REVIEW ARTICLE

Deliberative processes for comprehensive evaluation of agroecological models. A review

Gianni Bellocchi · Mike Rivington · Keith Matthews · Marco Acutis

Abstract The use of biophysical models in agroecology has increased in the last few decades for two main reasons: the need to formalize empirical knowledge and the need to disseminate model-based decision support for decision makers (such as farmers, advisors, and policy makers). The first has encouraged the development and use of mathematical models to enhance the efficiency of field research through extrapolation beyond the limits of site, season, and management. The second reflects the increasing need (by scientists, managers, and the public) for simulation experimentation to explore options and consequences, for example, future resource use efficiency (i.e., management in sustainable intensification), impacts of and adaptation to climate change, understanding market and policy responses to shocks initiated at a biophysical level under increasing demand, and limited supply capacity. Production concerns thus dominate most model applications, but there is a notable growing emphasis on environmental, economic, and policy dimensions. Identifying effective methods of assessing model quality and performance has become a challenging but vital imperative, considering the variety of factors influencing model outputs. Understanding the requirements of stakeholders, in respect of model use, logically implies the need for their inclusion in model evaluation methods. We reviewed the use of metrics of model evaluation, with a particular emphasis on the involvement of stakeholders to expand horizons beyond conventional structured, numeric analyses. Two major topics are discussed: (1) the importance of deliberative processes for model evaluation, and (2) the role computer-aided techniques may play to integrate deliberative processes into the evaluation of agroecological models. We point out that (i) the evaluation of agroecological models can be improved through stakeholder follow-up, which is a key for the acceptability of model realizations in practice, (ii) model credibility depends not only on the outcomes of well-structured, numerically based evaluation, but also on less tangible factors that may need to be addressed using complementary deliberative processes, (iii) comprehensive evaluation of simulation models can be achieved by integrating the expectations of stakeholders via a weighting system of preferences and perception, (iv) questionnaire-based surveys can help understand the challenges posed by the deliberative process, and (v) a benefit can be obtained if model evaluation is conceived in a decisional perspective and evaluation techniques are developed at the same pace with which the models themselves are created and improved. Scientific knowledge hubs are also recognized as critical pillars to advance good modeling practice in relation to model evaluation (including access to dedicated software tools), an activity which is frequently neglected in the context of time-limited framework programs.

G. Bellocchi (✉)
Grassland Ecosystem Research Unit, French National Institute for Agricultural Research, 5 Chemin de Beaulieu, 63039 Clermont-Ferrand, France
e-mail: gianni.bellocchi@clermont.inra.fr

M. Rivington · K. Matthews
The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK

M. Acutis
Department of Agricultural and Environmental Sciences—Production, Landscape, Agroenergy, University of Milan, Milan, Italy

Contents

1 Introduction

Agroecological systems form the fundamental basis for the provision of food for society and thus underpin human well-being. Under the increasing pressures of food demand and declining supply of ecosystem services, their sustainable use and management are an essential research priority. Agroecological systems are ecosystems manipulated by anthropogenic modifications, based on species diversity (Fig. 1, actual system) in which plants, vertebrate and invertebrate animals, and microbial communities cohabit with agricultural uses (such as crops, livestock, and orchards) in the context of soils, and provide ecosystem services (Gliessman 2007). Agroecological models are representations of agroecological systems providing an integrated perspective in which the biotic elements (e.g., communities of closely spaced plants) dynamically interact with soil, weather, and management factors (Fig. 1, model inputs and outputs) to simulate system feedbacks and revise approaches (Fig. 1, modeling flow). Nowadays

agroecological models are software simulation tools summarizing our understanding of how systems operate. They include sophisticated routines for simulating soil water and nitrogen balances as well as plant phenology, biomass accumulation, and partitioning. The way how they evolve over time is challenging because which processes are incorporated into which detail in a model depends on the purposes of the model, which change as knowledge progresses and new questions arise. For instance, abiotic stresses affecting plant growth and productivity (temperature, water, nutrients, etc.) only received attention after the simulation of non-stress conditions was satisfactory (e.g., Tanner and Sinclair 1983; Steduto et al. 2009). Also, early models were not much concerned by the effect of elevated atmospheric CO_2 concentration on plant growth and transpiration and on the partitioning of dry matter. The global carbon balance has become an issue of great societal concern (and an important focus of scientific research) during the last decades, when the global emission of CO_2 has continued to increase together with its impact on climate (IPCC 2013). This has required modeling efforts, for instance, to represent mechanistically the response of leaf photosynthesis to CO_2 levels (e.g., Ethier and Livingston 2004) and so make the models responsive to the changing environmental conditions (Asseng et al. 2013; Bassu et al. 2014; Li et al. 2014). The impact of biotic stress factors (pests, diseases, and parasites) has also received great attention recently because their development can be modified by climate change (e.g., Maiorano et al. 2014). The control of these organisms forms an integral part of agricultural production and ecological processes and represents (together with the factors that may hamper soil fertility maintenance) a need in the development of modeling tools for the evaluation and design of sustainable agroecological systems (Colomb et al. 2013). The recent advances in scale change methodology (for scaling the information provided by models from the field up to the regional level) have helped building ranges of indicators for the assessment of the contribution of agricultural landscapes (mosaics of agroecological systems) to a sustainable development (Chopin et al. 2014).

