Ecological Modelling 280 (2014) 129–139

Contents lists available at ScienceDirect

Ecological Modelling

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

Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE

Volker Grimm[a,b,c,*], Jacqueline Augusiak[d], Andreas Focks[d], Béatrice M. Frank[e], Faten Gabsi[f], Alice S.A. Johnston[g], Chun Liu[g,h], Benjamin T. Martin[a,i], Mattia Meli[j], Viktoriia Radchuk[c,e], Pernille Thorbek[h], Steven F. Railsback[k]

[a] Helmholtz Centre for Environmental Research – UFZ, Department of Ecological Modelling, 04318 Leipzig, Germany
[b] University of Potsdam, Institute for Biochemistry and Biology, Maulbeerallee 2, 14469 Potsdam, Germany
[c] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany
[d] Wageningen University, Aquatic Ecology and Water Quality Management Group, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[e] Université Catholique de Louvain, Earth and Life Institute, Croix du Sud 2, 1348 Louvain-la-Neuve, Belgium
[f] RWTH Aachen University, Institute for Environmental Research, Worringerweg 1, 52074 Aachen, Germany
[g] University of Reading, School of Biological Sciences, Philip Lyle Building, Room 405, Reading RG6 6AS, UK
[h] Syngenta, Environmental Safety, Jealott's Hill International Research Centre, Bracknell, Berkshire RG42 6EY, UK
[i] University of California Santa Barbara, Department of Ecology, Evolution, and Marine Biology, Santa Barbara, CA 93106-9620, USA
[j] Roskilde University, Department of Environmental, Social and Spatial Change, Universitetsvej 1, P.O. Box 260, 4000 Roskilde, Denmark
[k] Humboldt State University, Department of Mathematics, Arcata, CA 95521, USA

ARTICLE INFO

Article history:
Available online 18 February 2014

Keywords:
Standardization
Good modelling practice
Risk assessment
Decision support

ABSTRACT

The potential of ecological models for supporting environmental decision making is increasingly acknowledged. However, it often remains unclear whether a model is realistic and reliable enough. Good practice for developing and testing ecological models has not yet been established. Therefore, TRACE, a general framework for documenting a model's rationale, design, and testing was recently suggested. Originally TRACE was aimed at documenting good modelling practice. However, the word 'documentation' does not convey TRACE's urgency. Therefore, we re-define TRACE as a tool for planning, performing, and documenting good modelling practice. TRACE documents should provide convincing evidence that a model was thoughtfully designed, correctly implemented, thoroughly tested, well understood, and appropriately used for its intended purpose. TRACE documents link the science underlying a model to its application, thereby also linking modellers and model users, for example stakeholders, decision makers, and developers of policies. We report on first experiences in producing TRACE documents. We found that the original idea underlying TRACE was valid, but to make its use more coherent and efficient, an update of its structure and more specific guidance for its use are needed. The updated TRACE format follows the recently developed framework of model 'evaludation': the entire process of establishing model quality and credibility throughout all stages of model development, analysis, and application. TRACE thus becomes a tool for planning, documenting, and assessing model evaludation, which includes understanding the rationale behind a model and its envisaged use. We introduce the new structure and revised terminology of TRACE and provide examples.

1. Introduction

Modelling is an iterative process. First model versions should deliberately be oversimplified to make the 'modelling cycle' (Grimm and Railsback, 2005; Fig. 1a) start as soon as possible. Discrepancies between model output and observations then guide developing the next model versions. Thus, during iterative model development, many alternative submodels or even overall designs are tested, modified, improved, or discarded. As a result, models are usually a patchwork of elements that entered model development at different stages.

For example, in population models simple phenomenological submodels describing mortality due to senescence might have been introduced early on and never required intensive testing. Other

* Corresponding author at: Helmholtz Centre for Environmental Research – UFZ, Department of Ecological Modelling, 04318 Leipzig, Germany.
Tel.: +49 341 235 1711; fax: +49 341 235 1479.
E-mail address: volker.grimm@ufz.de (V. Grimm).

0304-3800/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2014.01.018

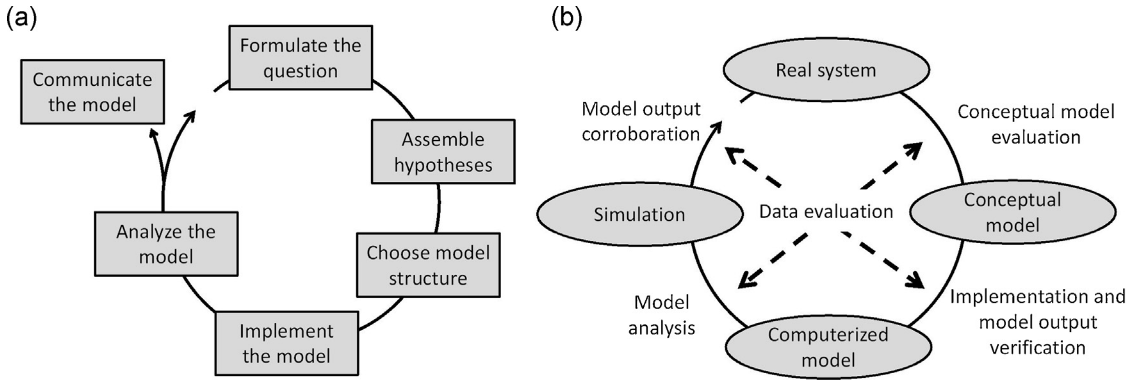

Fig. 1. Correspondence of iterative model development (the "Modelling Cycle", Grimm and Railsback, 2005) (a), and the elements of model evaludation as proposed by Augusiak et al. (2014) (b).

submodels representing key behaviours, for instance home range dynamics (Liu et al., 2013), habitat selection (Railsback and Harvey, 2002), or starvation (Martin et al., 2013), often require testing a large number of alternative versions, both in isolation and within the entire model.

When a model is finally published or made available to decision makers, most of the careful design and testing that went into the final model remains undocumented. However, without sufficient information about a model's design and testing it can be hard or even impossible to develop enough confidence to use it for supporting environmental decision making (Schmolke et al., 2010). This situation is similar to a laboratory buying an expensive new analytical instrument: how do the lab's owner and clients know that the instrument works correctly and produces reliable results, and exactly how to calibrate and use it to produce credible data? They would require that documentation of the instrument's theoretical basis, its detailed design, and how it has been tested be available somewhere. This documentation might not be read routinely, but they are key components of quality assurance: lab instrument manufacturers know that customers expect full documentation and that flawed or incomplete documentation might make them cancel their purchase.

Unfortunately, in ecological and environmental modelling there is not yet a generally established culture of documenting the scope, design, and tests of our virtual laboratories, i.e., our models. Without such a culture, three bad things can happen: modellers develop models without employing basic mechanisms of quality assurance, leading to poor model designs; decision makers might not consider a model even though it is well-designed and tested; or, vice versa, they might use a model to support decisions although the model has major flaws (Pilkey and Pilkey-Jarvis, 2007). Further, the modelling process itself becomes unnecessarily inefficient: analyses often must be repeated or revised because the original methods were not recorded; mistakes can be repeated; and unproductive approaches can be tried several times when the modeller does not document why they were unproductive the first time. Such a culture does however exist in other fields (e.g., in engineering and software development) that rely heavily on modelling and computation. There is in fact a vast literature on these topics (to get an idea of it, see the Wikipedia entries for topics such as software testing, software documentation, and software specification; and Augusiak et al., 2014).

