Progress in Biophysics and Molecular Biology 129 (2017) 13–19

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

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

A framework to establish credibility of computational models in biology

Eann A. Patterson [a, *], Maurice P. Whelan [b]

[a] School of Engineering, University of Liverpool, Liverpool, UK
[b] European Commission Joint Research Centre, Italy

ARTICLE INFO

Article history:
Received 16 April 2016
Received in revised form
18 July 2016
Accepted 1 August 2016
Available online 1 October 2016

Keywords:
Computational biology
Computational models
Credibility
Validation
In silico medicine
Biomedical modelling

ABSTRACT

Computational models in biology and biomedical science are often constructed to aid people's understanding of phenomena or to inform decisions with socioeconomic consequences. Model credibility is the willingness of people to trust a model's predictions and is often difficult to establish for computational biology models. A 3 × 3 matrix has been proposed to allow such models to be categorised with respect to their testability and epistemic foundation in order to guide the selection of an appropriate process of validation to supply evidence to establish credibility. Three approaches to validation are identified that can be deployed depending on whether a model is deemed untestable, testable or lies somewhere in between. In the latter two cases, the validation process involves the quantification of uncertainty which is a key output. The issues arising due to the complexity and inherent variability of biological systems are discussed and the creation of 'digital twins' proposed as a means to alleviate the issues and provide a more robust, transparent and traceable route to model credibility and acceptance.

Contents

1. Introduction

Whenever a model is developed, a primary concern of the modeller is the credibility of their model. Credibility has been described by Schruben (Schruben, 1980) as reflecting 'the willingness of persons to base decisions on information obtained from the model'. So, the issue becomes a matter of providing sufficient evidence of the model's fitness for purpose to induce this willingness. Rudner (1953) postulated that our judgement on the strength of the evidence depends on the importance or consequences of making a mistake, which implies that modellers need to consider the intended uses of their model when identifying the evidence required to underpin credibility.

Often in biology, as in other areas of pure science, the primary value of computational models is heuristic (Oreskes et al., 1994). They are representations of reality that are valuable for understanding and guiding further research or study. In these circumstances, when the role of the model is not associated with decision-making, its absolute accuracy is not the essential issue. Rather, it is more appropriate to consider computational models as the

* Corresponding author.
E-mail address: eann.patterson@liverpool.ac.uk (E.A. Patterson).

http://dx.doi.org/10.1016/j.pbiomolbio.2016.08.007

apparatus or environment in which simulations or *'in silico'* experiments are performed for the purpose of exploring hypotheses and revealing features of behaviour for which only sparse or no observational data is available (Winsburg, 2010). If the revealing of features is a sufficient outcome, then an adequate process of model validation to underpin credibility could be to simply ensure that the model is useful and functional in providing relevant insights. This approach has been employed, for example, in materials science and termed 'validation of phenomena' (Patterson, 2015).

Biology overlaps with engineering when it is used to create man-made components and products or when engineered products interact with human biology, such as in pharmacology and toxicology. In these circumstances, when models are used, it would be appropriate to adopt the level of rigour employed routinely by the engineering sector to demonstrate their credibility. Engineers use computational models to evaluate and refine the performance, reliability and safety of designs of engineered products. Hence for these models, which might be termed predictive rather than heuristic, the consequence of making a mistake will be typically measured in socioeconomic costs, often significant, such as loss of life or injury. This implies the need for strong evidence that the computational model closely reflects reality, and leads to the definition of validation as 'determining the degree to which a model is an accurate representation of the real world from the perspective of its intended uses' (ASME V&V 10-2006, 2006). The engineering community has developed a series of quantitative validation procedures, (e.g. in solid mechanics (Sebastian et al., 2013)), that allow the evidence to be assembled in a framework that is recognised by modellers and end-users, (e.g. for solid mechanics models (CWA 16799, 2014)), and supports the establishment of credibility and confidence.

