



Polycratic hierarchies and networks: what simulation-modeling at the LHC can teach us about the epistemology of simulation

Florian J. Boge^{1,2}  · Christian Zeitnitz^{2,3}

Received: 20 February 2019 / Accepted: 15 April 2020 / Published online: 14 May 2020
© The Author(s) 2020, corrected publication 2021

Abstract

Large scale experiments at CERN’s Large Hadron Collider (LHC) rely heavily on computer simulations (CSs), a fact that has recently caught philosophers’ attention. CSs obviously require appropriate modeling, and it is a common assumption among philosophers that the relevant models can be ordered into hierarchical structures. Focusing on LHC’s ATLAS experiment, we will establish three central results here: (a) with some distinct modifications, individual components of ATLAS’ overall simulation infrastructure can be ordered into hierarchical structures. Hence, to a good degree of approximation, hierarchical accounts remain valid at least as descriptive accounts of initial modeling steps. (b) In order to perform the epistemic function Winsberg (in Magnani L, Nersessian N, Thagard P (eds) *Model-based reasoning in scientific discovery*. Kluwer Academic/Plenum Publishers, New York, pp 255–269, 1999) assigns to models in simulation—generate knowledge through a sequence of skillful but non-deductive transformations—ATLAS’ simulation models have to be considered part of a network rather than a hierarchy, in turn making the associated simulation modeling messy rather than motley. Deriving knowledge-claims from this ‘mess’ requires two sources of justification: (i) holistic validation (also Lenhard and Winsberg in *Stud Hist Philos Sci Part B Stud Hist Philos Modern Phys* 41(3):253–262, 2010; in Carrier M, Nordmann A (eds) *Science in the context of application*. Springer, Berlin, pp 115–130, 2011), and (ii) model coherence. As it turns out, (c) the degree of model coherence sets HEP apart from other messy, simulation-intensive disciplines such as climate science, and the reasons for this are to be sought in the historical, empirical and theoretical foundations of the respective discipline.

✉ Florian J. Boge
fjboge@uni-wuppertal.de

¹ Institute for Theoretical Particle Physics and Cosmology, RWTH Aachen University, Aachen, Germany

² Interdisciplinary Centre for Science and Technology Studies (IZWT), Bergische Universität Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany

³ Physics Department, Bergische Universität Wuppertal, Wuppertal, Germany

Keywords Computer simulation · Hierarchy of models · High energy physics · Phenomenological models

Contents

1 Simulations at the LHC: what epistemologists of simulation can learn	446
2 From Suppes to Winsberg: hierarchies of models in experimentation and simulation	448
2.1 Simulation and hierarchies of models in high energy physics experimentation	448
2.2 Winsberg’s hierarchy and the epistemology of simulation	449
2.3 Models of phenomena versus phenomenological models	453
3 The plethora of simulation models used by LHC’s ATLAS experiment	454
3.1 Event generation	455
3.1.1 Hard process	455
3.1.2 Parton shower	457
3.1.3 Hadronization	458
3.1.4 Underlying event	460
3.2 Detector simulations	461
3.2.1 Particle decay	461
3.2.2 Bremsstrahlung	462
3.2.3 Nuclear interactions	462
4 Polycratic hierarchies, networks, and the justification of simulation-based inferences	462
4.1 Applying Winsberg’s hierarchy...with some ifs and buts	463
4.2 A network of simulation models	468
4.3 Justification (i): holistic validation	470
4.4 Justification (ii): model coherence	472
5 Conclusions: lessons for the epistemology of simulation	476
References	477

1 Simulations at the LHC: what epistemologists of simulation can learn

In July 2012, the ATLAS and CMS experiments at CERN’s LHC announced the discovery of a Higgs-like particle.¹ Further measurements of its properties indicated that the particle found could indeed be a Standard Model² (SM) Higgs,³ meaning a remarkable success for the fundamental theory of present day particle physics (the SM). LHC measurements and searches of this kind rely heavily on computer simulations (CSs), a fact that has recently caught philosophers’ attention (cf. Beisbart and Saam 2019; Karaca 2018b; Massimi and Bhimji 2015; Morrison 2015).

CSs are utilized ubiquitously in modern high energy physics (HEP), in the design and evaluation of its complex, large scale experimental procedures.⁴ On the other hand, none of the CSs would function appropriately (or could even *exist*) without input from both theory and (previous) experiment: calibration and benchmarking pro-

¹ Cf. <https://home.cern/topics/higgs-boson>, checked 02/18.

² In high energy physics, the ‘Standard Model’ is the union of electroweak theory with quantum chromodynamics.

³ Cf. <https://home.cern/about/updates/2015/09/atlas-and-cms-experiments-shed-light-higgs-properties>, checked 02/18.

⁴ Cf. Morrison (2015, Chap. 8) for an overview of different uses in the Higgs discovery.

cedures are used to secure and evaluate the overall appropriateness of complex CSs (cf. Winsberg 2010, p. 22), and many (parts of) CSs in HEP are direct implementations of numerical approximations to existing theory, or parametrizations of experimental data respectively.⁵

The implementation of any CS obviously requires appropriate modeling (e.g. Humphreys 2004; Morrison 2009; Winsberg 2010). The delicate involvement of the relevant models in knowledge generation and the experimental process itself raises a number of philosophical questions. For instance: How do they impact experimenters' inferential capacities, when it comes to the properties of a targeted system? How do models specific to simulation relate to theory or experimentation in detail? How do they relate to each another? Frankly, can (we as) philosophers retain certain pre-conceptions about the modeling steps involved in CSs in the simulation-intensive experimental environment that is the LHC?

Building on LHC's ATLAS experiment as a case study, "one of the largest collaborative efforts ever attempted in science",⁶ we will approach this complex of problems from the following angle: Following Winsberg (1999) and Karaca (2018a, b), we will investigate the concept of a *hierarchy of models* in the context of simulation. We will first recapture Karaca's main arguments in brief, whose criticism (and extension) of hierarchical accounts targets Suppes' (1962) seminal account of models of *experimentation*, but with a focus on simulation-intensive contexts.

Winsberg (1999), on the other hand, has offered a hierarchical account specifically designed to understand modeling and connected inferential steps in the context of *simulation alone*. Focusing on Winsberg's account, we will establish the following claims: (a) simulation models of individual *stages* of the processes investigated by ATLAS can be analyzed in terms of hierarchical structures like that suggested by Winsberg, albeit with some distinct modifications. Thus, at the level of these individual stages, Winsberg's account remains largely legitimate, at least as a *descriptive* account of *initial* modeling steps. (b) When it comes to the *purpose* of Winsberg's hierarchy, namely establishing a simulation's ability to promote genuine knowledge gain regarding the target system, a more complex, network-like structure of models is required, similar to what Karaca (2018b) suggests for simulation-intensive experimentation, and to the highly 'entrenched' structures recognized by Lenhard and Winsberg (2010, 2011) for climate science. (c) There are crucial empirical, theoretical, and even socio-historical differences between HEP and climate science though, and these allow HEP researchers a second angle on justification that is far less developed in climate science.

The paper is organized as follows: Sect. 2 discusses the relevant hierarchical accounts of models, including a brief recap of Karaca's criticism and suggested extension. In Sect. 3, we offer a pointed overview of the manifold simulation models used by ATLAS, allowing some insight into the complexities involved. This will be used in subsequent sections, but should also be considered as a general *resource* for philosophers of simulation, given the aforementioned recent interest in CSs in HEP and the fact that other expositions, such as Morrison (2015) or, more recently, Mättig (2019),

⁵ 'Parametrization' should be read here as 'parametric interpolation', which means the fitting of a (smooth) function to data by means of adjusting a number of free parameters (e.g. Behnke and Moneta 2013, p. 62).

⁶ <https://atlas.cern/discover/collaboration>, checked 02/18.

remain fuzzy on many details regarding modeling needs and connections between different models.

In Sect. 4.1, we then show how to apply Winsberg’s hierarchy to individual simulation stages or components, thereby also establishing some indicated modifications. In Sect. 4.2, we elaborate, however, how the intricate connections between these components spoil the *knowledge generating capacities* claimed originally for the hierarchy by Winsberg, and how an outright *network of simulation models* is required for that task instead. Sect.s 4.3 and 4.4 then establish, respectively, how justification of results is facilitated (i) by *holistic validation*, as similarly claimed for climate simulations by Lenhard and Winsberg (2010, 2011), and (ii) *model coherence*, which, we shall argue, sets HEP somewhat apart from climate science.

2 From Suppes to Winsberg: hierarchies of models in experimentation and simulation

It is a common assumption among philosophers (Winsberg 1999; Morrison 2009;⁷ Winsberg 2010; Beisbart 2012, 2014; Boge 2019) that models relevant for devising a CS can be somehow ordered into hierarchical structures.

The purpose of such hierarchical accounts is to understand the modeling process connected to a given CS, but, according to Winsberg (1999), also the way in which CSs can lead to *knowledge claims* regarding a target system. For simpler examples, it is unproblematic to apply such a hierarchical account. Can we, however, also use it to understand knowledge generation in such complex simulation-intensive environments as the LHC?

2.1 Simulation and hierarchies of models in high energy physics experimentation

A first reason to be skeptical comes from Karaca (2018a, b). Recall that the seminal account of hierarchies of models is Suppes (1962), adopted and developed, e.g., also by Mayo (1996) and Harris (1999), and concerned with the connection of a specific model of some theory to experimental data. It builds on Suppes’ semantic view of theories and aims to establish “whether or not an empirical substructure is isomorphic to a set of appearances” (Winsberg 1999, p. 266). Thus, it concerns the connection between theory and data in the context of experimentation, and not at all questions of simulation.

Karaca (2018a, b), however, argues that HEP *defies* Suppes’ hierarchical account, because it is an experimental context in which additional models are required that

⁷ An anonymous referee pointed out to us that it might seem odd that Morrison should defend a hierarchical account of models, given that she is an advocate of their partial autonomy (Morrison and Morgan 1999). Note, however, that standing in dependence relations is compatible with models’ *partial* autonomy, so long as the dependence in question is sufficiently weak. It will become clear below that this is the case with the models in hierarchies of simulation. Moreover, in cases where theory is at the top of the hierarchy, the models below that level can be seen to serve exactly the mediating function claimed by Morrison and Morgan (1999).

do not fit neatly into any place of the hierarchy—such as models of data acquisition (Karaca 2018a) or, precisely, *simulation models* (Karaca 2018b).

Suppes (1962), to recall, acknowledges a level of models of the theory, set theoretical structures that satisfy the axioms of the theory experimented on. Below that, there are models of the data, “designed to incorporate all the information about the experiment which can be used in statistical tests of the adequacy of the theory” (ibid., p. 258), and models of the experiment linking the two, which, in Mayo’s (1996, p. 133) words, “provide[...] a kind of experimental analog of the salient features of” a model of the theory. Suppes also considers models of experimental design as well as *ceteris paribus* conditions that lie even further down the hierarchy.

The linkage between these levels is specialization to the experimental context, and the hierarchy’s main purpose is, in the words of Winsberg (1999, p. 267), the establishment of “a model of the data, and [...] an isomorphism between the model of the data and a designated model of the theory.” Thus, through specialization steps, one comes to make a certain theory experimentally testable.

Karaca (2018a), however, claims that models of *data acquisition*, required to understand LHC experiments, fit nowhere into this hierarchical picture: On the one hand, data acquisition seems to be just less specific to the particular experimental context than *ceteris paribus* conditions and experimental design. So regarding specialization, it might be placed right below data models. On the other hand, there is a direct impact of existing theories on data acquisition procedures, and that could be understood as a specialization of the relevant theories to the experimental context (Karaca 2018a, p. 5450).

In a sequel paper (Karaca 2018b),⁸ Karaca has extended his negative verdict on the prospects of hierarchical accounts of models in the context modern day HEP experiments, by drawing attention to the delicate involvement of simulations therein.

Karaca suggests an uncontroversial distinction between simulation models of *instrumentation*, *target phenomena*, and *hybrid* simulation models that contain features of both. The target phenomena are the particles produced in LHC’s proton-proton scatterings; the instrumentation, of course, solely concerns the detector and associated electronics. He then outlines how all these simulation models relate in various ways to different theories and experimental data, leading him to propose a “‘network of models account’ [...] of scientific experimentation, which applies to experiments that involve theoretical, experimental and simulation models.”⁹ (Karaca 2018b, p. 18)

2.2 Winsberg’s hierarchy and the epistemology of simulation

Karaca’s approach constitutes an important step towards a more realistic and up-to-date account of physics experimentation, including the involvement of CSs. However, it thus rather contributes to the philosophy of *experiment* than to the philosophy of

⁸ The paper was presented at the 2017 conference of the European Philosophy of Science Association in Exeter and is at present not publicly available. We thank K. Karaca for providing us with a draft copy. Page references and quotes throughout the paper are to / from the draft copy.

