ORIGINAL RESEARCH

Traveling with TARDIS. Parameterization and transferability in molecular modeling and simulation

Johannes Lenhard[1] · Hans Hasse[1]

Abstract

The English language has adopted the word Tardis for something that looks simple from the outside but is much more complicated when inspected from the inside. The word comes from a BBC science fiction series, in which the Tardis is a machine for traveling in time and space, that looks like a phone booth from the outside. This paper claims that simulation models are a Tardis in a way that calls into question their transferability. The argument is developed taking Molecular Modeling and Simulation as an example. There, simulation models are force fields that describe the molecular interactions and that look like simple and highly modular mathematical expressions. To make them work, they contain parameters that are adjusted to match certain data. The role of these parameters and the way they are obtained is seriously under-appreciated. It is constitutive for the model and central for its applicability and performance. Hence, the model is more than it seems so that working with adjustable parameters deeply affects the ontology of simulation models. This is particularly crucial for the transferability of the models: the information on how a model was trained is like luggage the model must carry on its voyage.

Keywords Parameterization · Transferability · Computational modeling · Simulation · Molecular modeling

1 Introduction

Simulation models often can be successfully transferred to unexpectedly different contexts. This observation has motivated a growing discussion about the role computational templates, and other aspects of simulation modeling, play in model transfer (see Humphreys, 2019 for a starting point). This paper examines transferability, but looks at the matter from the opposite side, asking why and when transfer runs into

✉ Johannes Lenhard
johannes.lenhard@rptu.de

[1] Laboratory of Engineering Thermodynamics, RPTU, Kaiserslautern, Germany

problems. Our central claim is that predictive power of models as well as problems related to their transferability are linked by a common cause: the way how modelers employ adjustable parameters. Not only do these parameters play a major role in the modeling process, this role is seriously under-appreciated.

Models, such as those we study here, are based on relatively simple mathematical expressions that they can be written down in a few lines. From the outset, these expressions seem to be templates that are easily transferable to different applications. However, the picture changes when one moves to a finer level of analysis and accounts for the adjustable parameters these models contain. On this finer level, the models actually are much more complex entities and this affects how they can travel. The complexity of simulations has consistently been highlighted in the literature, with a special emphasis on black boxing and epistemic opacity. The present paper examines a class of cases where modelers work with parameters in a very open way, thus the terminology of black boxes and opacity does not seem appropriate. Philosophical terminology should reflect upon the difference that the level of examination makes. As a preliminary fix, we make use of popular culture. The English language has adopted a new word—Tardis—for something that looks simple from the outside but is much bigger and more complicated when inspecting it from the inside. The word comes from the BBC science fiction series Dr. Who, where a machine for traveling in time and space looks like a phone booth from the outside, but has spacious and complicated high-tech architecture on the inside. This paper shows in what sense molecular models are a Tardis.

We focus our examination on molecular modeling and simulation (MMS). It is widely used in many different fields of science and its features are typical for simulation modeling in general. Furthermore, it is suited to explain our arguments also to readers that are not experts in the simulation technique. This is of special importance as our argument involves the (relatively) fine level of parameterization. MMS is essentially about modeling properties of materials, which is done following a two-step recipe. The first step models the interaction of particles on the atomistic level via classical mechanics, i.e., by interaction potentials or (equivalently) force fields that cover relevant forces (like the van der Waals force). The second step consists in simulating a system containing many particles that interact as prescribed by these forces. Hereby, the number of particles can vary between a few hundred or thousand to many millions. In the simulation, it is observed how this many-particle system behaves and from that information its properties are derived.[1]

Two trends in MMS persist. The scope and precision of the simulation results improve permanently so that they often match data from precision measurements.[2] This made MMS a popular tool in science and engineering—with applications ranging

---

[1] MMS is used here as an umbrella term that covers a family of approaches that differ mainly in the way the second step, the simulation, is carried out. Popular techniques are Molecular Dynamics (MD) and Monte Carlo (MC) simulations. Both are linked to a statistical physics or thermodynamics point of view. In contrast, Molecular Mechanics (MM) originated from chemistry. MM is much more oriented at the structure of single molecules. However, using MMS as an umbrella term allows us to dispense with these differences.

[2] It is the success in matching data that we call the predictive power of a model. However, we are well aware that this use of the term prediction can be debated. In particular, the good fit can be a result of overfitting to the data at hand so that the promised predictive power is small or nil in other cases. Thus, what we call predictive power is shorthand for potential predictive power.

from studies of the strength of steel over the development of nanostructured adsorbents for more efficient gas cleaning to pharmaceutical research. At the same time, the community of users of the models is continuously increasing and so is the model traffic, to stay in the picture. Supposedly modular parts of one model are transferred to another, neighboring model—used and evaluated by another group.

Here is our argument on MMS in a nutshell. From the mathematical perspective, the models are formulated as a sum of potential functions, representing different forces. However, all these functions contain adjustable parameters, that are fitted to obtain good descriptions of certain target properties. Hence, their inner workings crucially depend on the choice of the form and the specification of the values for these parameters. When the modelers adjust parameters so that the models make good predictions, they create two sorts of dependencies. Firstly, one adjusted potential function (describing a certain type of interaction) gets dependent on the adjusted form of the other potential functions (describing the other pertinent types of interactions). Secondly, the entire force field gets dependent on the data that were used for the training, including the type of material and the conditions for which it was studied. Consequently, when different groups of modelers transfer (parts of) models and add more adjustments, this creates complicated path dependencies. All these dependencies tend to make the parameterized model a large and holistic object that is less easy to transfer than the modular math model promises. In short, the model is a Tardis.

The paper proceeds in four steps. The first step gives a primer to molecular modeling—without any technicalities beyond a description of how a simple potential might look like. The pivotal point of molecular modeling is the extra versatility that is added by adjustable parameters. Adjusting parameters, it is argued, takes many factors into account simultaneously: numerical approximation, compensation for missing details, and inaccuracies of assumptions about the relevant forces, among others. This versatility through parameterization is an extension of the classical (pre-computer-simulation) paradigm. Without this extension, one could hardly obtain good predictions.