In ecological modelling, however, we do not yet have a culture of documenting model development, testing, and analysis, because clients of models usually do not know what kind of documentation they should require, hence model developers do not know what clients expect. Here, clients include other scientists trying to learn from a published model and decision makers trying to use a model or its output to make better decisions. To establish a culture of comprehensive modelling documentation we thus need to establish clear expectations: model clients need to have clear expectations and model developers need to be aware of clients' expectations.

Establishing such expectations has worked before, often via standardization. For example, the structure of scientific articles – Introduction, Methods, Results, Discussion – reflects expectations both of the readers and writers. Or, for individual- or agent-based models (IBMs), modellers increasingly are using a standard format, the ODD protocol (Overview, Design concepts, Details; Grimm et al., 2006, 2010), for describing the model, thereby increasingly making readers expect that IBMs are described in this format, with certain kinds of information at certain places in the model description.

Thus, to help establish a culture of comprehensive modelling documentation, Schmolke et al. (2010) suggested a standard format and terminology referred to as "TRACE" (TRAnsparent and Comprehensive Ecological modelling documentation). The acronym also refers to the process of 'tracing' model development and testing by "going backward over the evidence step by step" ("trace" in Merriam Webster online dictionary). Schmolke et al. (2010) introduced the overall framework of TRACE but also made clear that subsequent work will be needed to establish TRACE and hence, the above-mentioned culture of modelling documentation: "The TRACE documentation framework can only become established as a standard if it is applied and refined by numerous projects" (p. 484).

Here, we report a first such refinement, based on using and discussing TRACE in the EU-funded project CREAM (Grimm et al., 2009) and three further modelling projects. In CREAM, ecological and organism-level models were developed to assess the effect of chemicals, in particular pesticides, on populations and individual organisms. Ultimately, the hope is that such models are used to make regulatory risk assessment of chemicals more ecologically relevant (Forbes and Calow, 2012). Nevertheless, although ecological risk assessment is a specific field of environmental decision making, the lessons learned about TRACE and how to actually use it are generic.

We first briefly summarize the original idea and structure of a TRACE document. Then we present the most important questions that came up in using TRACE, and from discussions at conferences, feedback from colleagues, and one publication (Wang and Luttik, 2012). We provide a practical answer to each question and then present, as a result, a revised TRACE format and brief guidance for writing and reading TRACE documents.

The basic idea of TRACE remains the same but we completely revise the structure and terminology of TRACE documents to clarify the purposes of TRACE documents: to help model clients understand the model and assess the quality of the model and hence the reliability of the results; i.e., to provide comprehensive model evaluation and validation. For this, we adopted the terminology proposed by Augusiak et al. (2014), in particular merging the terms 'evaluation' and 'validation' into the new artificial term 'evaludation', defined as: "The entire process of establishing model quality and credibility throughout all stages of model development, analysis, and application". TRACE thus becomes a tool for planning, documenting, and assessing model evaludation, which includes understanding the rationale behind a model and its envisaged use. TRACE is aimed at documenting model design and testing. Model application (i.e., the simulations carried out to answer specific environmental decision-making questions) will also need to be carefully documented. However, this is outside the scope of the TRACE documentation.

2. TRACE: the basic idea

The two basic tasks of using TRACE are: (i) keeping a modelling notebook in which you briefly document, preferably daily, what you did regarding model design, testing, and analysis, and what you learned from it; and (ii) using, in this modelling notebook, the standardized terminology used in TRACE documents.

The two basic ideas underlying TRACE are (i) TRACE and its terminology cover all elements of iterative model development, i.e., the modelling cycle; and (ii) by using a standardized terminology and document structure, readers and model users know exactly where in the TRACE document they can expect finding what kind of information. TRACE also lets clients quickly check whether all important steps of model development were documented and how carefully the model was designed, parameterized, tested, and analyzed.

To illustrate the potential benefits of TRACE documentation, imagine you developed an individual-based model of a small mammal which includes home range behaviour, similar to that of Liu et al. (2013). Due to the lack of appropriate data, you decided to use a phenomenological, not mechanistic, approach so that home ranges are more or less imposed rather than emerging from individuals making decisions in a heterogeneous landscape. You tested simple and complex ways of imposing home ranges and how they are related to habitat structure and then selected a rather simple approach which takes into account vegetation cover but not food resources.

If this model were just factually described in a publication, reviewers, readers, and potential users might consider the design of the home range sub-model ad hoc, unrealistic, and not good enough to make the entire population model reliable. By contrast, if each time you worked on the home range model you kept notes in the modelling notebook labelled, e.g., "home range model/purpose", "home range model/data", or "home range model/sensitivity analysis/alternatives", you can easily extract relevant information from your notebook and compile it in a TRACE document. This document then shows that for the purpose of the overall model, a mechanistic home range model was not essential, that no data existed for relating home ranges to resource availability in time and space, and that alternative simpler models created artefacts and more complex models did not improve usefulness and realism of the entire model. Reviewers, readers, and users of your model would understand that model design was not ad hoc but that the chosen design reflects the overall model purpose, data limitations, and careful selection of submodel structure.

The structure of TRACE documents proposed by Schmolke et al. (2010) (Table 1) reflects all elements of model development, testing, and analysis. In their review of literature on good modelling practice, Schmolke et al. (2010) found that virtually all authors agreed that quality assurance of models should address all elements of modelling, not only verification and validation (see also Augusiak et al., 2014). For the terms used for the different elements of TRACE documents, Schmolke et al. (2010) had to make choices, as terminologies vary considerably within and across disciplines.

Schmolke et al. (2010) concluded from their review that most elements of good modelling practice have long been known but never got established. The main challenge thus is not so much to define good modelling practice but to get it established and widely used. Since producing TRACE documents requires additional effort, it is unlikely that TRACE will be used if it does not provide direct benefits to the modeller. Schmolke et al. (2010) therefore suggested linking TRACE documentation to keeping modelling notebooks. Such notebooks have direct benefits to the modeller because they help organize and document the complex task of developing, testing and analysing a model. Extracting a TRACE document from a notebook requires much less effort than producing it from scratch, after model development has ended.

3. How to actually use TRACE? Questions and answers

How well did the ideas underlying TRACE work? In contrast to the ODD protocol for describing individual-based and agent-based models (Grimm et al., 2006, 2010), which was immediately used by modellers in ecology and other fields and is becoming a widely accepted standard, no independent use of TRACE has yet been published, although it is being used for models for decision support (P. Thorbek, *pers. communication*). This suggests that TRACE as presented by Schmolke et al. (2010) was not yet ready to use. The challenges in actually using TRACE became apparent when trying implementing it in 10 modelling projects. To summarize these challenges, we list and address the most frequently asked questions about TRACE.

How much detail should be in TRACE documents? – The modelling notebook should preferably be updated every day while working on a model. For complex models, which can take a long time to develop and test, this means that even if the entries in the notebook are concise, hundreds of pages of text could easily accumulate, in addition to sketches, figures, tables, links to program versions and data and model output files, etc. Of course, only a small proportion of all this should go into the TRACE documents.

The purpose of a TRACE document is to provide convincing evidence that a model was thoughtfully designed, correctly implemented, thoroughly tested, well understood, and appropriately used for its intended purpose. Readers will first want to see an overview and only then decide whether and where to go into more detail. Thus, to allow for hierarchical reading and to keep TRACE documents concise and readable, it is critical to start the entire document and each of its sections with an executive summary.

