



Simulated Data in Empirical Science

Aki Lehtinen¹  · Jani Raerinne² 

Accepted: 18 October 2023 / Published online: 6 November 2023
© The Author(s) 2023

Abstract

This paper provides the first systematic epistemological account of simulated data in empirical science. We focus on the epistemic issues modelers face when they generate simulated data to solve problems with empirical datasets, research tools, or experiments. We argue that for simulated data to count as epistemically reliable, a simulation *model* does not have to mimic its target. Instead, some models take empirical data as a target, and simulated *data* may successfully mimic such a target even if the model does not. We show how to distinguish between simulated and empirical data, and we also offer a definition of simulation that can accommodate Monte Carlo models. We shed light on the epistemology of simulated data by providing a taxonomy of four different mimicking relations that differ concerning the nature of the relation or relata. We illustrate mimicking relations with examples from different sciences. Our main claim is that the epistemic evaluation of simulated data should start with recognizing the diversity of mimicking relations rather than presuming that only one relation existed.

Keywords Data · Computer simulation · Computation · Monte Carlo models · Mimicking

1 Introduction

The use of simulated data has become ubiquitous in the sciences, but extant philosophical accounts of simulation are incapable of accounting for the diverse scientific uses of simulated data. We provide the first systematic epistemological account of the uses of simulated data in empirical science. The typical reason for using simulated data is that existing empirical datasets, research tools, or experiments are deficient, and simulated data are used to enhance their epistemic or evidential relevance. We aim to show the conditions under which simulated data can be used to provide better data-to-phenomena inferences.

Antoniou (2021) posits that a dataset is epistemically reliable if it provides correct information for the physical phenomenon it represents. Similarly, one can ask: When is

✉ Jani Raerinne
jani.raerinne@helsinki.fi

Aki Lehtinen
aki.lehtinen@helsinki.fi

¹ College of Philosophy, Nankai University, Tianjin, China

² Faculty of Social Sciences, University of Helsinki, Helsinki, Finland

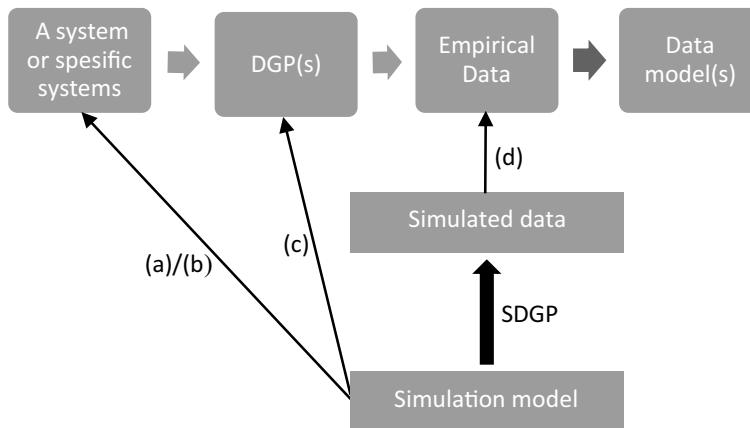


Fig. 1 The thin arrows and letters (a–d) indicate the relevant mimicking/representational relations and relata. The thick vertical black arrow indicates that the simulation model generates simulated data using a *Simulated Data Generating Process* (SDGP). While empirical data are generated from a system via a DGP or DGPs (light grey horizontal arrows), data models provide a representation of empirical data (the dark grey color of the rightmost horizontal arrow indicates this difference). A data model is needed for making empirical data usable for evidential purposes. Simulations can themselves be data models, test or correct data models, or take data models as input. (Color figure online)

a simulated dataset epistemically reliable? Answering this question requires defining what simulated data are, distinguishing them from empirical data, and then providing an account of what is epistemically relevant about simulated data. This is what we will do in this paper. However, we could not possibly provide an epistemology that covers the whole epistemology of simulation (see e.g., Winsberg, 2010; Beisbart & Saam, 2019; Durán, 2018).¹ Our definition posits that simulations² mimic something. We are thus providing answers to a more limited question: When does mimicking support the epistemic reliability of simulated data?

Why is an epistemology of simulated *data* needed in addition to the already existing literature on the epistemology of simulation *models*? This is because evaluating the reliability of simulated data requires paying attention to the variety of relevant representational relations in generating simulated data. Importantly, we will discuss an example in which the simulated *data* mimics empirical data. Such examples cannot be analyzed with accounts presupposing that a simulation *model* mimics something. Although using simulated data serves various methodological functions, there is always some mimicking relation that is relevant for their epistemic evaluation. An epistemological investigation should begin by asking: What is the relevant representational relation, and is the mimicking successful? We identify four such relations (see Fig. 1). A simulation *model* may aim to *mimic* (a) one system, (b) several specific systems, or (c) *represent* a Data Generating Process (DGP). *Data* produced by a simulation *model* may aim to mimic (d) existing empirical data.

A DGP consists of all the processes that are responsible for producing empirical data about a real-world system. Although the term is commonly used in statistics, we use it here in a broader sense in that DGPs can include instruments, experiments, measurements,

¹ We also do not discuss whether simulations are best characterized as experiments or arguments (see Morrison 2015; Lenhard 2016; Imbert 2017 for overviews).

² Although analog simulations mimic, we only discuss computer simulations.

and data-collection methods. Modelers typically talk about DGPs when they study some particular problems with data, and the problems are different for different variables. This explains why the content of a DGP may vary from one study to another even when they concern the same system: the DGP in a model abstracts from some properties of the system, and it is study-relative.³

The epistemically relevant mimicking relation depends on the function of the simulation and the case at hand. We discuss cases from several disciplines to show that the reasons for using simulated data, and the relevant mimicking relations vary. In sum, then, we need an epistemology of simulated data because extant accounts do not pay attention to the mimicking relations that are crucial in evaluating those data.

Mimicking relations are diverse because empirical data and the methods of analyzing them are problematic in different ways. Depending on the case at hand, researchers may need to *correct* empirical data concerning biases, *filter* the signal from noise, *add* or *augment* data when data are insufficient for the researchers, *enhance the interpretability* of ambiguous data (cf. Bokulich, 2020), or *aggregate* data when researchers endeavor to obtain an overall view of some empirical literature. Moreover, simulated data are used to test the performance of research tools that researchers utilize in correcting, estimating, or collecting empirical data.

What are simulated data? Barberousse et al. (2009) provide an origin-based definition: Simulated data have their origin in a computer simulation model. This account does not provide a complete definition without a specification of what a computer simulation model is. Unfortunately, the extant definitions of simulation models (see Sect. 5) restrict their targets to physical systems.

Any method that exploits computer-generated randomness in calculating some results is a *Monte Carlo* method. Monte Carlo models are often used in generating data that modelers claim to be simulated, and some of our examples involve Monte Carlo models. A well-recognized philosophical problem of the extant definitions of simulation is that they fail to accommodate Monte Carlo models (Winsberg, 2015). A related problem of origin-based definitions is that distinguishing between *simulated* and *empirical* data requires one to be able to separate between the computational processing of empirical data and using empirical data to calibrate simulation models (Arnold, 2013).

We provide a revised definition of simulation that can accommodate Monte Carlo models. It amends the extant ones by expanding the range of targets of simulation to include data. Moreover, we argue that to account for the diverse uses of simulated data, the traditional requirement that computational *models* must mimic something must be relaxed. Instead, it is sufficient if the *data* produced by the model mimic. We distinguish between simulation and mere computation⁴ by claiming that empirical input data remain empirical if their computational processing does not include mimicking. This philosophical machinery is required to successfully solve the problem of distinguishing between empirical and simulated data, but it is also necessary for paving the way for an account of the epistemology of simulated data: what is crucial is that simulated data either themselves mimic something or derive from a model that aims to mimic something.