⁹ The precise relations between the different models are not characterized by Karaca in detail; but the foregoing considerations show that they will generally be such that they facilitate inferences from simulation to experiment, or experiment to simulation, or from theory to simulation, and so forth.

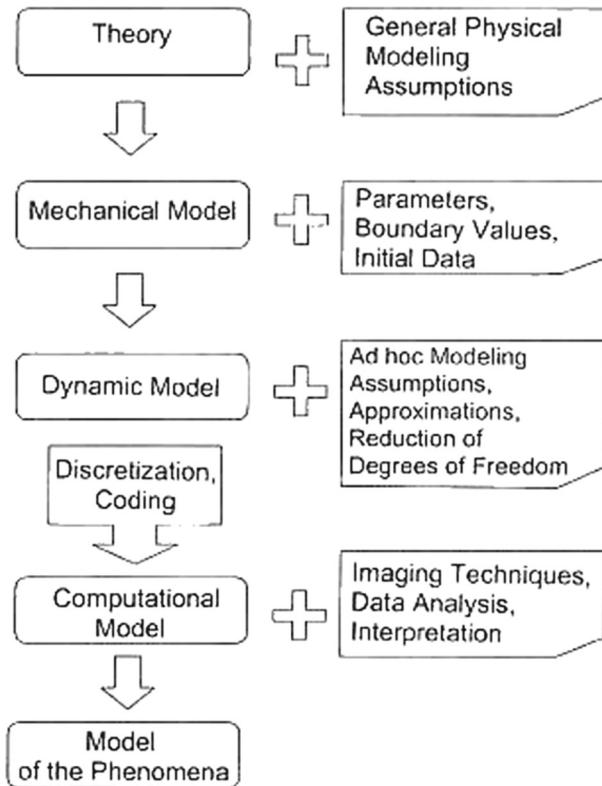


Fig. 1 Winsberg's hierarchy of models in simulation. (Reprinted by permission from Springer Nature: Springer Science+Business Media, LLC, *Model-Based Reasoning in Scientific Discovery* by Lorenzo Magnani, Nancy J. Nersessian, and Paul Thagard, eds., © 1999 Springer Science+Business Media New York (1999))

simulation, and misses out on questions of the knowledge-generating capacities of CSs *themselves* in the given context.

In rough analogy to Suppes' account, Winsberg (1999) has proposed an analysis of the models involved in a CS in terms of an equally hierarchical, linear structure with directed dependencies (Fig. 1). As was pointed out above, similar analyses can be found throughout the philosophy of simulation literature, so the account continues to enjoy a certain *prima facie* plausibility.

By 'mechanical model', Winsberg (1999, p. 258) means "a bare bones characterization of a physical system that allows us to use the theoretical structure to assign a family of equations to the system." This model, for Winsberg (1999, *ibid.*), "provides a foothold for the application of a theory to a set of real world problems[...]."

The name 'mechanical' is clearly inadequate for the context of HEP, but we maintain that the *slot* in Winsberg's hierarchy can be meaningfully filled here as well. Let us hence use the less committal term *simulation model* to refer to the kinds of models we

are interested in: physics models specific enough to serve as a *basis* for an algorithm to be executed on a computer.¹⁰

The dynamical model, for Winsberg (1999, p. 258), means a specific instance of the dynamics contained in the simulation model by means of boundary conditions. It will hence be more specific, and may imply the need to appeal to data, or to fix the values of free parameters on physical grounds.

To obtain from this a computational model in Winsberg's sense—which corresponds to Morrison's (2009, p. 44) simulation program—one not only needs to approximate the continuous equations of the dynamical model by discrete algebraic ones: a computationally tractable algorithm will also require a bunch of simplifying assumptions, including, e.g., the reduction of degrees of freedom or the neglect of certain mathematical substructures, possibly with a replacement by so called 'fudge factors', empirically based *ad hoc* structures that make up for the neglect of the more involved, theoretical ones (cf. Winsberg 1999, p. 259). This may include quite creative replacements of parts of the dynamical model, in case these do not allow for a translation into computer code for deeper reasons (cf. *ibid.*).

Finally, all sorts of representational means and additional mathematics may be used to interpret the data resulting from the simulation in terms of another model; a model of the phenomena, by which he means "a manifold representation that embodies the relevant knowledge, gathered from all relevant sources, about the phenomena. It can consist of mathematical relations and laws, images, both moving and still, and textual descriptions." Most importantly, such a model provides an *interpretation of data*, which, in the present context, means data produced as the result of a particular CS (cf. *ibid.*).

What is the purpose, for Winsberg, of this hierarchy? How are the relations characterized? In fact, Winsberg (1999, p. 263) describes "[e]very step down the hierarchy [...] [as] a transformation of knowledge in one form into another form, to be used either as a tool, or as a representation." Thus, models further down the hierarchy will in some sense be dependent on the models further up—though generally not in the sense of deductively valid inferences. This will allow the simulationist to approach the targeted phenomenon even if an established theory does not provide the desired information (e.g., for reasons of tractability). That the steps are often "patently *not* deductive" (Winsberg 1999, p. 260; orig. *emph.*) is reflected by the fact that "[e]ach transformation introduces the possibility, or even the likelihood, of introducing artifacts into the results." (*ibid.*, p. 263)

As for the purpose, Winsberg (1999, p. 260; *emph. added*) believes that it reflects a genuine "epistemology of simulation", (henceforth EoS) by which he means "the study of the means by which we *sanction* belief in the outcome of simulation studies

¹⁰ Note that this usage of 'simulation model' differs from that of, e.g., Morrison (2009, p. 44), who means by this "discrete algebraic equations" that constitute "a specific approximation" to a typically continuous model of the target system, devised "in order to make it amenable to numerical solution" so that it "can then be expressed as a sequence of computer instructions (an algorithm) that yield the simulation programme." Our usage is closer to that of Karaca (2018b, p. 12; *emph. added*), merely meaning "a mathematical model [...] numerically *solvable* through a program (called a 'simulation program') executed on a computer." Simulation models, as we use of the term here, need not be fully numerical or discrete, but may comprise discrete and continuous elements. This is a helpful terminological choice, but must not be over-interpreted, as we do, of course, recognize that implementation ultimately requires discretization at some point.

despite their motley methodological structure.” Moreover, the hierarchy *as a whole* will facilitate inferences to a targeted system: “*Via these steps*, the simulationist hopes to infer, from existing theoretical knowledge, new knowledge about the system being simulated.” (ibid., *emph. altered*)

There are then two distinct aspects to Winsberg’s EoS that we want to focus on here: the *justification* of inferences from simulation to a target and their very *facilitation* by means of the steps involved in the hierarchy. We take it for granted that accepting inferred information as new *knowledge* requires some sort of justification in this context. Hence, the points are not separate but connected: only if simulationists have appropriate *means* for sanctioning their beliefs in simulation results will they be able to acquire new knowledge through the inferences facilitated by CS-relevant models.

The standard repertoire for justification is usually referred to as ‘verification and validation’; the former meaning that one ensures that CSs reproduce theoretical results (insofar as available), the latter that they reproduce empirical findings (Morrison 2015, Chap. 7, for an overview). Winsberg (2010, pp. 22 ff.), however, points out that neither is usually really practicable in isolation, and that justification often comes from integrated ‘benchmark’ procedures, matching a CS to what is known about the target from whatever sources (theory, data, other simulations...).

As for the facilitation, Winsberg thinks of the transformative steps as “a case of theory articulation in the spirit of Kuhn[,] [...] a nontrivial process of bringing a theoretical structure into resonance with some phenomena that are ‘implicitly’ in its domain.” (Winsberg 1999, pp. 261–2) However, this points us to the fact that there is some amount of idealization and simplification involved in Winsberg’s account, because CSs can also be used where there is really nothing worthy of the name ‘theory’ available.

Winsberg (cf. 1999, p. 255) is certainly aware of this, as he *consciously* confines his attention to cases in which simulation is used to handle a theory that is (partly) intractable. It is a rather benign modification, though, to replace the top level by some sort of more general model, as is done also by Morrison (2009). An equally benign concession is that the modeling steps involved provide a kind of ideal limit, and should

not be over-interpreted as depicting any actual temporal sequence: The process of devising a CS is typically not as linear [...]but will contain multiple loops. For instance: once the output has some recognizable flaws, there are multiple junctions at which revisions are possible. (Boge 2019, p. 3; similarly Lenhard 2016)

Hence, there are some immediate, minor restrictions to Winsberg’s account as it stands, and we will establish further modification when we apply it to the case of HEP below. However, given that it is possible to so apply Winsberg’s account, we will, in a first step, indeed *vindicate* the account in a certain sense.

But in a second step, we will then demonstrate that, when it comes to the overall epistemological *purpose* of simulation modeling, i.e., inferring new *knowledge* of the target, HEP requires us to rethink the connections between models involved in CS far more drastically. This will lead us to a *network* account of models in HEP simulations, in loose analogy to Karaca’s (Karaca 2018b) network of models in (HEP)

experimentation, and closely related to features recognized for simulations in climate science by Lenhard and Winsberg (2010, 2011).

2.3 Models of phenomena versus phenomenological models

Before proceeding towards the analysis of ATLAS' simulation models, we would like to clarify another terminological subtlety. Phenomenological models, as the term is used in physics, are typically not models of *observable* phenomena, at least not if the term is construed narrowly (cf. van Fraassen 1980 vs. Shapere 1982). In referring, e.g., to quarks and gluons as phenomena, we here effectively embrace a notion of 'phenomena' along the lines of Bogen and Woodward (1988), whereby these "in most cases are not observable" (p. 306), and whereby it is rather (experimental) *data* "which play the role of evidence for the existence of phenomena" (p. 305) and "for the most part can be straightforwardly observed." (ibid.)

Indeed, Cartwright (1983, p. 1) famously warned us to "not be misled" by the use of 'phenomenological' in physics:

For the physicist, unlike the philosopher, the distinction between theoretical and phenomenological has nothing to do with what is observable and what is unobservable. Instead the terms separate laws which are fundamental and explanatory from those that merely describe. (ibid., pp. 1–2)

According to Bokulich (2011, p. 44) models phenomenological in *this* sense are "[o]ften—though not exclusively—[...]constructed via an ad hoc fitting [...] to the empirical data."

But, more importantly, they need not even be models of phenomena in Bogen and Woodward's sense: Often times, they are merely *oriented on* them. Suárez and Cartwright (2008, p. 70), for instance, distinguish three uses of 'phenomenological', as (a) representational of phenomena, (b) not theory-driven in construction, and (c) derived purely from measurement. Like them, we mostly have (b) in mind when we use the term 'phenomenological model' here.

Despite the fact that phenomenological models, as we here understand the term, are hence neither fundamental nor theory-driven, they usually "*mimic* many of the features of the theory but are much *easier to handle*." (Hartmann 1999, p. 327; *emph. added*) This view is shared by Karaca (2013, p. 100; *orig. emph.*), who finds phenomenological models to "function independently of theory, and yet this independence is only *partial*; because [their] construction [...] and their applicability to natural phenomena can be reliant on some important features of the fundamental theory [...]."

In sum, we here mean by 'phenomenological models' models that are (i) connected to but not derived from theories, (ii) non-fundamental, (iii) oriented on but not necessarily descriptive of phenomena, and (iv) mostly connected to observed data via free, fitable parameters.

