Simple is Beautiful

John Doherty and Catherine Moore

Preface

This short text summarizes a much larger document delivered through the "Smart Models for Aquifer Management" project undertaken by GNS, New Zealand. That document, Doherty and Moore (2017), provides a theoretical basis for the discussion that is presented herein. It can be downloaded from:

https://www.gns.cri.nz/Home/Our-Science/Environment-and-Materials/Groundwater/Research-Programmes/Smart-Aquifer-Models-for-Aquifer-Management-SAM/SAM-discussion-paper

Introduction

We commence with an apology to E.F. Schumacher, the author of the much acclaimed book "Small is Beautiful". The book's subtitle is "a study of economics as if people mattered". To cement our plagiarism we should subtitle the present text "a study of environmental modelling as if decisions mattered".

The subject of this paper is the construction, calibration and deployment of models that are built to support environmental and resource management. Nowadays, it is common practice to build models for this purpose. We submit that many of these models fail to deliver the support which they promise because they are too complex and insufficiently agile to use in the decision-making process. But first we discuss the role that models can play as an environmental decision support tool, and then we explain why they so often fail to provide that support.

Models and Decisions

The contribution that numerical modelling should make to the environmental decision-making process is succinctly described by Freeze et al (1990). Though often cited, the arguments presented in this landmark paper are mostly ignored by the modelling community. Freeze et al (1990) point out that modelling introduces to the decision-making process the vital ingredient of risk. Risk can be roughly equated to the probability of a bad thing happening as a consequence of a particular decision multiplied by the cost associated with its occurrence. It follows immediately that if a model is deployed to support environmental decision-making, its predictions of environmental behaviour under different management options must be accompanied by estimates of the uncertainties associated with those predictions.

Moore and Doherty (2005) show that model predictions of groundwater behaviour can be accompanied by large uncertainties, even where these predictions are made by a "calibrated model" i.e. by a model whose parameters have been adjusted to ensure that its outputs match historical measurements of system state (for example heads in wells and flows in streams). They show that, to the extent that a prediction is sensitive to components of the calibration null space, the uncertainty of that prediction is not reduced through the calibration process at all. (The "calibration null space" refers to model parameters, and/or combinations of parameters, which are uninformed by historical measurements that comprise the calibration dataset.) Such predictions, therefore, tend to be those that are sensitive to hydraulic property detail and hydraulic property variability that exists at small spatial scales. They include (but are not limited to) movement of a contaminant through a heterogeneous subsurface, nuances of groundwater/surface water interaction, and the response of a natural system to extreme climatic events, and/or to large changes in its management regime. Unfortunately, many decision-critical predictions fall into one or more of these categories.

Doherty and Simmons (2013) and Doherty and Vogwill (2015) extend the concepts introduced by Freeze et al (1990) to define failure of a modelling enterprise in lending support to the decision-making process. Based on the premise that decisions are taken to avoid the occurrence of an unwanted outcome (i.e. a "bad thing"), a model fails in its decision-support role if its predictive uncertainty margins underestimate the probability of occurrence of that bad thing.

Examined from a frequentist perspective, decision-making requires exploration of the hypothesis that a bad thing will happen if a certain management strategy is adopted. (This strategy can include so-called "adaptive management", where mitigation actions are prescribed if certain monitoring thresholds are crossed.) Ideally, an environmental model should be employed to implement the scientific method, whereby rejection of the "bad thing hypothesis" is attempted by demonstrating incompatibility of its occurrence with information about the system that is encapsulated in the model. This information is comprised of expert knowledge of system properties and processes on the one hand, and measurements of the historical behaviour of the system on the other hand. A type 2 statistical error occurs if the hypothesis of a bad thing happening is falsely rejected; we define this as failure.

Armed with a definition of failure, a firm conceptual reference point exists for setting specifications for any model that is built to support environmental decision-making. In the present discussion, specifications of most interest are those pertaining to the level of model complexity.

Another criterion of relevance to model design is that of usefulness. If the uncertainty bounds calculated by a model for a decision-critical prediction are too broad, then the bad thing hypothesis can never be rejected. Failure to reject a false hypothesis is referred to as a type 1 statistical error. A model's usefulness increases to the extent that it can reduce the uncertainty of a decision-critical prediction to the lower limit set by availability of current information pertaining to that prediction.