The second step provides a brief account of the methodology of simulation modeling, with a particular focus on a feedback loop during the modeling process. Employing adjustable parameters thrives on this feedback loop. Hence simulation modeling fosters the significance of adjustable parameters.

The third step brings together MMS and simulation methodology. On the level of forces, everything is supposedly modular, for instance, the van der Waals force and the electric force caused by dipole moments of particles can be captured in separate and additive modules. However, things are different on the level of modeling practice because parameter adjustment renders these modules interdependent. The data used for the adjustment create further dependencies. This provides the core of our argument why models created by MMS are a Tardis.

In the fourth step, we enrich the picture by studying the social organization of the field and argue that the travel patterns of software changed. In a phase up to around 1990, researchers travelled to hardware and to expertise in software. After this time, the pattern inverted. Now software travels to the researchers who become users of software packages. We argue that this regime change in social organization brought under scrutiny the transferability of models, fostering our claim that models have become hard-to-specify, holistic objects.

The conclusion wraps up what to learn from the MMS case study and the Tardis argument. *What* travels? Not the elegant mathematical formulation of the model. Such kind of theoretical model does not make the predictions. In fact, adjustable parameters (and how their values are specified) have to be included. In sum, the ontology of simulation models differs profoundly from that of theoretical mathematical models. Furthermore, we argue, questioning transferability means challenging reproducibility—a topic that calls for substantial further research. Finally, the investigation sheds new light on the rationale of mathematization, reflecting a tension between the universality vs. particularity of the world.

2 Molecular modeling and simulation

Molecular modeling elaborates on the basic idea that macroscopic behavior of matter results from the interaction of small particles. Since Greek atomism, this idea found varying expressions, often characteristic of the scientific state of the art. According to Laplace, to name one influential example, the physical dynamics of the universe can be fully predicted if only one knows the starting conditions and is able to integrate over all interactions between the particles, i.e. one is able to solve (integrate) the Newtonian equations for all particles of the universe simultaneously. Of course, he was fully aware that the necessary mathematical capabilities were beyond the reach of human beings, but called for a superhuman power, the so-called Laplacian demon.

Molecular interactions emerged as subjects of scientific theory later in the nineteenth century when scholars like Boltzmann (1844–1906), Maxwell (1831–1879), and Gibbs (1839–1903) framed statistical thermodynamics. In a way, this theory reformulates the mathematical problem, replacing the gigantic number of molecular interactions by statistical summary. Importantly, this theory is able to interpret measurable, macroscopic values as statistical functions of (then unobservable) particles and their interactions.

The space for molecular modeling was further defined when in the early twentieth century quantum theory made it clear that the interaction of smaller entities, like electrons forming a bond, cannot be described by classical forces. However, the quantum theoretical treatment of systems with many molecules remains largely intractable even with the computational power that is available today. Molecular modeling occupies the space in between (sub-)atomistic quantum mechanics and continuum mechanics where the discrete nature of the molecules can be neglected. The recipe remained surprisingly stable: one models the interaction of the particles via classical forces and then computes the resulting behavior by numerically solving a large number of differential equations. In a way, MMS employs the computer to emulate the Laplacian demon.

Let us start a more detailed discussion by considering a simple case: the molecular modeling of argon. The argon atoms are spherical and the only relevant forces between them are those resulting from repulsion and dispersive attraction.[3] In principle, all argon atoms in a many-particle system interact, but modeling this turns out to be

---

[3] This attractive force is also called the van der Waals force.

Fig. 1 Graph of the Lennard–Jones potential function: Intermolecular potential energy as a function of the distance of a pair of particles. The graph shows the "potential well", i.e., a favored distance between two particles where attracting and repelling forces are in balance. As the particles also have kinetic energy and are not locked in this position, they move continuously, which is known as Brownian motion
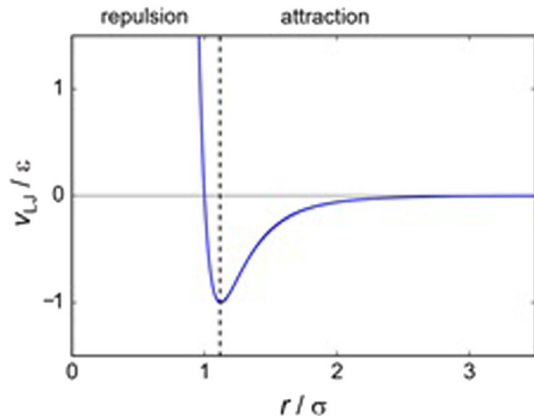
basically infeasible. A common simplification is to assume that the interactions in the system can be represented by pair-interactions, i.e. that it is sufficient to consider only interactions between two partners (which are then assumed to be independent of what the other atoms do).

However, what is the adequate mathematical form of the pair potential?[4] Finding suitable forms is far from trivial, even for the simple example of argon that is considered here, the ansatz can be formulated in various ways; but all of them contain parameters that have to be fitted to data. The most popular ansatz for doing this is the Lennard–Jones potential, named after the pioneer of computational chemistry, Sir Lennard–Jones (1894–1954) and is given in Eq. (1). This potential consists in the superposition of two exponential terms, the one (with the exponent m) controls how quickly the repulsive force rises when bringing two particles closely together, the other term (with the exponent n) expresses how quickly the attracting force decreases with increasing distance between the particles (r denotes the distance between them).

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^m - \left(\frac{\sigma}{r}\right)^n\right] \qquad (1)$$

Lennard–Jones proposed the exponents $m = 12$ for the repulsive term and $n = 6$ for the attractive term as adequate choices (Lennard–Jones, 1931). Basically, he chose $n = 6$ because Fritz London (1930) had calculated this exponent from quantum theoretical considerations. Given this choice, $m = 12$ (2*6) is simply convenient when using logarithmic tables for computation.[5] Having made these choices, the Lennard–Jones (12,6) potential has two remaining adjustable parameters (ε and σ), see Fig. 1. The parameter σ in the repulsive term can be considered as the diameter of the particle and the parameter ϵ in the attractive term describes the strength of the attractive interactions. These parameters have physical meaning, but this meaning is not independent from the parameterization scheme. The parameter values are fitted to thermodynamic

---

[4] Usually, the potential energy is modeled and the force is obtained by derivation. Hence the pair potential gives the force acting between these pairs.

