

Scientific, Research, and Medical (SRM) Objective

MSMs of mammalian systems (healthy and morbid) that, in the face of biological variability and considerable uncertainty, are mechanistically explanatory and suitable for 1) discovering and validating plausible, causal linkages in healthy and morbid systems, and 2) improving existing and discovering new therapeutic interventions as well as predicting intervention consequences. Clearly, such MSMs will be perpetual works in progress capable of evolving and accommodating new knowledge concurrently in order to continue being scientifically useful.

The discussion spans continuous, equation-based models to discrete event, object and agent oriented models (including agent-based (ABM) and actor models [1,2]).

Introduction

The units, dimensions, and/or objects to which a variable or model constituent refers establish groundings. Absolute grounding: variables, parameters, and input-output (I/O) are in real-world units. Each term is foundational and maps to a tacit thing having a conceptual or biological meaning. Relational grounding: variables, parameters, and I/O are in units defined by other system components. Variables and components may map to counterparts outside the system or have meanings that are unrelated to the real world. Biology uses relational grounding.

Grounding issues may not pose problems when a model is narrowly focused on a single aspect of a system (e.g., a pharmacokinetic or gene network model). However, when a model aims to describe multiple aspects (i.e., different phenotypic attributes), grounding problems emerge. A spectrum of MSM classes and methods is needed: absolute grounding occupies one extreme; relational grounding occupies the other. Model classes that use relational grounding are essential for achieving the SRM objective.

Grounding decisions influence model flexibility, adaptability, and thus reusability

Grounding to metric spaces and real world units within inductive mathematical models provides simple, interpretive mappings between output, parameter values, and referent data. Absolute grounding issues must be addressed each time one of the following occurs: expand the model to include additional phenom-ena; combining models to form a larger system; and/or model context changes. Expansions are challenging [3], even infeasible, when center-left in Fig. 1. A PBPK model, for example will have limited reusability when ental conditions change or when an assumption made in the original model formulation has been falsified. Reusability is hindered in part because two of the models conflated to form a PBPK modelthe mechanistic model and the in silico-to-referent mapping modelhave different uses.

A PBPK model can be made dimensionless to improve flexibility and reusability by multiplying a dimensioned variable by a constant having reciprocal units. Doing so relies on (or assumes) a constant part of a particular system. However, identifying a constant part of a mammalian system is problematic. The components and processes in discrete event, object and agent oriented, biomimetic analogs need not have assigned units [4-9]. Relational grounding enables synthesizing flexible, easily adapted, extensible, hierarchical analogs of the systems they mimic.

Measures, uncertainty, and system information influence grounding choices

Being on the right if the Fig. 1 scales favors reliance on inductive reasoning and developing inductive models: the generators of phenomena are well understood, and precise knowledge about mechanisms is available at most scales. Ample validation data is available. As one moves left, uncertainties from multiple sources increase. Conceptual mechanisms are less validated (less trustworthy) and more hypothetical. Reliance on inductive models requires conceptual simplification and idealization. Networked assumptions accumulate, some of which may be abiotic. Those assumptions are woven into a MSM by reliance on absolute grounding. Difficulties testing mechanistic hypotheses increase dramatically center-left.

MSM grounding issues: one main points; four sub-points

Grounding relates to the Degrees of Freedom within the MSM, the model's free, or non-local, variables. In this context, a degree of freedom is any property that is not dependent on other variables. Where are the model's free variables?

- (e.g., depth, flatness; variable locality)
- What are the free variable data types?
- (real, integer, alpha-numeric, etc. metric or hyperspatial)
- What is the distance from any given variable to a free variable? When absolute grounding is complete, the distance is always 0. Relational grounding implies that distances can be > 0.
- How many degrees of freedom does the model contain?
- Models grounded absolutely typically have fewer.

Scaling between species (or between different sets of individuals)

Consider scaling metabolic clearance of a particular drug in mice (ml/min/g) to enable human prediction when clearance is grounded absolutely on concentration and time. Scaling to human clearance values requires applying mass, volume, and time scaling factors to all parameters simultaneously, knowing that each scaling factor is imprecise and uncertain. When scaled predictions deviate significantly from observed values (common), there is no way to ascertain which scaling factor(s) and/or which scaled parameter(s) is problematic. Consider a relationally grounded clearance model. Mass and volume scaling can be done separately and validated independently. Setting the scale for one variable in accordance with trusted validation data can help set the scale for other, related variables. Time scaling is more complicated. It can be accomplished by finding an appropriate timescaling factor for each probability parameter, individually. In time, we can achieve automated scaling of models grounded relationally.

