

# Why the outputs of models should be interpreted as predictions

**F. Boschetti**<sup>a,c</sup> and **J. Symons**<sup>b</sup>

<sup>a</sup> CSIRO CMAR, Email: [Fabio.Boschetti@csiro.au](mailto:Fabio.Boschetti@csiro.au)

<sup>b</sup> Department of Philosophy, University of Texas at El Paso, El Paso, Texas 79968, USA

<sup>c</sup> School of Earth and Geographical Sciences, The University of Western Australia

**Abstract:** Physicists and engineers are accustomed to seeing the output of a computational model as representing a prediction. In this context, predictions are understood to be informed approximations of how a system's dynamics may evolve as a result of the modelled assumptions. In what is generally called inverse modelling or optimization, applied mathematicians employ the ability of models to reproduce observed behaviours in order to deduce causal relations or to retrodict the system's past history. However, in the modelling of ecological and social processes, scientists (including modellers themselves) tend to be much more pessimistic concerning the ability of models to provide reliable predictions. Because of this scepticism prediction is sometimes explicitly excluded from the list of useful model purposes (1). The four principle reasons for this scepticism are: a) computational models have a very poor prediction track record; b) most model predictions are not testable because of their conditional nature; c) despite the appearance of objectivity, model outcomes reflect the modelers' subjective beliefs and assumptions and d) the view that some scientific activities, including computational modelling of social and ecological phenomena are not designed, and should be expected to, provide predictions.

While we acknowledge that the modelling of physical vs ecological and human processes are different (2-4), we suggest that the output of any type of model which is employed as part of a decision-making process should be interpreted as a prediction. Our claim is based on the following reasoning. First, prediction should not be understood as a forecast of a precise event, instead it should be understood as an estimation of a probability distribution which provides bounds on the likelihood of sets of future events. Second, such predictions are indispensable for decision making, since they provide the basis upon which available alternative options are evaluated and chosen. Third, the predictive power of experts is known to be less reliable in certain contexts, than numerical models. Given that prediction is necessary for non-arbitrary decision making, it is useful to refocusing the question from whether models provide an accurate prediction to whether computational model can outperforms humans as predictors. Fourth, the ability to compare prediction vs observation is at the core of the scientific method and dismissing the predictive capacity of models prevents blocks the possibility of assessing the relative scientific merits of distinct models. Fifth, disregarding the predictive power of computational models prevents their use in inverse modelling. Other familiar uses of computational modelling, like deducing causal relations and past system behaviour are also logically denied. Sixth, further commonly recognized purposes of numerical modelling, like leaning and communication, also rely on the ability of models to predict within acceptable limits: it is pointless to learn from a mistaken teacher. Seventh, acknowledging the role of prediction in assessing a model's scientific validity and its impact on decision making forces modellers to accept the responsibility of providing all the necessary details of the model so that the reliability of such prediction can be estimated.

Below we expand on these points. In an attempt to reconcile different views on models, predictions and their merit in decision making, we conclude by providing an alternative interpretation of computational models, according to which models can be understood as an extension of native cognitive capacities.

**Keywords:** *Modelling, Philosophy of science, Predictability, Forecasting, Decision making*

## **1. ABOUT PREDICTION**

Three concepts are fundamental to our discussion. First, for prediction we do not mean the specification of all the properties of some future behaviour or event. Instead, prediction involves estimation of a range of possible states. In other words, prediction should not be understood as prophecy. For example, while it is widely known that weather forecasts are not reliable past 5-6 days, no one would believe that the temperature in Darwin in summer could be 40°C or -40°C with equal probability; as a result no rational person would travel to Darwin in the summer with a heavy winter coat. The limited predictability past 5-6 days still allows a certain level of effective planning. Second, predictions are conditional (4): any prediction is carried out within a context and is valid only within that context. In the above example, the conditioning is given by our understanding of tropical climate; should this change, the prediction would no longer hold and would require updating. Finally, the effectiveness of a prediction is scale-dependent (5). For example, while the geophysical community is today sceptical about its ability to provide accurate prediction on where and when large earthquakes can occur, they are nevertheless able to predict the broad geographical areas in which large earthquakes can be expected. While this kind of predictability seems to offer little to planning (6), it still has considerable practical impact in deciding, for example, in which geographical areas expensive anti-seismic constructions methods are necessary and where they are not.

Once understood in these terms, prediction becomes an integral part of any decision making process: formulating a plan implies choosing among potential alternatives and envisaging (=predicting) which one is more likely to deliver desired outcomes.