For example, to document software testing the executive summary should describe the kinds of approaches and tools used to detect programming errors. This might include debug code that checks, while the program is running, that variables stay within meaningful ranges. The corresponding entry in the TRACE document might then read like: "The program includes 23 elements of debug code, which stop program execution and give an error message when a variable assumes values outside its meaningful range". Then a table might list all variables checked in this way plus their ranges. This information adds evidence that the software was thoroughly tested.

In general, summaries should always come first and details later; if details provide long and repetitive information, this information should be put into separate files, or moved to an appendix at the end of the TRACE document. For example, if a key submodel has been implemented independently in a spreadsheet, and its outputs were

Table 1
New structure, terminology, and contents of TRACE documents.

TRACE element	This TRACE element provides supporting information on:
1. Problem formulation	The decision-making context in which the model will be used; the types of model clients or stakeholders addressed; a precise specification of the question(s) that should be answered with the model, including a specification of necessary model outputs; and a statement of the domain of applicability of the model, including the extent of acceptable extrapolations.
2. Model description	The model, i.e. a detailed written model description. For individual/agent-based and other simulation models, the ODD protocol is recommended as standard format. For complex submodels, include concise explanations of the underlying rationale. Model users should learn what the model is, how it works, and what guided its design.
3. Data evaluation	The quality and sources of numerical and qualitative data used to parameterize the model, both directly and inversely via calibration, and of the observed patterns that were used to design the overall model structure. This critical evaluation will allow model users to assess the scope and the uncertainty of the data and knowledge on which the model is based.
4. Conceptual model evaluation	The simplifying assumptions underlying a model's design, both with regard to empirical knowledge and general, basic principles. This critical evaluation allows model users to understand that model design was not ad hoc but based on carefully scrutinized considerations.
5. Implementation verification	(1) Whether the computer code implementing the model has been thoroughly tested for programming errors, (2) whether the implemented model performs as indicated by the model description, and (3) how the software has been designed and documented to provide necessary usability tools (interfaces, automation of experiments, etc.) and to facilitate future installation, modification, and maintenance.
6. Model output verification	(1) How well model output matches observations and (2) how much calibration and effects of environmental drivers were involved in obtaining good fits of model output and data.
7. Model analysis	(1) How sensitive model output is to changes in model parameters (sensitivity analysis), and (2) how well the emergence of model output has been understood.
8. Model output corroboration	How model predictions compare to independent data and patterns that were not used, and preferably not even known, while the model was developed, parameterized, and verified. By documenting model output corroboration, model users learn about evidence which, in addition to model output verification, indicates that the model is structurally realistic so that its predictions can be trusted to some degree.

compared to output from the model's original implementation, the TRACE document would provide hyperlinks to the spreadsheet.

How are TRACE, the modelling notebook, and the model description, for example the ODD for individual-based models, related to each other? – The modelling notebook corresponds to lab journals or notebooks in laboratory research. It should be kept for its own, direct benefits, which include documentation of test procedures so tests can be replicated later, and supporting the careful planning, execution, and interpretation of simulation experiments. A notebook does not just document experiment design and factual results: "The act of writing in the notebook causes the scientist to stop and think about what is being done in the laboratory. It is in this way an essential part of 'doing good science"' (Kanare, 1985, p. 1).

The link between the modelling notebook and TRACE documents is established by using TRACE terminology to label the entries in the notebook. Notebooks do not necessarily have the structure of TRACE documents but most often follow a chronological order, with each entry identified by its date and a label following TRACE terminology.

TRACE documents should include a full model description, which, if the model is individual-based or agent-based (or any other kind of simulation model), preferably should use the standard format ODD (Grimm et al., 2006, 2010); ODD can be used for any kind of simulation model, just by leaving out those 'design concepts' that are specific to individual-based models (e.g., Meli et al., 2014).

Is there an overlap between model description (e.g., ODD) and the TRACE document? – There can be overlap between a TRACE document and one part of the ODD protocol, its description of the design rationale for model components and submodels. The ODD protocol includes design rationale because when it was developed TRACE did not yet exist, so that the justification of the model's biological background, structure, simplifying assumptions, and parameterization had to be in ODD. Now, if both an ODD model description and a TRACE document are provided, the model's underlying rationale would be described in both. This redundancy is unavoidable when an ODD must be complete by itself, for example in a journal publication. Moreover, due to space limitations, the explanations of the model's rationale will still be quite short in the ODD description, whereas in the TRACE document they can be more detailed and discuss alternative model designs that were tested and then discarded.

Who is going to read TRACE documents tens or hundreds of pages long? – TRACE documents are not designed to be read from cover to cover, but to provide additional information convincing users that all tasks of model development have been performed according to general good modelling practice and to provide all elements of a model's evaludation (for details on evaludation, see below). A TRACE document can be thought of as a reference manual where users can find particular details when necessary. Therefore, as explained above, a hierarchical structure in the entire document and each section and subsection is mandatory, with each unit starting with an executive summary.

Is TRACE a technical document, written by modellers for modellers (see Wang and Luttik, 2012)? – Definitely not. TRACE is supposed to cover all aspects of model development, testing, and analysis, not only the technical ones. For example, an overview of the biological literature and reasoning that has been used to design the model and get parameter values is an integral and important part of each TRACE document.

There will certainly be elements that are more technical, for example documentation of software testing. However, TRACE's hierarchical structure requires that such elements are also first summarized in a non-technical way suitable for all users.

Do the "Parameterization", "Calibration" and "Sensitivity analysis" elements overlap, making it difficult to decide where to put (or expect) what information? – These three elements certainly are related, but are also different enough to distinguish as separate elements within the modelling cycle. In the updated TRACE format (see below), we give them clearer definitions. We distinguish between "direct parameterization", obtaining parameter values directly from the literature or experts, and "inverse parameterization", obtaining parameter values inversely by calibrating the model to observations. Regarding sensitivity analysis, we now distinguish between "local sensitivity analysis", which is based on one parameter at a time, and "global sensitivity analysis", in which several or all parameters are varied over their whole ranges.

What about models developed before TRACE existed? – TRACE documents can of course be assembled even if no modelling notebook was kept. If no notes were made during model development, the corresponding analyses, reviews, and tests must then be performed and documented in retrospect. This effort can be substantial for complex models; it is also our main argument for keeping a modelling notebook. You should do these analyses anyway, so why not keep notes so that no analyses have to be repeated while putting the TRACE document together?

Many reviewers and readers never look at the Supplementary Material, so why should I produce the TRACE document if I don't get any credit for it, i.e., higher chances of getting the model published or used? – Just stating in an article or report that the Appendix or Supplement includes a TRACE document might not be sufficient to get credit for the work that went into producing it. We therefore suggest that, whenever a TRACE document has been produced, the main text or a printed appendix includes a "TRACE table" concisely summarizing the TRACE document (see example in Table 2). Providing such summaries will also help establishing the culture of model evaludation and its documentation: the more publications or reports submitted to decision makers include a TRACE table, the more often model clients will use it as a checklist for scrutinizing a model's evaludation. TRACE tables therefore could be critical to establishing the readers' expectation that we mentioned in the Introduction. Of course, once TRACE is more widely used, the credit for having provided a TRACE document will be immediate, as it will increase chances of getting published and used.