Representing uncertainties

In absolutely grounded ODE MSMs, variables and parameters are often expressed as precise mathematical values, although their networked uncertainties are usually significant. Examples include the physiological parameters in PBPK models (values for different parameters may come from different wet-lab experiments and/or may be based on different subject types, etc.). Representing uncertainty within a system of differential equations grounded absolutely is mathematically complex. Integrating models from different contexts can require that the whole model be refitted, which should involve re-examination of the cumulative consequences of the networked assumptions. In contrast, in a relationally grounded MSM, probabilistic functions can represent inherent uncertainties conveniently. Importantly, the causes for being unable to adequately match (or later falsify) a relationally grounded MSM are made more obvious by the explicit inclusion of probabilistic functions.

Linking differently grounded sub-models

Consider a discrete event ABM linked to a discrete time ODE model [10,11]. Biomimetic linkage requires methods for smoothing discrete ABM outputs and discretizing the smooth ODE outputs. Example: ODE outputs are real-valued fluxes; those values must be mapped into clusters ("bins") for ABM use. The ABM depends intricately on the assumptions (which may be hidden) made in the ODEs plus the discretization of its outputs. Conversely, discrete output from the ABM needs to respect the continuity of the ODEs. The two components may be linked by, for example, a common time scale, the choice of which will likely have important implications for interpreting results.

Groundings can influence sensitivity analyses

There can be difficulties with sensitivity analyses for multicomponent MSMs. Consider an ABM linked to ODEs, as above. To incorporate new knowledge, expand phenotype, enable revision following a falsification experiment the ABM must be changed. Such change will likely require re-parameterization or reformulation of the ODEs, and that may undermine the extensive sensitivity analysis already performed. To overcome such difficulties, one can follow an alternative MSM formulation: 1st, develop relationally grounded models of both components, link them and achieve some degree of validation. So doing facilitates component replacement, limiting any one component formulation solely to its coupling with the others. Any component can be replaced at will as long as the minimal I/O matching requirements are met. With an absolutely grounded MSM it is often very difficult to replace a single component. Starting with a relationally grounded models allows the modeler to iterate progressively from qualitative to quantitative validation.

Qualitative and quantitative validation issues

Simulation validation is based on similarity to a referent system. Similarity can be defined on a spectrum, ranging from qualitative to quantitative. With qualitative similarity, objects will either possess some quality, or they will not. Simulation and referent attributes are considered similar if they have (almost) the same qualities. With quantitative similarity, attributes are categorically the same but vary by magnitude and can be compared by some ordering relation (e.g. less than or greater than). Despite the spectrum used to define similarity, it is always the case that a qualitative description is a prerequisite for (and provides constraints for) quantitative descriptions, in the sense that any quantities defined must relate to one or more qualities. Validation (or lack thereof) should act as an important determinant of what type of model (and therefore, what type of grounding) to employ to model the system of interest.

UCSF BioSystems Group

Relational Grounding Facilitates Development of Scientifically Useful Multiscale Models

C. Anthony Hunt and Glen E.P. Ropella

Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA

Qualitative and quantitative prediction issues

Some degree of validation should precede prediction. If only qualitative validation and prediction are currently possible, use relational grounding. Having guantitative data that agrees with quantitative predictions provides a degree of validation, but no new knowledge is created. Qualitative falsification is more useful because we learn how and where the knowledge instantiated in the model is inaccurate.

Knowledge embodiment (12,13) requires models (synthetic analogs) that are relational

Absolute grounding of all MSM hooks model semantics to the conceptual models on which biomedical scientists rely. Such grounding is common because it makes the computational MSM purely mechanical, a black box function that blindly transforms input into output. Humans interpreting the I/O do all the semantic grounding manually, outside the computational framework. Hence, neither knowledge nor semantics, is embedded in the model. Various methods have been invented to improve that reality and enable knowledge to be embedded in the machine. Relational databases, XML ontologies, cellular automata, artificial neural networks, object orientation, agents, actors, etc. enable implicit embedding of different types of knowledge by mapping computational mechanisms to the hypothesized referent mechanisms. Knowledge can be embedded progressively into MSMs by using and making explicit relational grounding. To be scientifically useful, knowledge must be visible to all users, which is not the case now.



