



Milking spherical cows—Yet another facet of model complexity



M. Mateus

MARETEC, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

ARTICLE INFO

Article history:

Received 25 January 2017

Accepted 3 March 2017

1. Introduction

The debate is still ongoing on what life is. Work such as the groundbreaking volume of Erwin Schrodinger's 'What is Life?' (Schrödinger, 1944) have thoroughly addressed this issue, but as with other philosophical discussion in science, it's definition is bound to be in constant mutation. Such enduring change is a corollary of the continuous progress in all the disciplines studying life's many forms and scales. Irrespective of what we know and don't know about life and how to define it, there's one unquestionable fact: life is complex. Complexity is an intrinsic and defining characteristic of all living forms, and a recurrent pattern across all living systems, from the micro to the macro scales. From the functioning and organizational levels of sub-cellular apparatus to ecosystems, every component of cells and biomes, and the relations between them, are complex. The complexity of living systems is not a concept established by scientists nor a philosophical mannerism of modelers; nor its representation in conceptual models, schemes or simple verbal or written accounts. Much has been said about the complexity of living systems, and the topic is still far from exhausted (Egel, 2012; Keto and Annala, 2012; Thingstad and Lignell, 1997), and the same holds true for the task of modeling complexity, which remains one of the key challenges of the modeling community (Anderson, 2005; García-Callejas and Araújo, 2016; Mateus, 2017; Peck, 2004; Sun et al., 2016).

Recently Cartwright et al. (2016) have addressed the issue of model complexity from a rather different and pertinent angle, by suggesting some guidelines to modelers on how to communicate model complexity to non-scientists end user. But what's so special about Cartwright et al. work? First let's remember that while most modelers struggle to accept Anderson's (2005) recommendation to have 'a healthy dose of skepticism regarding model outcomes', non-scientists seem to have no problem with having generous doses of it. So in a time when the use of models is widespread and model

results are often pivotal in the decision-making process, Cartwright et al. (2016) suggestions may as well save modelers from seeing their greatest efforts to produce realistic models converted into a helpless struggle to use them in the real world. Interestingly, their message is not so much on how to refrain models from being complex, but rather on how to facilitate the acceptance of their complexity.

However, Cartwright et al. (2016) discussion on how to communicate complexity does not relegate to the background the issue of model complexity, but emphasizes the importance of keeping the discussion open and up-to-date. The rationale may be that it is impossible to communicate model complexity if the reasons for that complexity (or lack of it) is not understood in the first place. While Cartwright et al. (2016) summarize some key points on complexity, a quick recapitulation of some critical elements are stressed here, especially for the new generation of modelers that take on available models and have limited background on the epistemological facet of modeling natural systems.

2. I think, therefore I am (a modeler)

If thinking is a philosophical proof of existence, according to Descartes, then crafting models can be seen as a consequence of thinking. The notion that life is complex is as old as man itself. As are the models developed to understand nature. Assuming that any representation of the world is in itself a model, wildlife drawings of deer, oxen, and other creatures in a cave wall can be considered the primitive models. Whether cave dwellers use those models as simple ornament (the primeval wallpaper), a feeble record of their own history, or to study animal behavior and delineate better hunting strategies, we can only speculate. But we do know they are probably the oldest representations of natural elements, and of men interaction with them. So the first models used in the study of biological systems can be traced back to the caves where our ancestors lived. Out of curiosity or necessity, sometimes both, men has been creating conceptual models ever since. With increasing levels of complexity and details, they stand among science's most notori-

E-mail address: marcos.mateus@tecnico.ulisboa.pt

ous messages, from Copernicus representation of the solar system, to Mendeleev's periodic table, to Watson and Crick's DNA double helix.

Simply put, '*a model is a logical machine for deducing the latter from the former*' (Gunawardena, 2013), so both drawings and differential equations are just a way to express the ideas on which they are based. But what's so special about them? One can say that there is nothing really special about models, apart from the fact that they provide understanding, data and means to study the world that is impossible to achieve by the simple combination of analytical methods (Anderson, 2010; Davidson, 1996; Hofmann, 1997; Peck, 2004). Models try to capture the complexity of the subject they address and mechanistic models, for example, '*can be of great value in applied science, and offer many benefits over pure empiricism*' (Nisbet et al., 2000), and provide '*the attractive capability to extrapolate and generate the bigger picture*' (Blackford et al., 2010).