## **2. WHY IS PREDICTION NECESSARY**

Much criticism of the value of computer models as predictive tools is based on the assumption that a prediction is a desirable, but not a necessary result of scientific research. More specifically, the core claim is that prediction is an ideal or discretionary input to, rather than a requirement for, decision making. Our view is that prediction is an essential component to any planning and that it is implicit even in approaches which claim not to require it. When understood as described in the previous section, prediction becomes an integral part of any non-arbitrary decision making process, whether formally at an organisational level or informally at an individual level. Prediction has a role, implicitly or explicitly, in formulating a plan and assessing which avenues should be followed. Formulating a plan implies choosing among potential alternatives and envisaging (=predicting) which one is more likely to deliver desired outcomes. Similarly, a decision will inevitably involve commitment to the exclusion of a range of alternative possible futures. Even the most straightforward decisions exclude a vast range of irrelevant or unlikely alternatives. This exclusion involves judgments with respect to future states of affairs.

The same applies to the implementing a plan: effective planning requires some sort of expectation of a system's future behaviour and of the outcome of available actions. Even when facing a very complex problem, we are rarely in a state of complete ignorance: some expectation on system behaviour and on the level of risk arising from uncertainty is usually available and it is on this expectation that most decisions are taken. In everyday life, humans use models, which are mental or formal representations of reality, to generate these expectations.

If we were to accept that a prediction is essential to any decision making, then the crucial question shifts from 'can model predictions be trusted?' to 'how do model compare to other approaches to prediction?'

## **3. HUMAN VS MODEL PREDICTION**

We have shifted our attention to the core question of what tools provide the most reliable prediction given the problem at hand. Notice that this question is problem dependent, not only because different problems may require different approaches, but also because the most accurate prediction is not necessarily the most reliable. Together with using numerical models or other computational tools, predictions can be provided by experts, local knowledge, or participatory settings. It is thus important to compare the predictive performance of models against alternative approaches on the core items of criticisms discussed above: a) prediction track record; b) lack of testability due to their conditional nature and c) inherent subjectivity.

We are aware of only one study in which a comparisons of predictive accuracy of model vs alternative methods has been carried out (7). In this study, on average, statistical methods outperformed experts. Only a very small number of experts were able to provide reliable predictions. Importantly, their ability seems to stem from reasoning style and attitude, rather than mere knowledge and technical ability. The generally poor performance of experts as predictors is confirmed by the available literature on the logical and attitudinal

fallacies which even experts display for simple dynamical problems. This should warn us that it is probably unwarranted to expect humans to mentally predict the behaviour of highly complex systems in a consistent and reliable manner (8-15).

On the other hands, some evidence suggests that elders show a marked improvement in addressing problems as complex as social dilemma compared to younger (higher IQ) groups (16). This suggests that some skills can be learned even informally via life experience and naturally acquired wisdom, especially in regards to dynamical processes which are more closely related to social and interactional factors.

The core question underlying this work ('how do models compare to other approaches to prediction?) is ultimately an empirical one. The evidence from the cognitive psychology and the modelling literature, according to which models can help overcome some of the cognitive limitations which hamper human ability to predict the behaviour of complex dynamic systems (17, 18), should be tested experimentally on a large scale.

Human beings invest considerable effort in predicting the future and studying the past, but considerably less in comparing the two: in the rare cases in which predictions are validated against actual occurrences it is difficult to draw clear conclusions, given that it is not obvious what criteria should be employed to compare complex, chaotic dynamics. This suggests a promising direction of research, including for historical analysis and laboratory experiments.

#### 4. WHY A MODEL SHOULD BE BETTER AT PREDICTING THAN AT EXPLAINING

One counterintuitive result of our discussion is our claim that prediction in computational models is more reliable than retrodiction. On reflection, this is a straightforward result of the nature of computational models. However, recognizing this fact should cause us to think carefully about the explanatory value of the kinds of retrodictive accounts of complex systems which we derive from computational models.

To simplify our discussion, let's assume we have a model  $M$  and that the model is 'structural' in the sense that it represents our understanding of the dynamics of a process  $P$  (extension of the discussion to non structural models can be addressed by following the line of argument in (19)).

Models can be used in forward or inverse mode. Let's first consider the forward mode. At time  $t_0$  we collect some data  $D_{t_0}$  about  $P$  and we use  $M$  to assess what may happen in the future at time  $t_n$ . This model use is called forward since it respects our perception of the 'arrow of time' (causes lead to effects) and leads to a prediction of  $D_{t_n}$ . Alternatively, we may wish to assess what may have happened in the past at time  $t_n$ . Since it attempts to reverse the arrow of time, this use of modeling is often referred to as 'inverse' modeling (20, 21) and leads to a retrodiction of  $D_{t_n}$ .