Compare this, now, Winsberg's (1999, p. 259) notion of a model of the phenomena, as introduced above. Winsberg here obviously has in mind models that could count as phenomenological rather in Suárez and Cartwright's sense (a), which ties them much stronger to phenomena than sense (b). Moreover, models of phenomena may

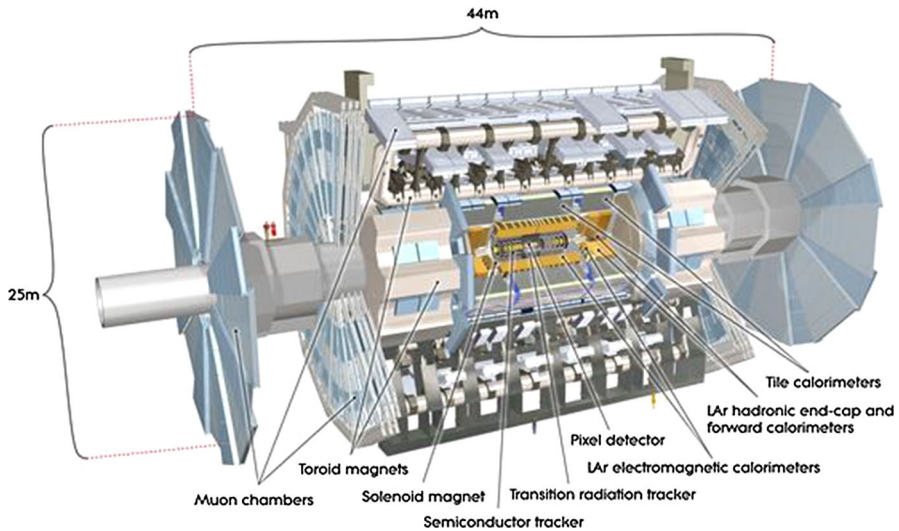


Fig. 2 Depiction of the ATLAS detector design (©CERN). Color figure online

be conceptually far richer, as Winsberg’s above list of methods for devising such a model shows. We will hence *not* treat ‘phenomenological models’ as synonymous with ‘models of phenomena’ here, a move that will prove fruitful in establishing various model-relations below.

3 The plethora of simulation models used by LHC’s ATLAS experiment

Standard accounts of the CSs used at ATLAS divide their totality

into three steps[...]: *generation* of the *event* and immediate *decays*[...], simulation of the *detector and physics interactions*[...], and *digitization* of the energy deposited in the sensitive regions of the detector into voltages and currents for *comparison* to the readout of the ATLAS detector[...]. (ATLAS Collaboration 2010, p. 834; emphases added)

An *event* here means the collision of two protons inside the beam pipe, including all interactions between their constituent partons (quarks and gluons) as well as the subsequent processes that ultimately lead to responses in the detector. The ATLAS detector (Fig. 2) is a “complex apparatus consisting of an inner detector, electromagnetic and hadronic calorimeters, a muon detector and four magnet systems, one of those[...] interleaved with detector components.” (Assamagan et al. 2005, p. 322) The inner detector itself comprises different technologies, such as pixel detectors, silicon micro-strip detection layers, and stacks of straw tubes (cf. *ibid.*). The detector hence consists of many discrete units, which implies the need for simulating digitization. We will, however, omit discussion of this step, as event generation and detector simulation already provide a rich basis for analysis.

From the (digitized) signals gathered from the detector, the event, possibly including a subprocess of interest, must be *reconstructed*. To a large extent, simulating events and detector responses serves the optimization of this reconstruction, i.e., increasing one's ability to determine what must have happened in the beam pipe and whether the process of interest has probably occurred.

3.1 Event generation

Event generation means the simulation of an event by means of Monte Carlo (MC) methods.¹¹ Generators can be grouped into *multi-* and *special purpose* ones, the most important multi-purpose ones being PYTHIA, HERWIG, and Sherpa, with PYTHIA and HERWIG having the longest tradition. New versions always go through a large amount of validation and benchmarking before being widely employed.

Special purpose generators include, e.g., Powheg for matrix element calculations at next to-next to leading order in the strong coupling (multi-purpose generators being restricted to next to leading order), or 'afterburners' like Photos for accurate simulation of photon radiation or Tauola for tau decays.¹² The final output are data sets in a standardized format, interpreted as 'stable particles' to be propagated through the detector simulation.

Any event will involve a large number of contributing processes. A fundamental assumption is that the totality of these can be *factorized*. 'Factorization' is really a technical term here that has little to do with, say, the factorization of functions for a separation of variables.¹³ It means "the ability to isolate separate independent phases of the overall collision[...] dominated by different dynamics[...]" (Mangano and Stelzer 2005 p. 556) These phases are: (i) hard process, (ii) parton shower, (iii) hadronization, (iv) underlying event, and (v) unstable particle decays (cf. Seymour and Marx 2015, p. 288).

To keep the amount of detail within sensible bounds, we focus on (i)–(iv). Moreover, to avoid confusion, we will from now on capitalize 'FACTORIZATION' wherever it means the technical notion in the sense of a separation of stages.

3.1.1 Hard process

The *hard process* usually is (or contains) the target phenomenon in Karaca's sense. It is the process with the highest momentum transfer in the entire event, an interaction between constituent partons of the colliding protons (cf. Seymour and Marx 2015, p. 288; Quadt 2007, p. 9). The central quantity to be determined in a MC simulation here is the *scattering cross section*, σ , which has the physical dimension of an area and provides information about the expected distribution of particles as a result of the scattering.

¹¹ 'Monte Carlo' is the label attached to "any technique making use of random numbers to solve a problem." (James 1980, p. 1147) An *MC simulation* is a simulation that relies on MC techniques. Often times *pseudo random numbers* are used in fact.

¹² We thank Dominik Hirschiühl for some updates on developments.

¹³ We thank an anonymous referee for pointing out that ambiguity may create confusion at this point.

It is generally assumed that a cross section $\sigma^{pp \rightarrow cd}$, for the collision of two protons (p) with resulting particles c and d , can be factorized into cross sections $\sigma^{ab \rightarrow cd}$ for elementary processes and additional information on the structure of the proton.¹⁴ Notably, this “factorization has only been proven in a couple of examples: inclusive deep inelastic [electron-proton—FJB & CZ] scattering (where one measures only the outgoing electron) and the Drell-Yan process (lepton pair production from pp or $p\bar{p}$ collisions).” (Schwartz 2014, p. 685) Another example are processes of the form “hadron A + hadron $B \rightarrow$ hadron C + X , where the X denotes ‘anything else’ in addition to the specified hadron C .” (Collins et al. 1989, pp. 1–2)

The elementary cross section includes a quantum field theoretical matrix element, so events simulated on the basis of it will be distributed according to perturbative calculations from quantum chromodynamics (QCD). This means that the matrix element will be expanded as a series in orders of the strong (QCD) coupling, higher terms being interpreted as smaller and smaller perturbations to a process without interaction. In fact, only a small number of contributions can actually be computed; to date next-to-next-to leading (i.e., third) order computations are state of the art. Such a perturbative treatment, moreover, is only possible for high momentum transfers, as quarks only exhibit asymptotic freedom over small distances (e.g. Barr et al. 2016, pp. 259 ff.).

According to the factorization assumption (and theorems), the aforementioned additional information on the proton is given by so called *parton density functions* (PDFs) $f_i(x, \mu_F^2)$, interpreted as a probability (density) for finding a parton of flavor i and (longitudinal) momentum fraction x in the proton when colliding it with another one at the (energy-momentum) scale μ_F^2 . μ_F^2 is the *factorization scale*, the scale at which the factorization may be assumed valid. Its value is not determined by theory, but has less impact on calculations with the inclusion of higher terms in perturbation series (Quadt 2007, p. 10).

The PDFs themselves cannot be computed in QCD and must be determined in ‘global fits’, i.e., by fitting a range of free parameters simultaneously to a large range of data (ibid.). Using these PDFs to select an elementary process, the cross sections $\sigma^{ab \rightarrow cd}$ remain to be determined by a MC simulation. Essentially, the simulation will randomly select values for the phase space variables (four-momenta) contained in $\sigma^{ab \rightarrow cd}$ (cf. also Baer et al. 2004, pp. 3–4, for a concrete example), which is considered the ‘generation’ of an elementary event.

Simply sampling the distribution given by the matrix element, however, will usually not produce *low-probability events*, which may be precisely the events of interest. One hence needs to ‘unweight’ the distribution by another MC technique (cf. Baer et al. 2004, p. 4), effectively discarding a large amount of (high-probability) events, or alternatively, use a *uniform* random number generator, likely to produce a sufficient amount of events in question. The latter will not reproduce the distribution predicted by the matrix element though, whence events so generated must be treated as having occurred with a relative frequency that approximates the matrix element at the respective phase space point.

¹⁴ This actually only holds term-wise; one needs to sum over different parton flavors and integrate for momentum fraction to obtain $\sigma^{pp \rightarrow cd}$.

3.1.2 Parton shower

The next part of the FACTORIZATION, the *parton shower*, means the simulation of additionally radiated gluons, in analogy to the emission of photons by accelerating and decelerating electrical charges.¹⁵ Gluons themselves, however, carry color charge, so they can also emit further gluons (cf. Seymour and Marx 2015, p. 290; Mangano and Stelzer 2005, p. 564).

In high energy collisions, a significant amount of such gluons is to be expected. Algorithms that simulate their emission implement *Markov chains*; step-wise evolutions of probability distributions, ‘forgetful’ in the sense that each step depends only on the previous one (e.g. Thijssen 2007, p. 299). In shower algorithms, Markov chains make use of two ingredients derived from QCD: *branching probabilities*, providing the probability that a parton of type i radiates a parton of type j with fraction z of i ’s momentum, and *Sudakov factors*, defining the probability of *no* emissions between the present and previous step. At any point, their product defines the probability of an emission at the given step (cf. Mangano and Stelzer 2005, pp. 565–6; Plehn 2014, p. 231; Schwartz 2014, p. 684).

A number of things are noteworthy. First, the evolution now *proceeds* in $\mu^2_{(F)}$, so it is not interpreted as a factorization scale anymore. Rather, it provides a *virtuality scale*, the intuition being that, as partons come closer and closer to the endpoints of Feynman diagrams, they must be ‘less and less virtual’.¹⁶ Effectively, μ^2 provides a time variable for the Markov evolution (cf. Schwartz 2014, p. 684).

Second, branching probabilities are derived in QCD in a *collinear approximation*, where the opening angle between radiated and radiating parton is much smaller than any angle in the hard process (cf. Seymour and Marx 2015, 293). Branching probabilities actually grow as μ^{-2} , however, and *diverge* for exactly collinear emissions, which obviously spoils their interpretation as probabilities and introduces the need for a *cut-off*.¹⁷ This is usually regarded unproblematic, “because *detectors* will not be able to *resolve* two partons very close together” (Mangano and Stelzer 2005, p. 565; *emph. added*). But considerations of *instrumentation* are thus invoked to resolve shortcomings on the side of theory. The relevant results also do not fix the physical dimension of μ^2 to a (four-)momentum; it could be “*any variable*[...] that becomes singular in the collinear limit [...]” (Schwartz 2014, p. 683; *emph. added*)

Third, Sudakov factors are an exponential in (negatively) integrated branching probabilities, but it is generally impossible to evaluate the integral. One hence either needs to resort to numerical approximations and store the results in lookup tables or integrate them ‘on the fly’ with a sophisticated MC method (cf. Baer et al. 2004, p. 27).

¹⁵ Gluons can be emitted in the form of *initial* or *final* state radiation, corresponding to Feynman diagrams with an additional vertex and (gluonic) leg on either of the incoming or outgoing partons from the collision respectively (cf. Seymour and Marx 2015, p. 290–1). Both processes are modeled essentially in the same way though.

¹⁶ Virtual particles correspond to (unobservable) *internal* lines of Feynman diagrams (cf. Griffiths 2008, p. 63). They are also ‘off shell’ and so do not satisfy the relativistic energy-momentum relation, whence they can only exist to the extent that the uncertainty relations allow violations of energy and momentum conservation (e.g. Schwartz 2014, p. 46).

¹⁷ They are also divergent for $z \rightarrow 0$ (called the ‘soft’ limit), but the cut-off for μ^2 implies a constraint on z (cf. Mangano and Stelzer 2005, pp. 565–6).

Moreover, Sudakov factors actually introduce a new variable, q^2 , that can be identified with μ^2 only up to some scaling function of the momentum fraction z , leaving room for several choices. There are also various possibilities for modifying the branching probabilities, e.g., by introducing non-vanishing quark masses on which they depend (Baer et al. 2004, p. 28).

The (highest) value of q^2 at which the Markov evolution begins needs to be chosen as somehow “characteristic of the *hard process*” (Baer et al. 2004, p. 27; *emph. added*). However, with radiative and virtual corrections to perturbative calculations of matrix elements, both parton shower and hard process may generate the same states, leading to a *double counting* (Mangano and Stelzer 2005, p. 564). Since fixed order matrix elements provide better predictions than showering algorithms in, e.g., searches for new physics, predictions of backgrounds, or precision measurements, but parton showers provide information about the internal structure of jets that is absent from matrix element calculations (Seymour and Marx 2015, p. 301), strategies to avoid the double counting while retaining the benefits from both are usually implemented, which go by the names of *matrix element matching and merging* (e.g. Höche 2016, p. 266 ff. or Mangano and Stelzer 2005, Sect. 5.2-3 for details).