A similar development took place in ecological modelling over the last several decades: 20 years ago, few publications included a sensitivity analysis, whereas in 2009–2010, 24% of all articles published in Ecological Modelling included some sort of systematic sensitivity analysis (Thiele et al., *unpubl. manuscript*). Thus, nowadays, most reviewers expect a sensitivity analysis; modellers are aware of this expectation so they just include the analyses as a normal part of publication.

Do I need a full TRACE document for every model application? – Different applications of the same model can refer to the same documentation of model development, but model analysis, which includes the description and justification of the scenarios explored, needs to be updated. A similar situation often arises with using ODD for model description: what if only one or a few elements of a model were changed? A technical solution for both ODD and TRACE is to re-use the original elements and track changes by crossing out deleted text and emphasizing new text by colour or bold fonts.

4. TRACE: a first revision and short guide

Schmolke et al. (2010) argue that TRACE documentation is critical in making a model fit for supporting environmental decision making. However, just the word 'documentation' does not convey this urgency. Therefore, we here suggest re-defining TRACE as a tool for planning, performing, documenting, and assessing a model's 'evaludation' (Augusiak et al., 2014).

4.1. Evaludation

Augusiak et al. (2014) review the terminology and ideas around the terms 'validation', 'verification', and 'evaluation', which all represent important elements of assessing whether a model is good enough for its intended purpose. The two main conclusions of their review are that (i) the term 'validation' is a catch-all term that has been given so many different, partly contradicting, meanings that it cannot be used for any practical purpose; (ii) comparing model predictions to independent, new data is neither sufficient nor necessary to make a model useful for, e.g., decision support. Rather, all steps of iterative model development have to fulfil certain quality criteria: a model can reproduce existing data or make even correct new predictions, while still based on biased data, unreasonable assumptions, faulty software, and excessive parameter and submodel tweaking. Quality assurance of models should therefore include all elements of iterative modelling development (Fig. 1a). Hence, Augusiak et al. (2014) suggest the new term 'evaludation' for this kind of comprehensive quality assessment.

Evaludation consists of six elements which largely correspond to the elements of the modelling cycle (Fig. 1b). These elements are (i) 'data evaluation', assessing the quality of numerical and qualitative data used for model development and testing; (ii) 'conceptual model evaluation', scrutinizing the simplifying assumptions underlying a model's design; (iii) 'implementation verification', checking the model's implementation in equations and software; (iv) 'model output verification', comparing model output to the data and patterns that guided model design and calibration; (v) 'model analysis', examining the model's sensitivity to changes in parameters and formulation to understand the model's main behaviours and describing and justifying simulation experiments; and (vi) 'model output corroboration', comparing model output to data and patterns that were not used for model development and parameterization.

4.2. A new terminology for TRACE

Since both TRACE and evaludation relate to the iterative steps of the modelling cycle, their elements can be easily linked (Fig. 2). Therefore we propose replacing the original TRACE terminology with the six elements of model evaludation, plus one element for problem formulation and one for model description (Table 1). By doing so, we also re-define the scope of TRACE from being a "standard format for documenting models and their analyses" (Schmolke et al., 2010) to being a tool for planning, performing, and documenting model evaludation. Accordingly, the "E" in the acronym TRACE changes from "Ecological modelling" to "Evaludation". TRACE thus now stands for "TRAnsparent and Comprehensive model Evaludation". The tasks documented in the original version of TRACE, including the documentation of scenarios tested with the model, remain largely the same, but have partly been renamed and re-grouped (Fig. 2). One original element of TRACE is no longer included: "Recommendation", because we believe that these are the main results of models for environmental decision making, so they should be presented in the main document.

4.3. An updated guide for using TRACE

A template for producing TRACE documents following the new structure and terminology defined in the previous section is provided in the Supplementary Material. The questions and checklists at the end of each of the eight elements should be helpful for compiling coherent, comprehensive, and concise TRACE documents. Here we give only a short characterization of the eight elements of TRACE, and their subsections. For more detailed discussion of the six evaludation elements, see Augusiak et al. (2014).

Each element should start with an executive summary, which can be a short narrative, a bullet-point list, or a table of contents of this element. The summary should include references to corresponding page numbers and hyperlinks for convenient navigation in the electronic version of the TRACE document.

1. *Problem formulation*. This element is largely unchanged from Schmolke et al. (2010). It should describe: the decision-making context in which the model will be used; the type of model clients, or stakeholders, addressed; the precise question(s) that should be answered with the model and the necessary model outputs; and the domain of applicability of the model, including the extent of acceptable extrapolations. For regulatory models

Table 2
Overview and comparison of the three example TRACE documents provided in the Supplementary Material. The example documents' summaries are presented for each TRACE element. Some of the original summaries include hyperlinks. References to figures and tables refer to the corresponding TRACE document.

	Focks et al. (2014) Integrating chemical fate and population-level effect models for pesticides on the landscape scale: new options for risk assessment.	Johnston et al. (2014) An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides.	Meli et al. (2013) Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: an individual-based model of springtails.
TRACE element			
1. Problem formulation	The MASTEP-regional model will be used by risk managers and scientists who are interested in effects of pesticides at the landscape scale. The model provides the possibility to assess the effects of multiple applications of a pesticide in space and time, as being typical for agricultural landscapes, on the population dynamics of aquatic species. The approach does not provide yet endpoints that are immediately operational for environmental risk assessment. Further research about what new indicators could look like is recommended. Extrapolations to other species and landscapes are intended and supported by the model.	The acquisition and expenditure of energy to life cycle processes depends on a combination of environment- and organism-specific conditions. In addition, exposure of individuals to chemical stress can alter a population dynamics via physiological pathways. To investigate the sublethal effects of pesticides we develop and evaluate an energy budget agent-based model of the earthworm *Eisenia fetida*. *E. fetida* is used as a model species here due to its recommended use in lower tier toxicity tests, and therefore ample quantity of literature data available for model development at the individual level.	Suitable habitat for soil organisms may be scarce, thus leading to locally high population densities, because soil, being more static than water or air, is heterogeneous: physical conditions often vary widely on a scale of a few centimetres. Moreover, toxic chemicals are likely to be unevenly distributed in the soil as well. The spatially explicit individual-based model presented in Meli et al. (2013) is developed to explore the consequences of these heterogeneities for the population dynamics of soil invertebrates, in particular the collembolan *Folsomia candida*. *F. candida* is a common arthropod that occurs in soils worldwide and is used as a standard test organism for estimating the effects of pesticides on non-target soil arthropods.
2. Model description	The MASTEP-regional model is described in detail. The model provides a framework that compiles the definition of a landscape structure, pesticide exposure time series and a population model (ODD format) into landscape-scale simulations. Concrete examples for the subparts of the MASTEP-regional approach are given. An overview about the modelling approach is given at first.	Here we present the complete model description. It is identical to the one given in Johnston et al. (2014), but nevertheless included here because it is relatively short and allows to keep all supplementary information in one document.	Here we present the complete model description following the ODD format.
3. Data evaluation	The model was not calibrated to experimental data. Overall, the available data for parameterization of the model parts was taken from peer-reviewed literature. For the parameterization of the pesticide fate model, several scientific publications were evaluated. The population model was parameterized based on a number of scientific publications that focused, however, primarily on size and fecundity related aspects. Previous applications of the population model indicate its reasonability. However, information on density dependence and dispersal parameters are scarce. The link between exposure and effects was parameterized based on an appropriate scientific publication.	Energy budget parameters for *E. fetida* have been directly derived from relevant literature data. As much of the data does not directly relate to energy equivalents, calculations were necessary to transform the literature data in to compatible units. The parameters represent energy acquisition and expenditure under optimal and constant environmental conditions. In suboptimal conditions, environmental variables (e.g., food availability) limit energy ingestion and subsequent allocation to life cycle processes. Methods used to parameterize the dose–response relationship between pesticide concentration and physiological parameters are also outlined below.	All life-cycle parameters of *Folsomia candida* used in the model, with the exceptions of those related to the energy expenditures, have been directly derived from empirical data published in the literature, as well as individual-level toxicity data for copper. Qualitative observed patterns were also used to design the overall model structure.
4. Conceptual model evaluation	The MASTEP-regional model builds on existing models whose model concepts make quite some simplifying assumptions. These simplifying assumptions are not discussed in this document. The concept for the landscape-scaled approach of the MASTEP-regional follows from embedding an already existing model into a spatially realistic landscape. Only a few simplifying assumptions had to be made and are discussed.	The conceptual model is represented in Figs. 1 and 2. The design concepts underlying model design are presented in Section 2, Model description. Further information regarding simplifying assumptions is presented in Section 3, Data evaluation.	The conceptual model is represented in Fig. 1. The design concepts underlying model design are presented in Section 2, Model description. Further information regarding simplifying assumptions is presented in Section 3, Data evaluation.
5. Implementation verification	In addition to standard verification tests such as code check being performed for compilation, two main approaches were followed to ensure a correct implementation of the MASTEP-regional upscaling approach. A species balance calculation is performed for each time step to ensure that individual processes are correctly linked to the upscaling framework. Specific test simulations using manipulated code ensured further integrity of the model code.	In order to ensure that the computer code implementing the model works according to its specification in the ODD model description, a series of tests has been performed. These tests included syntax checking of the code, visual testing through NetLogo interface, the use of print statements and spot tests with agent and patch monitors to check against calculations in Excel, stress tests with extreme parameters values and environmental variables, chemical exposure and concentrations and independent code reviews.	In order to ensure that the computer code implementing the model works according to its specification in the ODD model description, a series of tests has been performed. These tests included syntax checking of the code, visual testing through NetLogo interface, print statements, spot tests with agent monitors, stress tests with extreme parameters values, test procedures and test programs, and code reviews.