Models as Receptacles for Information

It is apparent from the above discussion that a model's support for the decision-making process rests on its ability to provide receptacles, or containers, for information about the system that is undergoing management. As stated above, this information falls into two broad categories, namely expert knowledge on the one hand, and the historical behaviour of the system on the other hand. Expert knowledge is expressed through the construction details of the model, its boundary conditions, and its parameterization (i.e. how hydraulic properties of the simulated system are represented in the model). Conceptually, all of these (especially its parameterization) must be stochastic, as that is the nature of expert knowledge as it pertains to environmental systems. Information that is encapsulated in the historical behaviour of the system is introduced to the model through the history-matching process. This information constrains parameters to a narrower stochastic range than that based on expert knowledge alone. This is because parameters employed by a model must be such as to allow the model to replicate that behaviour. These two types of information constitute the prior probability and likelihood terms that appear on the right side of Bayes equation; see any statistical text for more details.

Bayes equation states that the outcomes of the history-matching process are probabilistic, this pertaining to parameters employed by a model, and to predictions made by a model. At first glance it would appear that the concept of the "calibrated model" which forms the basis for widespread, model-based decision-making is at odds with this fundamental premise. This is indeed the case. However, if model calibration is considered as the first step in a two-step process of:

- finding a solution of minimum error variance to the inverse problem posed by fitting model outputs to the calibration dataset, and then
- quantifying the error in predictions made by the thus calibrated model,

then the precepts of Bayes equation can be followed in a way that is far easier to implement than Bayes equation itself, with outcomes that reduce the probability of occurrence of a type 2 statistical error to an acceptable level. See Doherty (2015) for further details.

Instead of being considered as devices through which the scientific method can be introduced to the decision-making context, models are often construed (especially by non-modellers who pay for them) as (presumably accurate) simulators of environmental behaviour at a particular study site. However even the most complex model can only be approximate in its attempted simulation of local environmental processes. Nevertheless, the ability to simulate, even in an approximate fashion, processes that are operative at a location of interest endows a model with the information receptacles that it requires to fulfil its decision-support role. Importantly, it is these receptacles which are of primary importance. In contrast, exact (and elusive) replication of environmental processes is of secondary importance, for the role of simulation is to serve the greater task of providing receptacles for information that resides in local knowledge of the system and in measurements of the behaviour of that system. Furthermore, exact replication of environmental behaviour may not even be required if the occurrence of a bad thing can be relegated to a low level of probability using only a fraction of the information that is available at a particular study site.

Complex Models

Presumably, the more complex is a model, the greater is the amount of information for which it can provide receptacles.

Complex models provide more receptacles for expert knowledge because of their "realistic" representation of real-world conditions, and because of their "physically-based" simulation of real-world processes. Local measurements of system attributes, and of hydraulic properties, can thus be transferred directly to the model. These can then exert constraints on properties assigned to parts of the model domain where no such measurements have been made, using correlations that have emerged from local site characterization.

Complex models provide receptacles for information encapsulated in historical measurements of system statebecause they can support many adjustable parameters, and because these parameters embody realistic descriptions of system heterogeneity. It follows that a good fit between model outputs and historical measurements should be readily achievable. This does not, of course, result in parameter uniqueness. However, it does result in tighter constraints on parameter variability than that which is permitted on the basis of expert knowledge alone.

In theory, environmental decision-making is therefore well supported by a complex model. In practice, this is rarely the case. Complex models have long run times. Often they are plagued by numerical instability. Furthermore, the more detail that a model can express, the greater is the necessity for that detail to be expressed stochastically (i.e. probabilistically). Stochasticity is required because neither direct measurements of system properties at a discrete number of locations, nor inferences of system properties made through history matching, can yield unique estimates of that detail throughout a modelled area. Unfortunately, however, the long run times of complex models preclude running the model the number of times required for proper stochastic analysis. These same run times, accompanied by a penchant for numerical instability, often make the task of history-matching very difficult indeed – if not impossible. A simplistic parameterisation scheme is often therefore draped over the domain of a complex model to ease the burden of history matching. The model then becomes simple again – losing the benefits of complexity while eschewing the quick run times that model simplicity can bring.

