Dynamical symmetries and model validation

Benjamin C. Jantzen

Abstract

I introduce a new method for validating models – including stochastic models – that gets at the reliability of a model's predictions under intervention or manipulation of its inputs and not merely at its predictive reliability under passive observation. The method is derived from philosophical work on natural kinds, and turns on comparing the *dynamical symmetries* of a model with those of its target, where dynamical symmetries are interventions on model variables that commute with time evolution. I demonstrate that this method succeeds in testing aspects of model validity for which few other tools exist.

1 Introduction

Scientists of all stripes are in the business of building models as tools for predicting, controlling, and explaining phenomena. For each of these purposes, it is generally not sufficient that a model merely "save the phenomena." That is, it's not enough that a model successfully summarize the data already in hand. Rather, the model builder wants some sort of assurance that the model accurately represents the world, at least with respect to those features pertinent to her epistemic goals. The most common approaches to establishing such a warrant of reliability focus on comparing features of the solutions or predictions of a model and states (or time-series of states) of the world. For example, one would typically validate a regression model that predicts lifetime earnings on the basis of socioeconomic factors like education by comparing the model's predictions against a collection of fresh data not used in its construction. The more such predictions match, the more confident we are in the model, at least as a tool for prediction.

While the bulk of statistical tools are designed with such comparisons in mind, there is often value in comparing relations amongst accessible states (or relations amongst time-series of states) rather than states themselves. That is, it can be helpful to ask whether the *change* in a model's prediction given a *change* in input or initial conditions matches the change in the target system given a corresponding change in it's initial conditions or external inputs. This sort of relation is exactly what one needs to know if a model is to be used for control. That is, if one wants a model to reliably reflect causal relations amongst variables – as opposed to offering purely correlative predictions – it is essential to verify that its gets these relations right.

In this paper, I introduce a new approach to validating dynamical models – including stochastic models – using 'dynamical symmetries'. This method is focused not on static features of states or time-series, but rather on relations amongst such things under changes in input or initial conditions. This is a tool for checking the causal information explicitly or implicitly contained in a model, and is therefore useful for validating models for control as well as for prediction. My narrow aim is to argue that this method is, in many circumstances, an

1

unusually powerful tool for model validation that gets at causal structure in a way most statistical methods do not. More broadly, I want to suggest that the success of this method is evidence of the practical, methodological relevance of philosophical work on natural kinds. Consequently, the technical results presented here amount to a sort of advertisement for the mutual benefits philosophers, applied mathematicians, and data analysts can offer one another.

To meet these aims, the rest of the essay is laid out as follows. In Section 2, I clarify the problem of model validation, and present a concise but somewhat more detailed overview of standard methods that focus on comparing static properties of model solutions or predictions with single measurements of the target system. This is then contrasted with what I call "structural approaches" that consider relations amongst model predictions. I summarize a variety of methods that are plausibly viewed as structural. In Sections 3 and 4, I introduce the theory of dynamical symmetries, and present previously published methods for comparing them for different systems given empirical data. In Section 5, I outline the way in which comparison of dynamical symmetries can be used as a powerful tool for model validation, and illustrate the method in a variety of contexts with concrete examples. Finally, I conclude with a discussion of the scope and limitations of this new method.

2 The problem of validation

2.1 Verification and validation

Any model – if it is to be useful for predicting or controlling its target system – needs both verification and validation. Verification is the process of assessing whether a given model possesses the intended properties. That is, does the actual instrument or mechanism for generating predictions instantiate that which was intended; do the outputs instantiate the intended mapping from inputs? Verification is not a significant concern in the case of analytic models, since there is generally little doubt that a set of equations is in fact the set intended. It becomes critical, however, when numerical approximations are used in extracting predictions or solutions from the equations, and even more pressing in the case of complex computational models. It is not at all obvious that a program correctly implements the numerical integration of a set of equations, or represents the intended set of functional relationships between variables. It is even less clear whether a multi-physics or agent-based model captures the intended set of approximations of law-like interactions amongst constituents. Verification poses a fascinating collection of epistemic problems, and there exist large literatures on model verification in engineering, software development, and mathematics.[1] However, for the purposes of this essay, I'll set aside the problem of verification, and assume that models are correctly implemented.

Validation concerns the accuracy of the model in representing the intended aspects of the target. There are three principal senses in which a model can be accurate. First, it can more or less successfully describe the target system. That is, it can reproduce the known data with varying degrees of fidelity. Second, it can more or less accurately predict passively observed features of the target system at later times, or for different boundary conditions or inputs.

---

[1] For an influential engineering perspective, see (Balci 1994). For a recent and comprehensive overview of both software and systems modeling aspects from the National Research Council, see (Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification 2012). For a pithy and very current overview of verification in the world of software design, see (Wilcox 2018). Finally, for an accessible and illuminating discussion of the state of the art from the perspective of applied mathematics, see (Fillion 2017).

And finally, it can more or less accurately predict the behavior of the target system under interventions or alterations of the environment, boundary conditions, or input. The first two tend to be the focus of many standard statistical methods. The novel method described below concerns the third.

There are, of course, many kinds of models with distinct epistemic aims, and the suitability of a particular method of validation will depend on the kind of model under consideration. The method below is appropriate for models in which each variable varies continuously, possibly as a function of other variables in the system. Of particular interest are dynamical systems, i.e., systems that change through time. The bulk of the examples below will deal with systems for which the values of variables change continuously over time.

A brief word on terminology is also important. The term "model" tends to be used in subtly different ways in different disciplines – ecologists tend to use the term differently than machine learning practitioners, and the same group tends to use the term differently depending on context. For clarity, I will use the term as follows: a *model* is the specification of a class of mappings from input to output.[2] The mappings may be via an explicit function or set of equations (as in differential models of thermodynamic phenomena) or via simulations of varying complexity. To produce an output, a model requires two things: (1) a set of parameter values, and (2) a set of inputs. Parameters are understood to represent features of the target system that may vary from system to system but do not vary for a given system. The intrinsic growth rate of a population or the Young's modulus of a given material are examples of parameters. Sometimes these can be measured independently, but often have to be estimated from data about other properties of the system that depend upon these parameters. Inputs are a set of initial or boundary conditions that can differ across time or contexts for a given system. The temperature of a reaction vessel or the current population size are typical inputs for chemical engineering or ecological models, respectively. I will refer to a particular output for a given input and choice of parameter values as a *solution* of a model.

2.2 Static fit approaches to validation

There are a wide variety of methods for model validation that appeal to a single output of the model and one or more datasets. If the model is designed to make point estimates – e.g., of a particular property of the target system such as the ionization energy of a molecule, the vibration frequency of a nano-beam, or the biodiversity of a region – then a wealth of standard statistical hypothesis testing methods are available. In the classical mode, these involve choosing a test statistic (e.g., χ^2), computing the distribution this statistic should have under the hypothesis that the model is correct, and then deciding whether the value of the test statistic for a set of validation data is sufficiently unlikely to reject the adequacy of the model. The procedure and interpretation of results are a bit different from a Bayesian point of view, but the emphasis remains on comparing a single model output with a validation data set. See (Ling and Mahadevan 2013) for a recent overview of both approaches to validating models in the case of point estimates. What's relevant to the discussion here is that no consideration is given to the answers the model *would* give under different inputs. That is, the model typically has parameter values and boundary conditions estimated from one data set, and a single point

---

[2]Note that this terminology is at odds with machine learning, where each specific set of parameter values constitutes a model. What I'm calling a model is, in the context of machine learning, or statistical learning theory a space of hypotheses or class of models.

estimate is tested with a second dataset from the same target system in the same configuration. There is no attention paid to how the model's estimate varies with variations in properties or initial states of the target system.