Table 2 (*Continued*)

	Focks et al. (2014) Integrating chemical fate and population-level effect models for pesticides on the landscape scale: new options for risk assessment.	Johnston et al. (2014) An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides.	Meli et al. (2013) Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: an individual-based model of springtails.
6. Model output verification	In this study, no calibration of model parameters was executed in the sense of optimizing parameters to a given data set. Information on how well model simulations match observations are presented in Model output corroboration.	In this section it is described how many and which parameters were inversely determined via calibration. As the energy budget parameters in Table 1 were all directly calculated from literature data sources, information on these parameters are confined to Section 3, data evaluation. Here, details on the modelling of the toxicity submodels are presented. To inversely determine the most plausible toxicity submodel (by altering physiological parameters according to the dose–response relationships in Section 3, Data evaluation), we set up the model as in the corresponding empirical study and evaluated the model output against several patterns observed in the respective laboratory populations (following "Pattern-Oriented Modelling (POM)" and "Akaike Information Criterion (AIC)".	In this section it is described how many and which parameters were inversely determined via calibration. To inversely determine the values of these parameters we made the model reproduce several patterns observed in laboratory populations at different scales and levels of biological organization ("pattern-oriented modelling"; Grimm et al., 2005).
7. Model analysis	A comprehensive sensitivity analysis of the MASTEP-regional model is due to the relatively high computation times not possible. However, the sensitivity of the model outcomes was evaluated in a set of simulations covering a wide range of pesticide toxicity and persistence. The simulation results indicate a reasonable and meaningful response of the model.	The sensitivity of the model to the values of its parameters is presented in Table 9. The model was run with the parameter values of Table 1 ($N = 100$) and again with parameter values increased one at a time by 10% ($N = 100$) and changes in model outputs (adult biomass, juvenile biomass and cocoon production per adult) are shown in Table 9. Also shown in Table 9 are the sensitivity of the model to the baseline values of the environmental variables varied individually; these were soil temperature: 25 °C; soil moisture: 60%; and food density: 20 g per patch.	A sensitivity analysis was performed to explore the behaviour of the model in response to variations in the values of parameters that were not directly determined from the literature. Two different model outputs, final population size and average weekly population growth rate, have been used in this sensitivity analysis. Results are shown in Table 5.
8. Model output corroboration	Given the spatial dimension and resolution of the landscape-scale simulations, data that can be used to corroborate model results is hard to find. We used data from field monitoring campaigns in the Netherlands to corroborate at least the undisturbed population dynamics as simulated with a local MASTEP population model.	A number of patterns on the individual life cycle processes and population dynamics of *E. fetida* have been identified as reproducing well the available literature data. The studies used to evaluate model output use variable laboratory conditions (e.g., temperature, food density). The energy budget model is parameterized with data relating to optimal environmental conditions, and so good model fits to variable conditions show our model to realistically represent *E. fetida* physiology. Good model fits to sublethal effects of the pesticides copper oxychloride and chlorpyrifos further show the methods for identifying how chemicals achieve their effects. At the population level, good fits to population density, biomass and stage structure show the potential of the model to extrapolate to more natural conditions. Simulation details of all results are available in Johnston et al. (2014).	Three patterns have been identified from the literature, which have been numbered 3–5 to distinguish them from the patterns used for calibration (1–2).

intended for comparison to field observations, what statistical measures of field data will the model be compared to? The assessment of the availability of knowledge and data included in this element by Schmolke et al. (2010) is now covered by the third TRACE element, "Data evaluation".

2. *Model description*. TRACE documents should include a complete model description that allows users to easily and fully understand a model and, in principle, replicate it independently. We recommend using the ODD protocol (see above, Section 3). The ODD description should include a table with all model parameters, their meaning, units, reference values, range, and data source (if taken from publications, page numbers should be included). Parameters that were determined inversely via calibration should be clearly identified. Direct parameterization is documented in detail in the third TRACE element, 'Data evaluation', and inverse parameterization is documented in the eighth element, 'Model output verification'.