[5] Rowlinson (2002) provides a wealth of original literature on the development.

data for argon, more precisely: they are chosen according to the overall fit to training data. The ansatz only becomes a model of argon by fitting its parameters to data for argon. Consequently, the resulting numbers will depend on the choice of that data and the way the fit is carried out.

In MMS, potentials are used like building blocks. There are different model building blocks that are put together to create a model of a complex structure. Lennard–Jones put together just two blocks to create his famous model, one for repulsion and one for dispersion. Add a dipole, and you will get another, more complex model, known as the Stockmayer potential, which has three parameters (σ, ε and the dipole moment). It is also common to combine several Lennard–Jones sites to describe chain-like molecules. However, for modeling more complex molecules, other types of interactions may become important. In general, one distinguishes between intermolecular interactions (between different molecules) and intramolecular interactions between the atoms inside a molecule, e.g., different types of vibration such as stretching, bending, or torsion. All these interactions are usually described by their potential energy, i.e., described by a potential.

$$V = V_{\text{intra}} + V_{\text{inter}} \qquad (2)$$

$$V_{\text{intra}} = V_{\text{stretching}} + V_{\text{Bending}} + V_{\text{Torsion}} + \ldots$$

$$V_{\text{inter}} = V_{\text{repulsion}} + V_{\text{Attraction}}$$

$$V_{\text{Attraction}} = V_{\text{VanderWaals}} + V_{\text{Polar}} + V_{\text{H-Bond}} + \ldots$$

The different contributions are then simply summed up—building the force field——to yield the total potential energy of the system. Being the sum of potentials, the force field looks perfectly modular and promises the transferability of building blocks.[6] However, important for our discussion, for each of the different contributions there are contribution-specific adjustable parameters.

Once the potential is defined, the potential energy of the system can be calculated for any configuration of the particles, the forces acting on the particles and the resulting motion can be simulated. The main task of the simulation is to generate a sufficient number[7] of representative configurations of the system to enable a meaningful determination of average properties. Such properties then can be compared to measurable macroscopic properties.

---

[6] These building blocks can be categorized in different ways. For instance, as a reviewer pointed out, van der Waals attraction between non-bonded atoms of the same molecule might count as intermolecular interaction. Our argument is independent of the precise categorization.

[7] There are several ways to generate the large number of representative configurations of the system that are needed for a meaningful averaging: two main routes can be distinguished: the deterministic route, which follows the ideas of Laplace, and the stochastic route. The most widely used techniques are Molecular Dynamics (MD) simulations in the former group and Monte-Carlo (MC) simulations in the latter.
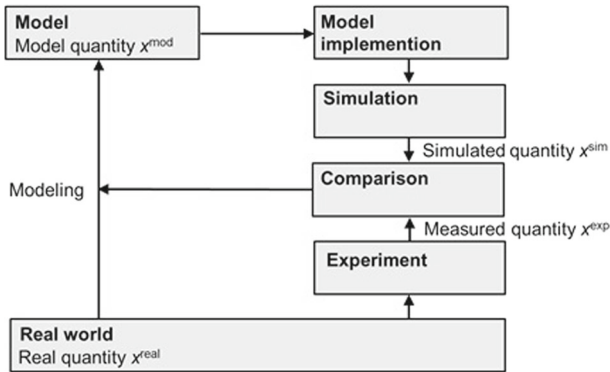
Fig. 2 Schema of simulation modeling, including a feedback loop. From Hasse & Lenhard (2017)

3 Adjustable parameters—the wild card of simulation

This section briefly recapitulates that simulation modeling is adding additional steps to the modeling process that make the relationship between simulated and target entities even more indirect. A feedback loop in the modeling process enables adjusting parameters, which play a key role.[8]

In the present account, the term model is used in the sense of a mathematical model that aims to describe certain aspects of physical reality, in particular measurable properties of a material like, e.g., density or viscosity. This model—think of a particular force field—is considered to be embedded in some kind of theory (here: thermodynamics, among others) that provides a frame not only for the modeling but also for carrying out physical experiments whose outcomes can be compared with those of the model. Figure 2 gives an overview of the picture on which the following discussion is based.[9]

We want to point toward a simple but momentous loop in simulation modeling. Models are usually imperfect and can get better through modification. Simulation modeling turns this caveat into an advantage by using this loop extensively.

According to our schema, researchers start with a quantity x_{real} they want to model, like the density or viscosity of argon. The corresponding entity in the model is x_{mod}. They might now choose to use the LJ (12, 6) model for this purpose. By this, we accept the host of simplifying assumptions on which that model is based. The parameters ε and σ are a priori unknown. For determining the property of interest x (say viscosity), they need to carry out molecular simulations: The model is implemented on a computer

---

[8] We follow Hasse & Lenhard (2017) who unfold a more detailed account of parameterization as "boon and bane". In general, the role of adjustable parameters is seriously under-examined in philosophy of science. Notable exceptions include Parker (2014) who focuses on the unavoidably incomplete representation in the context of climate modeling; Vincenti (1990) who identifies working with parameterization as a key element of engineering epistemology (though he is not arguing about computer models); and Kieseppä (1997) who argues about uses of parameters in statistics.