We adopt the following terminological conventions. When we use the term 'target,' we mean whatever the model or the data it generates aims to represent. A system,

³ Thus, for example, the DGP for the global mean surface temperature is different from the DGP for satellite data on temperature in the models that study these DGPs, even though the climate as a whole is ultimately responsible for both datasets.

⁴ In what follows, we use 'computation' or 'calculation' for 'mere computation/calculation'.

many specific systems, a DGP, or data may be ‘targets.’ Targets are abstractions over something (systems, phenomena, or data) in that the modeler selects particular features that must be accounted for (i.e., the target-defining properties), and ignores the rest (cf. Weisberg, 2013). The reason why we use the term ‘target’ rather than the more traditional ‘target system’ is that, while systems and DGPs qua targets are ‘target systems,’ some models aim to ‘target’ data, which is not a ‘target system’ because data, a (static) collection of properties or numbers that do not have interacting parts, cannot be referred to as a ‘system.’ We use the term ‘target system’ such that it only covers case (a) in which the ‘target’ is indeed a system. We will use the term ‘system’ when one can meaningfully identify a system that is ultimately responsible for generating data, even though that system as a whole is not the target of the investigation at hand. In case (b), we can say that the target consists of many specific systems.

We will speak about ‘mimicking’ whenever modelers *aim to represent the target accurately*, whatever accuracy means exactly in the context (see Contessa, 2007; Nguyen, 2020). In most cases, accurate mimicking means providing descriptions that are similar or isomorphic to the target, or true about it, but we do not rule out functional relationships (‘keys’) not captured by these concepts. When modelers are *not concerned with the accuracy of representing the whole target but aim to mimic only some parts of it*, we will say that they are ‘representing’ the target. ‘Representing’ typically involves strategic misrepresentations of targets. In ‘mimicking’, one aims to mimic accurately all properties which need to be represented in the first place (viz. the target-defining ones), while in ‘representing’, one does not aim to mimic accurately all properties that need to be represented. The reason for distinguishing between mimicking and representing is to show that mimicking relations differ not only with respect to their targets and relata (i.e., a system, DGP, data, and so on) but also with respect to the nature of the representational relation.

A mimicking model is not necessarily more accurate overall than a representing model. It is entirely possible to encounter two models: one aiming to mimic all properties of a target, while the other mimics only a subset. However, the latter may exhibit greater accuracy in representing the same target due to its less severely flawed idealizations or its superior description of specific target properties, including improved approximation of variable values. Instead of ranking representational relations in terms of accuracy, the point of this distinction is to pay attention to the fact that in some cases, inaccuracies are required to be able to test research tools.

Nonetheless, representing at least something accurately is typically important for the epistemology of simulated data. But why? It is because our examples concern cases in which empirical data are problematic. Given that simulations produce simulated data that are better than empirical data in our examples, they had better mimic the relevant aspects of their targets accurately.

Section 2 presents examples of the different mimicking relations. Section 3 explains how and why these mimicking relations must be distinguished from each other. Section 4 discusses how the difference between computation and simulation can be used to distinguish between empirical and simulated data. Section 5 provides a revised definition of computer simulation that can account for the diverse scientific uses of simulated data by accommodating Monte Carlo models.

2 Examples of Four Mimicking Relations

Understanding the epistemology of simulated data requires paying attention to the fact that simulations mimic something. Since it is not always the simulation model that mimics something, but rather the simulated data, one must pay attention to the variety of mimicking relations. However, demonstrating such variety is not sufficient because the previous conceptual frameworks are unable to accommodate some of our scientific examples. According to some extant accounts, some of our examples involve empirical rather than simulated data, some involve mere computation rather than simulation, and all except cases (a) and (b) are not simulations because they are based on Monte Carlo models. A credible philosophical account is thus needed for defining data and simulated data, for distinguishing empirical and simulated data from each other, and for defining simulation.

We will provide such an account in Sects. 4 and 5, but it is easier to understand what we are trying to do if we first present examples of different kinds of mimicking relations, and then provide a philosophical account that explains why each of the cases counts as an example of simulated data and explains why one must distinguish them from each other for epistemic reasons.

For now, it is important to bear in mind that the examples illustrate the variety of mimicking relations. The different examples we present provide the basis for a philosophical account of simulated data; they display the variety and properties of examples that a philosophical account must be able to accommodate. On the one hand, the definitions of simulation and simulated data must be able to include all the cases as examples of simulated data. On the other, they must be able to capture the differences in the representational and mimicking relations in the different examples. While we aim for our distinction between various mimicking relations to be comprehensive, the diversity of practices involving simulated data leads us to make this assertion with caution.

A simulation model generates simulated data using the *simulated data generating process* (SDGP). We will see that in cases (b–d) modelers exploit the fact that the SDGP is known. Since modelers know the SDGP, they also know what kind of simulated data it produces. The ‘benchmark property’ is a property of the SDGP that allows for assessing whether a research tool is reliable. Modelers know what the benchmark is because they put it into the programming code. Knowing the SDGP implies that the data it produces can constitute a benchmark against which the performance of research tools can be compared. We call cases (b)–(d) *benchmark data simulations* and the data generated in such exercises are *benchmark data*. Benchmark data are used to test the performance of research tools for correcting, filtering, imputing, or estimating empirical data. The point of benchmark data simulations is to test whether the research tool can track the benchmark properties correctly.

2.1 Mimicking a Target System (a)

Empirical raw data about the climate are problematic in various ways. There are temporal and spatial gaps in climate datasets, data suffer from discontinuities, and some of the data are biased, for example. *Data assimilation* refers to the whole process of integrating various data sets with global simulation models (see Edwards, 2010; Parker, 2017).

The justification for employing simulation in data assimilation is that it allows for the use of physical theory in determining the properties of the missing data rather than

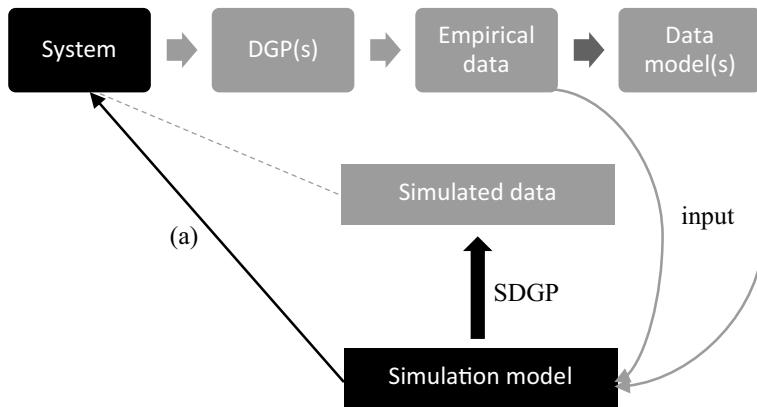


Fig. 2 A simulation model mimics a target system (a). The black arrow illustrates the mimicking relation. The dashed grey line describes the evidential role of the simulated data: If the model successfully mimics the target system, then simulated data may be used as evidence about the target system. The simulation model takes empirical data modified with data models as input (thin solid grey arrows), and yields simulated data that then replaces the empirical data as the best description of the climate system. (Color figure online)

adding artificial data points ad hoc. Constructing the reanalysis data with the help of global simulation models is an attempt to mimic the real processes in the target system that generate the needed missing climate data: In other words, the simulation model purports to mimic *a specific target system* (a), the climate system (see Fig. 2). The simulation does not purport to mimic the DGPs that produce empirical data about the climate, because the DGPs are known to produce erroneous and incomplete data. Instead, the purpose is to apply knowledge about the physical relationships governing the climate system to correct, fill and smooth the erroneous empirical data produced by the DGPs. Hence, the epistemic evaluation of models that generate reanalysis data hinges on whether they can mimic the target system successfully.