3.1.3 Hadronization

The next part of the FACTORIZATION is *hadronization*, the formation of hadrons¹⁸ from the bulk of partons produced in hard process and parton shower. The underlying idea is that, once the energy of individual partons is low enough, the confining effects of QCD become important (Mangano and Stelzer 2005, p. 568). At this stage, however, QCD-calculations break down and one has to find sensible phenomenological models. Two notable types are *string* and *cluster models* (e.g. Seymour and Marx 2015, pp. 305 ff.).

String models start from the assumption that, when a quark and antiquark move apart in space from a pair-creation point, “the self-interacting colour field between them collapses into a narrow ‘flux tube’ of uniform energy density per unit length, or string tension κ , estimated to be $\kappa \approx 1 \text{ GeV} / \text{fm}$.” (Dissertori et al. 2003, p. 162) This string-like color field can absorb the increasing energy as the quarks move apart (recall that QCD forces grow with distance). At some point the string may ‘tear’, meaning the creation of $q\bar{q}$ -pairs from the field energy.

In the so called *Lund model* (e.g. Dissertori et al. 2003, p. 162 ff.; Seymour and Marx 2015, p. 306 ff.), the probability density for obtaining a fragment with some fraction of the momentum of the original string obtains a concrete form on the basis of simplifying assumptions such as an average left-right symmetry for the breakup. Building on a result by Wilson (1974) also allows to relate the probability of a breakup to the area of the past lightcone of the breakup point (Dissertori et al. 2003, pp. 162-5).

The basic probability density of the Lund model has two free parameters, but in practice, algorithms introduce further ones. With the additional simplifying assumptions that a string cannot be excited longitudinally and that the string-tearing is indeed

¹⁸ *Hadrons* are particles made of quarks, which include mesons (two quarks) and Baryons (three quarks). *Leptons* like electrons are not made of quarks.

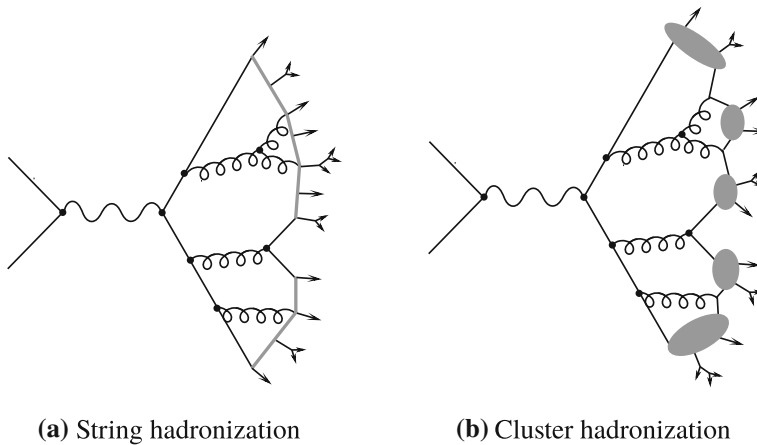


Fig. 3 Pictorial representation of string and cluster hadronization for the exemplary process $e^+e^- \rightarrow$ hadrons (cf. similarly Ellis et al. 2003, pp. 189–90)

pointlike, one of the two fundamental parameters can be related to the string tension κ for a first approximation (cf. Dissertori et al. 2003, p. 166). The resulting formula gains some support from an analogy with an approximate QED result, but the relevant approximation is rather poor (cf. Seymour and Marx 2015, p. 306; Dissertori et al. 2003, p. 166).

In some models, connected $q\bar{q}$ -pairs can ‘bounce back’, leading to so called ‘yo-yo modes’ (cf. Seymour and Marx 2015, p. 305); a model of an intermediate meson. Gluons, moreover, can be introduced as transverse excitations or ‘kinks’ in the string (cf. Fig. 3a), which leads to the qualitative prediction that hadrons will mostly be found in directions between quark and gluon or anti-quark and gluon, because they will be emitted in the direction of fragmented string segments.

This prediction can be investigated by appeal to data from two- and three-jet events (cf. Seymour and Marx 2015, p. 308), but the Lund model’s predictive power is undermined by the fact that it works in terms of unknown properties such as quark masses. The so called UCLA model at least partly compensates this by building on known hadron masses (Dissertori et al. 2003, p. 166).

A general reason for dissatisfaction is that string models contain a “smooth matching” (Seymour and Marx 2015, p. 308) with the parton shower, and so it is difficult to retain the information from the first two stages. *Cluster models* do not share this feature.

The starting point here is *preconfinement*, a result retrieved in certain QCD approximation, which means the organization of quarks into color singlets— $q\bar{q}$ -states of zero color also called ‘clusters’. Since this concerns only quarks, gluons from the parton shower have to be split into $q\bar{q}$ -pairs first, without reliance on QCD (Ellis et al. 2003, p. 190).

The resulting (mesonic) clusters exhibit a universal *mass distribution*, peaked at low mass and asymptotically independent of the scale of the hard process. It does, however, depend on the parton shower cutoff and the scale where perturbative QCD

breaks down (cf. Dissertori et al. 2003, p. 169; Amati and Veneziano 1979, p. 87). The universal low mass peak justifies the assumption of an isotropic two-body decay for the clusters (Seymour and Marx 2015, p. 309), but some clusters will be too heavy for this and first must be split into lighter ones (Bähr et al. 2008, p. 68). The final decay-products then represent the (known) hadrons to be registered in the detector.

Given its independence from many details of the other processes, the cluster-approach does not suffer from the washing out of information from earlier stages present in string models, but it generally does not produce as nice a fit to data. Older cluster models also had fewer parameters than string models, but did not produce enough hadrons with realistic energies and momenta (Seymour and Marx 2015, p. 309). In fact, “[t]he accumulation of more precise data”, and according modifications, “have led [string and cluster—FJB & CZ] models to converge to something more similar.” (ibid.)

3.1.4 Underlying event

The *underlying event* means taking into account further interactions between the ‘proton remnants’, the remainder of partons not involved in the hard process. Originally, models used to be *non-perturbative* parametrizations of data, but nowadays these are of little interest. Perturbative models usually assume multiple *independent* interactions across the proton remnant. This can be justified by considering that the average diameter of the remnants is ~ 1 fm whereas *hard* parton interactions are confined to much smaller distances, making it “very unlikely that there are additional partons exchanged between these multiple scatters.” (Gieseke and Nagy 2011, p. 117). Each interaction can be modeled in analogy to the hard process and requires its own parton shower accordingly.

In general, there are many open details still. For instance, the number density for scatters between partons from the remnants usually includes a PDF, but there is “no clear theoretical justification for using the PDFs in the same (or some modified) way as for the hard scatter, and therefore the existing event generators use different strategies to model these ‘remnant PDFs’.” (Gieseke and Nagy 2011, pp. 117–8)

Moreover, partons produced in hard process and underlying event are correlated in energy and momentum, as they stem from the same protons. Hence, if both processes are modeled independently, this can lead to nonsensical values of momentum fraction x . Simply discarding events with too high x would spoil the distribution predicted by the PDFs, so generators need to compensate this fact.

There is also the possibility of *color reconnections* between proton remnants and decay products from hard scatter and parton shower. Including corresponding color correlations will lead to a better fit with data from soft collisions, but it also “necessitates a retuning of the [...] hadronization parameters”, as these “are correlated with the reconnection probability[...].” (cf. Seymour and Marx 2015, p. 315) Finally, models that drop the (arbitrary) assumption that spatial and momentum information can be factorized in the number density “obtain equally good fits of the existing underlying event data but with significant differences in their extrapolation to higher-scale processes.” (Seymour and Marx 2015, pp. 316–7)

3.2 Detector simulations

A fundamental ingredient to all detector simulations is a model of the detector geometry and the physical properties of its parts. In ATLAS, this model appeals to three kinds of ‘volumes’: “*solids*, basic shapes without a position in the detector; *logical* volumes, solids with additional properties (e.g. name or material); and *physical* volumes, individual placements of logical volumes.” (ATLAS Collaboration 2010, p. 843) This has several advantages; for instance that one may use a single logical volume to create repeating shapes.

Simulating the detector geometry in this modular way, one encounters certain pitfalls. For instance, different volumes will be modeled as perfectly densely spaced. This can lead to *overlaps* between different (virtual) components, leading to *stuck tracks*, as the particle may happen to belong properly to neither volume (or both). These can be overcome by introducing small gaps between volumes, which means an extra step for each particle in the transition region (ATLAS Collaboration 2010, p. 844).

Another fundamental component to detector models is a map of the magnetic fields used to bend the trajectories of charged particles, thereby measuring their momentum. This map is generated from a mixture of computations based on the Biot-Savart law, numerical integration, and finite-element modelling (ATLAS Collaboration 2008, p. 30). Computing the field to desired accuracy, including local perturbations by magnetizable material, requires prior simulation and experimental spot-checks (ATLAS Collaboration 2008, p. 31). Taken together, geometry and material, field map, and some additional models of the electronics exhaust the (pure) models of instrumentation in Karaca’s sense.

In contrast to event generators, detector simulations do not factorize so neatly. Models of interactions between stable particles and the detector material, constitute the major part, i.e., hybrid models in Karaca’s sense. Reasonably precise modeling here requires input from atomic, nuclear, particle, as well as solid state physics.

The GEANT4 physics manual (GEANT Collaboration 2016) gives an impressive overview. Most models contained therein are phenomenological or even just parameterizations, tuned to data from well-controlled experimental setups (e.g. thin target data). The reasons reside either in an incomplete theoretical basis (e.g., in nuclear physics), or in the fact that the theory is only valid for a simplified process (e.g. scattering on a single atom vs. correlated effects in a solid of mixed atom types). We will sketch a few exemplary models below.

3.2.1 Particle decay

The sampling of decay events within the detector from decay rates is presumably the only process purely based on theory (cf. GEANT Collaboration 2016, pp. 13–9). The reason is that ‘decay’ here refers to the quantum mechanical possibility of a particle decaying without any external stimulus. Decay rates thus depend solely on the particle type and the particle’s four-momentum, not on the properties of the surrounding material.

3.2.2 Bremsstrahlung

The electromagnetic interaction between charged particles and screened nuclei or electrons in atomic shells results in acceleration and deceleration of the particle, associated with the emission of photons (*Bremsstrahlung*). On the atomic level, this effect can be calculated from first principles (e.g. Peskin and Schroeder 1995, Sect. 6.1). Within a material consisting of multiple elements with different densities, however, corrections and phenomenological assumptions are required in order to obtain a realistic photon spectrum (energy and angular distribution).

A particular example is the Seltzer-Berger model (cf. GEANT Collaboration 2016, Sect. 10.2), which splits the differential cross section for interactions leading to *Bremsstrahlung* into contributions by screened nuclei and electrons in atomic shells. The basis for these cross sections is an approximate theoretical result by Tsai (1974); but sampling this distribution “is complicated” (GEANT Collaboration 2016, p. 122), and so a simpler parametrization of the predicted distribution is used, “in good agreement with experimental data” (GEANT Collaboration 2016, p. 121).

3.2.3 Nuclear interactions

The most difficult processes to model are inelastic interactions of hadrons with atomic nuclei of the detector material. Due to the substructures of incident hadrons and target nuclei, the interaction is highly complex and includes all kinds of sub-interactions, ranging from elementary processes on the parton level down to nuclear and atomic excitation and evaporation. The list of existing simulation models is vast and it usually comprises also a range of different models for the same kind of interaction (cf. GEANT Collaboration 2016, Part V). Many of these are theory-inspired (e.g. parton interactions at high energies) but also contain simple parametrizations.

This is especially true for very low energies, at which modeling on the basis of existing theoretical knowledge is fully insufficient. Parametrizations here build on measured cross section data (as a function of energy) for different projectiles and different target nuclei. Final state particles are exhaustively characterized by energy, angular and multiplicity distributions, and excitation energies of the remnant nuclei.