Any verbal model description is likely to include ambiguities that prevent full replication. Therefore, the model's computer code should be provided as well, usually in a separate file. If it is not possible to provide the code, e.g., if it is proprietary, the executable program and all data and script files needed to run the model should be made available. Any other information needed to run the software (e.g., platform version, operating system limitations) should be provided. This material could be part of the Supplementary Material or made

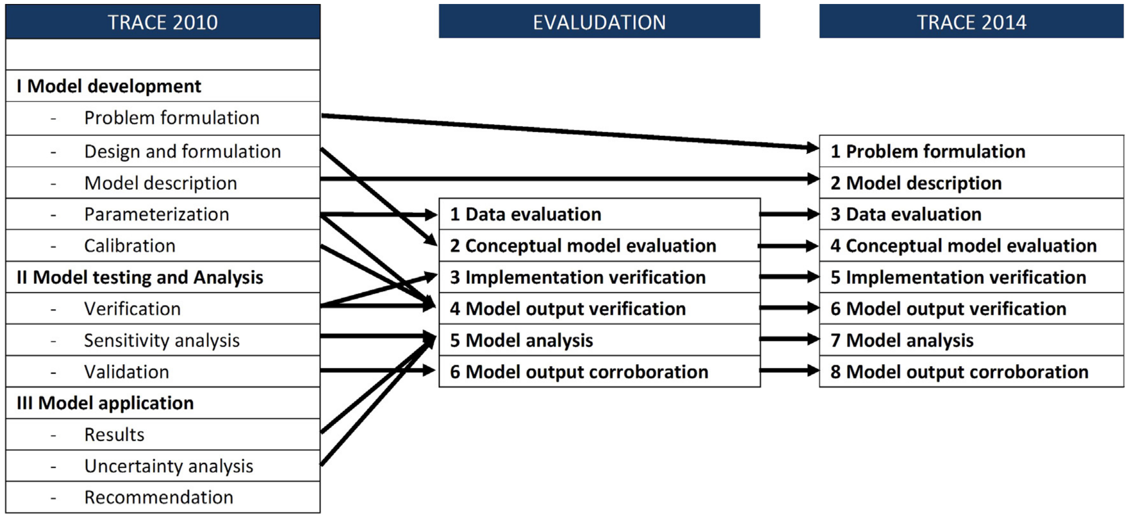

Fig. 2. Structure and terminology of the original TRACE format (Schmolke et al., 2010; TRACE 2010); model evaludation (Augusiak et al., 2014; EVALUDATION); and the new, updated TRACE format presented in this article (TRACE 2014).

available on permanent repositories such as the CoMSES Computational Model Library maintained by the OpenABM consortium (http://www.openabm.org/models).

3. *Data evaluation*. Augusiak et al. (2014) define 'data evaluation' as the "critical assessment of the quality of numerical and qualitative data used to parameterize the model, both directly and inversely via calibration, and of the observed patterns that were used to design the overall model structure". "Data" here refers both to quantitative data, taken from experiments, monitoring, or publications, and qualitative data, which often corresponds to expert knowledge. Data also include patterns in time, space, and organization, which are characteristic of the system to be represented ('pattern-oriented modelling'; Grimm et al., 2005; Grimm and Railsback, 2012).

   Concise text plus tables should summarize what data and knowledge went into the design and parameterization of the model, including references, data sources, and information about where and when those data were collected, and by whom. If possible, the reliability of the data used should be discussed, as data quality and ecological significance might be limited by measurement errors, inappropriate experimental design (e.g., number of replicates), and, in particular, the heterogeneity and variability inherent to environmental systems (Gass, 1983; Wang and Luttik, 2012). Likewise, expert knowledge and the detection of patterns are prone to bias and therefore must be treated with particular caution. The document should indicate which parameter values were used directly without calibration and which were determined inversely; the methods used for inverse parameterization will be described in the TRACE element, "Model output verification".

   The data description and evaluation allows model users to (1) see whether a model was mainly built on its authors' own data and knowledge, or on that of a certain expert or group of experts, or on a systematic evaluation of the literature, and (2) assess how uncertain the data are.

4. *Conceptual model evaluation*. This element is defined by Augusiak et al. (2014) as "the critical assessment of the simplifying assumptions underlying a model's design". The design of any mathematical or simulation model is based on a conceptual model which reflects our preliminary understanding and perception of the system to be represented in the model. For example, we may focus on nutrients and energy, species composition, or individual organisms. In this TRACE element, the underlying conceptual model should be described and its choice explained and justified. The evaluation applies to the overall model structure and sometimes to submodels, for example of metabolism, competition among individuals, movement, or the physical environment.

   In detail, this evaluation lists and explains the most important conceptual design decisions: spatial and temporal scales, selection of entities and processes, representation of stochasticity and heterogeneity, consideration of local versus global interactions, environmental drivers, etc. Moreover, conceptual models are often determined by certain theories, concepts, or, in particular, earlier models. Modellers should explain why they chose these elements and briefly contrast them, if applicable, to alternative conceptual designs that would have led to other model structures.

   Explaining and justifying conceptual models allows model users to understand that model design was not ad hoc but based on carefully scrutinized considerations. It makes users also aware that each model is only one of many possible ways to represent a certain system with regard to a certain question. Blind trust in a model can thereby be prevented, but so can blind distrust: even crude simplifying assumptions can be trusted if they are justified well.

5. *Implementation verification*. This term is defined by Augusiak et al. (2014) as "the critical assessment of (1) whether the computer code for implementing the model has been thoroughly tested for programming errors and (2) whether the implemented model performs as indicated by the model description". For instance, implementation verification might be conducted by peer-reviewing the code, i.e., other scientists thoroughly comparing it with the written formulation of the model, or by independently implementing submodels. This TRACE element provides evidence that the model software has been thoroughly tested and accurately implements the model description.

   A second component of implementation verification is documenting how the model's software has been designed to make it usable for the model's purposes. In addition to accurately implementing a model, its software often must also provide the graphical interfaces necessary to understand and test the model's behaviour (e.g., to see the behaviour of individuals in individual-based models), automate simulation experiments, be designed and documented to make modifications and maintenance easy, and be operable by clients.

6. *Model output verification*. Augusiak et al. (2014) define this element as "the critical assessment of (1) how well model output matches observations and (2) how much calibration and effects of environmental drivers were involved in obtaining good fits of model output and data". In developing any model, we try to make it reproduce some features or patterns of the real system before claiming that it is a good enough representation. In this TRACE element, we list the features we used plus the quantitative criteria for deciding whether a certain observation was matched by the model. Example features for population models include persistence, mean and standard deviation of population size, and metrics of size, age, or spatial distributions. The more observed features or patterns a model can reproduce simultaneously, the higher the chance that it has captured the internal organization of the real system sufficiently well ('pattern-oriented modelling'; Grimm et al., 2005; Grimm and Railsback, 2012).

   Output verification involves what often is referred to as 'face validation' and more formal tests. Face validation can be defined as: "All methods that rely on natural human intelligence" (Klügl, 2008, p. 39). Examples listed by Klügl (2008) include: "Structured walk-throughs, expert assessments of descriptions, animations of results". Klügl (2008) accordingly concludes that face validity shows that a model's processes and outcomes are reasonable and plausible within its theoretical basis and the knowledge of system experts or stakeholders. It should be noted, however, that system experts and stakeholders may disagree on the type of data and knowledge they have. Therefore more formal tests are required that are based on multiple quantitative criteria for a model matching data (e.g., Railsback and Grimm, 2012, Chapter 20.4.2).

   Evaluation of output verification needs to consider such concerns as over-fitting and extrapolation. The higher the proportion of calibrated, guesstimated, or uncertain parameters (see TRACE element 'Model analysis' below), the higher the risk that the model seems to work correctly (e.g., because it fits calibration data well) but for the wrong reasons, i.e., has not captured the mechanisms of the real system. Moreover, it is important to distinguish between system-level parameters and those related to lower level processes. A population model, for example, may be based on empirically determined demographic rates, but this restricts the scope of the model to environmental conditions under which those rates were determined. In contrast, if submodels, for example foraging or habitat selection, are parameterized for a wider range of environmental conditions, population-level phenomena are no longer imposed but emerge and the population model can be expected to predict responses to new conditions more reliably (Grimm and Railsback, 2005; Railsback and Grimm, 2012; Grimm and Martin, 2013).