Simulated data has brought about a clear improvement in data quality. As Parker (2016) notes, reanalysis data is not self-evidently less reliable than empirical data. This is a general feature of simulated data that holds in a variety of cases: Several recent contributions have argued against the idea that the epistemic reliability of data is highest when they are unprocessed (“pure” or “raw”) (Antoniou, 2021; Bokulich, 2021; Bokulich & Parker, 2021; Humphreys, 2014). Simulation is often used to generate better data models. Indeed, given that our examples included cases in which simulated data are superior to empirical data, whether a given dataset is simulated or empirical does not determine epistemic superiority (Parke, 2014; see also Massimi & Bhimji, 2015; Morrison, 2015).

The overall purpose of the reanalysis data is to provide a better description of the state of the climate than that produced from empirical data deriving from the DGPs. However, even if the *model* successfully mimicked the target system, this would not be sufficient for the data to do so successfully. This is because the reanalysis derives from a huge assortment of methods used to manipulate the empirical data over and above simulations of the target system. Without trying to describe these complexities, let us

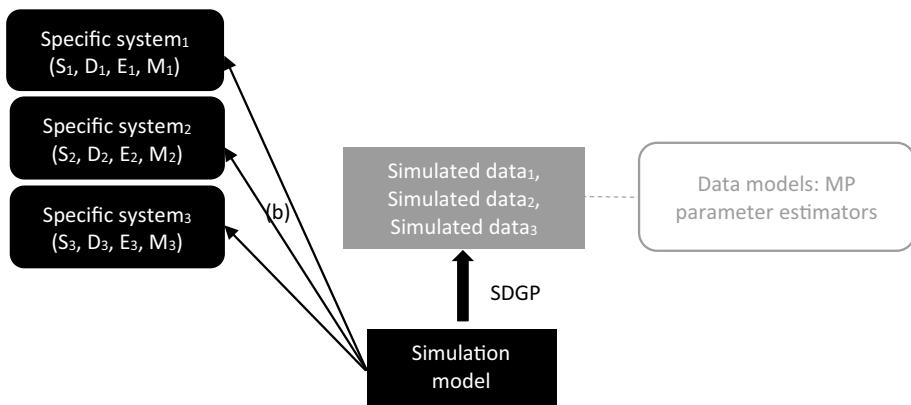


Fig. 3 A simulation model mimics several specific systems (b). The model contains a description of the metapopulation parameter estimators that are applied to the simulated datasets (the grey dashed line). (Color figure online)

merely note that additional mimicking relations are involved in generating reanalysis data.

2.2 Mimicking Several Specific Systems (b)

When scientists test the performance of research tools with the help of simulated benchmark data, robust performance under diverse simulated data is a frequently used reliability criterion. One strategy to produce diverse simulated data is to mimic several specific systems (b) in a simulation model.

Metapopulation (MP) models describe the colonization and extinction dynamics of a species over a spatial area (Levins, 1969). The species is described as one metapopulation. It consists of subpopulations living in separate areas. The spatial area consists of habitat patches and the habitat matrix, where subpopulations can and cannot persist. Subpopulations can migrate through the matrix to other habitat patches and colonize them. The extinction and colonization probabilities of a metapopulation depend on the size of the habitat patches, the distances and connections between them, and species-specific parameters concerning survival and migration ability.

Hanski's (1994) "incidence function" framework was commonly used for estimating parameters in MP models. The problem is that Hanski's method provides biased parameter estimates in many cases. Since Hanski's method lacked robustness, Moilanen (1999) developed an improved Monte-Carlo-based parameter estimation method. He used simulated benchmark data to test his and Hanski's methods. Given that Moilanen himself coded a description of the different metapopulation targets into the simulation model, this knowledge of the SDGP provided an objective criterion for evaluating the performance of the estimation methods.

In the simulations, Moilanen varied the sizes (=S) of the habitat patches and the distances (=D) between them, the migration abilities and extinction (=E) probabilities of metapopulations, and the minimum (=M) patch-size area required for subpopulations to persist. The different simulations created a diverse set of data concerning patch presences/absences and extinction as well as colonization turnover events in metapopulations of

different generic species with their varying extinction and colonization dynamics. These simulations thus mimicked a *diverse set of specific systems* (b) (see Fig. 3).

Moilanen (1999) tested the predictive accuracy and performance of parameter estimation methods on the above-mentioned simulated data. If a method captures the differences between various simulated systems from simulated data, then its performance is robust.⁵ The results indicated that the new Monte Carlo method not only produced more accurate parameter estimates of simulated metapopulation data than Hanski's regression method. It was also more robust in its performance under diverse metapopulation simulated datasets.

Whether the simulation model provides evidence of the performance of estimators depends on whether it correctly mimicked the relevantly diverse range of specific systems. Moilanen aimed to mimic various systems because the metapopulation models are meant to apply to a diverse set of real systems. If the set of simulated specific systems is not sufficiently diverse, modelers end up with an illusion of robust performance of the tested tools.

2.3 Representing a DGP (c)

Santer et al. (2008) use simulated data to criticize a data model proposed by other authors who claim to find phenomenon X from empirical data. Here, X is a discrepancy between the temperature predictions of global climate models and satellite data that climate skeptics (Douglass et al., 2008) claimed to detect, and the relevant data model is the statistical estimator d^* that they used. The point of using simulated data is to show that the data model used to arrive at X is faulty and hence they have failed to show that X is in the data.

The simulated data are generated in such a way that X is *not* part of the represented DGP. The main point is to show that the tested research tools (the estimator d^*) detect X under the simulated conditions even though X is not in the simulated DGP by design.

The simulated dataset represents accurately only some of the statistical properties of empirical data/DGP, but irrespective of whether X is to be found in empirical data, X is absent in the represented DGP.⁶ Yet, the epistemic credibility of these studies hinges on correctly mimicking other properties of the DGP. For example, Santer et al. calibrate a parameter (describing the intertemporal dependence of temperatures between the periods in the autoregressive dynamic stochastic DGP) with empirical data, to mimic the real DGP and the resulting data. Thus, in these cases, only certain parts of the DGP were accurately mimicked, whereas the DGP itself was only *represented* because it was crucial

⁵ The notion of *robust performance* is different from the more extensively discussed notion of *robustness of results*. Unlike the latter, the former is studied when the correct results are known, it does not require the sameness of any results, and the concept applies to a single research tool because the robust performance of a research tool can be established independently of other research tools. We use the notion of robust performance in the same way as the practitioners of benchmark data studies. We thank an anonymous reviewer for pointing out the need for clarification here.

⁶ In this case, it remains uncertain whether excluding X from the SDGP qualifies as an idealization. Santer et al. (2008) argue that X is absent in reality, which is irrelevant to the methodological foundation of their study. However, climate skeptics maintain the existence of X. Other studies, such as Hoover and Salyer (1998), which share the same structural framework as Santer et al. (2008), provide clearer instances of X as a distortion of reality. They demonstrate that an alternative data analysis method identifies technology shocks in the simulated data, even when such shocks are not explicitly included in the SDGP. In this context, the existence of technology shocks is more evident. The data analysis detects them even during periods when they do not occur. Hoover and Salyer (1998) explicitly acknowledge that they are not attempting to mimic every aspect of the Data Generating Process: "It is not the case that we have created an artificial world in which there are artificial analogues to true technology shocks" (p. 313).

