Reframing the Reliability of Models
Moving From Error to Quality for Use

Arthur C. Petersen[1,2] and Leonard A. Smith[2]

[1] Netherlands Environmental Assessment Agency (PBL)
[2] Centre for the Analysis of Time Series (CATS), London School of Economics and Political Science (LSE)

Draft Paper for ESF workshop
Exploring Epistemic Shifts in Computer Based Environmental Sciences
June 10–12th , Aarhus University, Denmark
(Contribution to session 3, ‘Standards of good science’)

1. Introduction

Scientific computer simulation—here defined as involving a mathematical model that is implemented on a computer and that imitates real-world processes—is portrayed by some philosophers of science as a new method of doing science, besides theorizing and experimentation (e.g., Rohrlich 1991; Humphreys 1994; Keller 2003). Science studies generally seems to support this conclusion from an historical or sociological perspective (e.g., Galison 1996; Dowling 1999). Two major reasons are typically given for why simulation should be considered qualitatively different. First, it is claimed that simulations make it possible to ‘experiment’ with theories in a new way (e.g., Dowling 1999: 271). Second, simulation enables us to extend our limited mathematical abilities so that we can now perform calculations that were hitherto unfeasible. Thus, we can both construct new theories using computer simulation and calculate the consequences of old theories.[1] An example of the former category is the application of cellular automata in

However, ‘deeper’ forms of uncertainty are often at play. These cannot be expressed statistically but can sometimes be expressed by a range. Such a range is then called a ‘scenario uncertainty’ range. Scenario uncertainties cannot be adequately methodological quality of the representation of a particular dynamic process that is thought to be of importance for its use, e.g., for modelling particular future changes.

Let us explore both notions of reliability a bit more, starting with reliability$_1$. For scientific simulation laboratory practice, 'reliability$_1$' is defined as follows: the 'reliability$_1$' of a simulation is the extent to which the simulation yields accurate results in a given domain. It is important here to distinguish between 'accuracy' and 'precision' (see Hon 1989: 474). Accuracy refers to the closeness of the simulation result to the 'true' value of the sought physical quantity, whereas precision indicates the closeness with which the simulation results agree with one another, independently of their relations to the 'true' value. 'Accuracy thus implies precision but the converse is not necessarily true' (Hon 1989: 474).[7] Traditionally, the distinction between 'systematic' and 'random' error is taken to correspond with the distinction between 'accuracy' and 'precision' (Hon 1989: 474). Since systematic and random error are both statistical notions, Petersen (2006: 55) proposed to dissociate these two dichotomies from each other, so that all sources of error may be assessed in terms of their impact on the accuracy and the precision of the results.[8]

There is an epistemological and practical problem with maintaining a strong focus on the statistical reliability of models, however. We know that models are not perfect and never will be perfect. Especially when extrapolating models into the unknown, we wish 'both to use the most reliable model available and to have an idea of how reliable that model is' (Smith 2002: 2491), but the statistical reliability cannot be established. There is no statistical fix here; also, we should not confuse range of outcomes of a diversity of models in ensemble projections, such as used by the IPCC, with a statistical measure of uncertainty. This does not mean that the information in models cannot be used, but it does imply that the 'reliability' of models needs to be assessed in the context of their use.

the impacts of different sources of uncertainty. Furthermore, disagreement (in distribution) between different modelling strategies would ague against the reliability of some, if not all, of them. ‘Reliability’ then will have to be defined in more pragmatic terms. In those cases, one may instead have recourse to *qualitative* judgments of the

References

Dowling, D. (1999), ‘Experimenting on theories’, *Science in Context* 12: 261–273.