In an ideal case, the retrodiction would be carried out with an inverse model  $M^{-1}$  such that  $M^{-1}(out)=in$ , where  $in$  refers to the input and  $out=M(in)$  to the output of  $M$ , respectively. Unfortunately, inverse models such as  $M^{-1}$  can be written explicitly for only a very small set of forward models  $M$ : most inverse engineering and scientific problems need to be solved by iterative methods in which  $M$  is run with sets of inputs  $in$  until a reasonable match between  $M(in)$  and the expected output,  $D_{t_0}$  in our case, is found. The procedure which allows us to recover  $D_{t_n}$  from  $M$  and  $D_{t_0}$  is called inversion, optimisation, or regression, depending on the discipline (20, 21). Here we will call it inverse modeling and will call this procedure  $M^{inv}$ .

Assessing the effectiveness of a computer model in predicting or retrodicting can thus be cast in terms of the reliability of the two processes  $M$  (prediction) and  $M^{inv}$  (retrodiction). By 'reliable', we mean the following: given the model output of the forward process (prediction,  $M_{t_n}$ ) and of the inverse process (retrodiction,  $M_{t_n}^{inv}$ ), we ask which one is likely to be closer to the states of the 'real' process  $D_{t_n}$  and  $D_{t_n}$ , respectively.

Let's begin by considering deterministic models. At time  $t_0$  we make a set of observations  $D_{t_0}$  and we use this information to parameterize our model  $M$ . The 'real' process  $P$  proceeds and at time  $t_n$  we make a new set of observations  $D_{t_n}$ . Because the model  $M$  is not exact, the prediction of  $M$  at time  $t_n$   $M_{t_n} \neq D_{t_n}$  and we call the prediction error  $E_p = |M_{t_n}, D_{t_n}|$ . Other runs of  $M$ , starting with different initial conditions in  $S_{t_0}$  ( $D_{t_0} \neq D_{t_0}$ ) generate different predictions in  $S_{t_n}$ .

Similarly, the inverse process  $M^{inv}$  allows us to retrodict from  $D_{t_0}$  in order to recover  $D_{t_n}$ . As explained above, this is an inverse process carried out by iteratively mapping  $S_{t_n}$  into  $S_{t_0}$  via  $M$ , until a satisfactory

match  $|M_{t_0}, D_{t_0}|$  is found. Ideally,  $M$  should map  $D_{t_{-n}}$  into  $D_{t_0}$  so that  $E_p = |M_{t_0}, D_{t_0}| = 0$ . Of course we cannot expect this match to be exact. The same approximations (or errors) which prevent  $M$  from modeling  $P$  exactly, and which are responsible for  $E_p = |M_{t_n}, D_{t_n}| \neq 0$ , are likely to imply  $E_p = |M_{t_0}, D_{t_0}| \neq 0$ . As a consequence, it is likely that a point  $D'_{t_{-n}}$  in  $S_{t_{-n}}$  ( $D'_{t_{-n}} \neq D_{t_{-n}}$ ) may generate a prediction at time  $t_0$  for which  $|M(D'_{t_{-n}}), D_{t_0}| < |M(D_{t_{-n}}), D_{t_0}|$ . The point  $D'_{t_{-n}}$  for which  $|M(D'_{t_{-n}}), D_{t_0}|$  is minimum will be chosen as retrodiction. The error in the retrodiction will then be  $E_r = |M(D'_{t_{-n}}), D_{t_0}| \neq 0$ . If  $M$  is non linear and 'complex', the magnitude of  $E_r$  and  $E_p$  may vary considerably as a function of the location of the parameterization in  $S_{t_{-n}}$  and  $S_{t_0}$ , but we have no a-priori reason to expect  $E_r < E_p$ . This is the crucial message of this work and we will address it again below. At this point, it is important to emphasize that  $E_r$  arises from the same process which generates  $E_p$  and that the relative magnitude of  $E_r$  and  $E_p$  cannot be deduced a-priori.

Two further problems, which affect any real world modeling exercise, complicate the inverse modeling  $M^{inv}$ : a)  $M^{inv}$  does not necessarily have a unique answer and b)  $M^{inv}$  can be computationally very expensive.

'Non uniqueness' or 'equifinality' is the property of a system wherein, under certain conditions, families of input parameters can produce the same model output. Non-uniqueness is a feature of inverse modeling but not of forward modeling, the outcome of which under ordinary circumstances is deterministic. Non-uniqueness affects retrodiction: even if an exact match  $|M_{t_0}, D_{t_0}| = 0$  can be achieved (that is even if the model allows to match the current observations perfectly), we are unable to differentiate among the (potentially infinite) number of solutions  $D'_{t_{-n}}$  which provide the match.