   Finally, a good match of model output to data can sometimes simply reflect the overarching influence of environmental drivers. For example, if the egg-laying rate of a honeybee queen follows uni-modal seasonal dynamics, colony size will vary accordingly and thus look realistic, but this does not indicate that all other processes included in a honeybee colony model have been captured realistically enough (Becher et al., 2013). Thus, example model runs should be presented along with time series of important environmental drivers.

   Model developers naturally often claim that their models are realistic enough for their purpose, but in this TRACE element they should summarize why they believe so, with supporting evidence. This information enables users to scrutinize the modeller's claim and to critically assess how well model output matches observations, the degree to which the match results from calibration and environmental drivers, and how much the model's reliability is limited by use of empirical parameters that reflect only a narrow range of conditions.

7. *Model analysis*. This element is defined by Augusiak et al. (2014) as "the assessment of (1) how sensitive model output is to changes in model parameters (sensitivity analysis), and (2) how well the emergence of model output has been understood". The purpose of the element is to prevent blind trust in the model output by asking "How did this output emerge?", and to challenge the model, which might look impressive, by asking "Does verification still look good if I change one or more parameters a bit?"

   Thus, foremost here we document how we made sure that we understood a model's main mechanisms. For example, if recovery after disturbance is strongly affected by a certain parameter and, thus, the processes the parameter represents, we should be able to explain why this parameter was so important. We can learn much about a model by performing controlled simulation experiments: keeping most parameters constant and varying one or a few over a wider range, and exploring the effect on one or more output variables. Simulation experiments should also include simplified model versions, in which the environment is made more homogenous and constant, system size is reduced, and certain processes are deactivated. Initial conditions and input data are other model components to which sensitivity should be analyzed.

   TRACE should not include details on all these experiments, but give an overview of what kind of experiments were performed and present results from experiments that significantly increased understanding but could not be included in the paper or report.

   Local sensitivity analysis is important for developing a first understanding of a model by evaluating how sensitive output is to small changes in one parameter at a time. The analysis can produce conclusions about model uncertainty: if the parameters to which the model is most sensitive are the most uncertain ones, the entire model will be quite uncertain. Moreover, such parameters indicate which processes are most important for certain model outputs.

   By varying more than one parameter at a time, local sensitivity analysis gradually becomes global analysis, which captures interactions among parameters by examining the entire parameter space, not only the local neighbourhood of a default parameter set. Run time, complexity, and stochasticity often limit global sensitivity analysis, but it should be performed for at least a subset of parameters. One way to summarize such sensitivity analyses is regression modelling, which quantifies the relative influence of parameters on model output. Uncertainty analysis can augment sensitivity analysis by demonstrating how uncertainty in model parameters translates into uncertainty in model output.

   Parameters often represent entire processes that the modeller chose not to represent explicitly. Submodels represent processes that are represented explicitly in more detail; therefore, sensitivity analysis should also be applied to important submodels by contrasting alternative submodels. For example, a submodel describing movement might be based on complex decision making, but contrasting this submodel with simpler, or even more complex, alternatives can provide insights into how important or useful it was to choose this very model design. This sensitivity analysis of submodels corresponds to what Railsback and Grimm (2012) refer to as 'pattern-oriented theory development': which submodel best causes the full model to reproduce a set of observed patterns?

   Model users learn from this TRACE element how the model works, i.e., which processes and process interactions are important and explain major behaviours of the model system.

Moreover, users learn how robust model results are to uncertainties in model parameters and submodel formulation.

8. *Model output corroboration*. This term is defined by Augusiak et al. (2014) as the "comparison of model predictions with independent data and patterns that were not used, and preferably not even known, while the model was developed, parameterized, and verified". Most scientists, in particular non-modelers, require this analysis, calling it 'validation': for a model to be trusted it should make predictions that are subsequently confirmed in empirical experiments. Indeed, we consider this the 'gold standard' for demonstrating that a model has captured the internal organization of a system sufficiently well. Corroboration is discussed in more depth by Augusiak et al. (2014).

   Model output verification always includes 'tweaking', i.e., we try to make a model reproduce certain observations by tuning parameters, environmental settings, and submodel formulation. Such adjustments are often necessary to compensate for processes not included in a model (due to insufficient information or to keep the model simple) but were important in the real system when the verification data were collected. Making a model simultaneously reproduce multiple observed patterns reduces the risk that the model is completely unrealistic, but does not eliminate this risk. Only when a model predicts phenomena that we even did not think about during model development and testing do we have the strongest indicator of its structural realism, because no tweaking could have been involved.

   However, achieving this standard is rarely possible with ecological systems because the empirical experiments are infeasible: we often build models to address questions such as response to climate change exactly because empirical experiments are impossible. Instead, we can directly test independent predictions of submodels. At the system level, we can identify characteristic patterns in model output that are robust and seem characteristic. Then, we can consult the literature or experts to find out how accurate these independent predictions are.

   Documenting model output corroboration provides model users evidence, in addition to model output verification, indicating the extent to which the model is structurally realistic so that its predictions can be trusted. The model's purpose should be a primary consideration in determining what model results need corroboration and how quantitatively and closely model results need to reproduce observations. If no corroboration was possible, the modeller should discuss here why, and why and to what degree the model still can be trusted. A classic example of model output verification that could be trusted was the structure of the DNA (a conceptual, not numerical, model), which Watson and Crick identified as a double helix because this structure was compatible with several patterns observed in DNA and its elements (Watson, 1968). This verification was convincing enough, even without independent predictions, which were made and tested only later.

5. Examples

Three example TRACE documents are in the Supplementary Material. They were not produced from scratch, but from existing TRACE documents produced according to the original TRACE format described in Schmolke et al. (2010); those documents are supplementary material to Focks et al. (2014), Johnston et al. (2014), and Meli et al. (2013). Table 2 summarizes the three TRACE documents.

6. Discussion and conclusion

Our experience producing TRACE documents following Schmolke et al. (2010) led to a revised terminology, structure, and rationale for TRACE. The most important new feature is the link to the framework of model evaludation (Augusiak et al., 2014).

We hope that the new format is easier to use than the original and that the resulting documents are more efficient to use and understand by model clients, so that clients can better assess whether or not a model is realistic and robust enough to let it influence decisions affecting the real world. The three example TRACE documents we provide follow the new format and terminology. However, they were compiled mostly from documents that followed the original TRACE format and terminology; TRACE documents following the new format and rationale from the start should be even more comprehensive and clear. Moreover, modelling notebooks following the new format should directly lead to more thorough model development, testing, and analysis because TRACE now provides a detailed checklist of all elements of modelling that have an influence on a model's credibility and usefulness.

In Table 2 we compiled the summaries of each of the eight TRACE elements. Similar tables might be a good way to summarize all the work that went into making a model fit for its purpose in the main text, or its printed appendix. However, Table 2 as well as TRACE in general do not provide criteria for when, for example, model output verification is good enough. TRACE by itself thus does not constitute good modelling practice. Nevertheless, the development of TRACE is the first step towards developing guidance and criteria for good modelling practice (Schmolke et al., 2010). It might be possible to provide more detailed guidance for at least some TRACE elements, for example providing a checklist for implementation verification, or calculating an index that quantifies the proportion of calibrated versus uncalibrated parameters. To try this, though, a critical number of TRACE documents are needed; for the ODD protocol, an update and more specific guidance became possible after the protocol had been used about 50 times (Grimm et al., 2010).