Since it is impossible to fully address the complexity found in the living world, given the significant amount of limitations, starting with knowledge itself, models only capture a part of it; the most relevant aspects, modelers like to believe. And for that, they sometimes leave out details that may well be irrelevant, even though they seem necessary, ending up with functional, yet not so elegant representations of the real world, such as legless spherical cows.

3. The spherical cow

The spherical cow is a modelers' tale that gives the title to a handbook in environmental modeling (Harte, 1988). The story, used as an introduction to the volume, provides an honest critic to science by satirizing scientists in general and modelers in particular, though physicists are used as a proxy to modelers that are not mentioned by name. For the argument sake, the story is outlined here. A farmer wants to know why his cows aren't giving milk, and goes to Loomis Lab seeking for a physicist. Once there he requests a theorist to come up with a solution to solve the problem of his cows not giving milk. The theorist takes some moments to reflect and then rushes to the blackboard, draws a circle and starts his explanation: "Consider a spherical cow. . ."

As with most jokes, the humor in the spherical cow is a vehicle for some elements of truth (Austin and Chan, 2003). All ecological models fit the comparison to the spherical cow in some way, irrespective of their degree of complexity. But besides the physiognomy of the cow (shape or size) and the complexity of its physiology (the internal processes) the attention is now also focused on their product. Following the symbolism of the metaphor, modelers have to find the right way to milk their spherical cows or, as Cartwright et al. (2016) put it, they have to demonstrate that something useful can be done with their results.

4. Let's consider. . . let's assume

The nuts and bolts of models must be clearly conveyed to non-scientists, as properly pointed out by Cartwright et al. (2016), as they express model complexity. The definition of a 'model' is not the same for everyone, and can mean either a precise reproduction or a broad depiction of reality, or something in between. The same applies to the perception of detail that can range from simplistic to overly complex. But frequently it's not just a matter of definition, but a matter of expectation. To most end users a model may as well look as suspicious and unrealistic as a spherical cow to a farmer. This has been a major shortcoming attributed to models by non-scientists who are frequently unaware of the starting point of a model, namely the baseline assumptions that determine its complexity. Prepositions, conditions and assumptions are part of every analytical method, and frequently a key to explain its results.

Assumptions stand as a simplification that facilitates the approach and study of processes and systems functioning, leading to solutions that otherwise would be unreachable. But they also can stand as a fatal flaw because they can undermine that same results.

Considering a spherical cow may imply a baseline assumption that shape or legs are irrelevant. Farmers know that cows are not spherical and have functional legs, nonetheless these assumptions may be realistic if the focus of the study is on milk production. Unless, of course, there is some kind of relation between milk and shape, or if the cow needs the legs to move to greener pastures in a way that is not possible by just rolling over. But if three-dimensional movements of the ruminant in a heterogenic space covered with random patches of grass are irrelevant, the legless assumption is viable. And the same applies for shape.

If correctly handled, assumptions are an effective way to balance model complexity. Unfortunately the same assumptions that are used to keep models 'as simple as possible' can doom the modeling exercise. By assuming a spherical cow and analyzing the problem, one can never reach the result that in fact the sphere is the proper shape to best represent a cow, i.e., every prediction of the model is based on the fact that the cow is spherical, so it's not something that the model attempts to validate. Unless, of course, if the model proves to be wrong. In that case the adequateness of the shape can be questioned.

5. How much is enough?

Developing a rigorous representation of the system end-to-end is considered the '*holy grail*' for modelers (Blackford et al., 2010; Fennel, 2010; Fulton, 2010). From this we can assume that complexity is a trend in modeling natural systems. Nonetheless, a key challenge to the modelling community is how to balance increasing model complexity with the resulting increase in uncertainty associated with parameterization (Blackford et al., 2007).