Relational grounding facilitates referent knowledge embodiment within computational mechanisms

Computational biology markup languages standardize relationships between the terms. To achieve the SRM objective, we must have explicit ontologies enabling specialization into technical programming and biomedical domain expertise. A biomedical domain expert must be able to examine a model and simulation events without needing computational expertise. Achieving the SRM objective will have parallels to the evolution of modern biomedical research. Biomedical scientists design and perform complex experiments without becoming experts in laboratory equipment/reagent design/production. Progressing from custombuilt experimental to standardized lab products allowed scientists to compress complicated methods. They were subsumed by engineering, production, and validation processes. That subsumption reduced experimental variability, facilitated experiment replication, lowered costs, which freed scientists to build increasingly more sophisticated experiments atop complicated equipment (e.g., a cell sorter, a confocal microscope, monoclonal antibodies, transfection reagents, automated DNA sequencer). The same progression must occur for the SRM objective to be achieved.

Guideline and recommendation

When validation data are not available, it is most reasonable to first build a qualitative, relational MSM, and then qualitatively validate and make predictions that will help design wet-lab experiments, the results of which may then be used to refine the qualities and relations in the MSM. This process should be iterated until quantitative data is available to help falsify the model, and then refined it to make precise predictions. In doing so, we are moving from left to right in Fig. 1.

A MSM will have groundings in both technical & biomedical ontologies; one's tool bias is a grounding issue.

Absolute grounding complicates combining models or modules to form larger models unless all are also grounded absolutely

Bassingthwaighte et al. [14] discuss the issues. Component and model integration issues balloon into important choices that must be made. To illustrate, consider:

 $dx/dt = f(x, u, t), x(t_0) = x_0; y = g(x, u, t), where,$

t is a Real number such that $t \ge t_0$;

x(t) is an n-dimensional real tuple representing the state of the system; u(t) is an m-dimensional real tuple representing the system input;

y(t) is an I-dimensional real output;

 x_0 is an n dimensional real vector representing the initial condition. Now consider another ODE that uses a different ordering parameter s, state description p(s), system input v(s), and output q(s): $dp/ds = f(p, v, s), p(s_0); q = g(p, v, s).$

To integrate dx/dt & dp/ds to form a sibling, the modeler must find mappings t \sim = s, p \sim = x, v \sim = u, and y \sim = q. To do so, the modeler typically finds expressions for each element in some real-world units, an absolute ground.

Integrating when the scales are different, even if the units are the same, presents technical choices the modeler must make. Those choices impact the behavioral solution the scientist sees. The typical solution is to ground the entire model in the same, low-level units, a "least common denominator." So doing effectively flattens the model. Note that there are modeling tools like Ptolemy II [15] that help the modeler make these decisions with much reduced effort [16]; but as made clear in [14], the decisions must still be made. Note: flattening a conceptually MSM will introduce abiotic features. Flattened models are immediately at risk of using abiotic simulation mechanisms, which may or may not impact model use.

Absolute grounding can simplify integrating sub-models by forcing common units, but context change may require model reengineering

Absolute grounding can make a MSM fragile to changes in context (context examples: particular experimental conditions; normal vs diseased; presence/absence of an intervention; different mouse strain). Biology adapts to a changed context. An absolutelygrounded MSM that has undergone validation against phenomena measured in context A may need to undergo considerable reengineering and reparameterization in order to validate against the altered phenomena in context B. Relationally grounded analogs can adapt easily to function in a new context, but the in silico-to-referent mappings will change.

Models grounded absolutely are implicitly multi-model. That reality needs to be made explicit

Most current MSMs are accretions of at least three different model types: 1: conceptual models of biological mechanisms, features, and aspects; 2: equations that describe temporal phenomena of (1); 3: measurement units that provide a quantitative 1:1 mapping from (2) to measures of biological phenomena. The conceptual models are grounded to the biology via the literature and expert opinion. Both (1) and (2) may contain bias &/or questionable assumptions. Good science requires all three be made specific and concrete. Such model accretion reduces flexibility and makes reuse problematic. When flexibility and reuse are required, there is an optional, threestep approach to achieving the SRM objective.

1) Map specific biological features to actual, concrete, relationally grounded software objects and spaces that have execution protocols. 2) Measure actual phenomena generated during executions.

3) Specify quantitative mapping models to relate measures of in silico phenomena to real measures of referent phenomena.

Such model separation increases flexibility and encourages reuse.



Multi-paradigm modeling requires relational grounding

To achieve the SRM objective, MSMs need the option for some internal mechanisms to be unrelated (can be independent of) to the internal mechanisms of others. Doing so will require relational grounding. Example problem: to integrate a discrete event submodel with a discrete time (ODE) sub-model, guantitative map-pings between components are needed. Use of absolute ground-ing provides a single paradigm (approximation to a continuous system) to which all components relate. However, within the software implementation, absolute grounding effectively flattens hierarchical models and prevents mechanism independence.