It is often the complex model itself, rather than the support that it must provide to management decisions, that constitutes the deliverable of a model construction exercise. This is based on the premise that a complex model can be built to support the making of many different decisions. To avoid failure (as defined above) of the time-consuming and expensive model-construction process, the model that emerges from this process must be capable of quantifying the uncertainties of many different predictions. Furthermore, for each of these predictions a guarantee must be provided that a type 2 statistical error has been avoided. Meanwhile model usefulness (as defined above) requires that predictive uncertainties be reduced to a level that is commensurate with all available information. This requires inclusion in the model of all parameters that <u>can be</u> informed by the calibration dataset. Meanwhile, avoidance of failure requires inclusion in the model of all parameters that <u>are not</u> informed by the calibration dataset but to which one or more predictions of management interest may be sensitive. Both sets of parameters must be capable of variation under calibration constraints, to support exploration of the uncertainties of all predictions that the complex model may be required to make. Where run times are high and where model numerical stability is questionable, this is not possible.

Meanwhile, the personal difficulties faced by a modeller who has been engaged to build a complex model are considerable. With his/her focus on the model as a deliverable, the modelling process descends into that of making the model “look good” by adding ever-increasing amounts of non-stochastic detail to the model, to forestall accusations of its inadequacy as an “accurate simulator” of reality. The decisions that the model must support are forgotten as the modeller nurses the model to a point at which it can be delivered to a manager or client, with the hope that the latter will not use it in ways that expose its numerical deficiencies.

Simple Models

Simple models can be built and calibrated more quickly than complex models. However, they must be used with caution. As is discussed extensively by Doherty and Moore (2017), the design, calibration and deployment of a simple model must be tailored to one or a small number of specific predictions to ensure its decision-support role is not corrupted. This is not necessarily a bad thing. The relationship between a simple model and the decision that it is built to support is thus clearly defined, right from the start of the model construction process. Furthermore, in deploying the simple model in support of the decision-making process, the modeller becomes an integral part of that process. The simple model is thus seen as an agent for better decision-making rather than as an end in itself.

In general, simple models do not provide good receptacles for the information contained in expert knowledge. They are often abstract in nature; their parameters are often “lumped”. It is therefore difficult for point measurements of real-world system properties to inform these parameters. Nor can the prior uncertainties of lumped parameters be readily established through site characterization studies. Hence if a prediction is sensitive to parameters that are not well informed by the calibration dataset, it is difficult to calculate the uncertainty of that prediction because the receptacles for expert knowledge (the only other source of information pertinent to the prediction) are not provided by a simple model. At the same time, the abstract parameter set employed by a simple model may not be capable of representing local heterogeneities in system properties to which a management-critical prediction may be sensitive. The assignment of too little or too much prior uncertainty to the parameters encapsulated in a simple model may, under these circumstances, lead to a type 2 statistical error in the first case or a type 1 statistical error in the second case.

A simple model may include too few parameters to support a good fit with a calibration dataset. Presumably this design defect would be rectified prior to its deployment, or an alternative model would be chosen. However, the ability of a simple model to fit a calibration dataset does not provide a guarantee that its predictions are without bias. Nor does it guarantee that the uncertainty interval ascribed to a prediction reflects the true uncertainty of that prediction, together with any "manufactured uncertainty" (i.e. bias) accrued through use of the simple model itself. Predictive bias can arise in two ways. It may be a direct reflection of simple model inadequacy as it pertains to that prediction. Alternatively, it may be incurred through the very process of attempted uncertainty reduction, that is, through the process of model calibration. Doherty and Christensen (2011), White et al (2015) and Doherty (2015) show how a simple model may be capable of making relatively unbiased predictions if parameter values are based on expert knowledge alone. However these same predictions may be accompanied by considerable bias once the model has been calibrated because of the surrogate roles that parameters must play to compensate for model defects as they are adjusted to fit the calibration dataset. These authors show that this applies to some predictions made by a simple model but not to others. For some predictions, model defects can be "calibrated out". For other predictions, the calibration process can actually thwart the model's ability to make an unbiased calculation.

Because simple models tend to employ fewer parameters than complex models, the null space exposed through the calibration process is generally of lower dimension for a simple model than it is for a complex model. A simple model may be designed in such a way, and calibrated in such a way (see below), as to reduce the chances of predictive bias. Furthermore, the calibration dataset may be rich in information pertaining to the prediction that a simple model is designed to make. However if the prediction is dependent on some null space parameter components, then these parameter components should be represented in the simple model, not because they are estimable, but precisely because they are inestimable; this avoids a type 2 statistical error when quantifying the uncertainty of the prediction. If the model is too simple to include these parameters, then their lack of representation must be accommodated through strategic inflation of the predictive uncertainty interval calculated by the simple model. Means through which this can be achieved will be context-specific, and will probably involve a degree of subjectivity.

