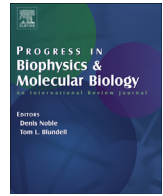




Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

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

Establishing the structures within populations of models



Philip M. Gemmell

Clyde Biosciences Limited, BioCity Scotland, Bo'Ness Road, Newhouse, Lanarkshire, Scotland, ML1 5UH, United Kingdom

ARTICLE INFO

Article history:

Received 7 March 2016

Received in revised form

10 March 2017

Accepted 20 March 2017

Available online 22 March 2017

ABSTRACT

As computational biology matures as a field, increasing attention is being paid to the relation of computational models to their target. One aspect of this is addressing how computational models can appropriately reproduce the variation seen in experimental data, with one solution being to use populations of models united by a common set of equations (the *framework*), with each individual member of the population (each *model*) possessing its own unique set of equation parameters. These model populations are then calibrated and validated against experimental data, and as a whole reproduce the experimentally observed variation. The primary focus of validation thus becomes the population, with the individual models' validation seemingly deriving from their membership of this population. The role of individual models within the population is not clear, with uncertainty regarding the relationship between individual models and the population they make up. This work examines the role of models within the population, how they relate to the population they make up, and how both can be said to be validated in this context.

© 2017 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	20
2. Model populations to represent variability	21
3. The nature of scientific paradigms	22
4. Model populations as scientific paradigms	23
5. Conclusion	23
References	24

1. Introduction

Computational models are becoming an increasingly common feature in scientific research, permitting hypotheses to be tested that would be either difficult or entirely impossible to answer using solely 'wet lab' experiments. As computational models move from the novel to the mainstream in scientific practice, they are employed for a wider range of tasks, and as available computational power increases, previous limitations on power and complexity are removed, allowing more detailed questions to be asked with fewer assumptions.

As is common with all new scientific practices, as the field matures the underlying philosophical basis for the field is examined: what the models represent and how they are related to their targets must be established to ensure that the conclusions reached can be trusted and considered valid. There is increasing research and discussion into the process of model validation and verification, and what these, and other terms, actually mean (MacLeod and Nersessian, 2013; Oreskes et al., 1994). This research is complicated by the use of computational modelling in inter-disciplinary research – different fields have different understandings of and requirements for verification and validation (Carusi, 2014). Despite this, significant advances have already been made to elucidate what computational models represent, how they relate to their target, and how they compare to the other, more established tools in the scientific arsenal (Carusi et al., 2013, 2012; Green, 2013; MacLeod and Nersessian, 2013). Of note is the concept of the *model-*

Abbreviations: PoMs, Populations of models; PDF, Probability Density Function; MSE system, Model-Simulation-Experiment system.

E-mail address: philip.gemmell@gmail.com.

simulation-experiment (MSE) system, which views the modelling process (the creation of the equations to describe the system), the simulation of the model and the experiments that produce the data as a continuous system, with no single part being fully isolated from the other (Carusi et al., 2012). No component of this system is privileged over the other – while the model may be considered to be designed to reproduce the results from the experiment, the experiment is designed to produce results that can be compared with the model, with this comparison only being possible thanks to successful simulation.

While this philosophical groundwork has been laid, computational modelling has expanded its horizons by looking to investigate the causes and consequences of experimentally observed variation. Several different methods are available to examine the dynamics underlying variability, with one of these being *populations of models* (PoMs). Instead of producing a single computational model to reproduce a given measure of experimental data, several models (potentially hundreds, thousands, or even more) are used to reproduce the complete set of experimental data, including the variability. The individual members of the population are united by the underlying equations (the so-called *framework*), and vary from each other by the parameter values used in the actualisation of these equations.

However, in extending from a single computational model to a population of models, the role of an individual model has become more uncertain. Previously, a single model was created to reproduce certain predetermined aspects of experimental data, providing a simple lineage from data to model to simulation (with interactions between these stages). The incorporation of variability complicates matters: individual measurements are arrayed to provide the experimental data which then exhibits variability. No individual measurement shows variation, in that each individual measurement consists of a single number with no ‘error bars’ associated with it. It is only when several measurements are combined that the underlying variability emerges, with the amalgamation of several different measurements permitting its evaluation. The population of models is constructed to reflect the population of measurements, and the potential significance of any given model within this population is no longer clear.