Triggered by the need to answer new scientific questions, but also to improve the accuracy of simulations, model improvement is expected to continue resulting from emerging new questions, better knowledge of physiological mechanisms, and higher accuracy standards (Lizaso 2014). However, the traditional paradigm in modeling for which modelers analyze the system and developers produce algorithms and programs that they believe would do the best job has weaknesses of its own. In so doing, in fact, scientists and experts often drive the research focus with a narrow or incomplete understanding of the information needs of end users, resulting in research findings that are poorly aligned with the information needs of real-world decision makers (Voinov and Bousquet 2010). The practice of stakeholder engagement in

Fig. 1 The elements of an agroecological system and a scheme of the modeling process. The basic research provides elements to model development to build formal equations (*1*), characterized by parameters (*2*), which are further translated into software code (*3*) to run simulations (*5*) under conditions represented by environmental factors (*4*). Model users perform evaluation of model adequacy (*6*), which is preliminary to the use of model outputs by third parties (*7*)
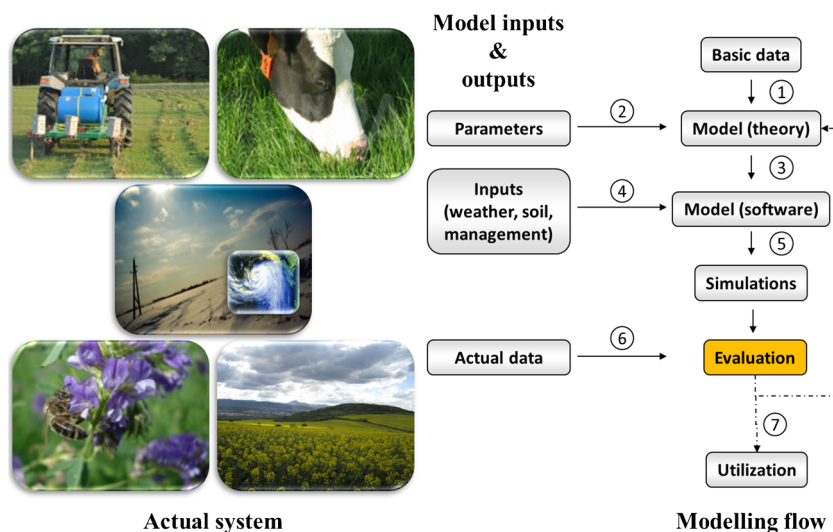
model evaluation seeks to eliminate this divide by actively involving stakeholders across the phases of the modeling process (Fig. 1, modeling flow) to ensure the utility and relevance of model results for decision makers (step 7). As such, stakeholder engagement is a fundamental, and perhaps defining, aspect of model evaluation (step 6, orange box). The evaluation of model adequacy is an essential step of the modeling process, either to build confidence in a model or to select alternative models (Jakeman et al. 2006). The concept of evaluation, in spite of controversial terminology (Konikow and Bredehoeft 1992; Bredehoeft and Konikow 1993; Bair 1994; Oreskes 1998), is quite generally interpreted in terms of model suitability for a particular purpose, which means that a model is valuable and sound if it accomplishes what is expected of it (Hamilton 1991; Landry and Oral 1993; Rykiel 1996; Sargent 2001) or helps achieve a successful outcome.

In this paper, model evaluation is discussed in its effectiveness to support modeling projects which are applicable across a broad range of subjects in agroecology. In particular, the role of non-numerical assessment methods, such as deliberative processes, is explored (Sect. 2). Deliberative processes are often used for implementing a participatory approach to decision making in natural resource management as described, for instance, by Petts (2001) for waste management and Liu et al. (2011) for managing invasive alien species. In general, the goal is enhancing institutional legitimacy, citizen influence and social responsibility, and learning. In model evaluation (the focus of this paper), the impact of stakeholder engagement depends on developing effective processes and support for the meaningful participation of stakeholders throughout the continuum of analysis, from setting priorities to study design, to research implementation, and the dissemination of model outcomes. The approach engenders a discursive, reasoning, discussing, double-side learning, and, consequently, influencing countenance, which should come to the fore in stakeholder-oriented evaluation. It can include a range of

different interests and concerns and allow for deliberation about "how to model this" and "what model works best" but also to understand "why this model works" or "how it could work better" (after Creighton 1983). Computer-aided evaluation may help integrate deliberative processes into the evaluation of agroecological models in a systematic way. The role of computer-aided support is considered (Sect. 3), with a focus on the integration of modular tools for evaluation within the overall modeling process. Two examples are provided in Sect. 4 to illustrate research projects with a clear involvement of stakeholders in the evaluation of agroecological models.