Despite all these shortcomings, there are some types of predictions that a simple model can make with impunity. These are predictions that are wholly dependent on the solution space of the "real world model" of which the simple model is an emulator. These predictions tend to be similar in nature to those which comprise the simple model's calibration dataset. White et al (2015) show that for predictions of these types, a sufficient and necessary condition for the making of an unbiased prediction of future system behaviour is that the model is able to replicate past system behaviour to a level that is commensurate with measurement noise. Because the uncertainties of these types of predictions are dependent only on measurement noise that accompanies the calibration dataset, and not on the nature and dimensionality of the calibration null space, a simple model is able to quantify their uncertainties with integrity.

Overcoming Simple Model Defects

The problems that beset complex models when used in the decision-support context are not about to disappear. Hence, despite the difficulties facing simple model usage that have been outlined above, ways must be found to build, calibrate and deploy them with integrity on a widespread basis as fleet-footed, flexible, prediction-specific replacements for the slow, prediction-general, complex models that are currently used on a widespread basis for decision support. Doherty and Moore (2017) outline ways in which this can be achieved. Some of these are now briefly discussed. Refer to the original text for further details.

Specifications of a simple model must be such that it is capable of making, with as little bias as possible, the prediction for which purpose it was built. Here it is worth noting that most models are better at predicting differences than absolutes. Even complex models are compromised in their ability to associate numbers with future system states. However, these compromises tend to be reduced for predictions that are defined as alterations to system state following alterations to current management practice. The same applies to comparative system states emerging from two competing management strategies. It follows that it is preferable to base management decisions on alterations to system states, or on comparative system states, rather than on the absolute values of system states.

The type of prediction required of a model sets a lower limit on the complexity of the model that is used to make it. This applies particularly to predictions that are sensitive to system and parameterization detail. The uncertainty associated with such predictions is often large. This follows from the fact that such detail may lie largely within the calibration null space. At the same time, representation and parameterization of such detail will probably be far more simple in the model than in the real-world; this may lead to under-estimation of predictive uncertainty especially if attempts are made to constrain inappropriately broad scale or lumped parameters through history-matching. In cases like these, the model's decision support role may be best served by not calibrating the model at all on the basis that the prior uncertainty of the prediction can serve as a useful (and not overly-conservative) surrogate for its posterior uncertainty. Evaluation of prior uncertainty is numerically easier, and far cheaper, than evaluation of posterior uncertainty. Methodologies such as multiple point geostatistics can be easily used to generate "realistic" random parameter fields that pay maximum respect for expert knowledge. Meanwhile, history-matching can be undertaken in a "stochastic" sense, by ensuring that model-generated counterparts to the calibration dataset calculated using these random parameter fields collectively encompass measurements comprising the calibration dataset.

The most difficult predictions to make with a simplified model are those whose uncertainties are partially reduced through history-matching, but which still retain a significant amount of null space sensitivity. These are the predictions that are most likely to incur bias as attempts are made to reduce their uncertainties through calibration; see White et al (2015) for details. Nevertheless, because of their partial solution space dependency, considerable reduction in predictive uncertainty can be achieved through the history matching process; hence calibration must be undertaken to promulgate model usefulness. Doherty (2015) and White et al (2015) show that it is possible to reduce the potential for calibration-induced predictive bias through formulation of a multi-component objective function that incudes observations and corresponding model outputs that have been specially processed prior to matching. This processing is designed to isolate or "orthogonalize out" aspects of the calibration dataset that are likely to entice parameters to adopt surrogate roles in order that model outputs can match that dataset. In many cases, it is not a difficult matter to design appropriate model and data processing schema to achieve this aim. It may only require that vertical, horizontal and temporal differences, as well as the individual measurements themselves, be included in the calibration dataset. The objective function components that emerge from introduction of these "differences observations" must then be weighted for visibility in the overall calibration objective function.

Finally, the prior uncertainties that are ascribed to the lumped and averaged parameters with which a simple model is endowed may require special consideration to avoid underestimation of posterior predictive uncertainties. These prior lumped parameter uncertainties may need to be wider than the uncertainties associated with their spatially distributed real world counterparts. This ensures that any prediction-salient details (for example continuous zones of high or low hydraulic conductivity) that may be "hidden" in these parameters because they are undetectable through the calibration process, can nevertheless find expression when evaluating the uncertainties of these predictions.