Deciding how much detail (or complexity) to add in a model, and how much abstraction to accept, involves tradeoffs (Peck, 2004). This is probably a common dilemma to all modelers: what to add, what to keep and what to leave out of a model. A model is by definition a small object, usually built to scale, that represents another object that is often much bigger and more complex. In that sense, a model is always a simplified representation of reality. Following this logic, a model is not reality at a smaller scale, but rather a small part of reality; the model only addresses some of the original features of the real object. So, there is an obvious question that haunts modelers: how much is enough? The specter of complexity is a constant presence because there is not a straightforward answer for the question of how much reality we must embed in our models, so to achieve any degree of realism.

For ecological modelers, for example, this question can have a slightly different form: how many parameters do I need to describe a particular process in an ecosystem model? The question may have different answers, depending on the processes and ecosystem, but it still is a fundamental step on the evaluation of any model and also a good starting point for any model developed from scratch. Clearly, the number of parameters required in a particular model varies with the purpose of the model, that is, the intended degree of realism and expected end products. Some authors say that the challenge is to describe biological processes and interactions adequately with a minimum number of additional parameters (Geider et al., 1998). And that seems to be the rule. Unfortunately this rule does not necessarily comes with a guide for its implementation!

5.1. When more means less

It has been argued that simple models of complex systems are sometimes better than complex models of the same complex systems (Forster and Sober, 1994). On a different record Thingstad and Lignell (1997) have noted the potential of complex models to be more realistic by producing time-dependent curves of variables that fit better to the variation observed in nature or in experiments. The downside, they emphasize, is that our ability to understand the link between assumptions and results decreases with model complexity. Some disadvantages are usually associated with the increasing levels of complexity within ecosystem components, the most obvious being that more knowledge is needed to set the proper parameters for the processes (Hofmann, 1997). So, having more realism with additional complexity is not necessarily the best choice, if the validation of the model proves to be unfeasible, for example. Additional complexity should only be an option when the underlying mechanisms are accurately represented (Los and Blaas, 2010; Ward et al., 2010), and there is adequate data to calibrate and validate models.

Another clear relationship seems to exist between the usefulness of complexity and the availability of data. For complex ecological systems, simple models tend to be closer to the truth when data are limited, than complex models, as argued by Mikkelsen (2001). In contrast, when more data are available, more complex models may be better. Verisimilitude, or proximity to the truth thus seems to be a good indicator of the value of a model. But such observation implies that we know what truth is.

5.2. When less means nothing

A general observation about complexity is that complex models may not behave as the sum of their component parts (Davidson, 1996), a fact long known by the Greek philosopher Aristotle in his treatise *Metaphysics*: 'The whole is more than the sum of its parts.' So the challenge is, according to Davidson (1996), to 'produce a realistic simulation while keeping the model simple enough that one can hope to analyze and understand the reason for a particular model prediction.' The most effective way to minimize complexity in models is by reducing the number of compartments, hoping that the model still performs reasonably well. Assumptions are thus the tool to unbolt complexity, something akin of throwing out pieces of a plane in mid-flight, expecting the plane to stay aloft.

Simplicity, however, may come with limitations and an illustrative example can be easily drawn from ocean modeling. If we rely on the common and verified models using N, P and Si as currency to simulate primary production in the south pacific, then we were bound to a spectacular failure. And the reason would not be the lack of detail, but the lack of the right ecological features; while these models consider nutrient limitation for three different elements, the limitation by the micro-nutrient Fe is absent. The South Pacific region is known as high nutrients, low chlorophyll (NHLC), because of the shortage of Fe, a necessary element in the cellular machinery to process NO₃ (Dugdale and Wilkerson, 1998; Pitchford and Brindley, 1999). From this follows that sometimes a complex model is not the solution, if the right elements are missing (Edwards et al., 2004; Gervais et al., 2002; Mongin et al., 2008; Moore et al., 2004).

Minimizing complexity can be a goal, but it must be pursued with caution. There is, sometimes, an unconscious assumption of the 'survival of the fastest' principle, i.e., faster models are preferable. This somewhat common sense is an incident fallacy amongst the modeling community, especially under the pressure of end user waiting on results for multiple scenarios or long term simulations.

A model can be made simpler, but it must not become simplistic. Despite the line that separates a simple model from a simplistic models is frequently not clear, the difference between both is sig-

nificant. Simple models can be useful; simplistic models are, by definition, worthless. Modelers have the responsibility to explain this difference to non-scientists, as it is part of communicating complexity.