4 Polycratic hierarchies, networks, and the justification of simulation-based inferences

The above gives a flavor of the complexity involved in HEP simulations. Before we can address the applicability of Winsberg’s hierarchical account to this study case, let us once more confront a terminological subtlety. Recall that Karaca distinguishes between models of instrumentation, of target phenomena, and hybrid models. However, the LHC was built to find and measure elementary particles like the Higgs or from physics beyond the SM. These are only to be expected at very high energies, i.e., as decay products from the hard process. Hence, most of the phenomena modeled in generators and detector simulations are not really the target phenomena of the (overall) simulation at all, i.e., not the phenomena of *primary interest*. They are of interest only for the

purpose of being able to infer the presence and properties of sought-for particles from characteristic patterns in the detector. Call these *secondary* phenomena, in contrast to the (primary) phenomena truly targeted by HEP simulations.

4.1 Applying Winsberg's hierarchy...with some ifs and buts

To see how Winsberg's hierarchy can be plausibly applied in HEP, consider now the parton shower. The basis for any CS will here be approximate QCD results, namely branching probabilities and Sudakov factors. Their application as a model of the showering process will involve general physical modeling assumptions, such as the negligibility of higher terms in the strong coupling. Moreover, pragmatic *choices* of the concrete form of branching probabilities, the evolution variable, as well as its connection to the variable of the Sudakov factors, will create a plethora of different simulation models. Still, these are all within the scope of QCD results, and so resulting simulation models may be thought of as *theory driven*.

To accurately model *initial state* radiation (gluons emitted before the hard process), PDFs from global fits are needed for showers as well. However, unlike in the hard process, they should here be understood as (providing information on) initial data, since they will be solely used to achieve accurate normalization in the backwards evolution from the initial state (Seymour and Marx 2015, p. 297). A similar function attaches to the highest value of the Sudakov evolution variable, q^2 , gathered from one's model of the hard process, as well as the lower cutoff, gathered from one's knowledge of the detector resolution. Both of these deliver boundary values for the evolution. In this way, one obtains a dynamical model in Winsberg's sense.

Resulting such models are then implemented in the form of a Markov chain. The approximation of the integrals involved in Sudakov factors means a discretization, and the assumption of (the very existence of) a cut-off for the evolution variable as induced by experimental constraints may be considered part of the ad hoc modeling assumptions mentioned by Winsberg.

This is a straightforward application of the steps involved in Winsberg's hierarchy to a HEP example. However, taking a closer look, we can already see some limitations: Sudakov factors *themselves* already amount to a discretization of the probabilistic dynamics into (*time*) *steps*. This means that some amount of discretization already figures at the conceptually prior level of devising a simulation model, and that discretization and coding can fall apart. Moreover, somewhat ad hoc pragmatic considerations, such as the instrumentation-based definition of a cutoff or a choice of ordering (angular vs. momentum) of the evolution steps, will too feature already in the formation of the relevant simulation model. They do not occur only as part of the coding process or, more generally, in the process of devising a computational model. Hence, we suggest to adapt Winsberg's hierarchy to the case of parton showers as displayed in Fig. 4.

These adaptations, like those discussed already in Sect. 2.2, are so far rather benign, but they do underscore how simulation modeling is not easily fit into a tight, preconceived format. Moreover, HEP simulations may also urge some less benign adaptations, as can be seen on the example of string models of hadronization.

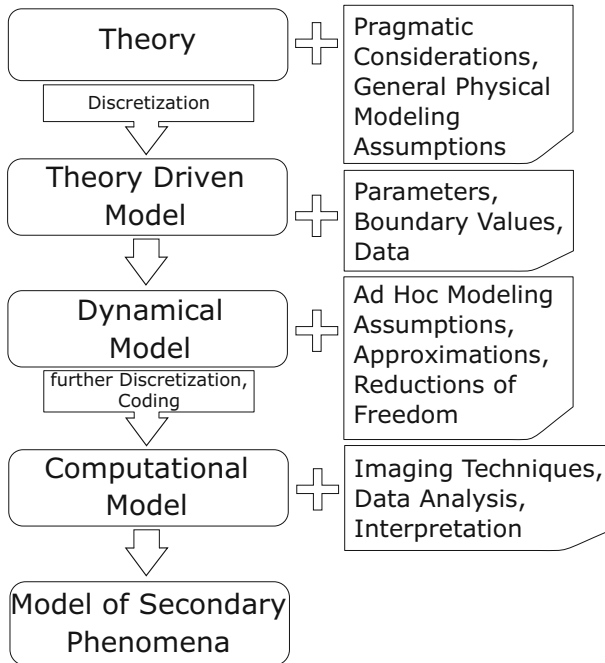


Fig. 4 Hierarchy of models in the simulation of parton showers

String models of hadronization are generally based on an interpretation of the data retrieved from showering algorithms in terms of the behavior of a (semi-classical) stretchable, tearable string. Basically all string models are based also on a particular mathematical relation, the area-decay law, but the specific implementation of this law will ultimately depend on the precise assumed properties of the strings. Quite usually, these will be conceptualized in terms of diagrams and illustrations depicting the underlying physical intuitions in the modeling process, or even by means of qualitative textual descriptions. In so far as this is part of the concrete physics modeling, basic string models hence amount to *models of phenomena*.

However, only the inclusion of further theoretical knowledge will lead to workable models that can be translated into an algorithm. This will include all sorts of relations oriented on features of QCD, but not provide a *fundamental* treatment of hadronizing partons. Selecting, for instance, a precise form of the decay law by appeal to the relation between quark and hadron masses as given by theory (cf. Dissertori et al. 2003, p. 167), one retrieves an enriched, workable model that can serve as the basis for an implementable algorithm and is phenomenological in the sense established in Sect. 2.3: non-fundamental, not too closely connected in construction to either theory or observation, displays exploitable mathematical relations and includes fitable parameters. Unlike a proper model of the phenomena, it will be deprived of other conceptual devices, such as illustrations or descriptions.

Hence, it seems incorrect to say that we straightforwardly obtain a simulation model from QCD, together with a set of general physical modeling assumptions. Rather, the

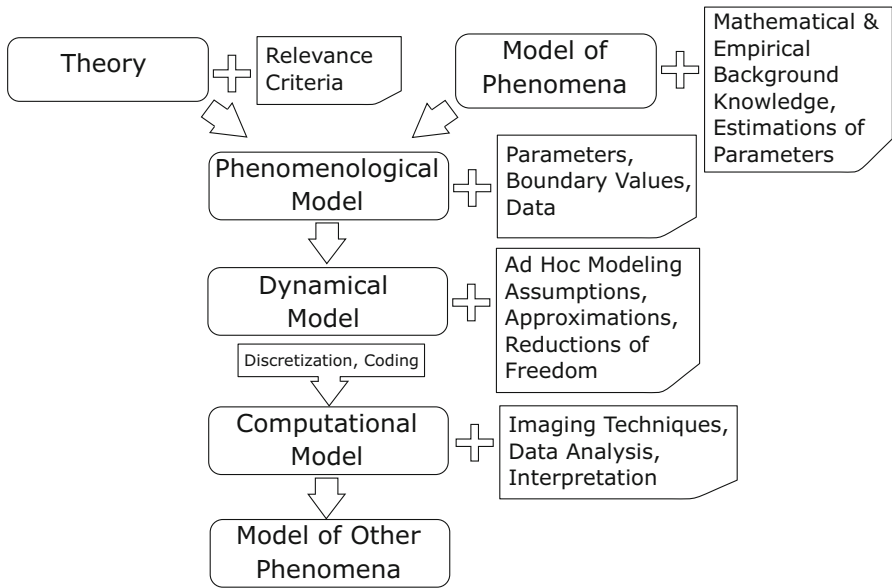


Fig. 5 Hierarchy of models in the simulation of string hadronization. All phenomena in question are secondary

top level of our hierarchy should be populated by *two* entities here: the underlying theory, as well as a genuine model of (secondary) phenomena.

To retrieve from these a properly so called phenomenological simulation model requires taking into account also *relevance criteria* for the selection of particular elements of QCD, empirical (prior data sets and measurement results) and mathematical (combinatorics, probability theory) background knowledge, as well as estimates of parameters (e.g. $\kappa \approx 1\text{GeV}/\text{fm}$).

From thereon, Winsberg’s hierarchy continues: The free parameters of a given string model have to be fitted to data, and the entire model must be discretized and translated into computer code for the execution of a simulation. Interpreting the output, one obtains a model of further secondary phenomena: *hadrons* that are either stable or decay into ‘jets’.

The structure adequately representing the relations between models for CSs of hadronization is thus still hierarchical, as displayed in Fig. 5. However, in contrast to Winsberg’s proposal, theory is here not the only entity on the top of the hierarchy: the top level is populated by theory together with models of phenomena. The resulting hierarchy is thus *poly-* rather than *autocratic*: Several distinct entities jointly govern the subsequent modeling steps.

A similar diagram can be drawn for the hard process, but the joint governance will here be different. Recall that the PDFs, which could not be obtained from theory, constitute a *fundamental* ingredient for modeling the relevant scatterings in the hard process. Hence, it is not a model of the phenomena but rather the PDFs gathered from fitting to a large range of experimental data that, together with QCD and (in most

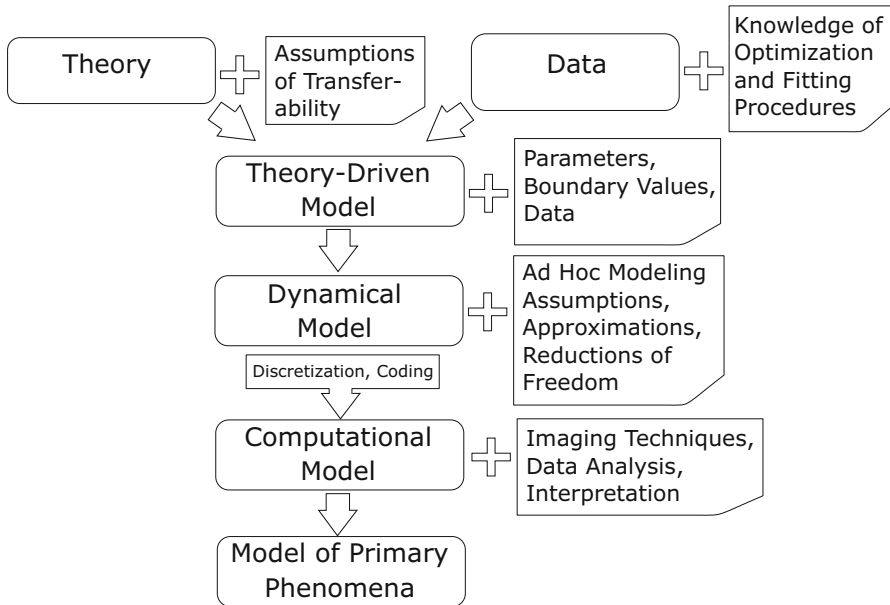


Fig. 6 Hierarchy of models in the simulation of the hard process

cases) the *assumption* of a factorization into an elementary cross section plus PDFs, give rise to a simulation model (cf. Fig. 6). Since the *existence* of PDFs and perturbative calculations of elementary cross sections come directly from QCD, theory should still be considered the driving force. Everything else then repeats as above.

Quite different modifications of Winsberg’s hierarchy can be gathered from hybrid models in the detector simulation. Consider, for instance, models of nucleon-nucleon interactions. Here, one proceeds quite directly from existing data (from the so called SAID database; cf. GEANT Collaboration 2016, p. 255), interpolated in the most convenient way relative to one’s needs of fit and speed, to a basic simulation model. Geant4 references a *linear* interpolation in both the values of the cumulative distribution retrieved from the cross-section data as well as the energies. This is the simplest interpolation method available, and its use here is presumably connected to reduced computation times.

Because data points are being interpolated, it makes sense to speak of a (primitive) model at all, not just of a data set called in the overall simulation program. Since data are the driving force here, moreover, we may call this simulation model *observation driven*. From this, one then proceeds *directly* to a computational model whose output can ultimately be interpreted in terms of a model of the (secondary) phenomena occurring inside the detector (Fig. 7).¹⁹ No dynamical model will be created; there simply is no model of the dynamics and the whole situation is effectively (though by no means explicitly) treated as one of instantaneous transition from interacting input nucleons

¹⁹ Note that ‘boundary values’ in Fig. 7 does not refer to the boundary conditions of a differential equation, but rather to the range of values for which data exist and wherein the interpolation is valid (10–1200 MeV for the SAID data).