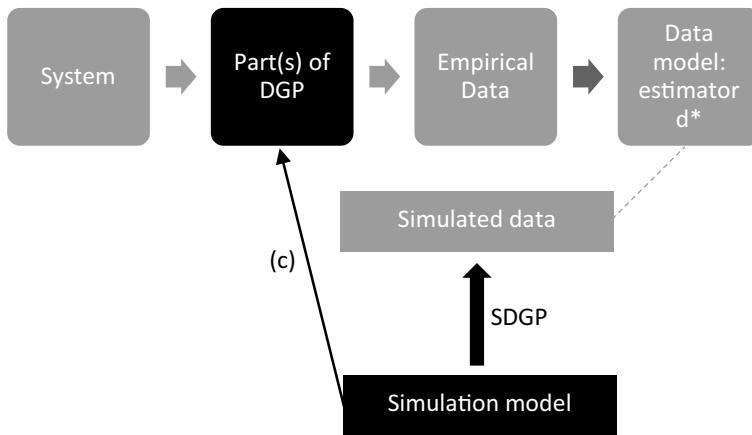


Fig. 4 A simulation model represents a DGP (c). Simulated data are used as evidence for the improper functioning of a research tool (the grey dashed line). (Color figure online)

to demonstrate the improper functioning of research tools. The simulated DGP did not include X but the research tool found it notwithstanding. Figure 4 illustrates the case (c).

Philosophers have described other cases in which the computer simulation mimics a part of the DGP (Lusk, 2021; Massimi & Bhimji, 2015). While the simulation generates a benchmark in these cases as well, the difference to the case above is that the simulation model must mimic the real DGP more extensively. Simulated benchmark data are used to develop research tools (e.g., the ‘spectrum’ in Lusk, and the ‘background’ in Massimi & Bhimji) that help interpret features in empirical data as evidence of detecting a phenomenon. In these cases, the research tools themselves are part of the DGP. The accuracy requirements of mimicking are more stringent because these studies aim to demonstrate the proper rather than improper functioning of the research tools.

2.4 Simulated Data Mimicking Empirical Data (d)

The normative voting theory aims to determine which voting rules are the best in different circumstances concerning the numbers of voters and candidates, the range of options, and the institutional setting. There are several criteria for the performance of voting rules, but they all reflect the idea that the rules should select the best candidate, and they should not exhibit various sorts of paradoxical changes in the selected candidates when the electorate changes slightly.

Plassmann and Tideman (2014) sought to evaluate how often various voting rules generated problematic outcomes. Given the rareness of many of these problems, however, they needed to study millions of elections, but empirical data are available for only a few thousand elections. Their solution was to use a statistical Monte Carlo model that was calibrated with empirical data on votes. This allowed them to construct an amplified simulated dataset of a million simulated elections that preserved and replicated the relevant aspects

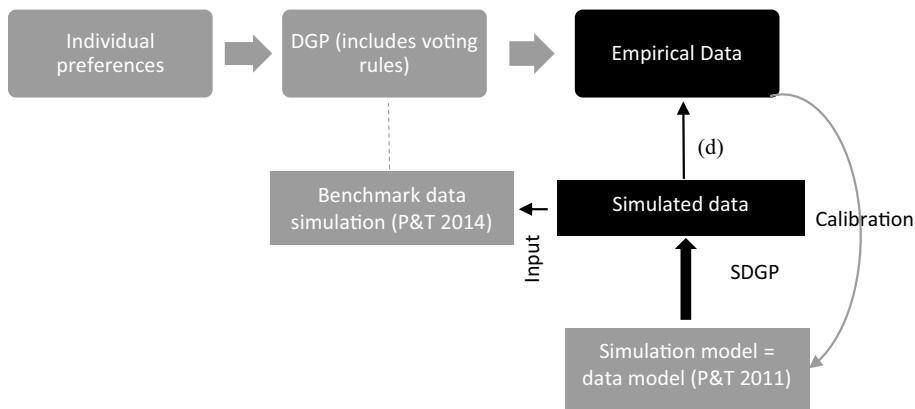


Fig. 5 Simulated data mimic empirical data (d) in Plassmann and Tideman (2011). Empirical data are used to calibrate the parameters of the simulation model. The SDGP consists of a Monte Carlo model that parameterizes a ‘spatial model’ expressed as a set of probability distributions. The simulated data are then used as input in the benchmark data study by Plassmann and Tideman (2014). The aim of this benchmark data study (indicated by the dashed line) is to evaluate which voting rules provide the best performance in translating individual preferences to voting results

of empirical data, especially the distributional properties.⁷ In this context, the relevant distributional properties of the data refer to the proportion of elections in which there is one clear winner, two close contestants, or three genuine competitors; and how often is there a consensus candidate who is not the best choice of many voters but who is the second-best for many.

The amplified datasets are similar to each other and empirical data in terms of their distributional properties, but they differ in detail. The amplified simulated data must mimic the relevant distributional properties of empirical data because the authors evaluate how often various problems are likely to occur under different voting rules.

One may think of individual preferences as the underlying system that generates empirical data here. The DGP could be taken to consist of those preferences, but filtered with the strategic behavior of voters,⁸ voting rules used in collecting the empirical data, as well as the selection procedures for putting together datasets from different elections.⁹ Mimicking the preferences is difficult if not impossible: if we knew the preferences of individuals, we would not need voting rules in the first place. However, Plassmann and Tideman (2011) did not try to mimic any other properties of the DGP either. Instead, Plassmann and Tideman (2011) *sought to mimic* the distributional properties of the existing *empirical data* on voting (d) (see Fig. 5). Constructing such simulated datasets yields information about how slightly different yet realistic datasets might affect the performance of various voting rules.

⁷ Their calibration procedure is described in an unpublished paper: Plassmann and Tideman (2011): “How to Predict the Frequency of Voting Events in Actual Elections”, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1911286.

⁸ Voting strategically (i.e., misrepresenting one’s preferences) has two important implications. First, it makes it impossible to find out voters’ preferences reliably. Second, even if one knew the preferences, they could not be aggregated directly into voting outcomes.

⁹ For example, some rules only collect information on the best candidates, but to be included in the empirical data requires information on the full rankings for three-candidate elections.

Although the simulation model includes a SDGP, it does not aim to mimic the real one, not merely because it does not aim to account for strategic voting or the selection procedures in the empirical dataset, but also because it represents the DGP as if it were random, even though the real DGP is not random in this sense. Note also that the sense in which Plassmann and Tideman (2011) ‘calibrate’ their model with empirical data is different from that of Santer et al. (2008). While the latter insert an estimated parameter value directly into the simulation model, the former use the similarity of the empirical and simulated datasets (measured with the Kullback–Leibler divergence) as the performance criterion for the Monte Carlo model. Several parameters of the Monte Carlo model are then selected based on their performance. In other words, here the distributional properties of the empirical data function as the benchmark. Unlike Santer et al.’s parameter, these parameters do not have a counterpart in the empirical data, and we do not even know how they would be empirically interpreted. This is why they cannot be said to mimic the DGP.

3 Diversity of Mimicking Relations and Relata

Simulated data are always generated by representing something and assessing whether such data are reliable requires evaluating whether that part of the representational relation that requires accurate mimicking is successfully mimicked. Table 1 summarizes the relevant mimicking relations/cases.

We now explain why this fourfold division helps in understanding the epistemology of simulated data. This, in turn, requires clarifying the differences in the mimicking relations. The criterion for distinguishing between the four different cases is that their mimicking relations or relata are different.

In general, the term ‘target system’ refers to whatever a model is about, it being implied that it is in some sense ‘out there’ (Elliott-Graves, 2020). Recall that we use the broader notion of a ‘target’ because some models are not about one system (b) and some are not about systems at all (d). Yet, even when data constitute the target for a model, they are ‘out there’ in the sense that they are what the simulated data aim to mimic and what the model is about. Consider first the difference between mimicking a system (a) and mimicking several specific systems (b).