TRACE is not intended to establish, in the end, good modelling practice that corresponds to Good Laboratory Practice (GLP). GLP is a formalized means for ensuring a defined quality of chemical tests by standardizing every single step of analysis. Ecological models, however, are completely different from chemical analyses (Wikipedia Contributors, 2013); they are scientific tools, and as such not amenable to something like GLP. Standardizing the documentation of model development, testing and analysis does not mean standardizing models; likewise, standardizing the structure of scientific articles or of the description of individual- or agent-based models using the ODD format does not impose any restrictions on scientific creativity.

Nevertheless, the purpose of TRACE is to establish a culture of model development, testing, and analysis more likely to produce models that are useful, in particular for supporting environmental decision making. TRACE thus is intended to establish expectations of what modellers should clearly communicate when presenting their model, for example a clear model description, sensitivity analyses, and a detailed description of the empirical information that went into the model's design and testing. "Culture" here means that you just do all these things as well as you can because you know that peers and model clients are expecting you to; there is no point any more in complaining about "additional effort" for these things.

In empirical sciences, results cannot get published until methods – including quality control – are fully described. Laboratory experiments require evaluation of instrument error, reagent reliability, etc.; field experiments require evaluation of observation error; and data analysis typically requires comparison of alternative models and evaluation of error and uncertainty. Similarly, the culture of good modelling practice mentioned above already exists in many fields. As ecological modelling matures as a scientific (and regulatory) approach, we must expect the same kind of scrutiny of methods as clients become more sophisticated and more demanding of careful practice. In fact, standards for publishing models and accepting their results have increased by several leaps already since the beginning of computer modelling. Further increases in the sophistication with which clients scrutinize models must be expected as models are used for increasingly high-consequence assessments such as predicting effects of climate change and pesticides. As a guide to how to model (beyond its role in documentation), TRACE should be especially valuable for ecologists and other scientists who are self-taught or otherwise lack training in modelling skills such as software testing and model analysis.

We hope that the new TRACE format presented in this article will be widely used, so that it can further be developed and refined. To facilitate TRACE's refinement, the template provided in the Supplementary Material should be used unchanged. Furthermore, in parallel to TRACE, establishing a culture of keeping modelling notebooks (Grimm et al., *unpubl. manuscript*) that use TRACE terminology will also improve the culture of ecological modelling.

Acknowledgements

We would like to thank N. Galic and two anonymous reviewers for helpful comments on earlier drafts of this manuscript, and all participants of the CREAM project for their comments on TRACE, in particular R. Ashauer, K. Kułakowska, T. Kułakowski, A. Nyman and J. Stadnicka. This work has been financially supported by the European Union under the 7th Framework Programme (Project acronym CREAM, contract number PITN-GA-2009-238148).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2014.01.018.

References

Augusiak, J., van den Brink, P.J., Grimm, V., 2014. Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. Ecological Modelling 280, 117–128.

Becher, M.A., Osborne, J., Thorbek, P., Kennedy, P.J., Grimm, V., 2013. Towards a systems approach for understanding honeybee decline: a stock-taking and synthesis of existing models. Journal of Applied Ecology 50, 868–880.

Focks, A., ter Horst, M., van den Berg, E., Baveco, H., van den Brink, P.J., 2014. Integrating chemical fate and population-level effect models for pesticides at landscape scale: new options for risk assessment. Ecological Modelling 280, 102–116.

Forbes, V.E., Calow, P., 2012. Promises and problems for the new paradigm for risk assessment and an alternative approach involving predictive systems models. Environmental Toxicology and Chemistry 31, 2663–2671.

Gass, S.I., 1983. Decision-aiding models: validation, assessment, and related issues for policy analysis. Operations Research 31, 603–631.

Grimm, V., Railsback, S., 2005. Individual-based Modeling and Ecology. Princeton University Press, Princeton, NJ.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jorgensen, C., Mooij, W.M., Muller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Ruger, N., Strand, E., Souissi, S., Stillman, R.A., Vabo, R., Visser, U., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198, 115–126.

Grimm, V., Ashauer, R., Forbes, V., Hommen, U., Preuss, T.G., Schmidt, A., van den Brink, P.J., Wogram, J., Thorbek, P., 2009. CREAM: A European project on mechanistic effect models for ecological risk assessment of chemicals. Environmental Science and Pollution Research 16, 614–617.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F., 2010. The ODD protocol: a review and first update. Ecological Modelling 221, 2760–2768.

Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a "multi-scope" for predictive systems ecology. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 367, 298–310.

Grimm, V., Martin, B.T., 2013. Mechanistic effect modeling for ecological risk assessment: where to go from here? Integrated Environmental Assessment and Management 9, e58–e63.

Johnston, A.S.A., Hodson, M.A., Thorbek, P., Alvarez, T., Sibly, R.M., 2014. An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides. Ecological Modelling 280, 5–17.

Kanare, H.M., 1985. Writing the Laboratory Notebook. American Chemical Society, Washington, DC.

Klügl, F., 2008. A validation methodology for agent-based simulations. In: SAC Symposium, Advances in Computer Simulation Track, Ceara, Brazil, March 2008.

Liu, C., Sibly, R.M., Grimm, V., Thorbek, P., 2013. Linking pesticide exposure and spatial dynamics: an individual-based model of wood mouse (*Apodemus sylvaticus*) populations in agricultural landscapes. Ecological Modelling 248, 92–102.

Martin, B.T., Jager, T., Nisbet, R.M., Preuss, T.G., Grimm, V., 2013. Predicting population dynamics from the properties of individuals: a cross-level test of Dynamic Energy Budget theory. American Naturalist 181, 506–519.

Meli, M., Auclerc, A., Palmqvist, A., Forbes, V.E., Grimm, V., 2013. Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: an individual-based model of springtails. Ecological Modelling 250, 338–351.

Meli, M., Palmqvist, A., Forbes, V.E., Groeneveld, J., Grimm, V., 2014. Two pairs of eyes are better than one: combining individual-based and matrix models for ecological risk assessment of chemicals. Ecological Modelling 280, 40–52.

Pilkey, O.H., Pilkey-Jarvis, L., 2007. Useless Arithmetic: Why Environmental Scientists Can't Predict the Future. Columbia University Press.

Railsback, S., Grimm, V., 2012. Agent-based and Individual-based Modelling: A Practical Introduction. Princeton University Press, Princeton, NJ.

Railsback, S.F., Harvey, B.C., 2002. Analysis of habitat selection rules using an individual-based model. Ecology 83, 1817–1830.

Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V., 2010. Ecological models supporting environmental decision making: a strategy for the future. Trends in Ecology and Evolution 25, 479–486.

Wang, M., Luttik, R., 2012. Population level risk assessment: practical considerations for evaluation of population models from a risk assessor's perspective. Environmental Sciences Europe 24, 3 http://www.enveurope.com/content/24/1/3

Watson, J., 1968. The Double Helix: A Personal Account of the Discovery of the Structure of DNA. Atheneum, New York.

Wikipedia Contributors, 2013. Good Laboratory Practice. Wikipedia, The Free Encyclopedia http://en.wikipedia.org/w/index.php?title=Good_Laboratory_Practice&oldid=583104663 (retrieved 07.12.13).