2 Deliberative processes for comprehensive model evaluation

While the methodologies reviewed in other papers (e.g., Bellocchi et al. 2010; Richter et al. 2012) represent the state of the art in terms of structured numerically based evaluation, it cannot be assumed that this analysis is all that is required for model outputs to be accepted particularly when models are used with and for stakeholders. The numerical analysis may provide credibility within the technoscientific research community. Yet, while necessary, this may be insufficient to achieve credibility with decision makers and other stakeholders. Possibly, a real test of model value is whether stakeholders have sufficient confidence in the model to use it as the basis for making management decisions (Vanclay and Skovsgaard 1997). Figure 2 is representative of the role that science-stakeholders dialogue may play in a comprehensive model evaluation.

The workflow is centered on the evaluation of agroecological models (crop and grassland models are taken as examples) by using data from the experimental and observational research, as well as socioeconomic and climate scenarios (considering that the problem of global changes has generated

Fig. 2 Model evaluation with deliberative process. Evaluation involves the gathering of information from a wide variety of sources, generally including consultation with people outside the modeling team. It also weights the information and facts obtained with a view to evaluating a model or reflecting upon the reasons underlying any particular model realization. As well, it may refer to the way decisions are taken based on model outcomes
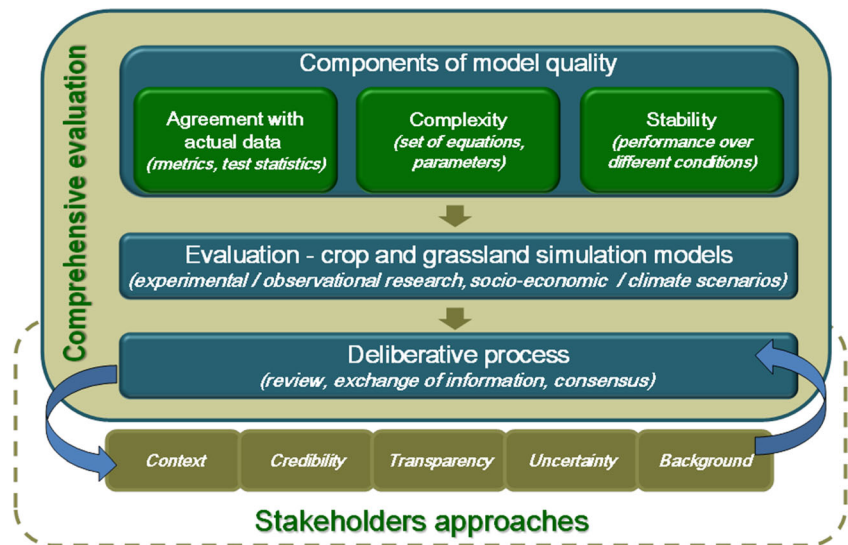
much interest in the use of model estimates for decision support and strategic planning). This step should integrate different components of model quality, not only the agreement between model outputs and actual data (based on metrics and test statistics), but also elements of model complexity (essentially related to the number and relevance of parameters in a model, which is pertinent for model comparison, Confalonieri et al. 2009), as well as an assessment of the stability of model performance over a variety of conditions (Confalonieri et al. 2010). The deliberative process consists for a group of actors (stakeholders) to receive and exchange information, to critically examine an issue, and to come to an agreement which will inform decision making (after Fearon 1998). The definition of group of actors is intended to be broadly applicable to all stakeholder groups (presented in Box 1) that may be involved in model evaluation in the agroecological domain, understanding that their information needs and eventual utilization of model results may differ.

Box 1

Definitions for stakeholder and stakeholder engagement in the context of model evaluation
• Stakeholders
  ◦ Individuals, organizations, or communities that have a direct interest in the model outcomes
• Stakeholder engagement
  ◦ An iterative process of actively soliciting the knowledge, experience, judgment, and values of individuals selected to represent a broad range of direct interests in a particular issue, for the dual purposes of the following:
    ▪ Creating a shared understanding of model outputs
    ▪ Making relevant, transparent, and effective decisions based on model results

---

A related foundational step is to also gain agreement regarding the general categories of stakeholders that are essential to the evaluation of agroecological models. Stakeholder

categories that should be considered for inclusion in model evaluation with brief descriptions and examples are listed in Box 2, recognizing that these categories may not be exhaustive or mutually exclusive. For example, environmental sciences industry stakeholders have their own category but also frequently fund research activities (e.g., carbon trade companies invest in projects that reduce greenhouse gas emissions).

Box 2

Defining the stakeholders: examples from agroecological modeling

Types and role	Stakeholders	Potential interests
Farmers and their agents	Persons or groups which represent the producer perspective generally or within specific situations such as land owners, farm workers, unions, and farmers' associations	Fate of agricultural chemicals; risk factors assessment and reduction; risk response (e.g., loss of income)
Food distributors and processors	Individuals or groups which represent the agricultural marketing perspective such as food wholesalers and retailers and transport companies	Risk response (e.g., breaks in the supply chain)
Environmental sciences industry	Profit entities that develop and market environmental services (e.g., tradable carbon quotas) as measured through scientific studies	Risk management
National/ regional agencies	Bodies that act on behalf of, or an instrument of the State, either nationally or	Risk management, regulation, and communication