Of course, point estimates are just one special class of model output. Many models explicitly represent one or more functional relations amongst variables. The simplest such model is a regression curve (a functional form fit to a single dataset), but one could include simulations and agent-based models in this category. For models like these, it is typical to construct the model using a single data set describing the target system under one set of initial conditions. For example, one could measure the growth of bacteria on a petri dish over time, and then use that data to assign values to parameters in a model that proposes an exponential functional relationship between time and population. Once the parameters of the model have been set, the predicted curves relating variables of the model are compared, either to the original dataset or to a validation dataset taken under identical conditions (often, one simply splits the original dataset into training and testing pieces).

There are a variety of approaches to comparing the curve predicted by a model with data from a target system. Commonly, simple measures of agreement, such as the Sum of Squared Errors (SSE) or the coefficient of determination (R^2) are deployed to assess how well the model captures the variation in the data. For example, Fujikawa, Kai, and Morozumi (2004) use the Mean Squared Error (MSE) – given by SSE/n where n is the number of samples – as a measure of goodness of fit for their growth model.

By themselves, these measures of fit only get at how well the model describes the target system (in one particular context). This is the first, descriptive sense of validity I mentioned above. To get a sense for how well the model is likely to generalize (how reliable it will be for prediction) we need other tools. Analysis of residuals is one such tool. More specifically, the distribution of the errors (the differences between values predicted by a model and the actual data) can tell one a lot about whether there is systematic error in the model of the sort that would impugn its ability to make accurate predictions outside the original data. Methods of residual analysis include hypothesis tests for bias (the errors all tend to be in one direction), skew (there is a trend in the errors, even though there may be no bias), and curvature (there may be no bias or skew, but the envelope of the errors exhibits curvature around the correct values) (see (Rhinehart 2016, , ch. 16) for a concise overview).

When we turn our attention to stochastic models, validation gets more complicated, at least insofar as we continue to directly compare particular model outputs with acquired data. This class of model has received less attention with respect to methods of validation, and the literature on modeling across disciplines harbors a consensus that it's difficult, particularly when data is limited. This difficulty seems to have stymied the emergence of anything that would be considered a standard method. As McCarthy and Broome (2000, p600) put it, "there are no established methods for validating stochastic population models, but useful methods are required."

Some useful methods include generalizations of those described above for the deterministic case. In the most straightforward approach, one would compare predicted distributions of the dependent variables (e.g., population size) for each value of the independent variables (e.g., time) with the observed data. But this requires many replicates of the target system so that such distributions can be estimated. That sort of data is usually not forthcoming. Typically, all one has for comparison is one or a few series of measurements, with at best a handful of measurements for each value of the independent variable. But a variety of approaches have

been proposed to overcome this difficulty. For example, Sokal and Rohlf (1994) propose a method of 'standard deviates' for assessing whether the stochastic variation predicted by a model coincides with a dataset. This is, in fact, the method used by McCarthy and Broome (2000) for a model of population viability (i.e., of the risk that a population will go extinct). The salient point is that this method and others like it still focus on features of a single output of a model.

This is not quite the case for cross-validation, a powerful tool for validating both deterministic and stochastic models. This technique iterates partitioning of the data into training and testing portions in order to estimate the error of a model on unobserved data.[3] Roughly, cross-validation provides an estimate of the generalization error of a model[4] by assessing how well a model, after being fit to a sample of data from a system, will do in predicting unseen data. Cross-validation thus does not focus on a single solution of a model, but rather the reliability of the model (and the method for setting its parameter values). Nonetheless, it is indifferent to the way in which model solutions relate to one another. Similarly, the method of "active nonlinear tests" (Miller 1998) amounts to probing the space of parameter values and inputs to assess the robustness of features of a model's output to variations in parameter values, and the model's stability and plausibility for inputs not observed. Here again, there is no attention paid to the details of how solutions relate to one another, only how robust a given solution is to variations of model features.

Why is this problematic? If the aim is prediction, it's not a problem at all. These are all effective approaches to predictive validation. But if one wants to be confident that a model which fits a given data set will get its predictions right when someone intervenes and changes the boundary conditions or inputs, more is needed.

2.3 Structural approaches to validation

Some models do aspire to capture more about a target system than is necessary to predict its behavior under passive observation. In particular, some models are intended to capture something about the structure of a target system, and the extent to which they do so has been called the "structural validity" of a model (see, e.g., Zeigler, Praehofer, and Kim 2000, ch. 2).[5] In the dynamical systems, engineering, and operations research literatures, the notion of structural validity seems to have a rather narrow and stringent sense. A model is only structurally valid if the structure of the model is isomorphic to that of the target system. As Zeigler, Praehofer, and Kim (2000, p31) put it, saying that a model is structurally valid "...means that the model not only is capable of replicating the data observed from the system, but also mimics in step-by-step, component-by-component fashion the way in which the system does its transitions."

This sort of validity can be assessed in a variety of ways.[6] What Barlas (1996) calls

---

[3]Most textbooks on machine learning include descriptions of cross-validation. An especially lucid presentation can be found in (Flach 2012, ch. 12).

[4]The estimate of the generalization error of a model is biased for cross-validation, but in the direction of over-estimating the error (see Hastie, Tibshirani, and Friedman 2009, ch. 7.10).

[5]Attention to structural validation is curiously discipline dependent. Concepts (such as those pertaining to testing "white-box" models in systems engineering) seem to have relatively little penetration in other fields such as ecology. This is probably partly due to the quantity and precision of data available in these different fields. Structural tests tend to be data-hungry or to require manipulations of the target system that are not available to, e.g., field ecologists.

[6]See (Barlas 1996) for a widely-cited review.

"structure-oriented behavior tests" include a variety of comparisons of those qualitative features of a model thought to be tied to its structure with those of the target system. For example, one can assess how well a model captures temporal patterns in the target system such as the period, phase, and amplitude of oscillatory behavior, or the presence of trends (Barlas 1989). A failure of the model to generate periodic behavior of approximately the right frequency, for instance, might suggest that a feedback in the structure of the model is incorrect. Another sort of structure-oriented behavior test involves assigning extreme values to model inputs or parameters and comparing the resulting output to the behavior of the target system under correspondingly extreme conditions.[7] Note that to conduct this sort of test with respect to parameter values, the parameters must be meaningful (and both measurable and manipulable) outside of the model.

In "direct tests", one attempts to establish the accuracy of structural components of the model (e.g., the existence and values of certain parameters) by directly testing hypotheses about these components against experiments on the target system, or even established knowledge in the relevant field. For example, one might attempt to empirically ascertain whether the form of the equations in an equation-based model match the functional form of the relations among variables in the target system (Barlas 1996).

Whether by direct or indirect approaches, structural validation in the narrow sense is often the wrong epistemic goal. Narrow-sense structural validity is frequently more than one needs to meet the epistemic aims of modelers. That is, there is a broader sense of structural validity that gets at what a model needs in order to accurately characterize the behavior of system under intervention or manipulation, and nothing more. In this broader sense, a model is structurally valid if it correctly characterizes the *change* in a behavior of a system under changes in inputs or boundary conditions. Whether the model does so in the same way the target system does is irrelevant.

Presumably, if a model is structurally valid in the narrow sense, then it is structurally valid in this broader sense as well. But the broader sense is easier to satisfy in that it doesn't matter how a model captures this information, only that it does. Consequently, tests that reject this sort of validity rule out a bigger class of potential models in one go. And yet, so far as I can tell, it is largely neglected in the model validation literature. Of course, the entire field of causal discovery is concerned with methods for *building* models that capture structure in something like this broad sense (see, e.g., Spirtes, Glymour, and Scheines 2000). But those models tend not to capture the sort of fine-grained temporal detail that engineers or dynamical systems folks are interested in. Nor do methods of causal discovery directly help us to validate existing models that use, e.g., differential equations or complex agent-based computations. When I say that scant attention is paid to broad-sense structural validity, I mean there are few if any tools in the modeling literature for validating models of arbitrary structure – especially dynamical models – with respect to counterfactual behavior. No one looks at which *changes* in model behavior follow from changing conditions or inputs, and whether this pattern of change (reflective of causal structure) matches the world. This is, however, exactly what a comparison of dynamical symmetries can do for us.