This paper will investigate a scheme by which the relationship of individual models to a broader population is clarified, and how these are related to the experimental data. The paper will start with a brief overview of the design of model populations, including details of how these populations are calibrated. A brief account of some of the benefits of model populations will be given. With this background, we will explore the nature of paradigms in the philosophy of science, before drawing analogies with PoMs to offer a mode of thinking about the role of individual models within a population.

2. Model populations to represent variability

Model populations have been used in many different fields, with a long history ranging from climate modelling (Epstein, 1969; Leith, 1974) to neurophysiology (Goldman et al., 2001; Prinz et al., 2003; Taylor et al., 2009) to cardiac electrophysiology (Britton et al., 2013; Gemmell et al., 2016, 2014; Sarkar et al., 2012). This latter field shall be used as the exemplar for this section, but the themes are universal for all PoM approaches. The initial step for PoMs is an underlying set of equations (the *framework*): these equations describe the dynamics of the model, without necessarily indicating the magnitude of the dynamics. In a cardiac electrophysiology model that is biophysically detailed, the equations represent the biophysical processes, such as the dynamics of the cellular ion channels, which model the flow of ions through various ion channel

gates, with these gates opening and closing in response to the cellular environment (and other potential inputs, such as time). To generate a population, the same framework is used with several parameters of the equations being varied. For our biophysically detailed model, these parameters would describe such factors as the maximum conductance or activation/inactivation time for a given ion channel. In the terminology of this paper, each individual instantiation of a framework, with a specific set of parameter values, is a *model*, with the collection of these models being the *population*.

Model populations can be thought of as existing within a given parameter space. Due to the continuous nature in which parameter values can be chosen, but the discrete choice of parameters required for a model to be instantiated, a given population can only ever properly be a sample of the parameter space. The uses of model populations depend on how this sample is constructed. If the population is derived by uniformly, systematically and extensively sampling from the space, one can investigate the inter-dimensional relationships that exist between different parameters, how these relationships change depending on their location in the parameter space, and how these affect the observable output of the cell model (Gemmell et al., 2014). However, it is often either impractical or not necessary to simulate and analyse a parameter space in as detailed a manner as this, and it is more common to instead sample from the parameter space in a less complete way: one common method is to use Latin Hypercube Sampling, which ensures that while the sample is taken randomly from the space, the samples are evenly distributed throughout the space (McKay et al., 1979). Through generating a population in this manner, it is possible to assess the relative contributions of individual parameters to individual outputs (Sarkar et al., 2012), and generate populations for further calibration and validation (Britton et al., 2013).

PoMs have been applied directly to analysis of variation and variability through comparison to experimental data that exhibit variability. After initially producing a ‘progenitor’ population, the membership of the population is then refined by comparison to experimental data – all those models that produce an output that does not match experimental data are removed from the population (Britton et al., 2013; Gemmell et al., 2014). The resulting populations then reproduce the observed experimental data by design. These populations are not sterile reproductions of the training experimental data, and have been shown to have some predicative power, for example recapitulating the effect of drug block on the population dynamics and variability (Britton et al., 2013).

It should be noted that it will rarely be the case that the population after calibration will be uniformly distributed across the parameter space, and similarly there is no requisite need for the progenitor population to be uniform across the parameter space. The distribution of the models across the parameter space can be considered to correspond to a probability density function (PDF) of the parameters. By this mode of thinking, the uncalibrated PoM distribution of parameters can be considered a prior PDF, which is then adapted to a post-calibration distribution which corresponds to a posterior PDF using Bayesian inference. These PDFs could then be postulated to reflect (a) the uncertainty inherent in ascribing parameters to the models (Epstein, 1969; Leith, 1974); (b) variation in the physical reality that the parameters seek to reproduce; (c) a certain ‘sloppiness’ in how accurately one can ascribe a precise value for the given parameter given the training data (Gutenkunst et al., 2007); (d) a combination of the three. This mode of thinking is hugely valuable, and provides a means to understand the relationship among individual sample points, and between them and the PDF. However, this provides a methodology to understand the models and their relation to each other – this paper looks instead to focus on the over-arching relation between model, framework and population.

Download English Version:

<https://daneshyari.com/en/article/5519836>

Download Persian Version:

<https://daneshyari.com/article/5519836>

[Daneshyari.com](https://daneshyari.com)