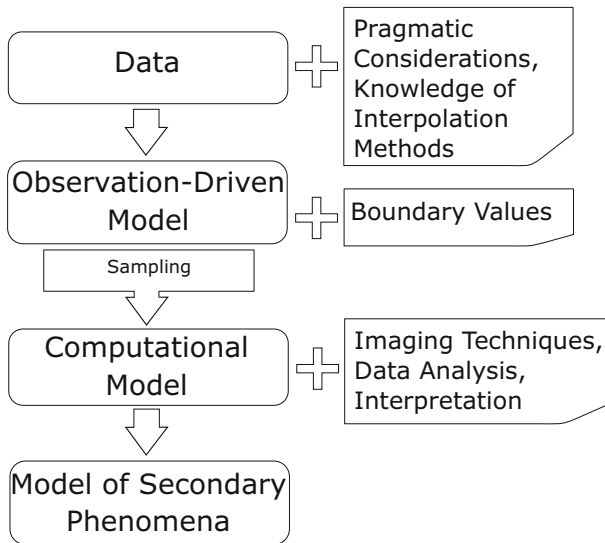


Fig. 7 Hierarchy of models in the simulation of nucleon-nucleon interactions

to output particles. Thus, Winsberg’s hierarchy can be fitted to this example only with the omission of one intermediate step.

We admit that one may question the appropriateness of the term ‘simulation’ for this part, due to the term’s dynamical connotation (cf. Hartmann 1996). Moreover, simply sampling some distribution according to some MC prescription is obviously borderline between ‘simulation proper’ and ‘mere computation’. However, Hartmann (1996, p. 83) emphasizes in particular the dynamical character of the *simulation step* itself, and the execution of the sampling by a computer is, of course, still dynamic. In other words, one might equally take it that the sampling prescription contained in the computational model constitutes a ‘covert’ dynamical model, and that two steps have here been merged into one.

Consider also what it would mean to think of this part as not being a simulation. It would mean that, when a detector simulation is run and nucleon-nucleon interaction models are being called up, the simulation *pauses* to make room for a computation. This is highly counter-intuitive, and conveys another *prima facie* reason to consider the term ‘simulation’ appropriate in this context.

More importantly, though, whether the *hard process* contains a dynamical model properly so called is debatable on essentially the same grounds: one here too simply samples a differential cross section; even if provided by the theory in this case. Given that the matrix element models the probability of a *transition* between two well-defined momentum states, and given that the kinds of constraints recognized by Winsberg will have to be applied here in order to create individual events from this matrix element, we think that a dynamical model *can* be said to exist in the latter case. However, the main difference between both kinds of sampling thus largely resides in the (un)availability of an underlying theory.

This seems like an ill-posed criterion for distinguishing simulation from mere computation, and so the boundary between both examples, when it comes to their being simulations or not, is fuzzy rather than well-defined. If one disputes, in other words, that nucleon-nucleon interactions are properly simulated this immediately raises doubts about other parts being properly simulated as well. For these reasons, we think that nucleon-nucleon interactions remain a valid example of simulation without a proper (or at least overt) dynamical model.

4.2 A network of simulation models

The previous section demonstrates that even in the case of CSs used by large-scale HEP experiments such as ATLAS, hierarchical accounts can, with some modifications, be successfully applied. These modifications include (i) the replacement of theory by data on the highest level (cf. Fig. 7), (ii) the omission of intermediate steps such as a proper dynamical model (cf. *ibid.*), (iii) the distribution of crucial techniques such as discretization across several consecutive (re)modeling steps (cf. Fig. 4), and, most importantly, (iv) the weakening of autocratic hierarchical structures to polycratic ones, wherein (e.g.) theory and data figure jointly on the top level (cf. Figs. 5 and 6).

There are lessons to be drawn from this already. For instance, the fact that data alone can populate the top level of the hierarchy impairs Winsberg's point that simulation modeling is "a case of theory articulation in the spirit of Kuhn." However, we already conceded that Winsberg is concerned with the special case in which CSs are used to cope with the intractability of theory. It is hence rather an extension than a refutation of his account to say that simulation modeling can also be an exercise in *data articulation*: It can allow us to tickle out the information contained in a set of data, by embedding them into a larger context and 'bringing them alive'.

Point (ii) was sufficiently addressed in Sect. 4.1, and there is no need to repeat the arguments here. Points (iii) and (iv), on the other hand, may be judged to reflect what Winsberg (1999, p. 266) calls the "motley character of simulation inferences"; that

our theoretical knowledge is just one of several ingredients that simulationists use to produce their results. All of these sources and their influences need to be considered when justifying simulation results. If the influences and possible pitfalls of each element are not properly understood and managed by the simulationist, they may potentially threaten the credibility of simulation results. Doing so, moreover, requires reliance upon an equally diverse range of sources of knowledge and skills. A great deal of this knowledge is not contained in the theoretical knowledge that formed the original basis for the simulation. (Winsberg 2001, p. S448)

As we can see, the inferences in question turn out to be even more motley than reflected in Winsberg's account. Sometimes extra-theoretic knowledge will not just contribute to the articulation of some theory in terms of a CS: It may be so important as to co-contribute, together with theoretical results, to the very *basis* of a simulation model. Moreover, the same kind of extra-theoretic knowledge may have to be used and *re-used* at multiple junctions in the modeling process.

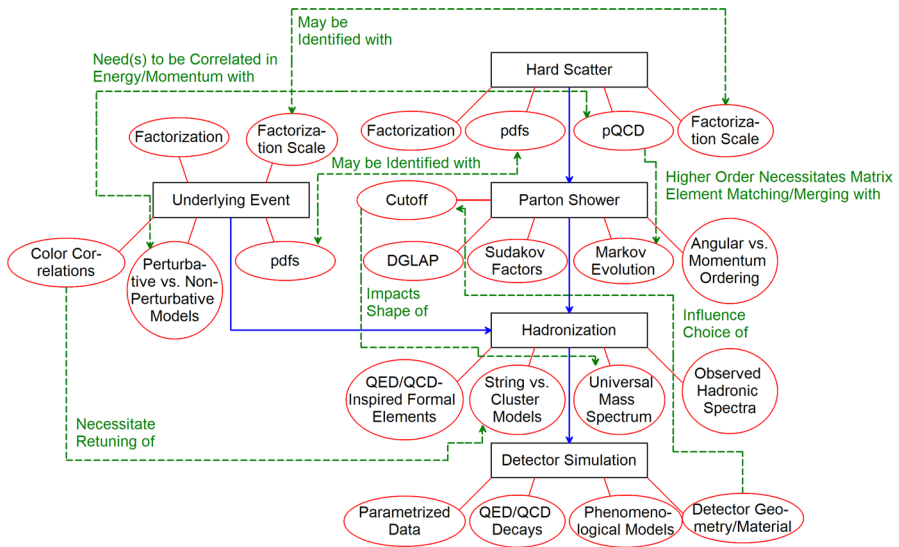


Fig. 8 ATLAS’ network of simulation models. Depicted are a few characteristic relations (dashed green arrows) gathered from the discussion. Solid (blue) arrows indicate the flow of data. ‘pQCD’ is short for perturbative QCD. Color figure online

All of this really *extends* Winsberg’s account, instead of impairing it; but the examples discussed in the last section were all concerned with only *one part* of the overall simulation. MC simulations actually used in experimental procedures in HEP will typically make use of the whole palette, or otherwise patch the linkage between primary phenomenon and reconstructed data by means of calculations and subsidiary data. So what happens when everything is chained together?

Building on the discussion in Sec. 3, we have isolated a number of relations pertaining between different simulation models. The result is an outright *network of simulation models*, depicted in Fig. 8. It is clear that these constitute only a proper subset of all relations between the component models, so an even more complex network may be assumed in fact.

Given the details provided in Sect. 3, the diagram should be largely self-explanatory. Still, a few details are worth discussing. The diagram, for one, includes a number of deliberate simplifications: for instance, details of models going into the detector simulation are not resolved in the same way as those of the event generator. Similarly, partons created in the underlying event will shower off further partons before contributing to hadronization, so the solid blue arrow going from underlying event to hadronization includes a ‘tacit’ parton shower.

Another issue is the apparent *loop* between the experimental constraints implied by the detector and models of the parton shower in generators with cluster models of hadronization, mediated by the influence of the cutoff-value on the mass spectrum of the clusters.

Per se, loops are nothing worrisome in simulation modeling. For instance, Lenhard (2016, p. 728) stresses that, when implementing a model on a computer, “iterations can

be performed in great numbers and at high speed[...]. Modeling hence can proceed via an iterated feedback loop according to criteria of model performance.” Loops of this kind, as similarly recognized by Boge (2019), concern modifications of a single implemented, computational model, according to its performance in comparison to benchmarks. The loop discerned in the diagram, however, concerns a modificatory feedback between two *different*, coupled simulation models, prima facie independently of comparison to experiment. One might hence suspect a *vicious* circularity between the models here, ultimately leading to the models’ sanctioning themselves.

Fortunately, this would be mistaken: First off, similar constraints on the possibility of resolving two showered partons would be implied by *any* detector model. A given cutoff should hence not be interpreted as strictly reflecting a feature of the *specific* model of ATLAS’ geometric properties and material composition, but rather a generic feature that transpires from our best understanding of present day detector technology. Secondly, geometry and material models *are* fully empirical and only include minor ad hoc corrections for simulation artifacts (e.g. stuck tracks due to infinitely dense spacing). Instead of entering into a vicious circle, empirical information about the experimental setup thus effectively enters in two places: in defining the models used to simulate the detector (response) itself and in defining what sorts of information from the phenomena of (secondary) interest can possibly influence the observable behavior of the detector in response to the interaction.

4.3 Justification (i): holistic validation

What is the significance of our network for the EoS? Recall the basic purpose that Winsberg had originally assigned to his hierarchical account: Through (re)modeling steps that preserve some (though usually not all) of what is known about a targeted system, but are also creative enough to provide new kinds of access, we come to obtain genuinely new knowledge of the given target. Moreover, because of their largely non-deductive nature, the steps need to be executed in such a way that we can *justify* simulation results as providing genuine knowledge.

However, it appears that this justification is not possible in HEP CSs, for most of the individual hierarchies recognized above *do not concern the target phenomena*: the LHC was designed to find the Higgs, determine its properties as well as those of certain quarks, and possibly gather evidence of particles beyond those of the SM. These are all particles reliably created only in hard (high energetic, elementary) scattering processes. But only if it is known what happens inside the beam pipe *beside* an elementary scattering process, and only if one understands the role of the detector in *creating* the characteristic experimental signatures, can one infer the existence and properties of specific kinds of particles from these signatures.

In other words: We do not deny that it would be possible to draw *some* inferences from the simulation of, e.g., a hard process alone; but these inferences would not reach beyond the beam pipe. There would hence be no possibility of directly *validating* corresponding simulation results, as our empirical access is inevitably mediated by the experimental setup, including the detector.

Compare our findings to those of Lenhard and Winsberg (2010, 2011), who recognize a *kludgy* methodology and a *messy* overall structure to the simulation modeling in *climate science*. Salient features that Lenhard and Winsberg (2010, 2011) highlight for climate simulations are a fuzzy modularity, a generative entrenchment of the different modules, the inclusion of kludges, and, related to all this, a *holism* regarding validation.

For now, let us focus on the ‘fuzzy modularity’, by which Lenhard and Winsberg (2010, p. 256) mean the following: “normally, modules are thought to stand on their own. In this way, modularity should have the virtue of reducing complexity.” However, in contexts such as climate science, “the modules [...] are interdependent and therefore lack this virtue” (ibid.).

As we have seen, this is certainly also true to a large extent of HEP. The assumed FACTORIZATION of the event, as well as the neat separation between events in the beam pipe and in the detector, have turned out to be necessary but ultimately idle simplifications: It is impossible to strictly separate all these different phases, as vividly illustrated in Fig. 8. Hence, a hierarchical account cannot possibly deliver the ways in which HEP researchers justify their simulation-driven inferences to target systems.

This holistic aspect is actually well-known to physicists. For instance, (Dissertori et al. 2003, p. 172; *emph. added*) write:

Monte Carlo event generators[...] *inextricably couple* the perturbative parton shower and non-perturbative hadronization model. Furthermore, both components contain free parameters which can be simultaneously tuned to the data. As a consequence, it is often difficult to correlate the inherent properties of the hadronization model with the quality of a Monte Carlo’s description of the data.

Another reference discussing the subtleties of validation in HEP in a philosophical context is Mättig (2019), who lays some emphasis on comparison of elements from generators and detector simulations to data taken *across* different experiments (i.e., other than those at the LHC). But Mättig (2019, p. 647) really only invokes these cross-checks as a means for “avoiding circularity”, and so does not touch on the holistic aspect.