---

[7]Balci (1994) calls this "stress testing."

3 Dynamical symmetries

3.1 Theory

As I indicated above, the new approach to validation described here is focused on the structure of a model or, more specifically, on the relations among solutions of a model that are implied by its structure. The important set of relations are what I previously dubbed *dynamical symmetries* (2014). Qualitatively, a dynamical symmetry is an intervention on one more variables in a system that commutes with the incrementation of another variable in the system. More precisely, I define a dynamical symmetry as follows (Jantzen 2017):

Definition 1 (Dynamical symmetry). *Let V be a set of variables and Ω be the space of states that can be jointly realized by the variables in V. Let $\sigma : \Omega \to \Omega$ be an intervention[8] on the variables in $Int \subset V$. The transformation σ is a dynamical symmetry with respect to some index variable $X \in V - Int$ if and only if σ has the following property: for all values x_i and x_f of X and for all initial states $\omega_i \in \Omega$, the final state of the system $\tilde{\omega}_f \in \Omega$ is the same whether σ is applied when $X = x_i$ and then an intervention $\Lambda_{x_i,x_f} : \Omega \to \Omega$ on X makes it such that $X = x_f$, or the intervention on X is applied first, changing its value from x_i to x_f, and then σ is applied. This property is represented by the following commutation diagram:*

$$
\begin{array}{ccc}
\omega_i & \xrightarrow{\ \sigma\ } & \tilde{\omega}_i \\
\Lambda_{x_i,x_f} \downarrow & & \downarrow \Lambda_{x_i,x_f} \\
\omega_f & \xrightarrow{\ \sigma\ } & \tilde{\omega}_f
\end{array}
\tag{1}
$$

For example, suppose we have a pressure tank full of fluid and attached to a pump that can increase or decrease the pressure in the tank. Inside the fluid-filled pressure tank, there is a vertical rail on which is mounted a pressure gauge. Initially, this gauge is at the top of the tank where the pressure is P. If we use h to represent the depth of the gauge relative to the top of the tank and p to indicate the pressure read by the gauge, then at the outset, $h = 0$ and $p = P$. Now consider two different sequences of interventions on this system. In the first, we leave the gauge where it is, and then turn on the pump until the pressure at the gauge is $P + c$. Then we lower the gauge until it is a distance h_f below the top of the tank. At that point, it reads a pressure of $P + c + \rho g h_f$, where ρ is the density of the fluid in the tank and g is the gravitational constant (9.81 ms^{-2}). This sequence of manipulations and results is summarized in Table 1.

Now suppose that we start over with our tank in the same initial state, and reverse the order in which we manipulate the pump and the gauge. That is, suppose we first lower the gauge so that its depth relative to the top of the tank goes from 0 to h_f and then turn on the pump to increase the pressure at the gauge by an amount c. As Table 2 indicates, we end up in exactly the same final state after performing these actions. Thus, increasing the pressure at the gauge by an additive constant is a dynamical symmetry with respect to the index variable h. Note, however, that scaling pressure by a multiplicative constant (i.e., an intervention of the functional form $\sigma(P) = kP$) is *not* a dynamical symmetry. The result of

---

[8]As indicated in (Jantzen 2014), I am using the term "intervention" in its technical sense as it appears in the literature on causation. In this context, "... an intervention on X (with respect to Y) is a causal process that directly changes the value of X in such a way that, if a change in the value of Y should occur, it will occur only through the change in the value of X and not in some other way"(Woodward 2001).

Table 1: Sequence of states when pressure is adjusted by an additive constant first and then the gauge is lowered a vertical distance h_f.

p	h
P	0
$P + c$	0
$P + c + \rho g h_f$	h_f

Table 2: Sequence of states when the gauge is first lowered a vertical distance h_f and then the pressure is adjusted by an additive constant.

p	h
P	0
$P + \rho g h_f$	h_f
$P + c + \rho g h_f$	h_f

applying transformations of this sort in either order with respect to moving the gauge is shown in Tables 3 and 4. Unlike in the additive case, the bottom rows of these two tables are not the same.

Since many models of interest are models of dynamical systems in the more restrictive sense of variables that evolve through time under a fixed law, I offer the following definition of a special dynamical symmetry (Jantzen 2017):

Definition 2 (Dynamical symmetry with respect to time). *Let t be the variable representing time, and let V be a set of additional dynamical variables such that $t \notin V$ and Ω is the space of states that can be jointly realized by the variables in V. Let $\sigma : \Omega \to \Omega$ be an intervention on the variables in $Int \subseteq V$, and Λ_{t_0,t_1} the time-evolution operator that advances the state of the system from t_0 to t_1. The transformation σ is a dynamical symmetry with respect to time if and only if for all intervals Δt and initial states $\omega_i \in \Omega$, the final state of the system $\tilde{\omega}_f \in \Omega$ is the same whether σ is applied at some time t_0 and the system evolved until $t_0 + \Delta t$, or the system first allowed to evolve from t_0 to $t_0 + \Delta t$ and then σ is applied. This property is represented by the following commutation diagram:*

$$
\begin{array}{ccc}
\omega_i & \xrightarrow{\ \sigma\ } & \tilde{\omega}_i \\
\Lambda_{t_0,t_0+\Delta} \downarrow & & \downarrow \Lambda_{t_0,t_0+\Delta} \\
\omega_f & \xrightarrow{\ \sigma\ } & \tilde{\omega}_f
\end{array}
\tag{2}
$$

Table 3: Sequence of states when pressure is adjusted by a multiplicative constant first and then the gauge is lowered a vertical distance h_f.

p	h
P	0
kP	0
$kP + \rho g h_f$	h_f

Table 4: Sequence of states when the gauge is first lowered a vertical distance h_f and then the pressure is adjusted by a multiplicative constant.

p	h
P	0
$P + \rho g h_f$	h_f
$k(P + \rho g h_f)$	h_f

For example, consider a microbial population whose growth is governed by:

$$\frac{dx}{dt} = rx \left(1 - (x/k)^2\right). \tag{3}$$

Such a population exhibits a whole family of dynamical symmetries with respect to time. Specifically, if we take an initial population of x_0 and add or subtract enough microbial stock to raise the population to

$$\frac{k e^{pk^2} x}{\sqrt{k^2 - x^2 + e^{2pk^2} x^2}}, \tag{4}$$

for any real value of p, and then allow the colony to grow for an hour, we would end up with the same final population size as if we allowed the population to grow for an hour starting from x_0 and then added (or subtracted) enough to scale the result according to Equation 4 (with the same value of p).