The previous argument can easily be extended to non-deterministic models. Non-determinism in  $M$  implies that the output  $M_{t_n}$ , obtained by running  $M$  initialized with  $D_{t_0}$ , is not unique. A non-deterministic outcome is obtained by using random perturbation in the input parameters, thereby generating artificial non-uniqueness in the forward modeling. As a result, the set of input  $D'_{t_{-n}}$  in  $S_{t_{-n}}$  leading to acceptable measures of  $|M_{t_0}, D_{t_0}|$  adds to the uncertainty resulting from non-uniqueness.

Finally, as we mentioned above,  $M$  can be computationally very expensive. In many real world applications, this implies that  $M$  cannot be run as many times as the iterative process  $M^{inv}$  would require. This, in turns, adds further errors to  $|M_{t_0}, D_{t_0}|$  and, as a result, potential further errors in  $E_r$ .

If we accept that  $E_p$  represents the error in prediction and  $E_r$  the error in retrodiction, the previous analysis suggests that a)  $E_r$  is inextricably related to  $E_p$ , b) there is no reason to assume that in general  $E_p > E_r$  and c) in practice, it is more likely that  $E_r > E_p$  as a result of non-uniqueness and the computational effort which may prevent the inverse process  $M^{inv}$  to run to completion. This leads to the unintuitive conclusion according to which, in the absence of additional information, we should trust a model prediction more than a model retrodiction.

We can rephrase the conclusions of this section. If we believe we should not trust a model to provide a reliable prediction, we should be even less willing to trust that model to give us a reliable explanation of past events or causal paths. This conclusion is widely accepted in engineering and applied mathematics: no practitioner would trust the result of an inverse modelling exercise carried out via an unreliable forward model. Curiously, in complex model applications, including ecological and social modelling, the opposite is believed, practitioners rely in the interpretation of past events obtained via the use of model for which they do not trust the forecast.

## 5. MODELS AS LEARNING AND COMMUNICATION TOOLS

Models can be used as learning tools. For example, modellers learn by reflecting on mapping relations between input parameters and outputs and by carrying out a sensitivity analysis of their models based on their experience and presumably via something like an unconscious inverse exercise. While experience and expertise are important, the accuracy of a researcher's deliberations, as we have seen in the previous section, depends crucially on the reliability of the forward model.

Not all learning that results from using computer models needs be so formal. Let's take an analogy often employed to explain the role of numerical models in complex processes: the flight simulator. Complex socio-ecological models, for example, offer decision makers the same opportunity offered by flight-simulators to

trainee pilots: they provide the opportunity to test policy initiatives in the safe world of virtual simulations. However, it is reasonable to expect that flight-simulators will provide effective training only in so far as the flight-simulator simulates well, that is, only in so far the flight-simulator effectively predicts how the real plane will behave under similar circumstances. We have no reason to believe that a pilot trained on an inaccurate flight simulation should learn how to handle a real plane in the real world. Similarly, there is no reason to believe that a decision maker should improve his/her ability to address a real world using a model which provides poor predictions with respect to the operation of the real world under the conditions embodied in the run of the model.

Some researchers have emphasized other roles for computational models, including communication between model developers, users, decision makers and the general public. However, just like the user of a poor flight-simulator model may misjudge the behaviour of a real plane and communicate this misjudged behaviour to others, there is no reason why useful, purposeful and pertinent communication obtained via the use of a model can happen which is independent of the model's reliability as predictor.

## **6. MODELLER RESPONSIBILITY**

Since computational models of complex systems are currently used to support decision making in a number of real world problems at the intersection of technical, economic, ecological and social issues, a considerable responsibility is placed on the modelling and complex system science community, since its work can impact policy making and thus the life of millions. Modellers themselves often eschew such responsibility by denying their models can provide reliable prediction, rather focusing on providing non committal descriptions of trends, scenario outcomes or relations between model conditioning and outputs.

Accepting the responsibility that a model provides a prediction, in our opinion, imposes a certain discipline both in the modelling exercise and in the delivery of the results, since it enforces an avenue for falsification and accountability. Both are inevitable steps in a scientific endeavour and too often modellers fall into the temptation to circumvent this responsibility.