[9] Figure 2 is a version of a not uncommon account, for instance agreeing with R. I. G. Hughes' DDI account (1997), though enriched for simulation modeling. The philosophy of simulation has examined extensively the stretch in between x_{mod} and x_{sim} (cf. Winsberg 2019 for an entry point into this discussion).

and simulations are carried out. This involves many more choices, such as the selection of a simulation method (MD or MC). These simulations then yield a quantity x_{sim} that can be compared eventually to the results of experimental studies x_{exp}. In general, neither x_{real} nor x_{mod} can be known; only the corresponding properties x_{exp} and x_{sim} are known and can be compared. There is no direct access to x_{real} or x_{mod}.[10]

There are two types of experiment in play. One from "below" that is the experiment in the classical sense and provides measured values; the other from "above" that provides simulated values. The latter are often called computer experiments or numerical experiments. However, we prefer using the term "simulation experiment" (or briefly just simulation) here. Such experiments are used to investigate the behavior of models. Importantly, relevant properties of simulation models can be known only through simulation experiments. They play a fundamental role in MMS, because most properties x of interest are unobservable otherwise.

The adjustment of parameters is of fundamental importance to our example. The adjustment of ε and σ requires iterated loops of comparison and modification. The assigned values determine the model's identity. Else, the LJ model does not describe argon.

Not all model parameters need to be adjusted in the control loop shown in Fig. 1, it is common to set some of them a priori. Consider the Lennard–Jones potential (Eq. 1) as an example. It has four parameters: m, n, σ, and ε. Two of them were set a priori by Lennard–Jones; $n = 6$ was set on physical grounds, whereas $m = 12$ was set on entirely pragmatic grounds, as explained above. In fact, it is known from quantum chemistry that $m = 12$ is a poor choice for describing repulsive interactions. Despite this, the Lennard–Jones (12, 6) potential has become the by far most popular molecular model building block and the choice $m = 12$ has survived, long after the pragmatic reasons for this choice have become obsolete. How can this be?

The answer highlights the compensating role of adjustable parameters. In the practical application, the physical flaws of this choice are compensated by the fact that the parameters ε and σ are fitted to data in the feedback loop. By this very procedure, the parameter adjustment compensates for all kinds of imperfections in the model (including mistakes), such as neglecting many-particle interactions, or simplifications made in the derivation of the ansatz for describing the pair-interactions, or, finally but not exhaustively: "With judicious parameterization, the electronic system is implicitly taken into account." (Bowen & Allinger, 1991, p.82).

That adjustable parameters compensate for all sorts of factors is of central importance to our paper. Finding values so that the model (more or less) fits to existing results indirectly deals with these unknowns. Parameterization schemes are architectures guiding actions of this type. They can be considered as a sort of formal

---

[10] Cf. Edwards (2010) on the "model-laden data", and the now classic volume on "Models as Mediators" (Morgan & Morrison, 1999). The second variety ("from below" in Fig. 2) is the experiment in the classical sense. When comparing simulations to their target system, such classical experiments will usually provide the data to compare with. However, the situation becomes complicated in an interesting way because of the growing influence of simulation on these experiments—and thus on measurement (cf. Morrison, 2009, 2014; Tal, 2013). This is underpinned by the fact that researchers in MMS begin using hybrid data sets, composed of experimental and simulation data for developing models (cf. Forte et al., 2019).

construction that is used intentionally to deal with missing knowledge and the inaccuracies of existing knowledge. The force field is *designed* to contain parameters that can be adjusted over the course of further development. Hence, parameterization schemes supply flexibility to a model.

What makes this parameterization issue so endemic and, in a sense, unavoidable? In general, any mathematical model presents an idealized version of a real-world target system. There is always greater abundance in the target system than in its mathematical model. Hence, there may be both known and unknown properties of the target system that should be, but have not been, included in the model. Even if all relevant inner workings of a target system were completely known, it might still be prohibitive to account for them explicitly as existing theories might be so complex that they would make the model intractable. Adjustable parameters are of prime importance in this context. They make it possible to use simplified but tractable models. Such models may be related only loosely to the target object and may be obvious over-simplifications. But leaving open some parameters in such models and using them for compensation can make the models work. This approach is at the core of MMS.

4 Tardis—holism and transferability

Based on the examination of MMS and simulation methodology, we are now equipped to detail the argument about transferability. We claim that models in MMS are a Tardis. In this context, we use the metaphor Tardis for characterizing simulation models, especially when they are exchanged between groups. The Tardis made it from a charming idea in a science fiction series into the Oxford English Dictionary (OED). It is an acronym for.

> *T*ime *A*nd *R*elative *D*imensions *I*n *S*pace"—the name in the science-fiction BBC television series Doctor Who (first broadcast in 1963) of a time machine outwardly resembling a police telephone box, yet inwardly much larger. In similative use, especially as the type of something with a larger capacity than its outward appearance suggests, or with more to it than appears at first glance. (OED, cited after Wikipedia)

In short, the Tardis is a device for traveling that looks simple from the outside, but is anything but simple on closer inspection. Let us now establish the analogy to MMS. From the outside, models in MMS work with a mathematically well-defined and modular (additive) force field (see Eq. (2)). However, the mathematically clean appearance is deceptive. On the inside, things are more complicated.

We want to highlight two reasons why models in MMS are a Tardis. Both arise from parameterization. Firstly, after parameter adjustment, the components of the force field build a network with complex interdependencies, hence the *mathematical descriptions* of the individual types of forces or potentials are not modular, although the forces themselves are additive. Secondly, the adjusted parameters are conditional on data and the adjustment strategy. Thus, the force field becomes a holistic entity, much larger than the "outward appearance suggests". In the following, we unpack the argument.

We distinguish two different types of parameter adjustments. Type 1 is adjusted a priori, i.e., before the simulation modeling feedback loop starts. An example is the exponent $n = 6$ of the attractive part of the Lennard–Jones potential. After different proposals had been made in the 1920s, Lennard–Jones settled with $n = 6$ in 1931, following London who had derived from quantum theoretical calculation that the attractive force decays with this exponent (Lennard–Jones, 1931, London, 1930, see Rowlinson, 2002 for much historical and technical detail). Thus, there is a good physical reason for choosing $n = 6$. However, the form of the LJ potential does not coincide with quantum theory, but rather is a mathematically feasible form that is good and adaptable enough. In other words, the physical interpretation of the parameter n is strong, but not determining the potential. Moreover, type 1 parameters do not necessarily have a strong physical interpretation. While physical theory can provide good reasons, computational feasibility can, too. The pragmatically chosen exponent $m = 12$ in the LJ potential illustrates the case.