Relational grounding will facilitate translational research and development of **MSMs that have long lifecycles**

Conceptual mappings from one wet-lab model to another or from a wet-lab system to humans are difficult to falsify, yet falsification, not validation, leads to new knowledge. One relationally grounded analog can be morphed into another. That morphing stands as a model of the conceptual mapping/translation. It can be directly challenged and falsified (or not). Such models must be adaptable and extensible. Those that are can have long lifecycles. Those features require relational grounding.

Exploring mechanisms of normal-to-disease transition requires model components that use relational grounding

Multi-attribute MSMs are needed to help achieve exploitable insight into normal-to-disease transitions and facilitate discovery of new treatment options. Normal-to-disease transitions will require change in how components at multiple levels interact. Alternative mechanistic scenarios will need to be explored and challenged. Having components grounded absolutely makes such exploration problematic. Reliance on relational grounding simplifies mechanism exploration and makes the process more intuitive.

Components in composite, multi-attribute, biomimetic modules and models need some autonomy: relational grounding facilitates providing component autonomy

All models have a degree of articulation, which is the extent to which the model consists of distinct parts, modules, or components. Articulation for biomimetic analogs extends beyond OOP encapsulation of state and behavior into activity. A component that initiates and maintains its own run-time is autonomous. Mammalian cells can be autonomous in vitro. Scientifically useful in silico analogs will likewise need autonomy.

Tissues and organs are highly articulated systems. The same can be true for their MSMs. When those components are quasi-autonomous, they can be effectively replaced by other components for which component I/O requirements are specified in ways consistent with biology. Model articulation issues are orthogonal to issues of absolute vs. relational grounding. The extent to which a component is autonomous is handled by the clear specification and maintenance of component use cases (aspects; phenotypic attributes) or, collectively, the component's phenotype. A model will lack any autonomy as long as there is only one use case for given component and a single use case for all connected components. Autonomy can be established regardless of how a module is grounded, but only when targeted phenotypic attributes are clearly defined.

Some ontologies (SBML or CellML) and common toolkits (JSim or FLAME) are moving us toward Relational Grounding

Increasing use of ontologies, which map model components to meaningful referents, is evidence of a trend toward relational modeling. The trend is less obvious in how the toolkits use relational grounding. JSim's [17] automated dimensional analysis feature helps the modeler think in terms local to a given sub-model without implicitly flattening every model into which the sub-model will be installed. Consider two other JSim's abilities:

1) call out to unspecified processes; 2) build hybrid models using MML events. These features allow JSim models to move incrementally from a completely flat, absolute ground, to a partially relational grounding. FLAME's architecture of communicating X-Machines [18] allows relational mapping from one X-Machine to another facilitating a system model where any given internal component may map only indirectly to a real world referent. Modelers are biased more toward absolute or relational grounding depending on which tools and paradigms they most commonly use and the domain in which they typically work. Tools that facilitate multi-paradigm or hybrid modeling, by forcing the modeler to design a model's grounding, mitigate against such bias.

Relational Grounding Significantly Enlarges the Modeler's Toolkit

The explicit upstream consideration of how a model is grounded facilitates the composition and integration of analogs with components from conflicting models of computation (MoC). Consider: 1) JSim's use of MML events to construct hybrid (continuous and discrete) models, and 2) reformulation of methods like CA or ABM into stream X-Machines (another MoC). Both relax requirements for flat, absolutely grounded models. So doing allows each component to be (partially) defined by the other components with which it interacts. So doing forces the modeler to encapsulate model logic within interfaces, thereby making grounding more explicit and often more relational. One of the most expressive of these standardizing toolkits is Ptolemy II, which provides for automated composition of models where each component can be implemented in a different MoC. Modeling with Ptolemy II forces the modeler, up front, to explicitly consider and design how the system and its components will be grounded according to the model's use cases.

Glossary of key technical terms

absolute grounding: variables, parameters, and I/O are in real-world units like seconds and meters

agent: an object within an OO program that can schedule its own events within an analog: it is quasi-autonomous; it senses and is part of its environment; it pursues and can revise an agenda within a larger script; it is identifiable by an observer as a cause of an effect; its attributes and actions may be designed to represent biological counterparts, whereas others will deal with issues of software execution

agent-based: something formulated with or built up from agents; [in agent-based modeling] a model designed for simulation in which quasi-autonomous agents are key components

analog: anything that is analogous or similar to something else, and that exists and operates in isolation even in the absence of a referent; a system that has aspects and attributes that are similar to those of a referent system; biomimetic model implemented in software that, when executed, produces phenomena that mimic those of the model's referent

articulation: the extent to which the model consists of distinct, interconnected parts or components: the extent to which components are encapsulated and their internal dynamics are independent of those of the other components

aspect: the perspective taken when an analog is observed; one of many functional effects that result and can be observed when an analog executes

biomimetics is the study of the structure and function of biological systems as models for the design and engineering of materials and machines, in this case computational models. It is often regarded as being synonymous with biomimicry, biomimesis, biognosis and biologically inspired design.