3.2 Motivation and generalization

The dynamical symmetries of a system depend upon and thus reflect its detailed causal structure. But dynamical symmetries are just one sort of feature of the causal structure of a model, and there are indefinitely many other features of causal structure that one could deploy for structural validation. So why focus on this one? There are at least three reasons to do so. The first is theoretical relevance. The notion of a dynamical symmetry is central to a general theory of *projectible kinds* (Jantzen 2014). Projectible kinds are categories or ways of binning portions of the world that are narrow enough that the members of a category share sufficient features in common to support generalizations of the sort we tend to call laws of nature, but broad enough to encompass sufficient variety in the world to make the law useful. In (Jantzen 2014), I propose that we use symmetry structures – collections of dynamical symmetries along with an algebra describing how these dynamical symmetries interact under composition – to pick out projectible kinds. Two systems belong to the same projectible kind (what I call a *dynamical kind*) just if they exhibit all of the same dynamical symmetries, and these dynamical symmetries compose with one another in the same way. The categories picked out by dynamical kinds align well with those carved out informally by scientific practice. For example, the categories corresponding to the order of a chemical reaction are also dynamical kinds. So all reacting systems that obey a first-order reaction rate law belong to the same dynamical kind. I argue in (Jantzen 2014) that recognizing dynamical kinds as the sort of projectible kinds scientists are after offers a variety of advantages for automated scientific discovery. In particular, systems can be sorted into kinds without first learning detailed models of their dynamics. Thus, one can learn how to delineate a new scientific domain pre-theoretically. The details are beyond the scope of our present concerns, but the point is that a focus on dynamical symmetries in

model validation is not arbitrary. Rather, it is motivated by a broader program in the logic of scientific discovery.

The second reason is the specific nature of the relation of dynamical symmetries to causal structure. In addition to the bare causal skeleton of which variable is a cause of which, dynamical symmetries are sensitive to the functional form of relations amongst variables. This makes them a discriminating tool for comparing models with target systems in a manner relevant to fine-grained prediction *and* control.

The final reason for emphasizing dynamical symmetries, and perhaps the most practically salient, is the extensibility of the concept. I'll focus on one particularly important extension of the basic notion of a dynamical symmetry: stochastic systems. As indicated above, stochastic models of (presumably stochastic) target systems are difficult to validate with respect to their predictive reliability. This is because there are more dimensions to a model's output – where before we had point values or trajectories of point values over time, now we have distributions characterized by indefinitely many non-vanishing moments (e.g., mean, variance, skew, etc.). Validating such models with respect to structure is even harder. But dynamical symmetries can be generalized to the stochastic case in a way that makes their application to validation straightforward.

So how do we extend the notion of dynamical symmetry beyond the deterministic case? In (Jantzen 2017), I provide one proposal. Specifically, Definition 5 of that paper shifts the focus from values of variables to distributions over variables. However, in hindsight it's clear that Definition 5 is ambiguous in important respects. I thus offer the following refinement:

Definition 3 (Dynamical symmetry). *Let V be a set of random variables, Ω the set of states that can be jointly realized by the variables in V, and Γ the space of probability distributions over Ω. Let $\sigma : \Gamma \to \Gamma$ be an intervention on the variables in $Int \subset V$. The transformation σ is a dynamical symmetry with respect to some index variable $X \in V - Int$ if and only if σ has the following property: for all initial joint distributions $\gamma_i \in \Gamma$ and marginal probability distributions f and g, the final joint probability distribution over V, $\tilde{\gamma}_f \in \Gamma$, is the same whether σ is applied when the marginal distribution over X is given by $p_x(x) = f(x)$ and then an intervention $\Lambda_{f(x),g(x)} : \Gamma \to \Gamma$ on X makes it such that $p_x(x) = g(x)$, or the intervention on X is applied first, changing its marginal distribution from $f(x)$ to $g(x)$, and then σ is applied. This property is represented in the following commutation diagram:*

$$
\begin{array}{ccc}
\gamma_i & \xrightarrow{\ \sigma\ } & \tilde{\gamma}_i \\
\Lambda_{f(x),g(x)} \downarrow & & \downarrow \Lambda_{f(x),g(x)} \\
\gamma_f & \xrightarrow{\ \sigma\ } & \tilde{\gamma}_f
\end{array}
\tag{5}
$$

Note that this definition captures the deterministic dynamical symmetries as a special case (at least insofar as one is willing to entertain degenerate probability distributions). As we'll see below, this more general notion of dynamical symmetry is useful because it allows us to check the causal structure of a model against that of a target system, even when the underlying dynamics is fundamentally stochastic.

4   Comparing dynamical symmetries

The dynamical symmetries of two systems can be directly compared without first learning a detailed model of how the variables of either system interact. The first published algorithm to implement such a test appears in (Jantzen 2017). To use the algorithm one must, of course, first obtain data about the dynamical symmetries to be compared. The most direct way to do so is to acquire two time series for System A (and two more for System B) starting at two different initial values. The initial values, let's call them x_0 and $\tilde{x}_0$, must be the same for A and B, though of course the rest of the time series may differ between them.[9] It is a consequence of the definition of a dynamical symmetry that, for systems that are deterministic, the function which maps the points of one time series to the points of the other time series corresponding to the same time is a dynamical symmetry. Furthermore, any two symmetry functions of a given system that agree on the initial values (any symmetry functions that map x_0 to $\tilde{x}_0$) must agree for the rest of the time-series.

The algorithm I reported in (Jantzen 2017) compares the dynamical symmetries exhibited by System A and System B using such pairs of time series. In broad strokes, the algorithm involves nested cross-validations. Cross-validation in general involves dividing the available data into training and testing portions. In 10-fold cross-validation, one partitions the data into 10 segments, nine of which are used for training and one of which is set aside for testing. With the training data, a particular solution of the model is fit. Then the fit model is used to predict the testing data, and the squared errors of these predictions are saved. Then the process is repeated using a different element of the partition as the testing data and the remaining 9 elements for training. After each of the 10 data segments has been used once as the testing data, the mean of the accumulated squared errors (MSE) is used as an estimate of the error of the model (or really, of the model plus the method used for fitting a solution).

In my algorithm, the outer cross-validation loop estimates the errors for two different models trained on data reflecting the dynamical symmetries exhibited by two systems of interest (call them A and B). The first model – called *sep* for "separate" – assumes that the data represent two different symmetries. That is, *sep* fits the data from A and B with two different and independent sets of parameters. The other model – called *joint* – assumes that the data from systems A and B derive from the very same dynamical symmetry, and fits a solution involving only a single set of parameters. The inner cross-validation loop is used for fitting polynomial models to the training data. Specifically, cross-validation is used to choose the order of the polynomial that should be fit to the data. Higher orders can fit a training set better but generalize poorly (in statistical parlance, they 'overfit' the data), and lower orders ignore salient variations (they are overly 'biased'). When the outer cross-validation is complete, the algorithm declares the symmetries to be different just if the MSE of the *joint* model is significantly larger than the *sep* model. That is, the dynamical symmetries are judged to be different if cross-validation estimates a higher error when the data are treated as coming from a single function than when they are treated as separate.

This algorithm was originally developed to compare symmetries of two physical systems. In the next section, I demonstrate how it can also be used to structurally validate a model by comparing the symmetries of the model with those of the target system.

---

[9]In principle, one could take a single long time series for each system and cut it in half to obtain two such curves, but for ease of exposition, I assume the time series are obtained separately.

5   Dynamical kinds and model validation

5.1   Growth models

To provide a concrete sense for how dynamical symmetries can contribute to model validation, I present three case studies in this section. In each case, the target system involves biological growth of a single species. More specifically, the models I'll consider are aimed at predicting population size – of mammals or microbes – as a function of time for a given environment. Perhaps the most influential model of this sort was published in the early 19th century by P.F. Verhulst (1838).[10] In Verhulst's "logistic model", the instantaneous rate of population growth is proportional to a quadratic function of the current population:

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K} \right).\tag{6}$$

The parameter r is generally interpreted as representing fecundity (or intrinsic growth rate) and K is viewed as the carrying capacity (the maximum sustainable population). The solutions of this equation are curves with a familiar sigmoid shape – they begin with a nearly exponential phase, pass through an inflection point, and level off in an asymptote to the carrying capacity. It is important to note that, although r and K can be given a biological interpretation, they are in general not directly measurable, and must be estimated by fitting one of these sigmoidal curves to the data.