In *in silico* biology, when computational models are used to reveal features of behaviour, even the 'validation of phenomena' can be challenging in the absence of reliable data from the real-world, which of course is often the reason for wanting to use a model in the first place. Some computational models of biological systems would appear to be untestable due to their complexity and the difficulty in acquiring reliable data from the biological system. It is tempting at this point, to trust to the judgment of the modeller and accept that the simulation will provide interesting information. However, Hughes (1999) has said that *in silico* experiments reveal information about three types of world: the actual world, possible worlds and impossible worlds; and that it is not possible to know which has been revealed without taking an extra step, such as some form of validation. So, it would be inappropriate to abandon some effort to test the reliability of computational biology models. Thus, our aim is to develop a framework for establishing the credibility of computational biology models that are classified according to our ability to test them and identify their epistemological foundations, to support the work of both modellers and those making decisions based on results from models.

2. Credibility matrix

Untestable models are employed in physics and, to a lesser extent, engineering. Tegmark has drawn an epistemological boundary between physics and metaphysics that is defined by whether or not a theory is experimentally testable (Tegmark, 2014). While for engineering models, Patterson (Patterson, 2015) has gone further and constructed a 2 × 2 diagram that identifies the appropriate approach to establishing the credibility of testable and untestable or meta models based on whether they are principled or unprincipled, i.e. whether the underlying physics is known or unknown. In Fig. 1, we have developed this approach for use in computational biology and *in silico* medicine.

Sober (Sober, 1993) has stated that there are no exceptionless laws in biology. Notwithstanding that some would point to the first law of biology being 'the tendency for diversity and complexity to increase in evolutionary systems' (McShea and Brandon, 2010), it is clear that it is difficult to identify universally accepted biological laws. Thus, the use of principled and unprincipled on the horizontal axis is potentially problematic when referring to biology. Instead, in Fig. 1 the more general terms 'known biology' and 'unknown biology' have been used. The allocation of a model between these two categories should be made based on whether or not its knowledge base is founded on one of the three types of scientific reasoning (Osimani and Mignini, 2015), namely (i) inductive reasoning from empirical data to a theory, (ii) hypothesis falsification through *modus tollens*, or (iii) explanatory reasoning. These modes of reasoning are generic, and in biology it would be appropriate to embrace Hill's criteria for causation (Villeneuve et al., 2014). Computational biology models are unlikely to be as readily categorised as implied above, so it is appropriate to include a transition zone between models based on known biology and those based on unknown biology, i.e. between principled and unprincipled. For example, a model of a biological system is usually constructed by combining models of its sub-systems, each perhaps reflecting different scales of biological organisation, and each based on varying degrees of phenomenological understanding. Such models would be located in this transition zone (i.e. the middle column in Fig. 1) especially when the linkages between the sub-systems are not understood.

In computational biology, at the boundary between testable and untestable models in Fig. 1, there will be another transition zone that originates from the difficulties in making quantitative observations of real-world biology, which leads to sparse or incomplete data. This is in part due to our inability to control the real-world, as observed by Viceconti (Viceconti, 2015).

The credibility of models that fall into the bottom left corner in Fig. 1 can be established using the type of quantitative validation procedures that are being codified by the engineering community

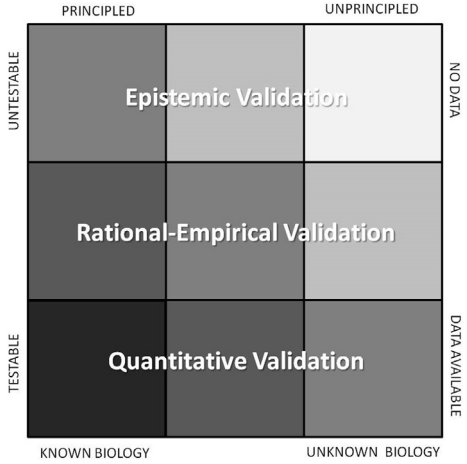

Fig. 1. a schematic diagram illustrating the relationship between testable and untestable models that are either based on known biology (i.e. principled) or unknown biology (i.e. unprincipled) together with the approaches to performing a validation and the resultant level of credibility that can be established indicated by the greyscale. Testable models are those for which it is possible to acquire measured data from real-world experiments, while untestable models are those for which it is not possible to make measurements corresponding to the model's predictions. Epistemic validation is based on the epistemic values of the model including simplicity, consistency and explanatory power; rational-empirical validation involves a series of three 'tests' using rationalism, empiricism and demonstration of predictive accuracy; while quantitative validation employs the rigorous methods described in engineering standards.