Type 2 parameters are adjusted in the loop, i.e. in the light of how the model behaves against the background of data.[11] This type might also have some physical interpretation—as once again the LJ illustrates. The parameters ε and σ are related to the strength of the attractive interaction, respectively the particle diameter. However, this relationship is indirect, mediated by the form of the model. The long-range behavior of the attractive force is governed by the exponent n, and the parameters ε and σ balance all other aspects, like the (in)adequacy of the restriction to pairwise interactions. Since the performance of the entire potential is evaluated, the evaluation may not reveal much about the contribution of individual terms.

The Stockmayer potential, mentioned above, provides further illustration. It models the dipole moment via a third parameter. However, when all three parameters are assigned over the feedback loop, their values reflect a balance over very different factors, including all kinds of modeling inadequacies. Consequently, the assignment of the dipole parameter—notwithstanding its physical interpretation—does not give a value for the dipole independently from the other adjusted parameters. Only the fully specified force field, with all parameter values assigned, matches the data in a meaningful way. Thus, over the course of parameter adjustments in the feedback loop, the model becomes a holistic entity.

Implicitly, the second reason for the Tardis was already involved. On the modeling side, parameters are adjusted. This process presupposes a benchmark against which the adjustment is evaluated and which is guiding further modification. Such benchmark includes a training set of data and a strategy of adjustment, including optimization criteria and the order in which adjustments are made. The data side is very relevant for transfer. Transferability can refer to using the model for making predictions for the same material and the same property, but different conditions as in the training data. This can be very challenging, when the conditions deviate extremely from those of the training data. Transferability can also refer to making predictions for the same material, but a property that was not included in the training set. This is no less challenging. Success after transfer is a litmus test for the physical content

---

[11] The borders between parameters that are set a priori and parameters that are adjusted in the feedback loop are often blurred. Sometimes parameters are fixed after some initial trials, while for other parameters the feedback loop is continued.

of the model—however indirectly mediated through parameter adjustments. There is no common standard that regulates which data to take. Furthermore, parameterization results depend on the strategy of adjustment, the optimization method, including the setting of boundary conditions, and more.[12] All these specifics matter and, moreover, are mutually interdependent. This interdependency is tantamount to models being a Tardis.

What does this imply for model transfer? The Tardis argument does not claim that traveling is impossible, rather it is more onerous than anticipated. Model transfer does not only involve the basic math expression. Beyond the force field, it involves the parameterization scheme, and all parameter values. Furthermore, transfer involves the specification of the training data and of the parameterization strategy—up to the test strategy, assuming that testing does not work with the training data. All this baggage is unpleasant and tedious to address in scientific publications, which misleads many authors to neglect the topic. Researchers who transfer an MMS model usually "buy a pig in a poke". Practitioners have noticed that the transferability of their models is in danger. Here is one example from molecular mechanics:

> This set of potential functions, called the force field, contains adjustable parameters that are optimized to obtain the best fit of calculated and experimental properties of the molecules, such as geometries, conformational energies, heats of formation, or other properties. The assumption is always made in molecular mechanics that corresponding parameters and force constants may be transferred from one molecule to another. In other words, these quantities are evaluated for a set of simple compounds, and thereafter the values are fixed and can be used for other similar compounds. It is not possible to prove that this is a valid assumption. (Burkert & Allinger, 1982, p. 3-4)

Our argument is well in line with this diagnosis. With growing success of MM a decade later, the problem was not solved but had aggravated:

> In general, parameters are not transferable from one force field to another because of the different forms of equations that have been used and because of parameter correlation" within a force field. That is, when one is carrying out parameterization, if one makes some kind of error, or arbitrary decision, regarding one parameter, other parameters in the force field adjust to minimize any error that would be caused. (Bowen & Allinger, 1991, p. 92/93)

This finding is not restricted to molecular mechanics, not even to MMS. It is endemic in many simulation models (cf. Hasse & Lenhard, 2017). What Bowen and Allinger describe points toward what we called holism. While it poses a threat to transferability, it is hard to avoid because it is rooted in the feedback loop for parameter adjustments, on which much of the success of simulation is built.[13]

---

[12] One anonymous reviewer aptly pointed out that the initial assignment of a parameter value might affect the outcome, too. Furthermore, a host of different considerations, from empirical results to theoretical calculations, can lead to this value assignment—fostering our argument.

[13] There is a small but relevant discussion on issues related to holism in the newer philosophy of science. Wimsatt (2007) has discussed the related issue of "generative entrenchment". Lenhard and Winsberg (2010)

The Tardis affects the ontology of models. The difficulty in transfer brings up the question of what it is that is (or should be) transferred. What travels is not the elegant mathematical formulation of the model, at least it should not travel alone. In fact, adjustable parameters (and how their values are specified) have to be included. In other words, the ontology of simulation models differs profoundly from that of theoretical mathematical models. The former can attain predictions, but are much bigger entities.

Let us briefly reflect upon the (dis)advantages of using the term Tardis. First of all, it originated from popular culture and, consequently, its import into philosophy of science cannot refer to already existing uses. Furthermore, the more established terms of black box and epistemic opacity seem to make Tardis dispensable. However, this is not the case because the existing terminology does not fit well. Black boxes originated from electrical engineering, meaning a functional unit whose input–output behavior is specified while the internal workings remain unknown. It is indeed an important aspect of simulations that software is often inaccessible to users (see Sect. 5 below). However, our argument about parameters creating relationships between different modules (thereby undercutting modularity) speaks against the character as a "box". It is not about containment, but about non-containment. Furthermore, the color black in the black-box terminology signals inaccessibility.[14] Indeed, epistemic opacity is an intensely debated feature of simulation.[15] In recent work, Beisbart (2021) provides an overview and advocates a broad understanding of epistemic opacity that centers on inaccessibility. However, our argument about adjustable parameters focused on a level where these parameters are very accessible. Hence, the Tardis problem is not congruent with the inaccessibility feature of black boxes and epistemic opacity.