	regionally such as public health institutions, food standards authorities, environmental regulatory authorities, occupational health and safety authorities, local/regional health boards and environmental health departments, ministries of agriculture and environment, local authorities	
Supranational and international agencies	Organizations that create, monitor, and oversee policies or regulations (policy makers) on agroecology-related issues such as European Commission (directorates for agriculture, environment, health, climate, energy, research), World Health Organization, and Food and Agriculture Organization	Risk management, regulation (e.g., prices of agricultural commodities), and communication (e.g., expected crop yield losses and level of agricultural stocks)
Non-governmental organizations	Organizations that are neither a part of a government nor a for-profit business such as environment action groups, organic farming groups, and animal welfare groups	Risk communication and regulation; lobbying for action
Research funders	Public or private entities that provide monetary support for research efforts (implying model development and model-based analyses) such as governments, foundations, and for-profit organizations	Scientific and experimental evidence
Others	People related to agricultural and rural development or providing advice and services such as rural residents, national and local media, and scientists (epidemiologists, toxicologists, environmental scientists, etc.)	Risk communication, analysis, and response

The implication of stakeholders of different nature in this process would expand the horizon of model evaluation (after Balci and Ormsby 2002) up to considering aspects related to

the specific context of research (Sect. 2.1), the credibility of model outputs when exploited for given purposes (Sect. 2.2), the transparency of the modeling process (Sect. 2.3), and the uncertainty associated with model outputs (Sect. 2.4), until a critical examination of the scientific background behind the models used (Sect. 2.5). Stakeholder approaches to model evaluation can also differ, as discussed in Sect. 2.6. These concepts, developed and set out to represent applications of environmental modeling (Matthews et al. 2011), have been clarified and put into the context of modern views on model evaluation, which opens up to the development of supporting and analytical tools (and the processes of using these tools) standing on the frontier of science and decision making (Matthews et al. 2013).

2.1 Dependence on the context

That the context within which models are used will affect the required functionality and/or accuracy is well recognized by model developers (French and Geldermann 2005). This is particularly apparent when comparing models developed to represent the same process at different scales and for which different qualities of input variables, parameterization/initialization, and data for evaluation will be available, for example, soil water balances at plot, farm, catchment, and region (e.g., Keating et al. 2002; Vischel et al. 2007). This has led to the development of application-specific testing of models and the idea of model *benchmarking*, by comparing simulation outputs from different models, where outputs from one simulation can also be accepted as a "standard" (based on previous evaluations, e.g., Vanclay 1994). Such approaches typically use multicriteria assessment (e.g., Reynolds and Ford 1999) with performance criteria weighted by users depending on their relative importance. Benchmarking tools are associated with alternative options for modeling (Hutchins et al. 2006).

Beyond the aspects of model performance covered by benchmarking, however, there are a range of factors that are increasingly being recognized as having a considerable effect on the use of models and their outputs and which mean that a case can be made for a wider consideration of how models are evaluated. The frequent failure for models and other model-based tools such as Decision Support Systems (DSS) to be seen as credible sources of information has been variously attributed to their lack of transparency, complexity, and difficulty of use (Rotmans and van Asselt 2001) and ultimately to the problem of implementation (McCown 2002a; Matthews et al. 2008b). Yet, despite advances in the documentation of modeling procedures such as Harmoni-QuA (http://harmoniqua.wau.nl) and model testing (Bellocchi et al. 2010) and the increasing sophistication of (and access to) human-computer interfaces and modeling tools (e.g., modeling platforms, http://www.gramp.org.uk; https://www6.inra.fr/record_eng/Download;http://mars.jrc.ec.europa.eu/mars/

About-us/AGRI4CAST/Models-Software-Tools/ Biophysical-Model-Application-BioMA), there still remains a significant distrust of model-based outputs with many stakeholders and decision makers.

2.2 Model credibility

One of the principles of evaluating models dictates that complete testing is not possible (Balci 1997); thus, to prove that a model is absolutely valuable is an issue without a solution. Exhaustive evaluation requires testing all possible model outputs under virtually all possible input conditions. Due to time and budgetary constraints, exhaustive testing is frequently impossible. Consequently, in model evaluation, the purpose is to increase confidence that the accuracy of the model meets the standards required for a particular application rather than establish that the model is absolutely correct in all circumstances. This suggests that the challenge for model evaluation is that in addition to ensuring that minimal (application-specific) standards are met, the testing should also increase the credibility of the model with users and beneficiaries while remaining cost-effective. As a general rule, the more tests that are performed in which it *cannot* be proven that the model is incorrect, the more confidence in the model is increased. Yet, the low priority typically given to evaluation in model project proposals and development plans indicates a tendency toward the minimum standard approach alone being adopted.

