Ecological Modelling 280 (2014) 117–128

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach

Jacqueline Augusiak[a,*], Paul J. Van den Brink[a,b], Volker Grimm[c,d,e]

[a] Wageningen University, Aquatic Ecology and Water Quality Management Group, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[b] Alterra, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[c] UFZ, Helmholtz Centre for Environmental Research – UFZ, Permoserstr. 15, 04318 Leipzig, Germany
[d] Institute for Biochemistry and Biology, University of Potsdam, Maulbeerallee 2, 14469 Potsdam, Germany
[e] German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

ARTICLE INFO

Article history:
Available online 29 November 2013

Keywords:
Model validation
Terminology
Decision support
Documentation
Ecological models
Risk assessment

ABSTRACT

Confusion about model validation is one of the main challenges in using ecological models for decision support, such as the regulation of pesticides. Decision makers need to know whether a model is a sufficiently good representation of its real counterpart and what criteria can be used to answer this question. Unclear terminology is one of the main obstacles to a good understanding of what model validation is, how it works, and what it can deliver. Therefore, we performed a literature review and derived a standard set of terms. 'Validation' was identified as a catch-all term, which is thus useless for any practical purpose. We introduce the term 'evaludation', a fusion of 'evaluation' and 'validation', to describe the entire process of assessing a model's quality and reliability. Considering the iterative nature of model development, the modelling cycle, we identified six essential elements of evaludation: (i) 'data evaluation' for scrutinising the quality of numerical and qualitative data used for model development and testing; (ii) 'conceptual model evaluation' for examining the simplifying assumptions underlying a model's design; (iii) 'implementation verification' for testing the model's implementation in equations and as a computer programme; (iv) 'model output verification' for comparing model output to data and patterns that guided model design and were possibly used for calibration; (v) 'model analysis' for exploring the model's sensitivity to changes in parameters and process formulations to make sure that the mechanistic basis of main behaviours of the model has been well understood; and (vi) 'model output corroboration' for comparing model output to new data and patterns that were not used for model development and parameterisation. Currently, most decision makers require 'validating' a model by testing its predictions with new experiments or data. Despite being desirable, this is neither sufficient nor necessary for a model to be useful for decision support. We believe that the proposed set of terms and its relation to the modelling cycle can help to make quality assessments and reality checks of ecological models more comprehensive and transparent.

"I assert that whenever a dispute has raged for any length of time, especially in philosophy, there was, at the bottom of it, never a problem about mere words, but always a genuine problem about things."

I. Kant (1786)

1. Introduction

Ecological models are increasingly used and needed for supporting environmental decision-making (Schmolke et al., 2010a). Often they are the only way to take into account the relevant spatial and temporal scales and the multitude of processes characteristic to ecological systems. Corresponding experiments can be impossible, and insights from descriptive studies do not necessarily provide enough mechanistic understanding to predict responses of ecological systems to new conditions.

Since models are simplified representations of real systems, a key challenge is, however, to show that the models are realistic enough to meet their intended purpose (Rykiel, 1996). Before we can transfer inferences from model results to the real world,

* Corresponding author at: Department of Aquatic Ecology and Water Quality Management, Wageningen UR (University and Research Centre), P.O. Box 47, 6700 AA Wageningen, The Netherlands. Tel.: +31 317 48 59 58; fax: +31 317 41 90 00.
E-mail address: jacquelinwe.augusiak@wur.nl (J. Augusiak).

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.11.009

we have to demonstrate that the model reproduces observations for the right reasons, not just because it has been tweaked via calibration to do the right thing. If models are in fact used without being carefully checked for their validity, they might lead to erroneous decisions. Pilkey and Pilkey-Jarvis (2007) call inappropriate models "useless arithmetics" and find that "these types of applied models are frequently detached from reality – built on oversimplified and unrealistic assumptions about natural processes". Thus, scepticism with regard to using ecological models to support environmental decisions is a healthy attitude. It is up to the modellers to provide evidence and indicators that their model is realistic enough.

An example field of decision making, where scepticism regarding ecological models so far has prevented the use of models, is ecological risk assessment of chemicals, in particular pesticides (Forbes et al., 2009, 2010; Thorbek et al., 2010). Ecological risk assessments are required for pesticides to minimise potentially negative impacts on non-target flora and fauna and, thus, on ecosystems in general. Regulatory decisions on whether or not a certain pesticide can be used are, at least in the lower tiers of the risk assessment, based on highly standardised schemes. They focus on effects on individuals of a set of standard species, observed under standardised conditions in the laboratory.