multi-paradigm model: A model that integrates more than one type of computational framework. (cf. http://en.wikipedia.org/wiki/Multiparadigm_programming_language) For example, when a model combines an expert system with several fluid dynamics models

relational grounding: variables, parameters, and I/O are in units defined by other components of the model. For example, if one component's output is in the set {form_lumen, elongate, bifurcate, branch, form_cleft} and a receiving component accepts elements in that set as its input

synthetic analog: an analog system constructed from extant, autonomous components whose existence and purpose are independent of the model they comprise; one formed specifically by combining elements, often varied and diverse, so as to form a coherent whole

use cases: the aspects of the referent that the model intends to mimic or represent; how and for what purposes the model will be used (simulation scenarios). A component's or model's phenotype: the set of all targeted attributes.

References

- Agha G, Thati P: An algebraic theory of actors and its application to a simple object-based language ("From Object-Orientation To Formal Methods"), Lec Notes Comp Sci 2004, 2635:26-57. doi: 10.1007/978-3-540-39993-3 4.
- Actor Model [http://en.wikipedia.org/wiki/Actor_model]
- Vlachos DG: A review of multiscale analysis: examples from systems biology, materials engineering, and other fluid–surface interacting systems. Adv Chem Eng2005, 30:1-61.
- Hunt CA, Ropella GEP, Lam TN, Tang J, Kim SHJ, Engelberg JA, Sheikh-Bahaei S: At the biological modeling and simulation frontier. Pharm Res 2009, 26:2369-2400.
- Kim SHJ, Park S, Mostov K, Debnath J, Hunt CA: Computational investigation of epithelial cell dynamic phenotype in vitro. Theo Biol Med Model 2009, 6:8. 6.
- Tang J, Hunt CA: Identifying the rules of engagement enabling leukocyte rolling, activation, and adhesion. PLoS Comput Biol 2010, 6:e1000681. 7 Park S, Kim SH, Ropella GE, Roberts MS, Hunt CA: Tracing multiscale mechanisms of drug
- disposition in normal and diseased livers. J Pharmacol Exp Ther 2010, 334:124-136. Engelberg JA, Datta A, Mostov KE, Hunt CA: MDCK cystogenesis driven by cell stabilization within computational analogues. PLoS Comput Biol 2011, 7: e1002030.
- Sheikh-Bahaei S, Maher JJ, Hunt CA: Computational experiments reveal plausible mechanisms for changing patterns of hepatic zonation of xenobiotic clearance and hepatotoxicity. J Theor Biol 265:
- Fallahi-Sichani M, El-Kebir M, Marino S, Kirschner DE, Linderman JJ: Multiscale computational modeling reveals a critical role for TNF-± receptor 1 dynamics in tuberculosis granuloma formation. J Immunol. 2011, 186:3472-83.
- . Marino S, El-Kebir M, Kirschner D: A hybrid multi-compartment model of granuloma formation and T cell priming in tuberculosis. J Theor Biol. 2011, 280:50-62. 2. An G: Closing the scientific loop: bridging correlation and causality in the petaflop age. Sci Transl
- Med. 2010, 2:41ps34. . An G, Bartels J, Vodovotz Y: In Silico Augmentation of the Drug Development Pipeline: Examples
- from the study of Acute Inflammation. Drug Dev Res. 2011, 72(2):187-200. . Bassingthwaighte JB, H.J Chizeck HJ, L.E Atlas LE: Strategies and tactics in multiscale modeling of cell-to-organ systems, Proc IEEE Inst Electr Electron Eng 2006, 94:819-830.
- . Ptolemy II [http://ptolemy.eecs.berkeley.edu/] 5. Liu J, Lee E: Component-based hierarchical modeling of systems with continuous and discrete dynamics. In Proceedings of the 2000 IEEE International Symposium on Computer-Aided Control System Design, CACSD 2000, pp.95-100.
- 7. JSim [http://www.physiome.org/jsim/] B. Holcombe M, Adra S, Bicak M, Chin S, Coakley S, Graham AI, Green J, et al. Modelling complex biological systems using an agent-based approach. Integr Biol (Camb) 2012, 4(1):53-64

2012 Multiscale Modeling Consortium Meeting October 22-23, NIH