## **7. CONCLUSIONS: AN ALTERNATIVE VIEW OF COMPUTATIONAL MODELS**

The foregoing discussion of the role and purpose of models may not match our intuition according to which highly complex processes are extremely difficult to understand. Also, our experience tells us that complex dynamics often appears to be controlled by surprises, rather than regularities. This has led many authors to claim that the use of computer modelling to study and predict complex processes is unwarranted. The authors themselves are well aware of the limitation of computer modelling in reproducing complex, novel and emergent processes (2, 3, 22-24).

Here we propose a way of reconciling this apparent contradiction. Numerical models can be seen as a formal implementation of the very mental models humans routinely use to represent reality. Seen in this perspective, a computational model is not an attempt to reproduce the dynamics of a physical, ecological or human system within a computer. Instead, it is a way to evaluate the consequences of our mental representations of such dynamics. Rather than highlighting the inadequacy of a numerical model in simulating reality, this view highlights the benefits inherent in carrying out such simulation in a computer rather than in our head: fast computation, check for consistency, circumventing of known human fallacies, explicit formalization of assumptions and unbiased presentation of results. This idea has been discussed within the application to climate change modeling in (25) and will be further developed in future work. In philosophy of science, this view has been defended by Paul Humphreys (26) and others. Humphreys argued that computational models are best understood as extensions of our native cognitive capacities. We are in broad agreement with this view and will develop the details further in future work.

## **REFERENCES**

1. Epstein J (2008) Why Model? *Journal of Artificial Societies and Social Simulation* 11(4):12.
2. Boschetti F & Gray R (2007) A Turing test for Emergence. *Advances in Applied Self-organizing Systems*, ed Prokopenko M (Springer-Verlag, London), pp 349-364.

3. Boschetti F & Gray R (2007) Emergence and Computability. *Emergence: Complexity and Organization* 9(1-2):120-130.
4. Boschetti F, Grigg NJ, & Enting I (2011) Modelling = conditional prediction. *Ecological Complexity* 8(1):86-91.
5. Israeli N & Goldenfeld N (2004) Computational Irreducibility and the Predictability of Complex Physical Systems. *Physical Review Letters* 92(7):074105-074101-074105-074104.
6. Pielke RA (2003) The role of models in prediction for decision. *Models in Ecosystem Science*, eds Canham CD, Cole JJ, & Lauenroth WK (Princeton University Press, Princeton and Oxford), pp 111-135.
7. Tetlock PE (2005) *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton University Press New Jersey).
8. Moxnes E (1998) Overexploitation of renewable resources: the role of misperceptions. *Journal of Economic Behavior and Organization* 37(1):107-127.
9. Moxnes E (2000) Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development. *System Dynamics Review* 16(4):325-348.
10. Moxnes E & Saisel AK (2009) Misperceptions of global climate change: information policies. *Climatic Change* 93:15-37.
11. Stanovich K (1999) *Who is rational? : studies of individual differences in reasoning* (Lawrence Erlbaum Associates).
12. Sterman JD & Sweeney LB (2002) Cloudy skies: assessing public understanding of global warming. (Translated from English) *System Dynamics Review* 18(2):207-240 (in English).
13. Sterman JD & Sweeney LB (2007) Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter. *Climatic Change* 80(3-4):213-238.
14. Tversky A & Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185:1124-1131.
15. Tversky A & Kahneman D (1983) Extensional versus intuitive reasoning: The conjunctive fallacy in probability judgment. *Psychological Review* 90:293-315.
16. Grossmann I, *et al.* (Reasoning about social conflicts improves into old age.).
17. Boschetti F, Hardy P, Grigg N, & Horwitz P (2011) Can we learn how complex systems work? *Emergence: Complexity and Organization* In Print.
18. Dorner D (1996) *The Logic Of Failure: Recognizing And Avoiding Error In Complex Situations* (Metropolitan Books, Ney York).
19. Suchting WA (1967) Deductive Explanation and Prediction Revisited. *Philosophy of Science* 34(1):41-52.
20. Parker R (1977) Understanding inverse theory. *Annual Review of Earth and Planetary Sciences* 5:35-64.
21. Tarantola A (1987) *Inverse Problem Theory* (Elsevier, Amsterdam).
22. Boschetti F (2010) Causality, emergence, computation and unreasonable expectations. *Synthese* 10.1007/s11229-010-9720-8.
23. Symons J (2008) Computational Models of Emergent Properties. *Minds and Machines* 18(4):475-491.
24. Symons J (Forecoming) Natural Ways to Emergence. *Systematic Thinking: Hintikka, Neville and The Future of Philosophy*, ed Anderson REAaDR).
25. Boschetti F (2011) A simple model of how people reason about climate change Submitted.
26. Humphreys P (2004) *Extending Ourselves Computational Science, Empiricism, and Scientific Method* (Oxford University Press, Oxford).