Where models are used for decision support or evidence-based reasoning, the credibility of estimates is a key to the success of the model. Credibility is a complex mix of social, technological, and mathematical aspects that requires developers to include social networking (between developers, researchers, and end users/stakeholders) to determine model rationale, aim, structure, etc. and, importantly, a sense of co-ownership. Drawing on experience within the agricultural DSS paradigm (McCown 2002a; McCown et al. 2005) and earlier research on the use of models within industrial manufacturing processes (McCown 2002b) have shown that the limited use of models is often due to their lack of *credibility*. One key component to this credibility is that the model should represent the *situated internal practice* of the decision maker. This means that models should, first of all, make available all the key management options that the decision maker considers important. Secondly, it should respond to an acceptable degree to management interventions in a way that matches with the decision maker's experience of the real system. In terms of models of natural process, management can be substituted with alternatives, such as external shocks, perturbations to the drivers of the system (e.g., climate change). The representation of the system, however, needs not be perfect, since decision makers are used to dealing with complex decisions in information poor or uncertain environments, but must not clash with established expectations (or clash only in specific intended aspects, Matthews et al. 2008b). It is further argued that credibility of model-based applications also depends on their ability to fit *within* and contribute to *existing processes* of decision making (McCown 2002a). This can be particularly challenging, as such processes may impose time constraints that may be difficult to meet and also require model developers or associated staff to be proactive in seeking application for their models in their validity domain (something that they may not be trained or indeed funded to do). Model developers may also find that decision makers are reluctant to concede agency (McCown 2002a) to software tools, however well evaluated, since their professional standing is at least partially based on their ability to make complex judgment-based decisions. The need to widen the consideration within evaluation merely from what the model does to how it will be used *by*, *with*, or *for* stakeholders is therefore essential.

2.3 Model transparency

While lack of transparency is frequently cited as the reason for the failure of model-based approaches, it is important to challenge some of the assumptions and conclusions that are drawn on how to respond to the issue of transparency. One response is to make models simpler, and hence, the argument goes easier to understand. Yet, while simplicity is in itself desirable (Raupach and Finnigan 1988) and the operation of simpler models may indeed be easier to understand, it may well be that the interpretation of their outputs is not simpler, and indeed, their simplicity may mean that they lack the capability to provide secondary data which can ease the process of interpretation. There is also a trade-off between simplicity and flexibility, and this flexibility may be a crucial factor in allowing the tools to be relevant for counterfactual analyses. For achieving a balance between simplicity and flexibility, within the model development process, the reusable component approach combined with a flexible model integration environment seems to be the most promising approach (van Ittersum 2006), which requires programing language and standards targeting at modularity and extension of solutions (Donatelli and Rizzoli 2007).

2.4 Model uncertainty

The problem of uncertainty contained in model outputs remains as a real challenge to the use of models as part of decision making or informative processes. In part, this is simply a matter of realizing that there are four kinds of issues (French and Geldermann 2005)—the *known* (for which models are irrelevant as decision makers already have the knowledge that they need), the *knowable with additional information* (a case where models may have a key role in providing or synthesizing data to provide information), the

complex (where models may have a similar role but where there is significant and irreducible uncertainty), and, finally, the *chaotic* where only short-term planning is possible and efforts need to be focused on adapting to unpredictable events. Uncertainty can arise from any number of sources (model structure, parameterization, input data, initialization), and how it manifests itself in the model estimates can be difficult to determine, but incorporation of expert stakeholder interpretation helps separate the *knowable with additional information* from the chaotic. An example is seen in climate change and impact projections, where differing levels of uncertainty exist across the multiple elements (biophysical, political, economic, etc.). Lack of certainty on the future is, first of all, in relation with prospective visions at various scales: global (socioeconomic scenarios), regional (land and soil uses), and field (agricultural systems, genotypes, practices). As well, the chaotic character of the climate system (e.g., interannual climate variability and effect of initial conditions) limits the reliability of climate projections. Here, for instance, in the case of crop response to CO_2 enrichment, the *knowable with additional information* from models indicates how crops would respond to a future climate, but large uncertainties exist in the climate model projections (complex) and scenarios which drive them (chaotic). In addition, another component of uncertainty is constituted by the extent of our imperfect knowledge of processes (and their interactions) imbedded in models (*epistemic uncertainties*). In climate modeling, this mainly refers to atmospheric and biosphere physics, ocean-atmosphere coupling, empirical relationships embedded, parameterization, and spatial resolution. At smaller scales, impact models (such as crop models) may suffer from omitted or disregarded climate-connected processes in the long term and their interactions (*structure*), as well as shifting of parameter values due to new climate conditions (*parameterization*). In such an example, it becomes possible to present possible outputs (e.g., estimates of future conditions) for review by stakeholders who are then able to assess the different levels of uncertainty across the multiple elements and estimate their own projections relevant to their own area of expertise. This has urged for benchmarking actions at an international level, where estimation of process-oriented epistemic uncertainties is done by running several models supposed to simulate the same reality (ensemble modeling) so as to generate an expanded envelope of uncertainty (Fig. 3), whereby uncertainty accumulates throughout the process of climate change prediction and impact assessment ("cascade of uncertainty", after Schneider 1983; Boe 2007).

In presenting future impact projections (Fig. 4), the deliberative process identifies possible aspirations (what stakeholders would like to happen), expectations (what they think will happen), and possible adaptations, i.e., to cope with climate change (what can happen), on the basis that stakeholders

acknowledge the uncertainty but can consider contingency plans (Matthews et al. 2008a).