As mentioned earlier, new releases will, in fact, be treated with much caution in the community, and not be applied in analyses until they have survived a whole host of benchmarking procedures over many years. But this of course means that *all* the (computational) models contained in the new release will have to undergo validation *in concert*, and possibly even in concert with other, already established models (e.g. in the case of special-purpose generators that cannot stand alone).

There is an important difference, however, between generators and detector simulation. After the hadronization stage, the remaining particles will all be known particles that can be produced and experimented on in contexts other than the LHC. Accordingly, the GEANT Validation Portal²⁰ includes the possibility to compare predictions from various models included in the simulation toolkit *individually* (and also in different versions) to reference data, produced in experiments that are relevantly similar to the respective conditions in the ATLAS detector. For example, thin foils of certain

²⁰ Cf. <https://geant-val.cern.ch/>; checked 02/20.

materials allow for a more direct assessment of the responses of individual materials used in the construction of ATLAS, because scattering hadrons and leptons on them produces data “without the effect of other processes like particle propagation or electromagnetic physics effects.” (Banerjee et al. 2011, p. 2)

In sum, a first result is that, while checks across experiments may safeguard at least against vicious circularities spoiling the *possibility* of validation, ‘fuzzy modularity’ is present in HEP as well, and it implies that justification by comparison to empirical data is in part possible only in a *holistic* fashion, much like in climate science. A difference arises between generators and detector simulation though, and this may somewhat lessen the impact of this kind of holism.

4.4 Justification (ii): model coherence

Another key difference between both fields lies in the details of how (and even why) models are *coupled* therein. In particular, the ‘generative entrenchment’ Lenhard and Winsberg (2010, 2011) recognize in climate science means

that climate models are, in interesting ways, products of their specific histories. Climate models are developed and adapted to specific sets of circumstances, and under specific sets of constraints, and their histories leave indelible and sometimes inscrutable imprints on these models. (Lenhard and Winsberg 2011, p. 116)

Certainly, there is some historical component to the development of HEP simulations as well and it does play a role in how simulation models and their implementations have played out. For instance, the GEANT detector simulation tool has been in continuous development since 1974, where it “initially emphasised tracking of a few particles per event through relatively simple detectors” (Brun et al. 1993, p. 5). Similarly, the need for matching and merging arose from developments in the ability to compute matrix elements to higher orders.

Various elements of detector simulations as well as such things as matching and merging algorithms, moreover, arguably fall under the category of *kludges*, held to be among the main reasons for entrenchment by Lenhard and Winsberg (2010, cf. p. 257).

By ‘kludge’, one means “an inelegant, ‘botched together’ piece of program; something functional but somehow messy and unsatisfying.” (Clark 1987, p. 278) One may dispute the ‘inelegance’ of matching and merging algorithms themselves, but it certainly seems unsatisfying that partons have to be identified and merged or removed *ex post*, and cannot be *produced* in a coherent fashion right away, when higher contributions from matrix elements are taken into account.

However, this is only part of the story in HEP, because many of the holism-promoting connections are established on purely *theoretical* grounds. For instance, asymptotic freedom implies that questions of hadronization can only play a role when the energy is sufficiently low, and so the dependence of the hadronization algorithm on the cutoff is at least partly theory-induced. Accordingly, it can be assessed also on theoretical grounds. Similarly, color reconnections come into play because of a more

detailed treatment of the proton remnants and the underlying event, and the expectation that they do play a role is hence rather theory-driven as well.

Certainly, some of the connections between climate simulation modules are in the same ways stimulated by theoretical considerations. But the general theoretical basis of climate science is far less crisp than that of HEP, where it consists mostly of our current QFTs (with the addition of atomic, solid state, and nuclear physics for detector simulation).

In particular, Lenhard and Winsberg (2010, p. 256) discuss general circulation models in climate science, which have been the primary source of theoretical ideas in earlier decades, and have as their “theoretical core [...] the so-called fundamental equations, a system of partial differential equations motivated by fluid mechanics and thermodynamics.” However: “Today, atmospheric [general circulation models] have lost their central place and given way to a deliberately modular architecture of coupled models that comprise a number of highly interactive sub-models, like atmosphere, oceans, or ice-cover.” (ibid.)

In a more recent assessment of climate science’s theoretical foundations, Katzav and Parker (2018) don’t even mention general circulation models anymore, but rather discern “a number of outstanding issues in the theoretical foundations of climate science”, such as “how to draw the boundaries of the climate system; whether to pursue fully reductive notions of Earth’s climate system and its states; whether climate states should be characterized statistically or in a combined physical-statistical way; which quantities [...] should be used in characterizing climate change” (ibid., pp. 7–8).

This observation makes for an important difference between both fields. Consider how CS results are sanctioned by the ATLAS group, apart from (holistic) validation. When a new piece of code is introduced, efforts will, of course, be undertaken to benchmark it against existing results (such as the predictions from the older code considered valid). In addition, however, a great deal of small- and large-scale *integration testing* will be performed, in order to ensure that the new code does not mess up the entire simulation’s ability to produce the desired results (cf. ATLAS Computing Group 2005, p. 78 ff.).

It is, in other words, well-acknowledged in the community that the knowledge-generating function associated with a given CS can only be established if the manifold connections among individual models are taken into account and everything is well calibrated also in the sense of *model coherence*, i.e., through integration testing and pre-integration assessment.

Assessments of integration are discussed in the environmental modeling-literature as well, but the challenges here appear to be very different (and arguably greater), and there also appears to be no generally accepted approach to pursuing integration tests. A review of existing approaches by Belete et al. (2017), for instance, considers a large (but non-exhaustive) list of integration frameworks (p. 51) and discerns (p. 61) “wide variation in strategies for implementing and documenting model integration across the development community”, as well as “wide variation in strategies for orchestrating execution of workflows within the frameworks, and preparing modeling components for assimilation into the frameworks.” In consequence, a number of recommendations for developing a unified approach are given in the paper, such as the inclusion

of an explicit interface for communication with other modules, or regarding dataset conversion, semantic mediation, error handling, and similar factors.

These factors appear to be pretty much under control in HEP experiments at the LHC. ATLAS, for instance, only has three testing frameworks, called AtNight (ATN), Kit Validation (KV) and Run Time Tester (RTT), with “differences in the intention as to how these frameworks are to be used” (ATLAS Computing Group 2005, p. 79):

ATN is run on the machines used to run the nightly builds. This limits the resources available for testing, and tests run under ATN are usually short. KV testing is usually run ‘by hand’ as part of kit building, and once again the tests are usually short. The RTT is currently run daily on a farm at UCL. With these resources, more extensive testing is possible. The full set of RTT results currently take a number of hours to generate. (ibid.)

Thus, in contrast to testing frameworks in climate science, the different frameworks used in HEP fulfill *complementary* roles, and together constitute a rather comprehensive means for testing the integration properties of new pieces of code.

There are also interesting differences on the *pre-implementation* level, as can be gathered from the following remark by Belete et al. (2017, p. 51):

On the science side, pre-integration assessment includes a problem statement; articulation of the purpose and goals of the science-based modeling effort; conceptualization of the relevant system (including major components and interactions); characterization of scenarios and use cases; analysis of alternatives; design of a science-based approach; and a statement of any project resource constraints and solution criteria.

These steps are all pretty much fixed in advance in HEP, in virtue of either the underlying theory (problems, use cases, science-based approach, alternatives and solution criteria) or the conditions of the experiment (goals, project resource constraints) or sometimes even both in concert (problems, use cases, science-based approach).

The bottom line is that HEP scientists seem to have a much better handle on establishing model-coherence than do climate scientists. But what (A) is the significance of that, and (B) what is this difference due to?

Regarding (A), we believe that model-coherence allows a respectable amount of (error-) attribution, which Lenhard and Winsberg (2010, p. 259) find to be a major problem for climate simulations. Testing the integration-properties of a certain piece of code—after having taken into account the connections that already exist across the underlying physics models (pre-integration assessment)—means checking whether including this code (or the model it represents) will spoil the validity established for the integrated whole. When this is the case, it is possible to attribute at least the *failure* to the new piece of code or to its connection to other pieces, and to re-assess it on the physics and implementation level.

As our discussion indicates, the solid theoretical basis in HEP often allows for the kind of “fundamental analysis of how the unexpected behavior occurred” Lenhard (2019, p. 955) finds to be sparse in software development. But it certainly also sometimes results in the “adapting [of] interfaces so that the joint model will work as anticipated in the given circumstances” which he finds to be far more common (ibid.).

Regarding (B), the *theoretical* differences between HEP and climate science were already outlined, and the far better understanding of relevant theory in HEP contributes to the possibility of re-assessment (and improvement) upon attribution of failure. However, this is clearly only part of the story, for why are integration frameworks in HEP complementary, whereas they appear to largely constitute rival approaches in climate science? The reason, we suspect, lies in the nature of the *subject matter*, and the related fact that HEP can be considered an *experimental* discipline, whereas climate science relies almost exclusively on *field observations*.

We do not have an elaborate account of experimentation in mind here, but we do maintain, with, e.g., Schurz (2014, p. 35) or Radder (2009, p. 3), that experiments crucially involve an element of *control* that is absent in field observations. This control is evident at the LHC, for instance, in the *preparation* of protons in bunches accelerated to a specified center of mass energy, and their colliding at pre-defined angles and in pre-defined interaction points inside the different detectors.²¹

Why does this difference between both fields exist? We believe that it directly relates to the nature of the subject matter and the different origins of (messiness-inducing) *complexity*. In HEP, the complexity is entirely conditioned on the *remoteness of the target phenomena*: quarks, gluons, or Higgs bosons are not directly detectable, in the same sense as electrons, photons, or neutrons. Moreover, they decay (or hadronize) so quickly that even, say, a hypothetical ‘quark detector’ would have a hard time catching free quarks (which is to say that this is pretty much inconceivable). The entire complexity of the simulation is conditioned on this and other (theoretically established) facts, such as the coloredness of gluons, sparking the parton shower, or even the (expected) rarity of relevant events, which requires detectors to be flexible enough for recognizing diverse and subtle signatures attached to the complementary decay channels for relevant particles.

This is very different from climate science where the target phenomena (local temperature, cloud patterns, oceanic streams, rainfall, and so forth) are, by and large, directly observable. But it is here totally out of the question to devise a controlled laboratory experiment that captures them all. Hence, in contrast to HEP, complexity is *inherent* in the subject matter of climate science.

As it turns out, the higher degree of control over observational conditions and the better theoretical understanding pave the way to better model coherence in HEP, as an additional means for justification. The theoretical understanding makes pre-integration assessment easier (maybe even largely superfluous), and experimental *design* can be such that it includes such a division of labor from the outset in which there are dedicated collaborations that assess the software-properties such as small- and large-scale integration according to highly organized schemes or even pre-defined strategies.

A *socio-historical* point surfaces in this context: whereas climate science had a continuous, integrated development as an entire *field*, in which some of the entrenchment is ultimately rooted (Lenhard and Küppers 2006; Lenhard and Winsberg 2010, 2011), experimental collaborations in HEP, in contrast, each have their own, somewhat secluded histories and social development. This allows HEP researchers to build on

²¹ E.g. <https://home.cern/news/news/accelerators/lhc-report-colliding-angle>; checked 02/20.

successes and failures of previous experiments, at other colliders, *from the very design stages on*.

5 Conclusions: lessons for the epistemology of simulation

In this paper, we have offered a detailed analysis of the simulation modeling that goes on in large-scale LHC experiments, and established three central results on the basis of that: (a) that modified hierarchical accounts are still valid and useful as an idealized model of initial working steps, (b) that a whole network of simulation models needs to be embraced for establishing justified claims to simulation-based knowledge, and (c) that for historical, empirical, and theoretical reasons, high energy physicists are in a far better position for establishing model coherence and error-attribution than are climate scientists.

Note that our point in (b) is not that the modeling steps necessary for establishing justifiable knowledge claims *do not in fact* order in a hierarchical fashion; we conceded that to be a quite generic feature of simulation modeling already in Sect. 2.2. We think that our analysis, similarly to that of Lenhard and Winsberg (2010, 2011), shows that, when everything is put together, they *cannot* be so ordered in contexts like HEP or climate science, not even in an ideal limit.