Mechanistic effect models have long been identified as potentially useful tools to extrapolate the limited findings from standard tests to more realistic conditions such as fluctuating exposure profiles, higher levels of biological organisation, and larger temporal and spatial scales, thus making risk assessments ecologically more relevant (Forbes et al., 2009, 2010; Galic et al., 2010; Pastorok, 2002; Thorbek et al., 2010). Mechanistic effect models comprise ecological and organism-level effect models. They are referred to as 'mechanistic' to clearly separate them from descriptive, or statistical, models, and as 'effect models' to separate them from physico-chemical models describing the fate and exposure of chemicals in the environment.

Despite the high potential of mechanistic effect models to improve the ecological realism of pesticide risk assessment, so far they have not often been used or accepted in regulatory risk assessments. A major obstacle is the doubt as to whether a given model represents the real world sufficiently well, which is reinforced by a lack of clear criteria for assessing a model's realism. Additionally, a comprehensive model assessment is often hampered by the ambiguous application of terminology within and between involved stakeholder groups. Academics, industry, as well as regulators each possess a different set of vocabulary, knowledge, and interests (Hunka et al., 2013; Jakeman et al., 2006), which interferes with both a more productive advancement and communication of methods, and with actually using models to support decision making.

Terminology regarding model assessment has in general proven to be a particular source of confusion (Oreskes et al., 1994a; Rykiel, 1996). To describe general tasks of quality assurance throughout a model's development and application, academics often use the term 'validation' more or less intuitively, due to a lack of a clear and unambiguous definition. Yet, academics are at odds with each other as to what 'validation' should mean in a modelling context, to which degree model validation would be generally feasible, and which methods or criteria should be applied to assess the compliance of a given model with its real counterpart.

This issue has been debated in the context of ecological modelling for the past 50 years and still no commonly accepted language and methodology could be agreed upon (see references in Rykiel, 1996). This makes it very hard to clearly assess and communicate the credibility of models, which in turn makes it difficult, if not impossible, for decision makers, who are usually not trained in assessing whether a model is good enough, to let models influence their decisions. Other domains, e.g. hydrology, economics, meteorology, or environmental engineering, where mechanistic models are being used as well to support decision making, are facing similar problems (Ferson et al., 2008; Gass, 1983; Hodges and Dewar, 1992; Oriade and Dillon, 1997; Refsgaard et al., 2005).

In this article, we review and evaluate the literature concerning the terminology and methodology regarding model validation. We focus predominantly on literature related to ecological models but draw relevant lessons from other scientific fields with relations to regulatory frameworks to provide a pragmatic solution to the above-mentioned challenges. According to the most dominant trends that we could identify, we will propose a common vocabulary for the evaluation of applied ecological models. This can for example assist the risk assessment process by introducing a structured system of language. In particular, we will suggest the new, artificial term, 'evaludation', which is a merger of 'evaluation' and 'validation'.

Evaludation consists of several elements, or steps, that correspond to the different stages of iterative model development forming the 'modelling cycle' (Grimm and Railsback, 2005). They thus serve as the main structuring elements for the suggested terminological system. The modelling cycle consists of the following elements (see also Section 3): formulation of the questions to be addressed; assembly of hypotheses that constitute our conceptual model of the system in question; choice of model structure, i.e. choice and representation of entities, state variables, and processes; implementation of the model via equations and/or a computer programme; model analysis; and communication of model output.

Based on this approach, we will demonstrate that validation is not a binary criterion that is determined once a model's development has been finished. Rather, overall model credibility arises gradually throughout the entire modelling cycle.

2. Terminology and concepts

Mechanistic modelling simplifies real-world processes to understand driving mechanisms well enough so that forecasts of a system's response to certain conditions become feasible. This simplification implies the risk that not all relevant factors were captured or that relevant data are missing. Investigating these deficiencies in detail is not always feasible due to monetary, time, or other constraints. For this and other reasons, models inherently possess a level of uncertainty.

To reduce the likelihood of a flawed decision due to an uncertain, simplified representation, decision makers usually demand that a model should be validated. Typically, they ask for a comparison of model output with new empirical data to determine whether possible discrepancies render the model too unrealistic for use. Many scientists argue (correctly in our opinion) on the contrary that this approach to validation is too limited for at least three reasons. First, agreement between modelled and empirical data does not necessarily imply that a model is 'correct', but could also result from a combination of 'wrong' input parameters and process representations (Oreskes and Belitz, 2001). Second, this kind of direct validation often is impossible to achieve because such data do not exist, which is rather the rule than the exception in ecological and environmental systems. In fact, this is the reason why models are needed for these systems in the first place. Third, the genuine meaning of the word "validation" does not fully match with the uses of the term in ecological modelling and is accompanied by philosophical discourses about its legitimate usage.

It seems obvious that validation should not be mistaken with 'truth', although the term certainly implies a strong sense of legitimation (Oreskes et al., 1994b; Rykiel, 1996). Decision makers would appreciate having some form of quantifiable certification