Verhulst's original model has spawned a menagerie of generalized, extended, or otherwise modified logistic models. The bulk of these can be gathered under a single class of models that Tsoularis and Wallace (2002) call "generalized logistic" functions.[11] These have the form,

$$\frac{dx}{dt} = rx^{\alpha} \left(1 - \left(\frac{x}{K} \right)^{\beta} \right)^{\gamma},\tag{7}$$

where the additional parameters α, β, and γ have no obvious biological interpretation. For parameter values not too far from 1 (e.g., $\alpha = \beta = 1$; $\gamma = 2$), the solutions of generalized logistic models are only subtly different in shape from the original Verhulst model, at least when one is free to choose values of r and K (see (Tsoularis and Wallace 2002) for a thorough review). This fact – coupled with the fact that r and K cannot be independently measured or estimated – leads to a profound underdetermination and a persistent problem for model validation. Which model is the right model of population growth for a given species in a given context? Lest the reader get the impression that this question is merely academic and this example merely a "toy", note that papers continue to be published in biological and industrial process journals addressing this question (Buchanan, Whiting, and Damert 1997; Fujikawa, Kai, and Morozumi 2004; Zwietering et al. 1990). Researchers actually want to know the answer so that they can not only predict but control and optimize the growth of, e.g., microbial stock species or virulent microbial contaminants. In the case studies that follow, I demonstrate how methods of assessing the sameness of dynamical symmetries can aid in model selection in the context of bacterial growth.

---

[10] For an English translation of the French, see (Vogels et al. 1975).

[11] Another equally old and venerable model is that of Gompertz (1825). This model also continues to be deployed for growth modeling.

5.2 Example: Deterministic generalized logistic models

In the first case, consider a simulated population whose actual growth is deterministic and dictated by a generalized logistic equation with $\alpha = 1$, $\beta = 3/2$, $\gamma = 2$. We can use this simulated population to generate data for which we know the ground truth. In Figure 1 (a), you can see two samples from this system – for two different starting populations – where Gaussian noise of standard deviation 0.3 has been added in order to accurately reflect the noise inherent in measurement.

Now imagine yourself as a researcher interested in learning the "right" model of population growth. For one reason or another, you've decided to consider two models: the Verhulst logistic equation (Equation 6) and a generalized logistic (Equation 7) for which $\alpha = 1$, $\beta = 2$, $\gamma = 1$. I'll call the latter the $\beta = 2$ model. Of course, neither of those reflects the true dynamics, but the scientist never gets to know this a priori (that would make inductive inference rather trivial). The point here is to examine what can be learned by different inferential methods in a realistic, relatively simple case where we happen to know the ground truth and can thus assess the performance of each method.

While there are myriad ways to fit and validate models of either sort, we'll follow a particularly simple procedure that exhibits the core features of most common statistical methods. In particular, we'll work with parameterized analytic solutions to the above differential equations. Specifically, Equation 6 has solutions of the form,

$$x(t) = \frac{K}{1 + \left(\frac{K}{x_0} - 1\right) e^{-rt}}, \tag{8}$$

while the $\beta = 2$ model has solutions of the form,

$$x(t) = \frac{K}{\left(1 + \left((\frac{K}{x_0})^2 - 1\right) e^{-2rt}\right)^{1/2}}. \tag{9}$$

We'll use one sample from our target system to fit parameters for each model. That is, we'll use one set of measurements to determine r and K using nonlinear least squares regression. We'll then use those fit parameter values to try and predict the data in the second set of measurements.[12] The sum of the squared errors (SSE) for the predictions made by each model can be used as a simple measure of goodness fit.

The results of carrying out this procedure are shown graphically in Figure 1 (a). The best fit of the Verhulst model (fit to the data of the lower curve) is depicted with a solid red line, and the best fit of the $\beta = 2$ model is shown with a dashed green line. Visually, it's clear that both models can be used to fit the initial curve very well. The SSE for the fit Verhulst logistic is 22.1, and 128 for the alternative model. The Verhulst has an advantage, but both do a decent job of at least summarizing the data. However, when we use the parameters from the first fit to predict the second data set, the $\beta = 2$ model clearly falls apart. The sum of squared errors are 95.9 and 310 for the Verhulst and $\beta = 2$ models, respectively. Note that the logistic is not merely better than the $\beta = 2$ model, but it does a compelling job of predicting the data. On the basis of this information – exactly the sort of information standard methods

---

[12]Note that the initial value of the population, x_0 is fit independently in each case. That's because, while the other parameters are presumed to be intrinsic features of the growing population, the initial population size is variable and assumed to have different (unknown) values in each case.
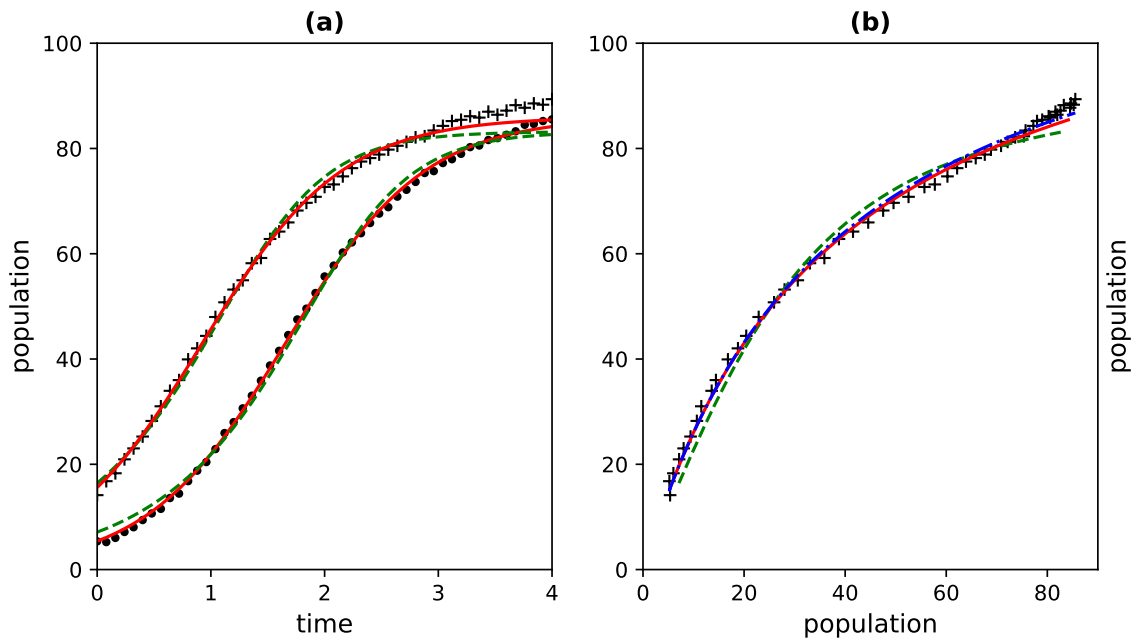
Figure 1: (a) Noisily sampled measurements for a simulated system governed by a generalized logistic equation starting from $x_0 = 5$ (black dots) and $x_0 = 15$ (crosses). The best fit Verhulst model is depicted with a solid red line and the best fit $\beta = 2$ model with a dashed green line. (b) The empirical symmetry function computed from the two trajectories in (a) is shown with black crosses. The theoretical dynamical symmetries implied by the Verhulst and $\beta = 2$ models are shown with solid red and dashed green lines, respectively.