Concluding Remarks

The history of model-based decision support is chequered. Too often, specifications for model construction and deployment have been based on an illusion that human beings can create a numerical surrogate for a complex environmental system. It is assumed that this surrogate can be used to predict the future of that system under any proposed management strategy before that strategy is implemented in the real world. Millions of dollars have been spent on models that were built in pursuit of this goal. Frequently, those who paid for models such as these were disappointed with what they received.

The authors of this paper suggest that models can provide better support for environmental decision-making if the premise of their construction is altered from that of a simulator of complex environmental processes to that of a tool for implementation of the scientific method. This method requires that hypotheses be tested, and maybe rejected, through testing their compatability with information pertaining to the nature and properties of the system whose management is being decided. In the decision-making context, hypotheses that require testing are clearly defined. They are the bad things that we seek to avoid if a certain course of management action is adopted. For each such bad thing, it should be possible to construct a model, specific to that bad thing, which can provide receptacles for information against which the probability of its occurrence can be tested. In all likelihood, such a model will be "simple", for its performance in carrying out the task for which it was built will be enhanced if it is unencumbered by unnecessary complexity, and is not required to provide receptacles for unnecessary information.

Reducing the probability of decision-support failure to an acceptable level requires a guarantee that the hypothesis of occurrence of a bad thing is not falsely rejected. Model-calculated uncertainty intervals must therefore accommodate contributions to predictive uncertainty that are an unavoidable outcome of the necessarily defective nature of any model that attempts to simulate a real-world system. Sadly, defects are the price that must be paid for an ability to quantify uncertainty. Pursuit of modelling "perfection" inevitably results in a level of numerical complexity that makes uncertainty analysis difficult, if not impossible.

With acceptance of the benefits of prediction-specific simplicity in simulation of real-world systems comes abandonment of the premise that a single model can be used as a basis for all environmental management within a certain study area. So too follows abandonment of the notion that model development should be separated from the decisions that a model is intended to support. Instead, construction, calibration and deployment of a model must all be done as part of a single, prediction-specific enterprise. This enterprise must result in demonstrably conservative quantification of the uncertainty of that prediction (thus avoiding a type 2 statistical error) while reducing that uncertainty to a level that reflects all available information (thus avoiding a type 1 statistical error).

Our industry is yet to acquire the skillset required for widespread construction of simple, prediction-specific models that maximize the support that model-based data-processing can provide to environmental decision-making. This is an area in which research is urgently needed. Paradoxically, it is often easier to build a complex model than a simple model. Construction of the latter requires a deep understanding of the system that is being simulated, founded on detailed data analysis. It also requires a deep understanding of the numerical strategies that are required for design of a tool that can provide receptacles for information from disparate sources through which specific hypotheses of management failure can be tested, and maybe rejected. Development of this skillset is fundamental to the future of model-based decision-making.

References

Doherty, J., 2015. Calibration and uncertainty analysis for complex environmental models. Published by Watermark Numerical Computing, Brisbane, Australia. 227pp. ISBN: 978-0-9943786-0-6. Downloadable from www.pesthomepage.org.

Doherty, J. and Moore, C.R., 2017. A Theoretical Analysis of Model Simplification. GNS. Downloadable from: https://www.gns.cri.nz/Home/Our-Science/Environment-and-Materials/Groundwater/Research-Programmes/Smart-Aquifer-Models-for-Aquifer-Management-SAM/SAM-discussion-paper

Doherty, J. and Christensen, S., 2011. Use of paired simple and complex models in reducing predictive bias and quantifying uncertainty. Water Resourc. Res doi:10.1029/2011WR010763.

Doherty, J. and Simmons, C.T., 2013. Groundwater modelling in decision support: reflections on a unified conceptual framework. Hydrogeology Journal 21: 1531–1537

Doherty, J. and Vogwill, R., 2015. Models, Decision-Making and Science. In Solving the Groundwater Challenges of the 21st Century. Vogwill, R. editor. CRC Press.

Freeze R.A., Massmann J., Smith L., Sperling T., James B., 1990. Hydrogeological decision analysis: 1 A framework. Ground Water 28 (5),738–766.

Moore, C. and Doherty, J., 2005. The role of the calibration process in reducing model predictive error. Water Resources Research. Vol 41, No 5. W05050.

White, J.T., Doherty, J.E. and Hughes, J.D., 2014. Quantifying the predictive consequences of model error with linear subspace analysis. Water Resour. Res, 50 (2): 1152-1173. DOI: 10.1002/2013WR014767