Submitted to Nature Precedings, 19 Jan 2007

Declarative modelling in the ecological and environmental sciences

Robert Muetzelfeldt
Honorary Fellow
School of Informatics
University of Edinburgh
r.muetzelfeldt@ed.ac.uk

The development of dynamic simulation models is a key component of ecological and environmental research. These models are used to check our understanding of ecosystem processes, to predict the future (for example, in climate change), and to explore possible management strategies (for example, in forestry or fisheries management). Many such models are quite complex, consisting of large numbers of equations and/or a high degree of disaggregation of model components.

The modelling process, as a scientific activity, involves many aspects, including running simulations, understanding other people's models, evaluating relationships in models against scientific knowledge, checking the structure of the model against sound modelling principles, and comparing alternative models of the same system.

At the moment, only one of these – running simulations, and activities based on that - is supported by computer technology. Models are (in general) implemented in a conventional programming language: the only thing you can do with this is to execute the program, i.e. to simulate the behaviour of the modelled system. If someone else want to understand the model, they either have to wade through a computer program listing (which frequently runs to thousands of lines), or they need to read a related paper, with no guaranteed completeness or synchronisation with the current version of the program. On top of this, implementing a model as a computer program is a time-consuming and error-prone activity. It is also bad in principle: the representation of a model should be independent of the code used to simulate its behaviour.

Environemental modelling currently has the feel of a cottage industry. Individual projects will come up with their own, local solutions. Even the recent move towards the development of integrated modelling frameworks has done little to help this: component submodels are still implemented as conventional programs, and the existence of a large number of mutually-incompatible frameworks signals their failure as a common modelling platform.

This prevailing culture contrasts markedly with developments in other disciplines. For example, someone wanting to simulate the behaviour of an electronic circuit would not dream of writing a program to do this. Rather, a computer-aided design package would be used to capture the circuit design, with the package providing the built-in ability to perform the simulation (along with numerous other tools). System Biology has seen the recent development and rapid uptake of the Systems Biology Markup Language (SBML), an XML-based language which enables models of metabolic, signalling and regulatory pathways to be interchanged between a wide number of application programs.

These developments are characterised by the use of a declarative language for representing the components of the model and the relationships between them. This language does not say how to simulate model behaviour: it simply is a mathematical specification of the model itself.

The adoption of a common declarative modelling language radically changes the practice of modelling. The main reason for this is that the same model can be processed in a large number of ways: for example, to produce human-readable model descriptions (e.g. HTML), to enable one to search through libraries of models looking for all models that have a certain feature, or to generate executable code for a variety of simulation platforms. New possibilities will emerge: for example the ability to automatically transform a complex model into a simpler one with similar behaviour, and the ability to link model variables with standard ecological and environmental ontologies.

Why is there no such common language in the ecological and environmental sciences, nor even any indication that the community is giving any though to the development of such a language? In most cases, the answer seems to be a simple lack of awareness that it is even possible to represent models in a declarative language. There is also a certain amount of inertia: tomorrow's model is derived from today's; and it is easier to work with familiar approaches.

A valid concern is that the sheer complexity of ecological and environmental modelling - the types of problems that it addresses, the range in scales, and the range in modelling constructs – makes it impossible to come up with a single language with the expressiveness required. While this may be true in some absolute sense, it is also true that a very large proportion of models in these areas are "continuous-systems models with disaggregation": models based on differential/difference equations, with the subdivision of model variables into multiple values (soil layers, spatial units, multiple species, etc). There are good reasons for thinking that it is possible to design a declarative language with the required expressiveness.

There is no doubt that in the near future (say, 10 years from now), a large proportion of environmental models will be expressed in a declarative language. Technological developments (XML-related technologies) will pull in that direction, while the community will be looking for better ways to publish their models, and funding bodies will be insisting on more effective use of research funds. It is therefore time for the ecological and environmental research community to face up to the question of how their models should be represented in the future.

References

References provided by the author: http://www.connotea.org/user/robertm/tag/envirodecmod
References provided by the author and others: http://www.connotea.org/tag/envirodecmod

This paper was originally prepared for consideration as a Nature Commentary article.

19 July 2007