Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Multi-class and multi-scale models of complex biological phenomena

Jessica S Yu and Neda Bagheri

Computational modeling has significantly impacted our ability to analyze vast (and exponentially increasing) quantities of experimental data for a variety of applications, such as drug discovery and disease forecasting. Single-scale, single-class models persist as the most common group of models, but biological complexity often demands more sophisticated approaches. This review surveys modeling approaches that are *multi-class* (incorporating multiple model types) and/or *multi-scale* (accounting for multiple spatial or temporal scales) and describes how these models, and combinations thereof, should be used within the context of the problem statement. We end by highlighting agent-based models as an intuitive, modular, and flexible framework within which multi-scale and multi-class models can be implemented.

Address
Chemical & Biological Engineering, Northwestern University, Evanston, IL, United States

Corresponding author: Bagheri, Neda (n-bagheri@northwestern.edu)

Current Opinion in Biotechnology 2016, **39**:167–173

This review comes from a themed issue on **Systems biology**

Edited by **Mark P Styczynski** and **Fabian J Theis**

For a complete overview see the Issue and the Editorial

Available online 23rd April 2016

http://dx.doi.org/10.1016/j.copbio.2016.04.002

0958-1669/Published by Elsevier Ltd.

Introduction

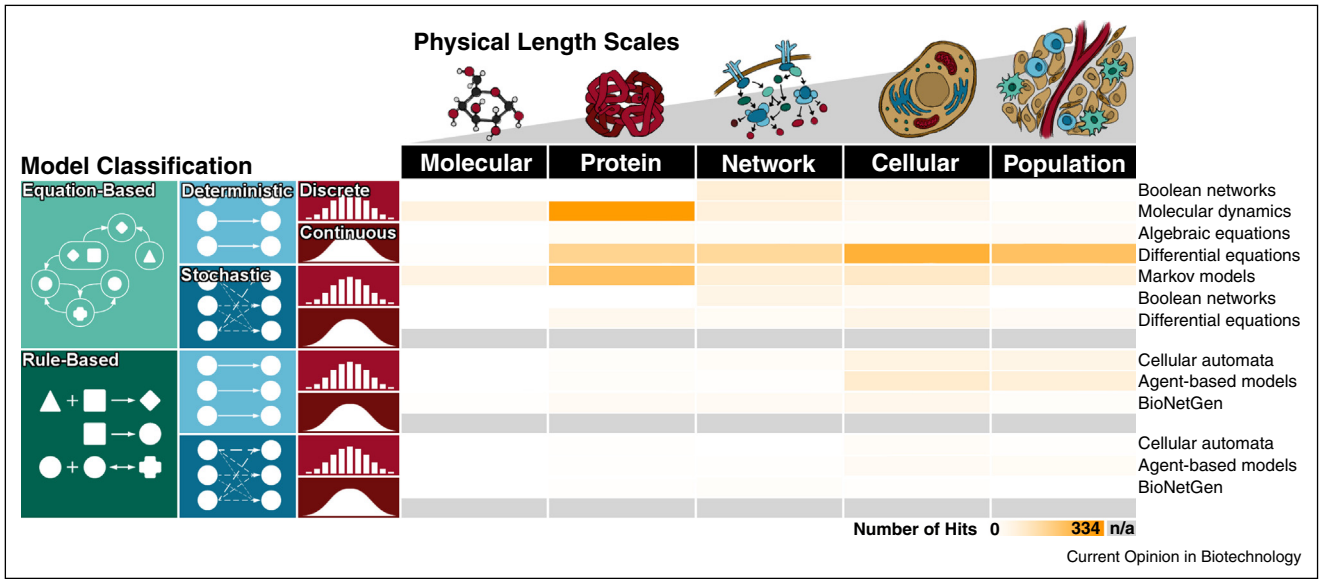
Computational modeling is a powerful tool that impacts science and society in remarkable ways. Major applications include accelerated design and development of drug therapeutics, as well as integration of theory and experiments to advance basic science. In drug therapeutics and in view of a rapidly approaching patent cliff for several major drugs, pharmaceutical companies have embraced computer-aided drug design through techniques such as binding pocket modeling, molecular dynamics simulations, and pharmacokinetic modeling [1–3]. Models of disease and organ level physiology help researchers predict system response to therapeutic perturbations and design clinical trials [4]. On a larger scale, computational models have also been used to investigate incidence, spread, and intervention for diseases such as Ebola and West Nile virus [5–8].