6. Final remarks

Model complexity will always be a topic for discussion amongst modelers; communicating their complexity will always be a challenge in the interaction between modelers and those relying in their results in the decision-making process. Cartwright et al. (2016) surely provide a good road map for modelers to go through that difficult, yet worthwhile, path. Their insightful recommendations may help to milk spherical cows, but modelers need to constantly keep in mind why their cows are spherical and legless, yet complex organisms.

End users stand as a reality check on complexity; their acceptance of models can be seen as the ultimate validation test. The outcome of this test can be determined not so much by the degree of complexity of the models, but rather on the ability of modelers to convey it to non-scientists. And that's why Cartwright et al. (2016) deserves serious attention from the modeling community.

References

- Anderson, T.R., 2005. Plankton functional type modelling: running before we can walk? *J. Plankton Res.* 27, 1073–1081, <http://dx.doi.org/10.1093/plankt/fbi076>.
- Anderson, T.R., 2010. Progress in marine ecosystem modelling and the unreasonable effectiveness of mathematics. *J. Mar. Syst.* 81, 4–11, <http://dx.doi.org/10.1016/j.jmarsys.2009.12.015>.
- Austin, R.H., Chan, S.S., 2003. Of spherical cows, cloudy crystal balls, and proteins. *Biochem. Biophys. Res. Commun.* 312, 215–221, <http://dx.doi.org/10.1016/j.bbrc.2003.09.177>.
- Blackford, J., Allen, I., de Young, B., 2007. Advances in marine ecosystem modelling research. *J. Mar. Syst.* 64, 1–2, <http://dx.doi.org/10.1016/j.jmarsys.2006.05.001>.
- Blackford, J., Allen, J.L., Anderson, T.R., Rose, K.A., 2010. Challenges for a new generation of marine ecosystem models: overview of the advances in marine ecosystem modelling research (AMEMR) symposium, 23–26 June 2008, Plymouth, UK. *J. Mar. Syst.* 81, 1–3, <http://dx.doi.org/10.1016/j.jmarsys.2009.12.018>.
- Cartwright, S.J., Bowgen, K.M., Collop, C., Hyder, K., Nabe-Nielsen, J., Stafford, R., Stillman, R.A., Thorpe, R.B., Sibly, R.M., 2016. Communicating complex ecological models to non-scientist end users. *Ecol. Modell.* 338, 51–59, <http://dx.doi.org/10.1016/j.ecolmodel.2016.07.012>.
- Davidson, K., 1996. Modelling microbial food webs. *Mar. Ecol.-Prog. Ser.* 145, 279–296, <http://dx.doi.org/10.3354/meps145279>.
- Dugdale, R.C., Wilkerson, F.P., 1998. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature* 391, 270–273.
- Edwards, A.M., Platt, T., Sathyendranath, S., 2004. The high-nutrient, low-chlorophyll regime of the ocean: limits on biomass and nitrate before and after iron enrichment. *Ecol. Modell.* 171, 103–125, <http://dx.doi.org/10.1016/j.ecolmodel.2003.06.001>.
- Egel, R., 2012. Life's order, complexity, organization, and its thermodynamic-holistic imperatives. *Life* 2, 323–363, <http://dx.doi.org/10.3390/life2040323>.
- Fennel, W., 2010. A nutrient to fish model for the example of the Baltic Sea. *J. Mar. Syst.* 81, 184–195, <http://dx.doi.org/10.1016/j.jmarsys.2009.12.007>.
- Forster, M., Sober, E., 1994. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *Br. J. Philos. Sci.* 45, 1–35.
- Fulton, E.A., 2010. Approaches to end-to-end ecosystem models. *J. Mar. Syst.* 81, 171–183, <http://dx.doi.org/10.1016/j.jmarsys.2009.12.012>.
- García-Callejas, D., Araújo, M.B., 2016. The effects of model and data complexity on predictions from species distributions models. *Ecol. Modell.* 326, 4–12, <http://dx.doi.org/10.1016/j.ecolmodel.2015.06.002>.
- Geider, R.J., MacIntyre, H.L., Kana, T.M., 1998. A dynamic regulatory model of phytoplankton acclimation to light, nutrients, and temperature. *Limnol. Oceanogr.* 43, 679–694.
- Gervais, F., Riebesell, U., Gorbunov, M.Y., 2002. Changes in primary productivity and chlorophyll a in response to iron fertilization in the Southern Polar Frontal Zone. *Limnol. Oceanogr.* 47, 1324–1335, <http://dx.doi.org/10.4319/lo.2002.47.5.1324>.
- Gunawardena, J., 2013. Biology is more theoretical than physics. *Mol. Biol. Cell* 24, <http://dx.doi.org/10.1091/mbc.e12-03-0227>.
- Harte, J., 1988. *Consider A Spherical Cow – A Course in Environmental Problem Solving*. University Science Books, Sausalito, CA, 283 pp.
- Hofmann, E., 1997. Models for marine ecosystems. In: Tuljapurkar, S., Caswell, H. (Eds.), *Structured-Population Models in Marine, Terrestrial, and Freshwater Systems*, vol. 18. Springer, US, pp. 409–432.