$$
\begin{array}{c}
\begin{array}{cc} t & x \end{array} \\
\begin{bmatrix}
t_0 & x_0 \\
t_1 & x_1 \\
t_2 & x_2 \\
\vdots & \vdots
\end{bmatrix}
\end{array}
\quad
\begin{array}{c}
\begin{array}{cc} t & \tilde{x} \end{array} \\
\begin{bmatrix}
t_0 & \tilde{x}_0 \\
t_1 & \tilde{x}_1 \\
t_2 & \tilde{x}_2 \\
\vdots & \vdots
\end{bmatrix}
\end{array}
$$

$$\Downarrow$$

$$
\begin{array}{c}
\begin{array}{cc} x & \tilde{x} \end{array} \\
\begin{bmatrix}
x_0 & \tilde{x}_0 \\
x_1 & \tilde{x}_1 \\
x_2 & \tilde{x}_2 \\
\vdots & \vdots
\end{bmatrix}
\end{array}
$$

Figure 2: Schematic showing how samples from two time-series are restructured to obtain an implicit model of the dynamical symmetry that maps one trajectory into the other. (Adapted from Figure 1 in (Jantzen 2017).)

provide the scientist trying to infer a model of growth – you might reasonably be inclined to conclude the Verhulst model is not just the best of the available options, but also a fairly reliable representation of the structure of the growth dynamics.[13]

But this would be a mistake. We know that the Verhulst model is wrong in this case, and that it will systematically lead us astray for growing populations not yet observed. Here is where attending to dynamical symmetries can help. To extract information about one of the dynamical symmetries (with respect to time) of a system from two trajectories of that system, one can simply build a new curve by matching each value of the variable of interest (x, or population size, in this case) in one trajectory with its contemporaneous value in the other. This operation is shown schematically in Figure 2. The resulting empirical curve ($\tilde{x} = \sigma(x)$) is shown by the black crosses in Figure 1(b).

To use this information about the dynamical symmetries of our unknown growth system, we need to compare the symmetry function predicted by each of the models we have already fit to the data. These predicted symmetries, computed numerically for the models in precisely the same way as for the experimental data, are shown as solid red and dashed green lines in Figure 1 (b). The comparison algorithm discussed above in Section 4 judges both theoretical symmetries to be significantly different from the empirical symmetry, and thus rejects the hypothesis that either of them accurately describes the structure of the target system. In other words, neither the Verhulst model (using the best-fit parameter values), nor the $\beta = 2$ model (again, using the best-fit parameters) accurately represent the target system. We have learned that they are both wrong.

But an even stronger result can be established. In general, one can consider the entire space of dynamical symmetries implied by a model, and ask whether whether there exist *any* parameter values that could account for the observed symmetry function, regardless of how

---

[13]This is the line of reasoning presented in (Zwietering et al. 1990), where the Gompertz model is favored.

well the associated solutions describe individual trajectories. In this case, it is possible to solve analytically for the set of all dynamical symmetries for each of the two classes.[14] For the Verhulst logistic model, the dynamical symmetries are given by

$$\sigma_p(x) = Kx / \left((1 - e^{-p})x + e^{-p}K \right), \tag{10}$$

where each real value of p corresponds to a distinct symmetry transformation. Using this analytic form, it's possible to search for a set of parameter values (including p) that best fit the empirical symmetry directly. The optimal fit can then be compared, via the comparison algorithm described above, with the empirical symmetry. Doing so in this case leads to a *rejection*. In other words, we can with confidence reject the claim that *any* parameterization of the Verhulst logistic model accurately represents the dynamics of the target system.

5.3   Example: Real populations

The procedure for checking the symmetries of a theoretical model against the empirical, measured symmetries of a dynamical system was demonstrated in the previous section for artificial data – data for which we know the ground truth about the governing dynamics. I crafted the artificial data to be as faithful to the messiness of real-world data as possible, but there is always a concern that a method will break down when confronted with real data. So let's take a look at an example of actual biological (or microbiological) growth.[15] Figure 3 shows two segments of data from a growth experiment. The experiment was designed to answer a question about the fitness of bacterial strains in a variety of environments. As such, it involved many populations of three bacterial strains, each tested in 3 distinct environments. But for our purposes, I have selected time-series measurements indicating the size of just one of these populations of bacteria growing on a microtiter plate.

In the interests of full disclosure, this particular population was not selected at random from the available datasets. Rather, I focused on this particular population because it was the one with a growth curve most plausibly described by one of the models considered above. In other words, it was chosen to maximize the difficulty of rejecting a logistic or $\beta = 2$ model. Other curves were clearly poorly fit by such models, and one would not have been inclined to try. It's also important to note that though two curves are shown in Figure 3 (a), there was really only a single measured time series. What I've done is to split the time series in half, and translate the time-values of the second half so that it begins at $t = 0$. The validity of such a procedure rests on the assumption that the dynamics is autonomous. When such an assumption is warranted, it means that a dynamical symmetry can be directly estimated from purely observational data, without any interventions.

With the pair of sampled time-series curves, we can proceed as before and use them to estimate a symmetry of the growing population. Figure 3 (a) shows the best-fit models as solid red and green lines for the Verhulst and $\beta = 2$ models, respectively. The fit models are nearly indistinguishable in terms of their SSE values, and it's clear from visual inspection that neither provides an exact fit. In fact, given the obvious curvature in the residuals, both models

---

[14]It's generally possible to determine and fit symmetries numerically, without an analytic, closed form solution. But since one is available in this case, I use it to simplify the analysis.

[15]This data was obtained from Connelly (*Data Set for 'Analyzing Microbial Growth with R'*) and is used here with permission (and gratitude). The dataset can be found at https://zenodo.org/record/1171129. I am specifically considering the sixteenth row of the table.
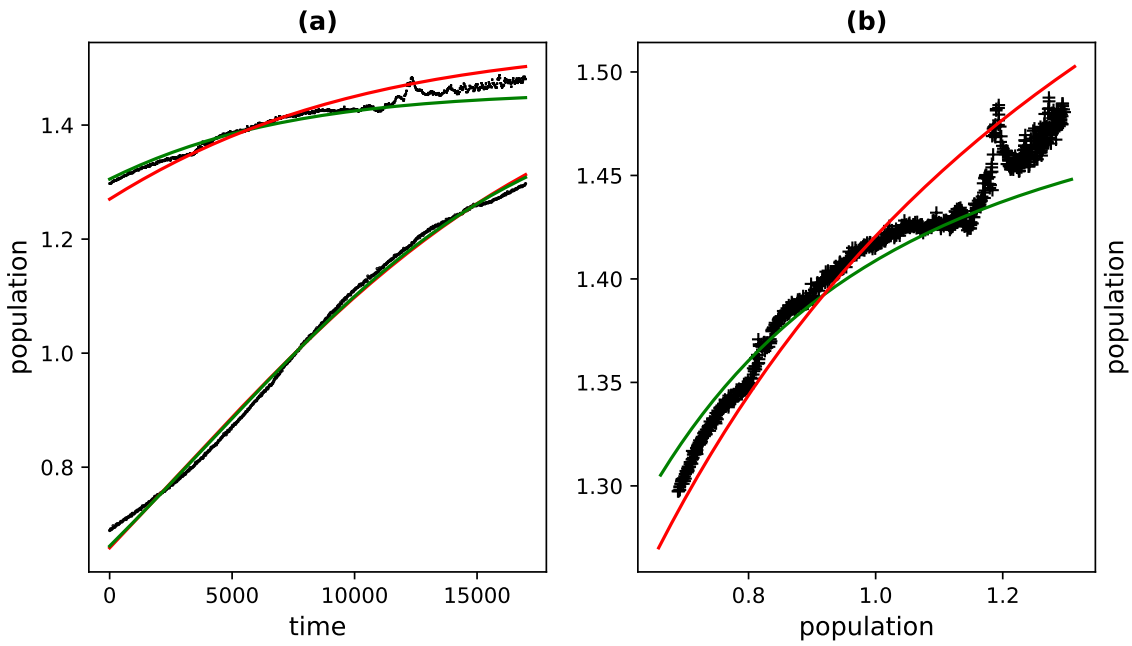
Figure 3: (a) Measurements of population size for a real bacterial colony grogin on a microtiter plate. The growth trajectory was divided in two to indicate how the population changes starting from two different initial conditions. The best fits Verhulst model is depicted with a solid red line and the best fit $\beta = 2$ model with a dashed green line for each measured curve. (b) The empirical symmetry function computed from the two trajectories in (a) is shown with black crosses. The theoretical dynamical symmetries implied by the Verhulst and $\beta = 2$ models are shown with solid red and dashed green lines, respectively.