A model applies *generically* to a system if it describes the properties that define the target and if the system possesses those target-defining properties (Lehtinen, 2021). Given that a target may consist of a set of similar but not identical systems, a model may also apply to other similar systems if it applies generically to a target. For example, the Lotka–Volterra predation model applies generically to sharks and sardines, but also to foxes and rabbits, in virtue of the fact that its target consists of predator–prey systems. A model applies *specifically* to a system if it describes some of its system-specific properties (Lehtinen, 2021). The Lotka–Volterra model does not apply specifically to sharks and sardines (unless the target populations are further specified to have specific features of sharks and sardines).

Let us now apply Lehtinen’s distinction to distinguish between (a) and (b). When a system is mimicked (a), we can call it a ‘target system’, and the simulation model may apply to it either generically or specifically. If a target system is mimicked and the simulation model applies to it generically, the modeler aims to correctly represent the target-defining properties. It follows that a model may apply to several systems that one can distinguish from each other with respect to their system-specific characteristics. If, in contrast, the

Table 1 Mimicking relations

What is mimicked	A system (a)	Several specific systems (b)	Part(s) a DGP (c)	Existing empirical data (d)
What does the mimicking Case	Model Data assimilation in climate science	Model Metapopulation parameter estimators	Model Climate skeptics' estimator d^*	Data Voting rules
Purpose	Produce simulated data mimicking one target system	Test research tools under simulated data conditions mimicking specific and diverse systems	Show the failure/success of a research tool or interpret empirical data under partially mimicked simulated data conditions	Create a research tool that can mimic and augment empirical data

simulation model aims to mimic the characteristics of a target that only encompasses one system, then the model applies specifically to that system. To put it differently, the model is required to mimic the properties of only one system. In contrast, in case (b), the simulation model must apply specifically to several systems because it must describe the system-specific properties of more than one system.

The difference between (a) and (b) does not lie in the *number* of mimicked systems because a model may apply to many systems generically in virtue of mimicking the target-defining properties. It is rather that, in case (b), the simulation model must mimic some *specific properties of more than one system*, that is, properties that are *not shared* by all the systems to which the model applies generically. The difference between (a) and (b) thus lies in the fact that evaluating whether mimicking several specific systems succeeds (i.e., b) cannot be conducted by comparing the simulated system with a single system because the simulation model must mimic the specific characteristics of several systems, characteristics that are not shared by all the systems (that constitute the target in virtue of sharing the target-defining properties).

Consider now the difference between *mimicking* one system (a) and *representing* a DGP (c). A given system may give rise to several DGPs. Therefore, it is possible to mimic a DGP without mimicking the system that gives rise to it, and vice versa. However, we do not use these differences in the relata in constructing our four-fold categorization of mimicking relations because there is no epistemic difference between cases that mimic a DGP (see Tal, 2011 for an example) and cases that mimic a system (viz. case (a)). This is why we have not treated mimicking a DGP as an independent mimicking relation. Since the only difference between the two is that they have different relata, mimicking a DGP can be treated as an instance of (a). Put in another way, in cases of mimicking a DGP or a system, traditional philosophical literature on modeling would posit that a 'target system is mimicked.' The difference between (a) and (c) is rather that, when a target system is mimicked (a), modelers are concerned with mimicking all the properties that define the target as *accurately* as possible. When they are *representing* a DGP (c), they only *aim to mimic a part or parts of the DGP as accurately* as possible, whereas other parts are represented less accurately and sometimes even misrepresented. Cases (a) and (b) thus have different representational relations: in (c) the target is represented (and partially mimicked) whereas in case (a) the target is completely mimicked.

In typical cases of benchmark data simulation, a research tool is applied to simulated data that a SDGP produces. The output of the research tool can then be compared to the benchmark data that the SDGP produced. The benchmark property in the simulation model may represent its target correctly, which holds for cases (b) and (d). However, in case (c), it mustn't, since it is crucial for the testing of a research tool that the target (i.e., the DGP) is "only" represented, especially if the aim is to show the improper functioning of a research tool.

In contrast to (c), in case (b), the tools are tested to find out how robustly they perform under simulated data conditions that mimic *diverse targets*. In typical benchmark data simulations, trying to represent one DGP (c) or mimic a target system (a) responsible for empirical data is beside the point because the performance of the research tools typically depends on their ability to track data produced by different DGPs or specific systems.

The tested tools should be robust with respect to finding the relevant phenomena in different simulated circumstances that cover diverse specific systems or many DGPs because researchers want their tools to function reliably in different empirical circumstances. Benchmark data simulations provide evidence about the reliability of research tools only if

the simulated specific systems or DGPs include the relevant characteristics of the empirical systems or DGPs to which the tools are applied. If the simulated specific systems or DGPs are not relevantly diverse, the good or robust performance of the tools in benchmark data simulations may be just an illusion, and the evidential support of simulated data can be questioned.

In case (d), the simulation model first (stage 1) generates simulated data that aims to mimic empirical data. Then (stage 2) the simulated data are used to study how data-aggregation mechanisms (voting rules) perform in analyzing the simulated data that now provide the benchmark. To evaluate whether stage 2 provides reliable information about the data-aggregation mechanisms (i.e., how well voting rules perform under simulated data conditions) requires evaluating whether the stage 1 simulation model mimics empirical data accurately. The target (in the sense of what one is interested in mimicking accurately) in the first stage is the empirical dataset, but in the second stage, the relationship between the data-aggregation mechanisms and the simulated benchmark data constitutes the target.

In cases (a–c), the crucial epistemic relation is whether the simulation *model* successfully mimics the target system, specific systems, or a part of the DGP. Iff this mimicking succeeds, then the simulated *data* could be said to successfully mimic targets. The fact that this is not true of cases, such as (d) explains why we need an epistemology of simulated *data* which goes beyond the traditional epistemology of simulation *models*.

4 Simulated and Empirical Data

According to extant accounts of simulation and simulated data, case (d) does not qualify as simulated data. While all extant definitions of simulation specify that simulations must represent something (see Sect. 5), none of them accommodates Monte Carlo models because these *models* do not necessarily represent or mimic. Given that case (d) involves a relevant mimicking relation, but the relation concerns simulated data rather than the model, we need a definition of simulation that can accommodate case (d). Moreover, in case (d), empirical data is used to calibrate a model that generates (simulated) data. However, some philosophers, such as Lusk (see below), could claim that these data are empirical. We thus have two philosophical problems to solve: distinguishing between simulated and empirical data in such a way that case (d) exhibits the former and providing a definition of simulation that can accommodate the Monte Carlo models in the cases discussed, including cases (c) and (d). These problems are interrelated because they both require distinguishing between computation and simulation.

Data are typically defined as something given (lat. *datum*) to us in observation, or as that which registers on a measurement or recording device in a form that is accessible to humans (e.g., Woodward, 1989, p. 394). Barberousse et al. (2009) define *empirical data* (' data_E ') as being of 'empirical origin, namely produced by physical interactions with measuring or detection devices' (p. 560).

The origin of *simulated data* lies in a computer simulation model. Consequently, simulated data are not what nature 'gives' us because they are the result of calculating the consequences of what is in the computer code (Beisbart, 2018). A simulation is not in causal interaction with whatever the simulated data are taken to represent.

The problem with origin-based definitions of data is that they do not always allow a clear differentiation between simulation and computation. It is difficult to distinguish

between a simulation that uses empirical data as input, and a measurement that involves the computational refinement of empirical data (Arnold, 2013). Calculating an average for a variable from empirical data with a computer does not change its empirical status, but data may be processed and transformed more extensively, which could even change what the data are about (Humphreys, 2013).