Both terms are certainly relevant to the problem how models that appear modular on one level, but are holistic on a finer level do (not) transfer. We see the present paper as a step that first diagnoses the problem with the help of an examination that runs deep enough to capture the finer level where modularity and transferability are questioned. For this goal, it seemed advisable to keep it simple and use a fresh and accessible terminology.

Footnote 13 continued

show the relevance of holism in the context of the validation of simulation models. Lenhard (2018) argues simulation modeling has a tendency to erode modularity.

[14] Alternatively, it can signal that access is dispensable as long as the box works as specified. We do not enter into this discussion. There is rich literature on black boxes and their relevance outside electrical engineering, see Wiener (1948) on cybernetics, or the investigation into types of boxes that Boumans (2006) provides.

[15] Here is a sample of works that shows how rich and diverse the discussion is. Humphreys (2004) coined the term and later reaffirmed epistemic opacity as a feature that makes simulation novel (2009). Lenhard (2019) places it in the history of mathematical modeling. Durán and Formanek (2018) discuss its consequences regarding trust in simulations, while Kaminski and Schneider (forthcoming) discern different types of opacity. See Beisbart (2021) for more references.

5 Software and travel patterns

In this fourth step, we add an argument from the role of software that fosters our claim on transferability. That software is an important part of simulation is a no-brainer. Nevertheless, investigations into how software and coding practices influence modeling are still rare exceptions. The work of Wieber and Hocquet is one[16]:

> Our claim is that parameterization issues are a source of epistemic opacity and that this opacity is entangled in methods and software alike. Models and software must be addressed together to understand the epistemological tensions at stake. (2020, p. 4/5)

This is well in line with our stance. The investigation of model transfer has to take into account that models are linked to software and computers. Software exerts influence in two directions. It works like a gate through which researchers can access computational power. At the same time, existing software packages inform researchers about what research projects might be feasible. Software, like hardware by the way, provides opportunities and sets limitations. These are not only a technical issue, but also a matter of social organization.

We focus on a major transformation that, we claim, happened in the 1990s and divides between what we call the old and the new regime. Notably, both regimes differ in the travel pattern of software. Furthermore, we claim, this pattern affects transferability. Again, MMS serves as our example.

In the old regime, users were also masters. Not that every single researcher was an expert in theory and software development simultaneously. Rather, research projects that used the computer were typically conducted by a group that covered this expertise. Software was developed for special tasks, often oriented at specific modeling tasks. Moreover, knowledge about the model, the access to a computer, especially a very fast one, and the knowledge about the software that ran on this computer, remained in close proximity. A striking example is the almost exclusive access that military-funded physics projects enjoyed for about two decades after the second world war. MMS, among other simulation approaches, originated from in this context.[17] However, what we call the old regime is relevant also for everyday practices that are only accidentally mentioned in publications. One example is Martin Neumann, who conducted a project in MMS at the University of Vienna, but had to realize that he needed more computing power than the local computers provided. Therefore, he collaborated with researchers from Edinburgh who had access to a distributed array processor (DAP), a type of parallel computer.

> The simulations reported in this paper have been performed on the ICL DAP installed at the Edinburgh Regional Computing Centre. They have been made possible through a study visit award granted by the Royal Society, London, and a travel grant from the Österreichische Akademie der Wissenschaften. (Neumann et al., 1984, p. 113)

---

[16] Two more exceptions that highlight the social dimension of networked computational infrastructure are Hocquet and Wieber (2021) and Saam et al. (forthcoming).

[17] For historical sketches of MC and MD, see Battimelli et al. (2020). For MM, the history looks different.
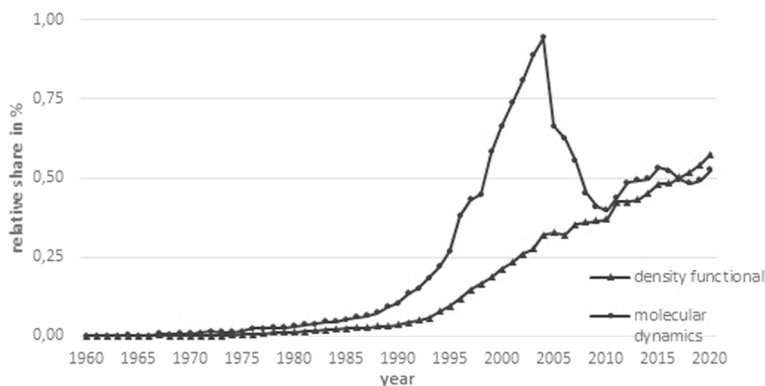
Fig. 3 The relative share of papers on molecular dynamics, resp. density functional theory, in all papers appearing in the databank Scopus. In absolute numbers, publications per year went up from below hundred in 1975 to nearly 20,000 in 2020, both MD and DFT. MD shows an additional "tooth" around 1995–2010

This accidental find illustrates that researchers travelled to the places where they had access to high performance computing power and, at the same time, access to the local expertise of how to utilize this computing power for their modeling task. Another piece of evidence about the old regime is that standard books on theory of MMS also include discussions of numerical recipes, like for instance Allen & Tildesley (1987), reflecting the interests of their readership.

In the new regime, users do not have to be—and commonly are not—masters. Allinger and colleagues were pioneers in software development for general use in MM—the MM1, MM2, and MM3 packages for molecular mechanics. They foresaw that the success of these software packages would likely start a new era in which users and developers become distinct (Burkert & Allinger, 1982). What Allinger predicted for MM early on happened in all fields of MMS (and beyond). The range of users grew far beyond the group of experts on molecular theory. In a striking way, this can be gathered from simple counts of published papers in this field. They virtually explode around the 1990s. Figure 3 displays what happened in the field of molecular dynamics (MD) and that the upswing is virtually in parallel with the electronic structure method DFT (outside the family of MMS). The turn that happened around the 1990s seems to happen in a variety of computational methods.