would likely be rejected by methods that focus on single trajectory analysis. Nonetheless, the $\beta = 2$ model provides a better relative fit, and might seem a reasonable approximation to the data. However, the failure to represent the target system is quite pronounced when we examine how well the symmetries of the theoretical models fit the symmetry estimated from the data. The latter is shown in Figure 3 (b), along with the theoretical symmetries implied by the best-fit logistic and $\beta = 2$ models. The decision procedure we've been considering strongly rejects the hypothesis that either theoretical symmetry is equivalent to that in the data. In other words, neither model accurately represents the causal structure of this growing population.

5.4   Example: Stochastic logistic models

As discussed above, stochastic models present special problems for validation. To demonstrate the efficacy of the symmetry comparison approach, I simulated a stochastic version of a generalized logistic equation. Specifically, I built a (simulated) target system governed by the following stochastic differential equation (SDE),

$$\frac{dx}{dt} = rx \left(1 - \left(\frac{x}{K} \right)^2 \right) + sx dW_t, \tag{11}$$

where W_t is a one-dimensional Wiener process, and s is a constant determining the amount of multiplicative noise. Solutions to this equation were generated numerically using the discrete equations derived in (Liu and Fan 2017) (based on the Milstein method mentioned in (Higham 2001)).

Figure 4 (a) shows data from times series measured for mulitple replicates of the target system. There are 10 replicates of the system for initial condition $x_0 = 5$, and 10 replicates for the initial condition $\tilde{x}_0 = 15$. That is, the target system was evolved through time for ten iterations starting from each of two initial population sizes. Data from all iterations and initial conditions are plotted together.

There are a variety of ways in which to fit a stochastic model to such data (assuming we don't know the ground truth of Equation 11). One might build a numerical simulation and then attempt to optimize parameters with respect to one or another of the measures of fit like those discussed above in Section 2.2. The generalized definition of dynamical symmetry (see Definition 3 above), however, suggests that we focus on the expected value of the population at a given time (for a given initial condition). Consider the two sets of replicates corresponding to the two distinct initial conditions. If one averages the values measured at a given time for each set, one obtains two curves indicating expected population as a function of time. The function mapping one of these curves into the other is a dynamical symmetry. The result of this procedure is shown by the black crosses in Figure 11 (b). For whatever our model is, we can similarly compute a dynamical symmetry and compare the two as before. A significant difference would allow us to reject the structural validity of our model. In this case, I have chosen to try to model the system with a stochastic Verhulst equation:

$$\frac{dx}{dt} = rx \left(1 - \left(\frac{x}{K} \right) \right) + sx dW_t, \tag{12}$$

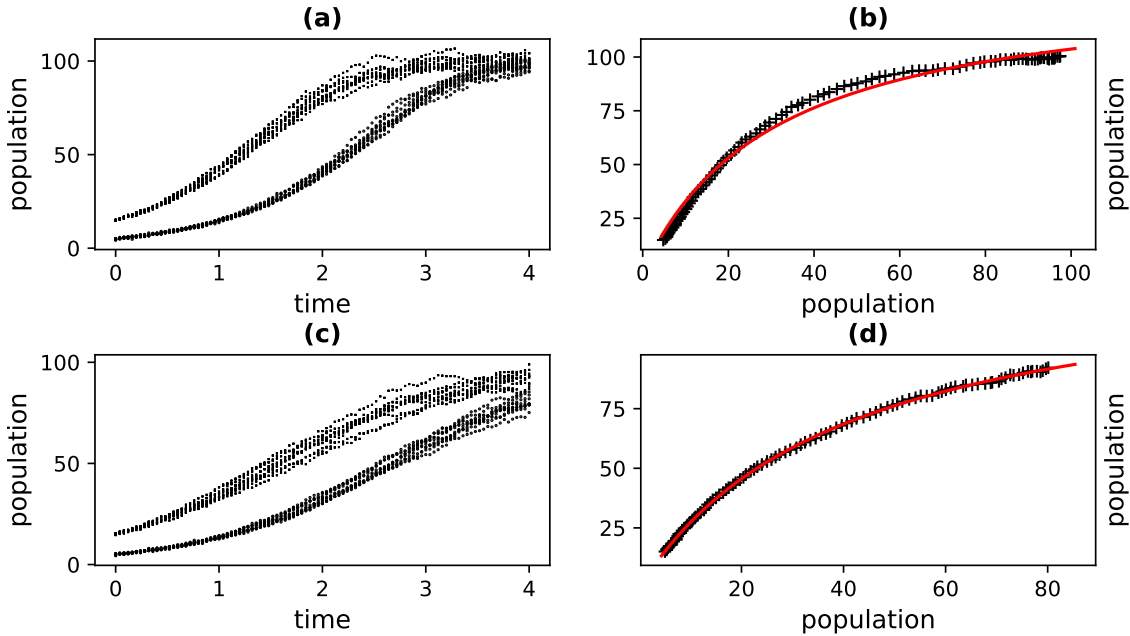For such a model, it is possible to find an analytic expression for the expected value of x as a

Figure 4: (a) Measurements for a simulated system governed by a stochastic generalized logistic equation for which $\alpha = 1$, $\beta = 2$, and $\gamma = 1$ (black crosses). The lower set of curves starts from $x_0 = 5$, and the upper from $x_0 = 15$. (b) The empirical symmetry function computed from the means of each of the two trajectories in (a) is shown with black crosses. The theoretical dynamical symmetry implied by the Verhulst model is shown with a solid red line. (c) Measurements for a simulated system governed by a stochastic Verhulst logistic equation (black crosses). (d) The empirical symmetry function computed from the means of each of the two trajectories in (c) is shown with black crosses. The theoretical dynamical symmetry implied by the Verhulst model is shown with a solid red line.

function of time. This is given by:

$$E[x(t)] = \frac{k}{1 + ((k/x_0) - 1)e^{(-rt)}},\tag{13}$$

where x_0 is the value of x at $t = 0$ (Skiadas 2010, Sec 4.4). This, as it happens, is exactly the form of the solutions of the deterministic Verhulst equation. Dynamical symmetries of the stochastic model are thus those of the deterministic model (see Equation 10). The red solid lines in Figure 4 (b) show the result of performing a least squares fit for symmetries of the Verhulst equation on the empirical data obtained from the target system. When the comparison algorithm is applied to this pair of dynamical symmetries – the one estimated for the target system and the best fit solution for our model – it rejects the model. In other words, even in the stochastic case, comparison of dynamical symmetries can lead to definitive rejection of a model.

Of course, this would be useless if the method also rejects the true model. In Figures 4 (c) and (d), results are shown for the same procedure carried out when the underlying system really is governed by a stochastic Verhulst equation. In this case, the comparison algorithm tentatively declares that the model and the target system share the same symmetry (and thus the model may be structurally valid).

6   Conclusion

I have argued above for the need to recognize a sort of structural validity for models that is broader and more forgiving than exact structural isomorphism of model and target (whatever that might mean), but that is nonetheless sufficient for establishing the reliability of a model's predictions regarding outcomes under a manipulation of inputs or boundary conditions. In other words, the sort of model reliability that is needed for the confident prediction and control of a target system is looser than the strict notion of structural validity that can be found in much of the scientific and engineering literature.