Arnold (2013, pp. 31–36) proposes that data are empirical if the output data from a simulation are interpreted to concern the same variables that are responsible for the empirical input data. Arnold's proposal fails because data assimilation for climate models (Section 3.1) takes empirical data about some variables as input, and then replaces the values of the same variables with simulated values (see e.g., Parker, 2017; Werndl, 2019). Even though the input data and the output data concern the same variables (and the same spatiotemporal units), the input data are empirical whereas the output data are simulated because the global climate model was used in generating them, and that model is a simulation according to all extant definitions.

Lusk (2016, pp. 148–151; see also Barberousse & Vorms, 2013) argues that data produced by a computer (simulation) model are always empirical as long as the model has any empirical data as input. Lusk does not try to distinguish between computation and simulation and does not limit the transformations of data in any way. He also argues that the epistemic properties of data ought to be the same irrespective of whether their initial collection and processing are included in the activity of simulation. We take him to mean that, for the epistemic properties of data, it does not matter whether the transformations are made inside a simulation that takes empirical data as input, or before entering the simulation when the empirical data were collected or processed. While we agree with Lusk that simulated data may derive their epistemic importance from empirical data from which they are generated, we do not think it makes sense to argue that those data (viz. the output data) are also empirical irrespective of what kind of transformations are conducted on the empirical data. When a simulation model is calibrated with empirical data, the empirical data are typically used in *modeling* the DGP as in case (c), the target system as in case (a), or simulated data as in case (d), whereas when empirical data are filtered or corrected, they typically retain their status as *evidence* about a target system. Lusk's proposal forces one to ignore this crucial epistemic difference.

Consider some characterizations of data to clarify the issue. According to the “representational” conception of data, data have representational content in the sense that they instantiate some of the properties of a target of investigation (Barberousse & Vorms, 2013; Leonelli, 2015, 2016, 2019). Leonelli argues that if both models and data represent, it becomes impossible to distinguish between models and data, arguing instead for a “relational” account of data: any object that can be used as evidence for a claim about the world can be taken as data. Data are thus defined by their *epistemic* properties as evidence for something. We agree with Leonelli that data do not always have a representational function. Neither, however, do they always function as evidence.

Consider the example we discussed in Sect. 2.3 to see the problem more clearly: Santer et al. (2008). The point of their Monte Carlo model was to provide an artificial DGP that aims to mimic some but not all aspects of the real DGP. The data were generated with a simulation model which is calibrated with parameters estimated from empirical data. This parameter *mimics* (and thereby also represents) the relevant distributional properties of the empirical DGP. The data generated by the Monte Carlo model could not be used as evidence about the target system, however, because the SDGP that generates them deliberately misrepresents some *other* aspects of the empirical DGP (see Sect. 2.3). The data which

were originally collected for *evidential* purposes have now assumed a *representational* function within the model.

We propose to distinguish between computation and simulation as follows (cf. Lehtinen & Kuorikoski, 2007). A model simulates if its epistemic credibility depends on successfully mimicking, and it calculates or computes if its credibility does not depend on mimicking. The epistemic credibility of a simulation always depends on whether it or the data it generates successfully mimics its target. A model is computational if it does not (aim to) mimic anything, and if the data it generates do not mimic empirical data. For example, calculating an average of empirical data counts as computation because its epistemic credibility does not depend on any mimicking relation. It depends, instead, on whether the mathematical calculation is correctly carried out. Empirical data thus remain empirical after being computationally processed with a computer if the processes do not include mimicking.

In contrast, the model of Santer et al. is a simulation because it aims to mimic part of the empirical DGP. It is just that the mimicking is done with the parameter estimated from empirical data. While the transformation from empirical data to an estimated parameter may well be a matter of mere computation, the data generated by Santer et al.'s Monte Carlo model calibrated with this parameter are *simulated* rather than empirical because, according to our distinction, this Monte Carlo model is a simulation. After all, it mimics parts of the DGP.

Calling the generated data empirical in virtue of the fact that the model takes empirical data as input via the estimated parameter would be misleading because the primary function of empirical data here is not to provide evidence about the climate. Although the empirical input data were originally collected to provide evidence, after the repurposing (Bokulich & Parker, 2021), they no longer have that role. The simulated output data are used as evidence about the performance of a data model (the estimator d^*), whereas the empirical input data had a representational rather than an evidential function because it was used for mimicking the empirical DGP.

However, it is worth bearing in mind that using some data as evidence for some real phenomena does not define those data as empirical or simulated. After all, there are plenty of simulation models that do not take any empirical data as input, but that aim to provide data that are to be used as evidence for empirical phenomena. They generate simulated data.

Our examples in Sect. 2 included Monte Carlo models. First, the Monte Carlo simulation *model* aims to *represent the DGP* to produce benchmark data (Sect. 2.3). Second, if there are some empirical data, but not sufficient for the purpose at hand, the *data* produced by a Monte Carlo simulation may mimic *the relevant properties of existing empirical data* (Sect. 2.4). The distinction between simulation and computation allows us to accommodate case (c) in Sect. 2.3. The distinction itself, however, does not yet accommodate case (d).

More importantly, it could be argued that case (d) does *not* qualify as an example of simulated data according to extant definitions because the mimicking is not done by the model, but rather by the data that the Monte Carlo model generates.

5 A Revised Definition of Simulation

We now provide a definition of simulation that can accommodate case (d). Recall that we require that the epistemic credibility of a simulation hinges on the success of a representational relation. The basic idea is to expand the scope of applicability of extant definitions by dropping the requirement that the representational relation must hold between the simulation model and the physical target system.

Hartmann's (1996, pp. 83–84) definition of simulation refers to imitating a process with another process, whereas Humphreys (2004, p. 110) expands the scope of this definition by taking into account the possibility that a simulation may mimic a static object rather than a process: 'System S provides a core simulation of an object or process B just in case S is a concrete computational device that produces, via a temporal process, solutions to a computational model [...] that correctly represents B, either dynamically or statically. If in addition the computational model used by S correctly represents the structure of the real system R, then S provides a core simulation of system R with respect to B.' The system S, a computer simulation model, is a program that runs on a computer (see also Winsberg, 2015).

Imbert's (2017, p. 739) working definition reads as follows: 'A computer simulation corresponds to the actual use of a computer to investigate a physical system S, by computationally generating the description of some of the states of one of the potential trajectories of S in the state space of a computational model of S.'

It would seem to follow that the simulation model must mimic the target system which is interpreted as a spatiotemporal object or process (cf. Peschard, 2019). There are two interrelated problems with these definitions. First, the definitions restrict the targets of simulations to physical target systems, but simulated data are often generated with models that do not aim to mimic a physical target system. What is being mimicked or represented instead can be a DGP, the data, a data model, or several specific systems. To accommodate them, the definition of computer simulation must be broadened to include such relata.

Second, some *models* commonly taken to be simulations do not aim to mimic their targets. Monte Carlo models are a common case in point because the randomness on which the method is based is not normally meant to be a claim about the object or process of interest (Beisbart & Norton, 2012). The randomness is produced with a pseudorandom number generator (PRNG). Although some philosophers (e.g., Hartmann 1996, pp. 87–88; Lenhard, 2016, p. 722; Morrison, 2015, p. 213) and many scientists refer to Monte Carlo models as "simulations," the above definitions cannot accommodate Monte Carlo models because the models themselves do not necessarily mimic. Moreover, Grüne-Yanoff and Weirich (2010, p. 30) argue that all Monte Carlo models lack the mimicking aspect and therefore should be counted as computations, not simulations.