The climate change literature widely discusses uncertainties and the need to build the computational ability of models, especially in relation to issues of local adaptation. Papers often reflect the institutional and political barriers presented by the divide between adaptation measures that focus on the role of agroecosystems (as identified by models) and those that support the role of communities (Girot et al. 2012). Public participation in decision making through the use of deliberative processes would enhance legitimacy of model-based advices (identification and implementation of adaptation measures), develop social responsibility, and learning how to make decisions and problem solving. Through the co-development of alternative future scenarios, it is possible to raise awareness of the issues, provide new information, influence attitudes, and begin to stimulate action, despite the inherent uncertainty. However, modelers need to be able to manage the expectations of stakeholders as to how likely it is that research will be able to provide an answer and how good that answer will be. Ultimately, there is a limit as to how valuable a model can be, after which point stakeholders must make their own evaluations on its utility. Climate change in the tropics and the role of agrobiodiversity for adapting to variability and sustaining local livelihoods are examples where the sole source of knowledge may reside among local users and managers (Mijatović et al. 2013). Knowledge systems of this type (locally or regionally maintained, adapted, and transmitted, e.g., Tengö et al. 2014) can help the science policy community to think beyond aspects that can be fitted into models, and the variables that continue to be refined by model improvement might well be what policymakers begin to identify they want refined.

2.5 Modeling background

The issues of transparency and uncertainty are, however, often conflated with situations where the research that forms the basis of the models is contested. It is important to recognize that the scientific understanding, while the best available, may still be contested or contestable; thus, criticism for lack of transparency may just be a screen for stakeholders with legitimate or selfish vested interest disagreeing with the outcomes of the model (e.g., Caminiti 2004). Many of the decisions in natural resource management have substantial normative components, e.g., the preferences for alternative outcomes expressed as minimum standards or thresholds or the ideal state to which a systems' management should lead (Rauschmayer and Wittmer 2006). In these cases, it may be that disagreement on the model outcomes is simply a means of delaying or preventing undesirable outcomes. In such circumstances, it is preferable that inclusive processes be undertaken. These issues have been explored further in the context of

Fig. 3 The cascade of uncertainty scheme (after Boe 2007) implied by coupling the probability distributions for emissions and biogeochemical cycle calculations to arrive at greenhouse gas (*GHG*) concentrations needed to calculate radiative forcing, climate sensitivity, climate impacts, and evaluation of such impacts into agroecological systems. Impact assessment is achieved through "top down" methods which focus on developing fine-scale climate data from coarse-scale global climate models. The resulting local-scale scenarios are fed into impact models to determine vulnerability
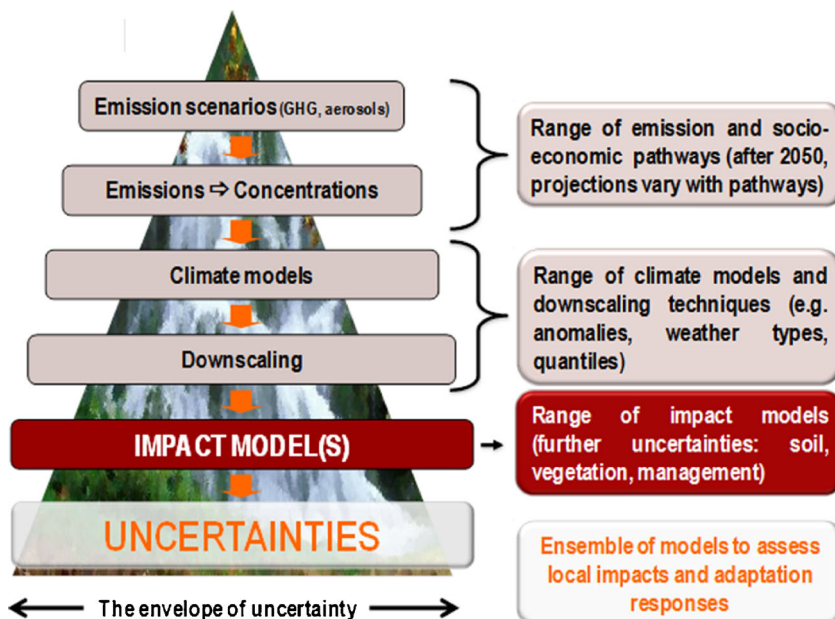
environmental modeling and policy by Kolkman and van der Veen (2006) who conclude that it is only through a process of identifying alternative mental models of the model developer and other interested parties (formulations of process deriving from the individuals' perspectives) that the true value of models may be determined.

2.6 Stakeholder approaches

Stakeholder-inclusive approaches to increasing the credibility and thus value and appeal of models can be conducted *ex ante* or post hoc. While the former is seen as the most desirable, since the stakeholders get to influence from the outset the content and assumptions within the model, such modeling can be expensive, can have difficulties in meeting the expectations of stakeholders, and maintaining their interest and involvement may be difficult within what can be a protracted development process. There is also the potential for impasse if there are conflicting interests within the stakeholder groups and issues of control between developers and stakeholders.