Furthermore, I've shown by way of a series of concrete examples how dynamical symmetries can be used to test for this broader sense of structural validity. Roughly, one compares the theoretical dynamical symmetries entailed by the model with estimates of real dynamical symmetries exhibited by the target system. This method, with a suitably generalized definition of dynamical symmetry, even applies to the case in which both the model and target system are stochastic. The method has important limitations. For one, it assumes that there are no latent variables driving the dynamics. Nonetheless, it represents a rigorous new tool in a field that is increasingly in need of new tools as computational models grow ever more complex.

While these results are, I think, important in their own right, I wish to draw out some implications of the mode of origination and practical success of these methods. In a sense, the development of this method of model validation is an exercise in applied philosophy. The notion of a dynamical symmetry derives from philosophical work on natural kinds (i.e., projectible kinds) (Jantzen 2014). A specific tool for model validation was derived from a very general answer to an epistemic puzzle fundamental to scientific inquiry: how do we recognize clusterings of systems or phenomena that are good candidates for instantiating scientific laws? This is obviously not the first philosophical project to contribute to scientific practice. But it is a reminder that philosophers can be helpful partners in developing well-motivated tools for empirical inquiry, not just sideline commentators.

But the converse is also true; the development of the philosophical idea into an operative tool has provided a variety of lessons that philosophers should heed. For example, the focus on approaches to narrow-sense structural validity to the exclusion of methods that would more effectively satisfy the aims of a modeler is largely a product of philosophical myopia, not a quirk of those working with models. Prominent authors in the literature on model validation frequently and explicitly take their cue from the philosophers. For instance, in speaking of methods for establishing structural validity, Barlas (1996, p 186) frames the project of structural validation this way:

> Validation of a system dynamics model is much more complicated than that of a black-box model, because judging the validity of the internal structure of a model is very problematic, both philosophically and technically. It is philosophically difficult, because, as we shall briefly review in the next section, the problem is directly related to the unresolved philosophical issue of verifying the truth of a (scientific) statement.

Thus, it is a philosophical lesson of the applied work presented here that there exists an important feature of models (and scientific theories) that sits between mere predictive success and perfect representational fidelity. Specifically, models can more or less reliably make judgments about what would be the case under intervention or manipulation without using a mechanism exactly isomorphic to whatever drives the real-world target system. This is an aspect of modeling that philosophers often ignore but shouldn't given its demonstrable utility in making sure models do what we need them to do. In other words, philosophers interested in foundational epistemic problems would do well to listen closely to their colleagues in applied math, data analysis, and statistics.

Acknowledgments

References

Balci, Osman (Dec. 1994). "Validation, verification, and testing techniques throughout the life cycle of a simulation study". en. In: *Annals of Operations Research* 53.1, pp. 121–173.

Barlas, Yaman (Sept. 1989). "Multiple tests for validation of system dynamics type of simulation models". In: *European Journal of Operational Research* 42.1, pp. 59–87.

— (1996). "Formal aspects of model validity and validation in system dynamics". In: *System dynamics review* 12.3, pp. 183–210.

Buchanan, R.L, R.C Whiting, and W.C Damert (Aug. 1997). "When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves". en. In: *Food Microbiology* 14.4, pp. 313–326.

Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification (2012). *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification.* Washington, UNITED STATES: National Academy Press.

Connelly, Brian. *Data Set for 'Analyzing Microbial Growth with R'.* doi:10.5281/zenodo.1171129.

Fillion, Nicolas (2017). "The Vindication of Computer Simulations". In: *Mathematics as a Tool: Tracing New Roles of Mathematics in the Sciences.* Ed. by Johannes Lenhard and Martin Carrier. Cham: Springer International Publishing, pp. 137–155.

Flach, Peter (Sept. 2012). *Machine Learning: The Art and Science of Algorithms That Make Sense of Data.* en. Cambridge University Press.

Fujikawa, Hiroshi, Akemi Kai, and Satoshi Morozumi (Oct. 2004). "A new logistic model for Escherichia coli growth at constant and dynamic temperatures". en. In: *Food Microbiology* 21.5, pp. 501–509.

Gompertz, Benjamin (Jan. 1825). "XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. F. R. S. &c". en. In: *Philosophical Transactions of the Royal Society of London* 115, pp. 513–583.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The elements of statistical learning.* Second. Springer Series in Statistics. New York: Springer.

Higham, Desmond J. (2001). "An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations". In: *SIAM Review* 43.3, pp. 525–546.

Jantzen, Benjamin C. (2014). "Projection, symmetry, and natural kinds". In: *Synthese* 192.11, pp. 3617–3646.

— (2017). "Dynamical Kinds and their Discovery". In: *Proceedings of the UAI 2016 Workshop on Causation: Foundation to Application.* arXiv: 1612.04933.

Ling, You and Sankaran Mahadevan (Mar. 2013). "Quantitative model validation techniques: New insights". en. In: *Reliability Engineering & System Safety* 111, pp. 217–231.

Liu, Meng and Meng Fan (Apr. 2017). "Permanence of Stochastic Lotka–Volterra Systems". en. In: *Journal of Nonlinear Science* 27.2, pp. 425–452.

McCarthy, Michael A. and Linda S. Broome (July 2000). "A method for validating stochastic models of population viability: a case study of the mountain pygmy-possum (Burramys parvus)". en. In: *Journal of Animal Ecology* 69.4, pp. 599–607.

Miller, John H. (1998). "Active Nonlinear Tests (ANTs) of Complex Simulation Models". In: *Management Science* 44.6, pp. 820–830.

Rhinehart, R. Russell (2016). *Nonlinear Regression Modeling for Engineering Applications: Modeling, Model Validation, and Enabling Design of Experiments.* John Wiley & Sons.

Skiadas, Christos H. (2010). "Exact solutions of stochastic differential equations: Gompertz, generalized logistic and revised exponential". In: *Methodology and Computing in Applied Probability* 12.2, pp. 261–270.

Sokal, Robert R. and F. James Rohlf (Sept. 1994). *Biometry: The Principles and Practices of Statistics in Biological Research.* English. 3rd edition. New York: W. H. Freeman.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines (2000). *Causation, prediction, and search.* 2nd ed. Adaptive computation and machine learning. Cambridge, Mass: MIT Press.

Tsoularis, A. and J. Wallace (July 2002). "Analysis of logistic growth models". In: *Mathematical Biosciences* 179.1, pp. 21–55.

Verhulst, P. F. (1838). "Notice sur la loi que la populations suit dans son accroissement". In: *Correspondence Mathematique et Physique* X.

Vogels, Maurice et al. (Dec. 1975). "P. F. Verhulst's 'Notice sur la loi que la populations suit dans son accroissement' from correspondence mathematique et physique. Ghent, vol. X, 1838". In: *Journal of Biological Physics* 3.4, pp. 183–192.

Wilcox, James R. (Feb. 2018). "Research For Practice: Highlights in Systems Verification". en. In: *Commun. ACM* 61.2, pp. 48–49.

Woodward, Jim (2001). "Law and Explanation in Biology: Invariance Is the Kind of Stability That Matters". In: *Philosophy of Science* 68. 1, pp. 1–20.

Zeigler, Bernard P., Herbert Praehofer, and Tag Gon Kim (2000). *Theory of Modeling and Simulation.* 2nd. San Diego: Academic Press.

Zwietering, M. H. et al. (June 1990). "Modeling of the Bacterial Growth Curve". en. In: *Applied and Environmental Microbiology* 56.6, pp. 1875–1881.