We propose, however, that the best way to account for why Monte Carlo models are often counted as simulations is because of their mimicking aspects. However, Grüne-Yanoff and Weirich are partially right because Monte Carlo models are a diverse collection of different approaches, and while some are simulations, some are computations. An updated definition of simulation is thus needed because the definition should allow for the judgment that some Monte Carlo models are simulations even though the model or its random components may not mimic any target. On the other hand, the definition should clarify why not all the Monte Carlo models are simulations because they lack mimicking aspects, thus explaining why philosophers have had divergent intuitions concerning the status of Monte Carlo models.

Although the data generated by PRNGs always mimic the corresponding analytical probability distributions, we do not appeal to *this* mimicking relation in arguing that some Monte Carlo models count as simulations. It is epistemically relevant only if it fails, and if the failure affects the relevant results calculated with pseudorandom numbers (see Lenhard, 2019). A Monte Carlo model may involve several different components, some of which aim to represent the target, while others like the PRNG, do not aim to represent. If a Monte Carlo *model* has a relevant mimicking relation, it holds between the broader simulation model that *embeds* the pseudorandom generator, and the target. The simulation model of Santer et al. (Sect. 2.3) is a case in point because the non-random properties of their model (specifically, the parameters calibrated with empirical data) mimic parts of the DGP. However, in some cases, it is the *data* produced by the Monte Carlo simulation that *mimic* the empirical data (Sect. 2.4). The target of such simulation models is a dataset. This is one reason why a definition of simulation cannot require that a simulation must represent a dynamic process: a dataset is a static object.

With the necessary components at hand, we can now provide a definition of computer simulation that can be applied to the relevant cases, including Monte Carlo models. Humphreys' definition applies to cases (a), (b), and (c), but it does not cover case (d).¹⁰ We take his definition as a starting point, but drop the part that refers to physical systems and simplify by substituting a target T for objects and processes:

(Def) System S provides a simulation of a target T just in case S is a concrete computational device that produces, via a temporal process, solutions to a computational model that aims to represent the target, or if T is an empirical dataset, then the data produced by the model aims to represent T.

This definition of simulation is not restricted to representing physical target systems, it accommodates static targets, including datasets, and it includes the requirement that a simulation—the model or data it produces—must aim to represent the target. This definition captures case (d) because the simulated data aim to mimic the empirical data. The Monte Carlo model in this case can thus be called a ‘simulation model’ even though the model does not mimic because the data the model generates mimic something. Moreover, this definition accommodates different degrees of accuracy of representation. Although the ‘target’ is expressed in singular in the definition, it is to be interpreted in such a way that case (b) can be accommodated; if one must mimic several specific systems, those systems constitute the target.

Let us finally return to the issue of distinguishing between empirical and simulated data. We have seen that when a simulation model is calibrated with empirical data, they are used in modeling something. When empirical data are computationally processed, they retain their status as evidence about the target system. However, even though this distinction is relevant for some models, there are cases in which this difference is insufficient for determining the status of the data that a model generates. Consider, for example, Bowden et al.’s (2006) Monte Carlo estimator. It corrects empirical data by taking them as input and then fills in data points randomly, but such that ‘publication bias’ is corrected. The data that this estimator generates retain their status as evidence about the target system, but it would not be altogether wrong to deny that they are empirical because they are generated with a random Monte Carlo process that is not in causal interaction with the target system.

¹⁰ The definitions of Hartmann (1996), Imbert (2017), and Parker (2009) do not apply to (b) because they require the targets to involve dynamic change.

Whether such computational data are appropriately called simulated or empirical is somewhat interesting but not particularly important for their epistemic evaluation. To see this, note that, for example, Stanley and Doucouliagos' (2017) estimator carries out a similar task as Bowden et al.'s estimator without generating any data before calculating its value. It is thus a clearly computational method that does not alter the status of empirical data. The reliability of the data generated by Bowden et al. does not derive entirely from empirical data, and the very reason for using such Monte Carlo methods is to correct problems of empirical data by generating data that takes the errors in the DGP into account. According to our definitions, Bowden et al.'s model is not a simulation (despite that they describe it as one) because it does not mimic the target, and the empirical dataset is not a target that the data aim to mimic.

We have supplied a definition of simulation models and simulated data to justify that our cases count as 'simulated data' in the first place. However, from the point of view of epistemology, what is more important than whether this or that dataset counts as simulated, is recognizing the relevant mimicking relation. But the example of Monte Carlo estimators shows that the epistemic importance of simulated data extends beyond cases in which simulated data appears! This is because benchmark testing with simulated data is commonly applied to computational methods. Even though Bowden et al. do not generate simulated data, the epistemic credibility of their estimator depends on an earlier benchmark test with simulated data. The relevant mimicking relation lies in whether this earlier benchmark data simulation correctly described the properties of the DGP to which the Monte Carlo estimator is applied.

6 Conclusions

We have argued that to account for the diverse epistemic and evidential uses of simulated data in contemporary empirical science, we should relax the requirement of traditional definitions of computer simulation that computational *models* must mimic something. Instead, it is sufficient if the *data* produced by such a model mimic something. This allowed us to accommodate some Monte Carlo models as simulations. Distinguishing between computation and simulation because the latter but not the former involves mimicking something enables defining simulated data as having its origin in a simulation model.

We considered various cases in which a simulation model or simulated data mimic something and identified four mimicking relations, differing in their relata or the representational relation. These relations are relevant for evaluating the epistemic credentials of simulation methods and the data produced. The common denominator in the cases was that the available empirical data, research tools, or experiments were deficient in some way, and simulation methods and the data they generated were employed to make them more reliable or relevant.

Given that no account studies the variety of mimicking relations, it seems to us that philosophers have taken for granted that simulation models—and scientific models in general—have one privileged mimicking relation that has the same representational relation and relata across different cases. Typically, they have focused on the idea that simulation models should correctly mimic their target system (a). We have suggested that this is not a privileged mimicking relation. Moreover, in case (d), it is not the simulation model but rather the *data* it produced that is intended to mimic some properties of existing empirical data. It is impossible to understand the epistemology of these cases without paying attention to the special features of simulated data.

Recognizing the relevant mimicking relation is the starting point for evaluating the evidential and epistemic relevance of the use of simulated data in empirical science. Mimicking matters but the fact that a given simulation *model* fails to mimic anything may be epistemically irrelevant.

Funding Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). This study is supported by Nankai University Arts and Humanities Development Funds (ZB21BZ0218) (Aki Lehtinen) and Suomen Kulttuurirahasto (00220848), (Jani Raerinne).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Antoniou, A. (2021). What is a data model? *European Journal for Philosophy of Science*, 11(4), 101.

Arnold, E. (2013). Experiments and simulations, do they fuse? In E. Arnold & J. Duran (Eds.), *Computer simulations and the changing face of scientific experimentation* (pp. 46–75). Cambridge Scholars Publishing.

Barberousse, A., Franceschelli, S., & Imbert, C. (2009). Computer simulations as experiments. *Synthese*, 169(3), 557–574.

Barberousse, A., & Vorms, M. (2013). Computer simulations and empirical data. In M. Juan (Ed.), *Computer simulations and the changing face of experimentation* (pp. 29–45). Cambridge Scholars Publishing.

Beisbart, C. (2018). Are computer simulations experiments? And if not, how are they related to each other? *European Journal for Philosophy of Science*, 8(2), 171–204.

Beisbart, C., & Norton, J. D. (2012). Why Monte Carlo simulations are inferences and not experiments. *International Studies in the Philosophy of Science*, 26(4), 403–422.

Beisbart, C., & Saam, N. J. (2019). *Computer simulation validation*. Springer.