In basic science, computational models predict complex behavior, elucidate regulatory mechanisms, and inform experimental design [9,10]. For instance, a quorum sensing model in *Agrobacterium* described complex, experimentally observed behavior of the organism and provided experimentally-testable hypotheses for the evolutionary significance of the sensing phenomenon [11]. A previously longstanding question in biology involving the regulatory mechanism driving *E. coli* chemotaxis was also resolved by iteration between a computational model of chemotaxis signal transduction and experimental validation with mutant strains [12]. Simulations of the *Drosophila* segment polarity network by von Dassow *et al.* revealed that no parameter sets produced the observed behavior, leading the authors to amend their understanding of the network and propose new candidate mechanisms for further experimentation [13].

With the rapid expansion and improvement of experimental techniques, scientists are generating unprecedented amounts of high-throughput, high-quality biological data [14]. This exponential growth of multi-dimensional biological data requires a parallel growth in quantitative modeling methods of such data to explain non-intuitive observations [14,15]. The appropriate model group and framework necessary for attaining biologically relevant insight depends on context and data. Commonly, these models are single-scale and single-class; Figure 1 outlines a high-level summary of these frameworks used at different biological length scales of interest. At one extreme, models considering the interactions of molecules and protein structure tend to be discrete and based on first principles. At the other extreme, models describing single cells and cell populations tend to be more continuous. Across these scales, rule-based models are less common than equations-based models.

While we acknowledge the existence of countless additional length-scales — including organs, organ systems, individuals, and populations — as well as countless time-scales, these layers of resolution are outside the scope of this review. Statistical learning theory also falls outside the scope of this review. In the following section, we describe fundamental attributes of computational models to establish a common vocabulary. Next, we highlight attributes of multi-scale and/or multi-class models (as well as combinations thereof) and their unique advantages for describing complex biological phenomena.

Figure 1



Number of PubMed 'hits' highlight trends in modeling framework selection across different scales. Representative model frameworks for each of the different model classes are used. The color bar highlights molecular dynamics models of proteins as one of the most published frameworks and rule-based approaches as one of the least explored. Search strings used to generate these numbers are provided on our website.

Classes and scales of computational models

A *model* represents a real-life system or phenomenon of interest [16,17]. Unlike physical models, composed of tangible parts, computational models are mathematical abstractions of a system [16]. *State variables* describe the model at an instance of a computational simulation and *parameter variables* characterize the model itself [16]. Consider a model of a single cell undergoing mitosis: a state variable could denote the phase of the cell cycle at a given time point, whereas the duration of the phase would depend on the value of a parameter variable.

Models can be classified according to two characteristics: (i) the evolution of state variables and (ii) the computational repeatability of the output, or response, trajectory. *Evolution* describes how state variables change: *continuous* variables take on any value, *discrete* variables take on integer values. *Repeatability* characterizes the expected output trajectories from a given set of inputs. *Deterministic* models produce identical output response profiles provided identical initial state and parameter conditions, whereas *stochastic* models allow random or probabilistic events such that a single simulation given identical initial state and parameter conditions can produce a family of response profiles [16]. We include a third characteristic of classification, *specification*, which identifies how the interactions among state variables are defined. *Rule-based* models use qualitative rules to describe interactions, while e*quation-based* models use mathematically formulated ones [18,19]. Thus, a model class is given by its evolution, reproducibility, and specification characteristics. We define models that incorporate multiple model classes in a single framework as *multi-class*; otherwise, the model is *single-class*.

In parallel, we define the *scale*, temporal or spatial, of the model. By definition, *multi-scale* models explicitly incorporate two or more levels of resolution [20,21]. Biological systems are inherently multi-scale: cell population dynamics rely on intercellular protein signaling, which relies on fluctuations of intracellular protein concentrations, and so on [21]. Existing reviews offer detailed discussion of multi-scale models for biological systems [20,22,23]. At the center of Figure 2, the four different groups of models are organized based on increasing *multiplexity* of class and scale.

When modeling complex biological systems, and regardless of approach, the multiplexity of a computational model should be defined by the context of the problem statement, the type of answer sought (i.e., qualitative vs. quantitative) and the experimental data. Disparities between the modeling framework and problem statement can limit predictive capacity and biological insight. A model is too detailed if the training data cannot fully support the model structure or parameter estimation, which can result in over fitting and unnecessary computational expense and be revealed through parameter identifiability or uncertainty analysis. A model is too simple if it is unable to capture the possible depth or breadth of understanding that a qualitative and/or quantitative data set provides. Different aspects of cancer, for