From the 1990s onwards, the travel patterns change. Now, there is a networked infrastructure and a quickly growing resource of software packages that are made by experts for users that are non-experts. In short, the software travels, not the researchers. In their review of molecular dynamics in engineering, Maginn and Elliott observe:

> There are probably hundreds of MD codes used and developed in research groups all over the world. A noticeable shift has occurred, however, toward research groups becoming *users* of a few well-established MD codes instead of developers of their own local codes. This is perhaps inevitable and follows the trend of the electronic structure community. (Maginn & Elliott, 2010, p. 3065, emphasis original)

Additionally, the growing need for expert knowledge on parallel computer architectures contributes to the re-organization of expertise: "As a result, the drive to abandon locally developed serial codes in favor of highly optimized parallel codes developed elsewhere became great." (Maginn & Elliott, 2010, p. 3065) Overall, the new regime is characterized by practitioners that are users rather than developers of software. While there is ample supply of software packages, user-friendliness—one does not have to know the details—distances the users from the models. Thus, the vastly increased usability and uptake of software goes hand in hand with increasing opacity for its users.

We argue that this regime change brought under scrutiny the transferability of models. Results should be invariant under exchange of software packages (as long as conditions are kept constant). But is this the case? Answering this question calls for an empirical investigation of research practice. There are first attempts in the community to start this kind of investigation. Loeffler and colleagues, for instance, examine whether a number of popular molecular simulation software packages (AMBER, CHARMM, GROMACS, and SOMD) give consistent calculations of free energy changes. They are clear that this kind of project is just about to start:

> In particular, we need to ensure reproducibility of free energy results among computer codes. To the best of our knowledge this has not been systematically tested yet for a set of different MD packages. However, there have been some recent efforts to test energy reproducibility across packages—a necessary but not sufficient prerequisite. Another study went further and also compared liquid densities across packages, revealing a variety of issues.[18] (Loeffler et al., 2018, p. 5569)

Thus, reproducibility works as an indicator for transferability. Loeffler et al. go on:

Nevertheless, it is critical that free energy changes computed with different simulation software should be reproducible within statistical error, as this otherwise limits the *transferability* of potential energy functions and the relevance of properties computed from a molecular simulation to a given package. This is especially important as the community increasingly combines or swaps different simulation packages within workflows aimed at addressing challenging scientific problems. (ibid., emphasis added)

Thus, the community is starting to recognize the problem. The difficulties to inspect the code, control for the behavior of users, and create a well-defined workflow all play together. There is not yet a consensus about the direction in which a solution is supposed to be found. Some advocate a rigorous standardization. "The Simulation Foundry (SF) is a modular workflow for the automated creation of molecular modeling (MM) data." (Gygli & Pleiss, 2020, p. 1922) Uniquely determining the workflow should guarantee replicability. However, such standardization approach is more eluding than tackling the problem of transferability because it does not address eroded

---

[18] Here, Loeffler et al. refer to the earlier study of Schappals et al. (2017) that pioneered the issue of reproducibility in molecular simulation. It reports on a round robin study that identified to a number of factors from simulation modeling that work against reproducibility.

modularity nor dependency from data. Any strategy that tries to overcome the problem of reproducibility will likely have work with a clear picture of what it is that is transferred. In current practice, publications often do not address how researchers dealt with parameters, parameterization schemes, or the options of software suites. As a rule of thumb, the higher the quality of a journal paper, the more information it contains about handling of parameters. However, providing full detail—including arcane information like about the compiler version—arguably is neither practical nor desirable. The connection to reproducibility makes transferability especially relevant.

6 Conclusion

The paper has identified and analyzed factors that impede a model's transferability. From a methodological and epistemological point of view, the models in MMS are large holistic entities because the dependency on data and the parameterization process is complicated. Additionally, the way development and use of software is socially organized contributes to, or aggravates, the problem of transferability. Overall, transferability poses a characteristic problem for simulation modeling because parameterization is a key factor that works toward conflicting goals. The use of adjustable parameters is key for predictive success but, at the same time, also for creating the Tardis problem of transferability. Our findings arguably hold for a far wider class of simulation models than the family of MMS. Since the use of parameterization is typical for large parts of applied science and engineering (see Vincenti, 1990, for a classic study), the dilemma between prediction and transferability might be even more general.

Our investigation into simulation modeling took a much more application-oriented viewpoint and looked at how mathematics works as an instrument in MMS. The result is without doubt horrible for a fan of abstract structures. Models are a Tardis, i.e., behind the elegant outside is lurking a complicated inside. Even worse, these complications are hard to get rid of because parameterizations are not a mere nuisance, but also the keys to applicability. This is a warning for the lover of clarity and elegance. The superhuman computational capabilities of the computer have been welcomed as a promising surrogate for the superhuman powers of Laplace's demon. Making computational power do work via simulation models creates a situation where the models themselves pose problems of complexity and opacity.[19] At the same time, our analysis harbors an insight on the metaphysical level, namely about the tension between universality and particularity or, rather, how computational science deals with this tension. Mathematized theories of great generality are surely an exquisite achievement of science. However, these theories do not make predictions in concrete situations and contexts. The world is a dappled world (as Cartwright (1999) put it) whose particularity cannot be adequately captured by general theories. Our point is that such position is not directed against mathematization. Quite on the contrary, the

---

[19] For a well-seasoned insight in the philosophy of simulation, see Humphreys (2004), Gramelsberger (2011), or Winsberg (2019).

study of mathematized and computerized modeling reveals this tension. Models work *because*, on closer inspection, they are a Tardis.[20]

Declarations

Conflict of interest There are no potential conflicts of interest.

References

Allen, M. P., & Tildesley, D. J. (1987). *Computer simulation of liquids*. Oxford University Press.

Battimelli, G., Ciccotti, G., & Greco, P. (2020). *Computer meets theoretical physics. The New Frontier of molecular simulation*. Springer Nature.