Bokulich, A. (2020). Towards a taxonomy of the model-ladenness of data. *Philosophy of Science*, 87(5), 793–806.

Bokulich, A. (2021). Using models to correct data: Paleodiversity and the fossil record. *Synthese*, 198, 5919–5940.

Bokulich, A., & Parker, W. (2021). Data models, representation and adequacy-for-purpose. *European Journal for Philosophy of Science*, 11(1), 31–31.

Bowden, J., Thompson, J. R., & Burton, P. (2006). Using pseudo-data to correct for publication bias in meta-analysis. *Statistics in Medicine*, 25(22), 3798–3813.

Contessa, G. (2007). Scientific representation, interpretation, and surrogate reasoning. *Philosophy of Science*, 74(1), 48–68.

Douglass, D. H., Christy, J. R., Pearson, B. D., & Singer, S. F. (2008). A comparison of tropical temperature trends with model predictions. *International Journal of Climatology*, 28(13), 1693–1701.

Durán, J. M. (2018). *Computer simulations in science and engineering*. Springer.

Edwards, P. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. MIT Press.

Elliott-Graves, A. (2020). What is a target system? *Biology & Philosophy*, 35(2), 28.

Grüne-Yanoff, T., & Weirich, P. (2010). The philosophy and epistemology of simulation: A review. *Simulation & Gaming*, 41(1), 20–50.

Hanski, I. (1994). A practical model of metapopulation dynamics. *Journal of Animal Ecology*, 63, 151–162.

Hartmann, S. (1996). The world as a process: Simulation in the natural and social sciences. In R. Hegselmann, U. Müller & K. G. Troitzsch (Eds.), *Modelling and simulation in the social sciences from the philosophy of science point of view* (pp. 77–100). Kluwer.

Hoover, K. D., & Salyer, K. D. (1998). Technology shocks or coloured noise? Why real-business-cycle models cannot explain actual business cycles. *Review of Political Economy*, 10(3), 299–327.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press.

Humphreys, P. (2014). X-ray data and empirical content. In P. Schroeder-Heister, G. Heinemann, G. Hodges & P.E. Bour (Eds.), *Logic, methodology and philosophy of science* (pp. 1–15). College Publications.

Humphreys, P. (2013). What are data about? In J. Durán & E. Arnold (Eds.), *Computer simulations and the changing face of experimentation* (pp. 12–28). Cambridge scholars publishing.

Imbert, C. (2017). Computer simulations and computational models in science. In P. Schoroeder-Heister, G. Hodges, G. Heinemann, & P. E. Bour (Eds.), *Springer handbook of model-based science* (pp. 735–781). Springer.

Lehtinen, A. (2021). The epistemic benefits of generalisation in modelling I: Systems and applicability. *Synthese*, 199(3–4), 10343–10370.

Lehtinen, A., & Kuorikoski, J. (2007). Computing the perfect model: Why do economists shun simulation? *Philosophy of Science*, 74(3), 304–329.

Lenhard, J. (2016). Computer simulation. In P. Humphreys (Ed.), *The Oxford handbook of philosophy of science* (pp. 717–737). Oxford University Press.

Lenhard, J. (2019). *Calculated surprises: A philosophy of computer simulation*. Oxford University Press USA - OSO.

Leonelli, S. (2015). What counts as scientific data? A relational framework. *Philosophy of Science*, 82(5), 810–821.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. The University of Chicago Press.

Leonelli, S. (2019). What distinguishes data from models? *European Journal for Philosophy of Science*, 9(2), 1–27.

Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15, 237–240.

Lusk, G. (2021). Saving the data. *The British Journal for the Philosophy of Science*, 72, 277–298.

Lusk, G. (2016). Computer simulation and the features of novel empirical data. *Studies in History and Philosophy of Science*, 56, 145–152.

Massimi, M., & Bhimji, W. (2015). Computer simulations and experiments: The case of the Higgs boson. *Studies in History and Philosophy of Science Part b: Studies in History and Philosophy of Modern Physics*, 51, 71–81.

Moilanen, A. (1999). Patch occupancy models of metapopulation dynamics: Efficient parameter estimation using implicit statistical inference. *Ecology*, 80, 1031–1043.

Morrison, M. (2015). *Reconstructing reality: Models, mathematics, and simulations*. Oxford University Press.

Nguyen, J. (2020). It's not a game: Accurate representation with toy models. *The British Journal for the Philosophy of Science*, 71(3), 1013–1041.

Parke, E. C. (2014). Experiments, simulations, and epistemic privilege. *Philosophy of Science*, 81(4), 516–536.

Parker, W. (2009). Does matter really matter? Computer simulations, experiments, and materiality. *Synthese*, 169, 483–496.

Parker, W. S. (2016). Reanalyses and observations: What's the difference? *Bulletin of the American Meteorological Society*, 97(9), 1565–1572.

Parker, W. S. (2017). Computer simulation, measurement, and data assimilation. *The British Journal for the Philosophy of Science*, 68(1), 273–304.

Peschard, Isabelle (2019): "Is Simulation an Epistemic Substitute for Experimentation?", in *Simulations and Networks*, edited by Sandro Vaienti. Paris: Hermann,

Plassmann, F., & Tideman, T. N. (2011). *How to predict the frequency of voting events in actual Elections*. Unpublished manuscript. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1911286

Plassmann, F., & Tideman, T. N. (2014). How frequently do different voting rules encounter voting paradoxes in three-candidate elections? *Social Choice and Welfare*, 42(1), 31–75.

Santer, B. D., Thorne, P. W., Haimberger, L., Taylor, K. E., Wigley, T. M. L., Lanzante, J. R., Solomon, S., Free, M., Gleckler, P. J., Jones, P. D., Karl, T. R., Klein, S. A., Mears, C., Nychka, D., Schmidt, G. A., Sherwood, S. C., & Wentz, F. J. (2008). Consistency of modelled and observed temperature trends in the tropical troposphere. *International Journal of Climatology*, 28(13), 1703–1722.

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, 8(1), 19–42.

Tal, E. (2011). From data to phenomena and back again: Computer-simulated signatures. *Synthese*, 182(1), 117–129.

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford University Press.

Werndl, C. (2019). Initial conditions dependence and initial conditions uncertainty in climate science. *British Journal for the Philosophy of Science*, 70, 953–976.

Winsberg, E. B. (2010). *Science in the age of computer simulation*. University of Chicago Press.

Winsberg, E. (2015). Computer simulations in science. In N. Z. Edward (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University.

Woodward, J. (1989). Data and phenomena. *Synthese*, 79, 393–472.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Aki Lehtinen is currently a Professor of Philosophy at Nankai University in Tianjin, China. He also has the title of docent in practical philosophy at the University of Helsinki, Finland. Professor Lehtinen received his PhD in 2007 at the Erasmus Institute for Economics and Philosophy, Rotterdam, The Netherlands. After completing his PhD studies, he has had various academic positions at different universities, such as the University of Helsinki, Finland, University of Bayreuth, Germany. His research interest include various topics, such as models and simulations, robustness analysis, confirmation theory, game theory and rationality, philosophy of economics, and voting and social theory. In addition, he has worked as a principal investigator for several research projects funded, for instance, by the Academy of Finland and University of Helsinki.

Jani Raerinne is currently a university researcher at the University of Helsinki, Finland. He received the PhD degree in Theoretical Philosophy at the University of Helsinki (Finland) in 2011. His research interests include various topics related to philosophy of science, such as scientific modelling, scientific explanation, progress in science, and androcentric biases in science. While his main research focus has been on philosophy of biology, he is also affiliated with TINT, Centre for Philosophy for Philosophy of Social Science <https://tint-helsinki.fi/>.