As pointed out in Sect. 4.3, the assumption of FACTORIZATION is an important heuristic, simplifying starting point for getting the modeling process going. At this point, modified hierarchical reconstructions of the modeling are still possible, as we have demonstrated in detail. But when attempting to find an overall model of the transition from colliding protons to measured stable particles that can be connected to empirical evidence and live up to prior theoretical knowledge, physicists need to embrace a back and forth between different models that entirely spoils the possibility of a linear working process. Since we have focused only on a number of exemplary such relations, one obtains, in the relevant limit, an *even more connected* network rather than a hierarchy.

We still consider (a) to be important, though, because hierarchical accounts of simulation modeling tend to reoccur throughout the literature, and it seems far fetched to claim that they are all in error. Rather, one needs to be conscious of what a hierarchical analysis can and cannot deliver, and it (repeatedly) turns out to be the wrong endpoint for establishing elements of a general EoS.

Maybe the most salient results are those that fall under (c), the differences between HEP and climate science. As we noted, the reasons for entrenchment are in part theoretical and not historical in HEP, and the theoretical basis is far better understood and less disputable. Moreover, the fact that the occurrence of complexity is conditioned on the remoteness of the phenomena in HEP, and not so much inherent in the subject matter, makes an important difference as to the possibility of establishing coherence.

As we saw, much of the pre-integration assessment is fixed from the outset in HEP, and so is the general framework for integration testing. This, in turn, is an expression of the fact that HEP is (by most standards) a properly so called *experimental* discipline, and that the conditions of long-term and large-scale experiments like ATLAS, regarding such things as division of labor and general approaches to securing methods

of analysis, are strategically planned and pre-conceived from the outset. Moreover, the possibility of validating at least individual components of detector models in isolation is a consequence of the experimental character of the discipline as well, for it is conditioned on the ability to reproduce relevant particles and stand-ins for detector materials in isolation and under controlled laboratory conditions.

Clearly, all of this is quite different from a discipline like climate science, which is almost entirely reliant on field observations and where both theory and techniques for observation have a strong component of historical development, including political influences (Lenhard and Küppers 2006).

Bottom line: Messiness and a kludgy methodology may be unavoidable for simulation whenever complexity comes into play. But depending on the conditions for that messiness (remoteness-induced vs. intrinsic) as well as the theoretical (well-understood vs. vague core) and empirical conditions (experiment vs. field observations) under which the resulting CSs are used, one may get a far better handle on these aspects, especially in terms of coherence and the (related) possibility of error-attribution.

Acknowledgements We thank Dominic Hirschi for help with some details concerning recent developments in ATLAS simulation, Robert Harlander for checking the theoretical physics, Michael Krämer for helpful comments at an earlier stage, and four anonymous referees for helpful additional comments. FJB also thanks Klaus Hamacher, Jan Heisig, Alexander Mück, Björn Sarrazin, and Christian Schwinn for discussions on the underlying physics.

Funding Open Access funding enabled and organized by Projekt DEAL. The research for this paper was conducted as part of the research unit *The Epistemology of the Large Hadron Collider* (German Research Foundation (DFG); grant FOR 2063).

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

- Amati, D., & Veneziano, G. (1979). Preconfinement as a property of perturbative QCD. *Physics Letters B*, 83(1), 87–92.
- Assamagan, K. A., Baranov, S., Boudreau, J., Tsulaia, V., Nairz, A., Lelchuk, I. T. M., et al. (2005). The description of the ATLAS detector. In A. Aïme, J. Harvey, & N. Knoors (Eds.), *CHEP'04: Proceedings of computing in high energy and nuclear physics* (Vol. I, pp. 321–324). Geneva: CERN.
- ATLAS Collaboration. (2008). The atlas experiment at the cern large hadron collider. *Jinst*, 3, S08003.
- ATLAS Collaboration. (2010). The ATLAS simulation infrastructure. *The European Physical Journal C*, 70(3), 823–874.
- ATLAS Computing Group (2005). Computing technical design report. Technical report, CERN. <https://cds.cern.ch/record/837738/files/lhcc-2005-022.pdf> (checked 11/18).
- Baer, H., & et al. (2004). Les Houches guidebook to Monte Carlo generators for hadron collider physics. *arXiv preprint arXiv:hep-ph/0403045*.

- Bähr, M., Gieseke, S., Gigg, M. A., Grellscheid, D., Hamilton, K., Latunde-Dada, O., et al. (2008). Herwig++ physics and manual. *The European Physical Journal C-Particles and Fields*, 58(4), 639–707.
- Banerjee, S., Folger, G., Ivanchenko, A., Ivanchenko, V. N., Kossow, M., Quesada, J. M., et al. (2011). Validation of geant4 hadronic generators versus thin target data. *Journal of Physics: Conference Series*, 331(3), 032034.
- Barr, G., Devenish, R., Walczak, R., & Weidberg, T. (2016). *Particle physics in the LHC era*. Oxford: OUP.
- Behnke, O., & Moneta, L. (2013). Parameter estimation. In O. Behnke, K. Kröninger, G. Schott, & T. Schörner-Sadenius (Eds.), *Data analysis in high energy physics: A practical guide to statistical methods* (pp. 27–73). New York: Wiley.
- Beisbart, C. (2012). How can computer simulations produce new knowledge? *European Journal for Philosophy of Science*, 2(3), 395–434.
- Beisbart, C. (2014). Are we Sims? How computer simulations represent and what this means for the simulation argument. *The Monist*, 97(3), 399–417.
- Beisbart, C., & Saam, N. (2019). *Computer simulation validation: Fundamental concepts, methodological frameworks, and philosophical perspectives.*, Simulation foundations, methods and applications Berlin: Springer.
- Belete, G. F., Voinov, A., & Laniak, G. F. (2017). An overview of the model integration process: From pre-integration assessment to testing. *Environmental Modelling & Software*, 87, 49–63.
- Boge, F. J. (2019). Why computer simulations are not inferences, and in what sense they are experiments. *European Journal for Philosophy of Science.*, <https://doi.org/10.1007/s13194-018-0239-z>.
- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, 97, 303–352.
- Bokulich, A. (2011). How scientific models can explain. *Synthese*, 180(1), 33–45.
- Brun, R., Urban, L., Carminati, F., Giani, S., Maire, M., McPherson, A., & et al. (1993). *Geant: Detector description and simulation tool*. Technical report, CERN.
- Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Clarendon Press.
- Clark, A. (1987). The kludge in the machine. *Mind & Language*, 2(4), 277–300.
- Collins, J. C., Soper, D. E., & Serman, G. (1989). Factorization of hard processes in QCD. In A. H. Mueller (Ed.), *Perturbative QCD* (pp. 1–91). Singapore: World Scientific Publishing Co.
- Dissertori, G., Knowles, I., & Schmelling, M. (2003). *Quantum chromodynamics: High energy experiments and theory*. Oxford: Clarendon Press.
- Ellis, R., Stirling, W., & Webber, B. (2003). *QCD and collider physics*. Cambridge: Cambridge University Press.
- GEANT Collaboration. (2016). Physics reference manual. *GEANT 4*, Release 10.4.
- Gieseke, S., & Nagy, Z. (2011). Monte Carlo generators and fixed-order calculations: Predicting the (un)expected. In I. C. Brock & T. Schörner-Sadenius (Eds.), *Physics at the terascale* (pp. 97–121). New York: Wiley.
- Griffiths, D. (2008). *Introduction to elementary particles* (2nd ed.). New York: Wiley.
- Harris, T. (1999). A hierarchy of models and electron microscopy. In L. Magnani, N. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 139–148). New York: Kluwer Academic/Plenum Publishers.
- Hartmann, S. (1996). The world as a process. In R. Hegelsmann, U. Mueller, & K. G. Troitzsch (Eds.), *Modelling and simulation in the social sciences from the philosophy of science point of view* (pp. 77–100). Berlin: Springer.
- Hartmann, S. (1999). Models and stories in hadron physics. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators* (pp. 326–346). Cambridge: Cambridge University Press.
- Höche, S. (2016). Introduction to parton-shower event generators. In L. Dixon & F. Petriello (Eds.), *Journeys through the precision frontier: Amplitudes for colliders. Proceedings of the 2014 theoretical advanced study institute in elementary particle physics*. New Jersey, London: World Scientific.
- Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford: Oxford University Press.
- James, F. (1980). Monte Carlo theory and practice. *Reports on Progress in Physics*, 43(9), 1145.
- Karaca, K. (2013). The strong and weak senses of theory-ladenness of experimentation: Theory-driven versus exploratory experiments in the history of high-energy particle physics. *Science in Context*, 26(01), 93–136.
- Karaca, K. (2018a). Lessons from the Large Hadron Collider for model-based experimentation: The concept of a model of data acquisition and the scope of the hierarchy of models. *Synthese*, 195(12), 5431–5452.

- Karaca, K. (2018b). The interplay among experiment, simulation and theory at the Large Hadron collider: A network account of models in high energy physics experiments. Unpublished manuscript, presented at *EPSA17 Exeter*.
- Katzav, J., & Parker, W. S. (2018). Issues in the theoretical foundations of climate science. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 63, 141–149.
- Lenhard, J. (2016). Computer simulation. In P. Humphreys (Ed.), *The Oxford handbook of philosophy of science* (pp. 717–737). Oxford: Oxford University Press.
- Lenhard, J. (2019). How does holism challenge the validation of computer simulation? In C. Beisbart & N. J. Saam (Eds.), *Computer simulation validation* (pp. 943–960). Berlin: Springer.
- Lenhard, J., & Küppers, G. (2006). From hierarchical to network-like integration: A revolution of modeling style in computer-simulation. In J. Lenhard, G. Küppers, & T. Shinn (Eds.), *Simulation: Pragmatic construction of reality* (pp. 89–106). Berlin: Springer.
- Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 41(3), 253–262.
- Lenhard, J., & Winsberg, E. (2011). Holism and entrenchment in climate model validation. In M. Carrier & A. Nordmann (Eds.), *Science in the context of application* (pp. 115–130). Berlin: Springer.
- Mangano, M. L., & Stelzer, T. J. (2005). Tools for the simulation of hard hadronic collisions. *Annual Review of Nuclear and Particle Science*, 55, 555–588.
- 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.
- Mättig, P. (2019). Validation of particle physics simulation. In C. Beisbart & N. J. Saam (Eds.), *Computer simulation validation* (pp. 631–660). Berlin: Springer.
- Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.
- Morrison, M. (2009). Models, measurement and computer simulation: The changing face of experimentation. *Philosophical Studies*, 143(1), 33–57.
- Morrison, M. (2015). *Reconstructing reality: Models, mathematics, and simulations*. Oxford: Oxford University Press.
- Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators* (pp. 10–37). Cambridge: Cambridge University Press.
- Peskin, M. E., & Schroeder, D. V. (1995). *An introduction to quantum field theory*. Reading, MA: Perseus Books.
- Plehn, T. (2014). *Lectures on LHC physics*. Berlin: Springer.
- Quadt, A. (2007). *Top quark physics at hadron colliders*. Berlin: Springer.
- Radder, H. (2009). The philosophy of scientific experimentation: A review. *Automated Experimentation*, 1(1), 2.
- Schurz, G. (2014). *Philosophy of science. A unified approach*. New York, London: Routledge.
- Schwartz, M. (2014). *Quantum field theory and the standard model*. Cambridge: Cambridge University Press.
- Seymour, M. H. & Marx, M. (2015). Monte Carlo event generators. In *LHC Phenomenology* (pp. 287–319). Springer.
- Shapere, D. (1982). The concept of observation in science and philosophy. *Philosophy of Science*, 49(4), 485–525.
- Suárez, M., & Cartwright, N. (2008). Theories: Tools versus models. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 39(1), 62–81.
- Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science: Proceedings of the 1960 international congress* (pp. 252–261). Stanford: Stanford University Press.
- Thijssen, J. (2007). *Computational physics*. Cambridge: Cambridge University Press.
- Tsai, Y.-S. (1974). Pair production and Bremsstrahlung of charged leptons. *Reviews of Modern Physics*, 46(4), 815–851.
- van Fraassen, B. C. (1980). *The scientific image*. Oxford: Clarendon Press.
- Wilson, K. G. (1974). Confinement of quarks. *Physical Review D*, 10(8), 2445.
- Winsberg, E. (1999). The hierarchy of models in simulation. In L. Magnani, N. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 255–269). New York: Kluwer Academic/Plenum Publishers.

- Winsberg, E. (2001). Simulations, models, and theories: Complex physical systems and their representations. *Philosophy of Science*, 68(S3), S442–S454.
- Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: University of Chicago Press.

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