Beisbart, C. (2021). Opacity thought through: On the intransparency of computer simulations. *Synthese, 199*, 11634–11666. https://doi.org/10.1007/s11229-021-03305-2

Bowen, J. P., & Allinger, N. L. (1991). Molecular mechanics: The art of science and parameterization. In K. B. Lipkowitz & D. B. Boyd (Eds.), *Reviews in Computational Chemistry* (pp. 81–97). Wiley. https://doi.org/10.1002/9780470125793.ch3

Burkert, U. & Allinger, N. C. (1982). *Molecular Mechanics*. ACS Monograph 177. American Chemical Society

Cartwright, N. (1999). *The dappled world*. Cambridge University Press.

Cartwright, N., Hardie, J., Montuschi, E., Soleiman, S., & Thresher, A. (2022). *The tangle of science*. Oxford University Press.

Durán, J. M., & Formanek, N. (2018). Grounds for trust: Essential epistemic opacity and computational reliabilism. *Minds and Machines, 28*(4), 645–666.

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

Forte, E., Jirasek, F., Bortz, M., Burger, J., Vrabec, J., & Hasse, H. (2019). Digitalization in thermodynamics. *Chemie Ingenieur Technik, 91*(3), 201–214.

Gramelsberger, G. (Ed.). (2011). *From science to computational science. Studies in the history of computing and its influence on today's sciences.* Diaphanes.

Gygli, G., & Pleiss, J. (2020). Simulation foundry: Automated and F.A.I.R. molecular modeling. *Journal of Chemical Information and Modeling, 60*, 1922–1927.

Hasse, H., & Lenhard, J. (2017). Boon and bane. On the role of adjustable parameters in simulation models. In J. Lenhard & M. Carrier (Eds.), *Mathematics as a tool tracing new roles of mathematics in the sciences boston studies in the philosophy and history of science* (Vol. 327, pp. 93–115). Springer.

Hocquet, A., & Wieber, F. (2021). Epistemic issues in computational reproducibility: Software as the elephant in the room. *European Journal for Philosophy of Science*. https://doi.org/10.1007/s13194-021-00362-9

---

[20] In recent work, Cartwright et al. (2022) make a highly related claim that the working condition of science is that it is a "tangle" of various particular elements.

Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science, 64*(Proceedings), S325–S336.

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

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese, 169*, 615–626.

Humphreys, P. (2019). Knowledge transfer across scientific disciplines. *Studies in History and Philosophy of Science Part A, 77*, 112–119.

Kaminski, A. & Schneider, R. (forthcoming). Social, technical, and mathematical opacity: Computer simulation and the scientific work on purification. In: M. Resch, A. Kaminski & P. Gehring (Eds.), *Science and Art of Simulation (SAS) II*. Springer, London

Kieseppä, I. A. (1997). Akaike information criterion, curve-fitting, and the philosophical problem of simplicity. *British Journal for the Philosophy of Science, 48*(1), 21–48.

Lenhard, J. (2018). Holism or the erosion of modularity—a methodological challenge for validation. *Philosophy of Science, 85*, 832–844.

Lenhard, J. (2019). *Calculated surprises*. Oxford University Press.

Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Modern Physics, 41*, 253–262.

Lennard-Jones, J. E. (1931). Cohesion. *Proceedings of the Physical Society, 43*, 461–482.

Loeffler, H. H., Bosisio, S., Matos, G. D. R., Suh, D., Roux, B., Mobley, D. L., & Michel, J. (2018). Reproducibility of free energy calculations across different molecular simulation software packages. *Journal of Chemical Theory and Computation, 14*(11), 5567–5582. https://doi.org/10.1021/acs.jctc.8b00544

London, F. (1930). Über einige Eigenschaften und Anwendungen der Molekularkräfte. *Zeitschrift Für Physikalische Chemie, B11*, 222–251.

Maginn, E. J., & Elliott, J. R. (2010). Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Industrial & Engineering Chemical Research., 49*, 3059–3078.

Morgan, M., & Morrison, M. (1999). *Models as mediators*. Cambridge University Press.

Morrison, M. (2009). Models, measurement, and computer simulation. The changing face of experimentation. *Philosophical Studies, 143*, 33–57.

Morrison, M. (2014). *Reconstructing reality. Models, mathematics, and simulations*. Oxford University Press.

Neumann, M., Steinhauser, O., & Pawley, G. S. (1984). Consistent calculation of the static and frequency-dependent dielectric constant in computer simulations. *Molecular Physics, 52*(1), 97–113. https://doi.org/10.1080/00268978400101081

Parker, W. (2014). Values and Uncertainties in climate prediction, revisited. *Studies in History and Philosophy of Science, 46*, 24–30.

Rowlinson, J. S. (2002). *Cohesion*. Cambridge University Press.

Saam, N. J., Kaminski, A. & Ruopp, A. (forthcoming). Who is the agent? Epistemic opacity, code complexity, and the research group in advanced simulations. In: M. Resch, A. Kaminski & P. Gehring (Eds.), *Science and art of simulation (SAS) II*. Springer.

Schappals, M., Mecklenfeld, A., Kröger, L., Botan, V., Köster, A., Stephan, S., García, E., Rutkai, G., Raabe, G., Klein, P., Leonhard, K., Glass, C., Lenhard, J., Vrabec, J., & Hasse, H. (2017). Round robin study: Molecular simulation of thermodynamic properties from models with internal degrees of freedom. *Journal of Chemical Theory and Computation*. https://doi.org/10.1021/acs.jctc.7b00489

Tal, E. (2013). Old and new problems in philosophy of measurement. *Philosophy Compass, 8*(12), 1159–1173.

Vincenti, W. (1990). *What engineers know and how they know it*. The Johns Hopkins Press.

Wieber, F., & Hocquet, A. (2020). Models, parameterizations, and software: Epistemic opacity in computational science. *Perspectives on Science, 28*(5), 610–629.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. The MIT Press.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings*. Harvard University Press.

Winsberg, Eric (2019). Computer simulations in science. In: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2019 Edition). URL = <https://plato.stanford.edu/archives/win2019/entries/simulations-science/>. Accessed 1 July 2022.