Geophysical Research Abstracts
Vol. 13, EGU2011-13186, 2011
EGU General Assembly 2011

Statistical Issues of Managing Uncertainties in Environmental Models

Dan Cornford, Richard Jones, Alexis Boukouvalas, and Remi Barillec
NCRG and Computer Science, Aston University, Birmingham, United Kingdom (d.cornford@aston.ac.uk)

Increasingly both decision and policy making are framed in evidential, often model based, settings. This means that information we obtain from models, which are by definition approximations of the underlying systems they represent, needs to be treated with some care. A key issue in using models is therefore understanding their limitations. In this paper we argue that this is best achieved by including explicit representations of the model uncertainties. A natural framework for describing and reasoning about uncertainties in environmental models is Bayesian probability theory. Errors or uncertainties on the model outputs are likely to have two major sources:
a) uncertainties on the model inputs, for example because they are not precisely observed, or are produced as outputs of other models, and;
b) uncertainties on the model itself, due to incomplete knowledge of the underlying system, inability to computationally represent the complete system, etc..
In this paper we address how it might be possible to consider both sources of uncertainty. Note that we do not claim this is an exhaustive or complete list of the sources of uncertainty in models, simply that these are often the most significant sources of error in many environmental models.

Dealing with uncertainties on inputs to the models, and propagating these through the models is reasonably well understood. In general most techniques use some form of Monte Carlo analysis based on random samples, although systematic sampling strategies such as unscented methods are possible. Other approaches include linearisation and adjoint methods for sensitivity analysis. Assuming that the uncertainties on inputs to the models can be well characterised in terms of probability distributions then the main challenge in terms of undertaking an uncertainty analysis is computational, since typically the model has to be evaluated for every sample from the distribution of inputs. Within the FP7 UncertWeb project (http://www.undertweb.org/) we are exploring workflows of multiple models and it is likely that performing uncertainty analysis of these models could be extremely computationally demanding. Here we propose a solution based on the concept of emulation, that is replacing the evaluation of the model, with a surrogate statistical representation of the model. This emulator, which has carefully validated (and minimised) emulator uncertainty, can then be used to undertake uncertainty or sensitivity analysis.

Construction of emulators remains a challenging task, since they require a number of judgements to be made, so here we propose the development of a series of tools to assist the user in creating the emulator. In particular emulators are not suitable for all types of models; we discuss the limitations and options for emulation here. We also emphasise the importance of user interaction in building and validating the emulator. Once the emulator has been constructed we have a very compact statistical representation of the model. This has a crucial benefit in the workflow context - we can now communicate the emulator (as a surrogate for the model) and evaluate the emulator on any computational resource. This potentially results in significantly reduced communication over the network, and makes uncertainty and sensitivity analysis feasible in reasonable time, even for computationally demanding models. It is also possible to emulate the workflow in its entirety if that is needed, or at a granularity that best suits the emulation methodology or problem being addressed.

In the second part of this paper we discuss the issue of model error, sometimes called model discrepancy. This can arise from a variety of causes such as incomplete knowledge of the phenomena being represented in the model, spatial or temporal discretisation within the model, numerical solvers used in the model and representations of unresolved processes (parameterisation) used in the models. All these factors together we typically call model error. Formally we can think about model error as representing the discrepancy of the model from reality. There are two main approaches to representing model error: as bias (deterministic) or as variability (stochastic). Here we argue that model developers should ideally address the issue of model discrepancy, which is likely to have a complex and model state dependant nature. Typically one would assume that any known biases would be removed by the model creators, meaning that the residual (irreducible, given current knowledge and resources) uncertainty would be stochastically represented. This would naturally lead to stochastic models. We discuss the main sources

of uncertainties and how these might be represented in environmental modelling.

Finally we consider the issue of using the emulator framework with stochastic models, which will typically include a description of the model error. We discuss options in terms of emulating the mean and variance only, of emulating certain indicator transforms and of more parametric approaches. We conclude with a discussion of the limitations and challenges of the methods we propose.