Tensions may emerge between modes of expression, for example, when members of the public use anecdotal and personal evidence while experts use systematic and generalized evidence based on abstract knowledge (Dietz et al. 1989). This requires the deliberative process to explore the hidden rationalities in the arguments of any party and thus avoid that the two modes remain in different corners. Despite these caveats, approaches such as mediated modeling have been in fact successfully used within the social sciences as part of processes to address (through the exercise of value-laden judgments) complex issues with conflicting stakeholder groups (Rauschmayer and Wittmer 2006).

For *post hoc processes*, a version of the model exists and is used within the process as a *boundary object* (that is a stand in for reality that allows contending parties to make their case by arguing through the model rather than directly at one another). The underlying principle of such processes is that conflicting views of the world are best resolved through *deliberation* (Dryzek 2000), that is reason-based debate, where evidence is presented and evaluated. The role of models in such
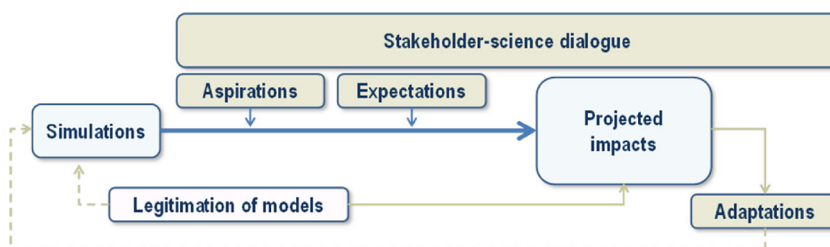

Fig. 4 Deliberative process in model-based climate change studies. The simulation process uses an integrated modeling framework to estimate the impacts. The deliberative processes allow the research team to elicit influential factors in determining alternative scenarios, such as

aspirations, expectations, adaptations, and practitioner interpretation. The primary outcome from the process is the generation and critical evaluation of scenarios that include anticipated decision maker adaptations and hypotheses to be tested

processes can be to assist as a common framework within which to compare and contrast alterative formulations. Such activities can also be useful in making assumptions and trade-offs explicit (Matthews et al. 2008b). Models in such a role need to be sophisticated and flexible enough for each interested party to be able to represent their strategies, the emphasis being on the ability of the model to adequately represent subjects under dispute with appropriate levels of transparency. Experience in the use of such approaches indicates that they can be successful not only in knowledge elicitation, but also in targeting and prioritizing model development of primary research (Matthews et al. 2006). There are also examples where models developed initially, as research tools have subsequently been used within a *Participatory Action Research* (*PAR*) paradigm (a mixture of the *ex ante* and post hoc cases above) (Meinke et al. 2001; Carberry et al. 2002; Keating et al. 2003). Such applications have helped both build the credibility of the models with stakeholders through collaborative use and adapted the form and content of models based on these interactions.

2.7 Summary

To evaluate simulation models is far more urgent, as many of the decisions in agroecology are based on model outcomes. Dealing with existing agroecological systems and designing new ones are a priority that deliberations about model evaluation contribute to accomplish in a more efficient (maybe more appropriate) manner, in any case with more awareness if genuine collective deliberations are possible. The central issue is to think and conceive model evaluation in a clear decisional perspective about type of model, operability, transparency, etc. As several models are at hand, "mod-diversity" imposes the analysis of case-by-case issues, while also integrating the specific context in a larger-scale perspective (in space and time).

3 Computer-aided evaluation techniques

Complex biophysical models are made up of mixtures of rate equations, comprise approaches with different levels of empiricism, and make use of partially autocorrelated parameters. These models aim to simulate systems which show a non-linear behavior, and they are often solved with numerical solutions, which are more versatile than analytical solutions (which typically apply to fairly simple situations). Modeling applications are therefore based on software (inherently complex and difficult to engineer), and it is the computer program (Fig. 1, right, step 5), including technical issues and possible errors, to be tested rather than the mathematical model (Fig. 1, right, step 4) representing the system (van Ittersum 2003).

Each version of a model, throughout its development life cycle, should be subjected to output testing, thanks to test scenarios, test cases, and/or test data. Applying the same test to each model release is repetitive and time-consuming, requiring the preservation of the test scenarios, test cases, and test data for reuse. Modelers are hardly capable of developing reasonably large and complex modeling tools and guaranteeing their accuracy over time. A disciplined approach, effective management, and well-educated personnel are some of the key factors affecting the success of software development. Modeling professionals in agroecology can learn a lot from software engineering, stakeholder deliberation, and other disciplines, in order to include the necessary knowledge to conduct successful model evaluation. To meet the substantial model quality challenges, it is necessary to improve the current tools, technologies, and their cost benefit characterizations (Sects. 3.1 and 3.2). The emergence of new technologies in simulation modeling has, in fact, fostered debate on the reuse and interoperability of models (Sect. 3.3). This has implications for the practice of model evaluation (Sect. 3.4) because the deliberative process may inform the selection of evaluation metrics and setting of thresholds and weights in computer-aided evaluation.

