. F. (2009). *Simply complexity: A clear guide to complexity theory*. Oxford, UK: Oneworld Publications.

Johnson, P., Nicholls, S., Student, J., Amelung, B., Baggio, R., Balbi, S., ... Steiger, R. (2017). Easing the adoption of agent-based modelling (ABM) in tourism research. *Current Issues in Tourism*, *20*(8), 801–808.

Knuth, D. (2011). *The art of computer programming* (3rd ed.). Reading, MA: Addison-Wesley.

Krieger, M. H. (2005). A 1940 letter of André Weil on analogy in mathematics. *Notices of the American Mathematical Society*, *52*(3), 334–341.

Lutz, M., & Ascher, D. (2013). *Learning python* (5th ed.). Sebastopol, CA: O'Reilly Media.

Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3), 151–162.

Magnani, L., & Bertolotti, T. (Eds.). (2017). *Springer handbook of model-based science*. Cham, CH: Springer.
Mariani, M., & Baggio, R. (2020). The relevance of mixed methods for network analysis in tourism and hospitality research. *International Journal of Contemporary Hospitality Management, 32*(4), 1643–1673.
Martin, R. C. (2008). *Clean code: A handbook of agile software craftsmanship*. Upper Saddle River, NJ: Prentice Hall.
McConnell, S. (2004). *Code complete: A practical handbook of software construction* (2nd ed.). Redmond, WA: Microsoft Press.
Merinero-Rodríguez, R., & Pulido-Fernández, J. I. (2016). Analysing relationships in tourism: A review. *Tourism Management, 54*, 122–135.
Minsky, M. (1968). Matter, mind and models. In M. Minsky (Ed.), *Semantic information processing* (pp. 425–432). Cambridge, MA: MIT Press.
Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace and World.
Nersessian, N. J., & MacLeod, M. (2017). Models and simulations. In L. Magnani, & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 119–132). Cham, CH: Springer.
Nicholls, S., Amelung, B., & Student, J. (2017). Agent-based modeling: A powerful tool for tourism researchers. *Journal of Travel Research, 56*(1), 3–15.
Olsen, W. (2004). Triangulation in social research, qualitative and quantitative methods can really be mixed. In M. Holborn (Ed.), *Developments in sociology: An annual review* (pp. 103–121). Ormskirk, UK: Causeway Press.
Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software, 79*, 214–232.
Robinson, S. (1999). Simulation verification, validation and confidence: a tutorial. *Transactions of the Society for Computer Simulation, 16*(2), 63–69.
Robinson, S. (2008). Conceptual modelling for simulation part I: Definition and requirements. *Journal of the Operational Research Society, 59*(3), 278–290.
Robinson, S. (2011). Choosing the right model: Conceptual modeling for simulation. *Proceedings of the 2011 Winter Simulation Conference (WSC), Phoenix, AZ, (December 11-14)* (pp. 1423–1435).
Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation, 7*(1), 12–24.
Scott, N., Baggio, R., & Cooper, C. (2008). *Network analysis and tourism: From theory to practice*. Clevedon, UK: Channel View.
Scott, N., Baggio, R., & Cooper, C. (2011). Network analysis methods for modeling tourism inter-organizational systems. In C. Megehee, A. Ogle, & A. G. Woodside (Eds.), *Advances in culture, tourism, & hospitality research. Vol. 5*. (pp. 177–222). Bingley, UK: Emerald - JAI Press.
Silvert, W. (2001). Modeling as a discipline. *International Journal of General Systems, 30*(3), 261–282.
Starfield, A. M., Smith, K. A., & Bleloch, A. L. (1990). *How to model it: Problem solving for the computer age*. New York, NY: McGraw-Hill.
Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., ... Buchmann, C. M. (2016). Simple or complicated agent-based models? A complicated issue. *Environmental Modelling & Software, 86*, 56–67.
Turner, J. (1955). Maxwell on the method of physical analogy. *The British Journal for the Philosphy of Science, 6*(23), 226–238.
Viglia, G., & Dolnicar, S. (2020). A review of experiments in tourism and hospitality. *Annals of Tourism Research, 80* (art. 102858).
Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics, 13*(1), 1–14.
Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. Cambridge, MA: MIT Press.
Wittgenstein, L. (1921). Logisch-philosophische abhandlung. In W. Ostwald (Ed.), *Annalen der naturphilosophie. 14.*.
Wolpert, D. H. (2001). The supervised learning no-free-lunch theorems. *Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications, On the Internet, WWW (September 10-24)*.
Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67–82.
van der Zee, E., & Vanneste, D. (2015). Tourism networks unravelled; a review of the literature on networks in tourism management studies. *Tourism Management Perspectives, 15*, 46–56.

Rodolfo Baggio is professor at the Master in Economics and Tourism and Research Fellow at the Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy and professor at the Tomsk Polytechnic University, Tomsk, Russian Federation. He teaches and researches on the use of information technology in tourism and on the interdisciplinary applications of complexity and network science methods to the study of tourism destinations.