- Keto, J., Annala, A., 2012. The capricious character of nature. *Life* 2, 165–169, <http://dx.doi.org/10.3390/life2010165>.
- Los, F.J., Blaas, M., 2010. Complexity, accuracy and practical applicability of different biogeochemical model versions. *J. Mar. Syst.* 81, 44–74, <http://dx.doi.org/10.1016/j.jmarsys.2009.12.011>.
- Mateus, M.D., 2017. Bridging the gap between knowing and modeling viruses in marine systems—an upcoming frontier. *Front. Mar. Sci.* 3, <http://dx.doi.org/10.3389/fmars.2016.00284>.
- Mikkelsen, G., 2001. Complexity and verisimilitude: realism for ecology. *Biol. Philos.* 16, 533–546, <http://dx.doi.org/10.1023/a:1011905829922>.
- Mongin, M., Molina, E., Trull, T.W., 2008. Seasonality and scale of the Kerguelen plateau phytoplankton bloom: a remote sensing and modeling analysis of the influence of natural iron fertilization in the Southern Ocean. *Deep Sea Res. Part II* 55, 880–892, <http://dx.doi.org/10.1016/j.dsr2.2007.12.039>.
- Moore, J.K., Doney, S.C., Lindsay, K., 2004. Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochem. Cycles* 18, <http://dx.doi.org/10.1029/2004gb002220>, n/a–n/a.
- Nisbet, R.M., Muller, E.B., Lika, K., Kooijman, S.A.L.M., 2000. From molecules to ecosystems through dynamic energy budget models. *J. Anim. Ecol.* 69, 913–926, <http://dx.doi.org/10.1111/j.1365-2656.2000.00448.x>.
- Peck, S.L., 2004. Simulation as experiment: a philosophical reassessment for biological modeling. *Trends Ecol. Evol.* 19, 530–534, <http://dx.doi.org/10.1016/j.tree.2004.07.019>.
- Pitchford, J.W., Brindley, J., 1999. Iron limitation, grazing pressure and oceanic high nutrient-low chlorophyll (HNLC) regions. *J. Plankton Res.* 21, 525–547, <http://dx.doi.org/10.1093/plankt/21.3.525>.
- Schrödinger, E., 1944. *What Is Life?: The Physical Aspect of the Living Cell*. The University Press, Cambridge, 91 pp.
- Sun, Z., Lorscheid, I., Millington, J.D., Lauf, S., Magliocca, N.R., Groeneveld, J., Balbi, S., Nolzen, H., Müller, B., Schulze, J., Buchmann, C.M., 2016. Simple or complicated agent-based models? A complicated issue. *Environ. Modell. Software* 86, 56–67, <http://dx.doi.org/10.1016/j.envsoft.2016.09.006>.
- Thingstad, T.F., Lignell, R., 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* 13, 19–27, <http://dx.doi.org/10.3354/ame013019>.
- Ward, B.A., Friedrichs, M.A.M., Anderson, T.R., Oschlies, A., 2010. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *J. Mar. Syst.* 81, 34–43, <http://dx.doi.org/10.1016/j.jmarsys.2009.12.005>.