3.1 Concepts and tools

Evolution in model evaluation approaches, also accompanied by the creation of dedicated software tools (Fila et al. 2003a, b; Tedeschi 2006; Criscuolo et al. 2007; Olesen and Chang 2010), has culminated in reviews and position papers (Bellocchi et al. 2010; Alexandrov et al. 2011; Bennett et al. 2013) with the aim of characterizing the performance of models and providing standards for publishing models in forms suitable for use by broad communities (Jakeman et al. 2006; Laniak et al. 2013). Several evaluation methods are available, but, usually, only a limited number of methods are used in modeling projects (as documented, for instance, by Richter et al. 2012 and Ritter and Muñoz-Carpena 2013), often due to time and resource constraints. This is also because different users of models (and beneficiaries of model outputs) may have different thresholds for confidence: some may derive their confidence simply from the model reports displayed, and others may require more in-depth evaluation before they are willing to believe the results. In general, limited testing may hinder the modeler's ability to substantiate sufficient model accuracy.

3.2 Model coding

A large number of existing agricultural and ecological models have been implemented as software that cannot be well maintained or reused, except by their developers, and therefore cannot be easily transported to other platforms (Reynolds

and Acock 1997). In order to include legacy data sources into newly developed systems, object-oriented development has emerged steadily as a paradigm that focuses on granularity, productivity, and low-maintenance requirements (Timothy 1997). While some research has been undertaken focusing on establishing a baseline for evaluation practice, rather less work has been done to develop a basic, scientifically rigorous approach to be able to meet the technical challenges that we currently face. This activity can be valuably supported by modular, object-oriented programing on both sides of modeling and evaluation tools, allowing consolidated experience in evaluating models to be formed and shared.

Software objects are designed to represent elements of the real world, and the focus needs to be on the development of consistent design patterns that encourage usability, reusability, and cross-language compatibility, thus facilitating model development, integration, documentation, and maintenance (Donatelli et al. 2004). In particular, component-oriented programing (combination of object-oriented and modular features) takes an important place in developing systems in a variety of domains, including agroecological modeling (Papajorgji et al. 2004; Argent 2004a, b, 2005). Although different definitions of "component" do actually exist in the literature (Bernstein et al. 1999; Booch et al. 1999; Szypersky et al. 2002), a component is basically an independent delivered piece of functionality, presented as a black box that provides access to its services through a defined interface.

The component development paradigm is to construct software that enables independent components to be plugged together. This requires an environment that implements the communication issues of the components' interaction. The platform-independent Java language (http://java.sun.com) and the .NET technology of Windows (http://www.microsoft.com/net), for instance, have emerged with the aim to support interoperability between different components and therefore facilitating their integration process. Some advantages of component-based development that can be realized for model application development are (Rizzoli et al. 1998; Donatelli and Rizzoli 2007) the following: reduction of modeling project costs in the long term, enhancement of model transparency, expansion of model applicability, increase of automation, creation of systematically reusable model components, increase of interoperability among software products, and convenient and ready adaptation of model components.

3.3 Modular simulation

The increasing complexity of models and the need to evolve interoperability standards have stimulated advanced, modular, object-oriented programing languages, and libraries that support object-oriented simulation. Various object- and component-oriented solutions have approached the issue of

agricultural and environmental modeling, such as maize irrigation scheduling (Bergez et al. 2001), multiple spatial scales ecosystems (Woodbury et al. 2002), greenhouse control systems (Aaslyng et al. 2003), weather modeling (Donatelli et al. 2005), households, landscape, and livestock integrated systems (Matthews 2006). In the same context of the agricultural and environmental modeling community, alternative frameworks have been made available to support modular model development through provision of libraries of core environmental modeling modules, as well as reusable tools for data manipulation, analysis, and visualization (Argent et al. 2006). There is some consensus (Glasow and Pace 1999) that component-based development is indeed an effective and affordable way not only for creating model applications but also for conducting model evaluation. Particular emphasis should be placed on designing and coding object-oriented simulation models to properly transfer simulation control between entities, resources, and system controllers. It is crucial, therefore, to consider the issue of model value when considering model reuse, as it needs to be a fundamental part of any reuse strategy.

3.4 Coupling between simulation and evaluation

Where complex models are to be evaluated, options are available to combine detailed numeric and statistical tests of components and subprocesses with a deliberative approach for overall model acceptance. Model development should aim to incorporate automated evaluation checks using embedded software tools, with the aim of achieving greater cost and time efficiency and to achieve a higher level of credibility. The distribution of already evaluated model components can substantially decrease the model evaluation effort when reused. A key step in this direction is the coupling between model components and evaluation techniques, the latter also implemented into component-based software. Figure 5 illustrates a possible coupling strategy.

The evaluation system stands at the core of a general framework where the modeling system (e.g., a set of modeling components) and a data provider supply inputs to the evaluation tool. The latter is also a component-based system as well, both communicating with the modeling component and the data provider and allowing the user to interact in some way to choose and parameterize the evaluation tools. The output coming out of the evaluation system can be offered to a deliberative process (stakeholder review) for interpretation of results (see Sect. 2). Without a procedure to reach consensus (deliberative process), an automatic process based on numerical tests would stand at the forefront of model evaluation leaving behind stakeholder assessment an interpretation. Adjustments in the modeling system or parameterization/initialization can be made afterward, if the results are assessed as unsatisfactory for the application purpose. A new evaluation-interpretation cycle
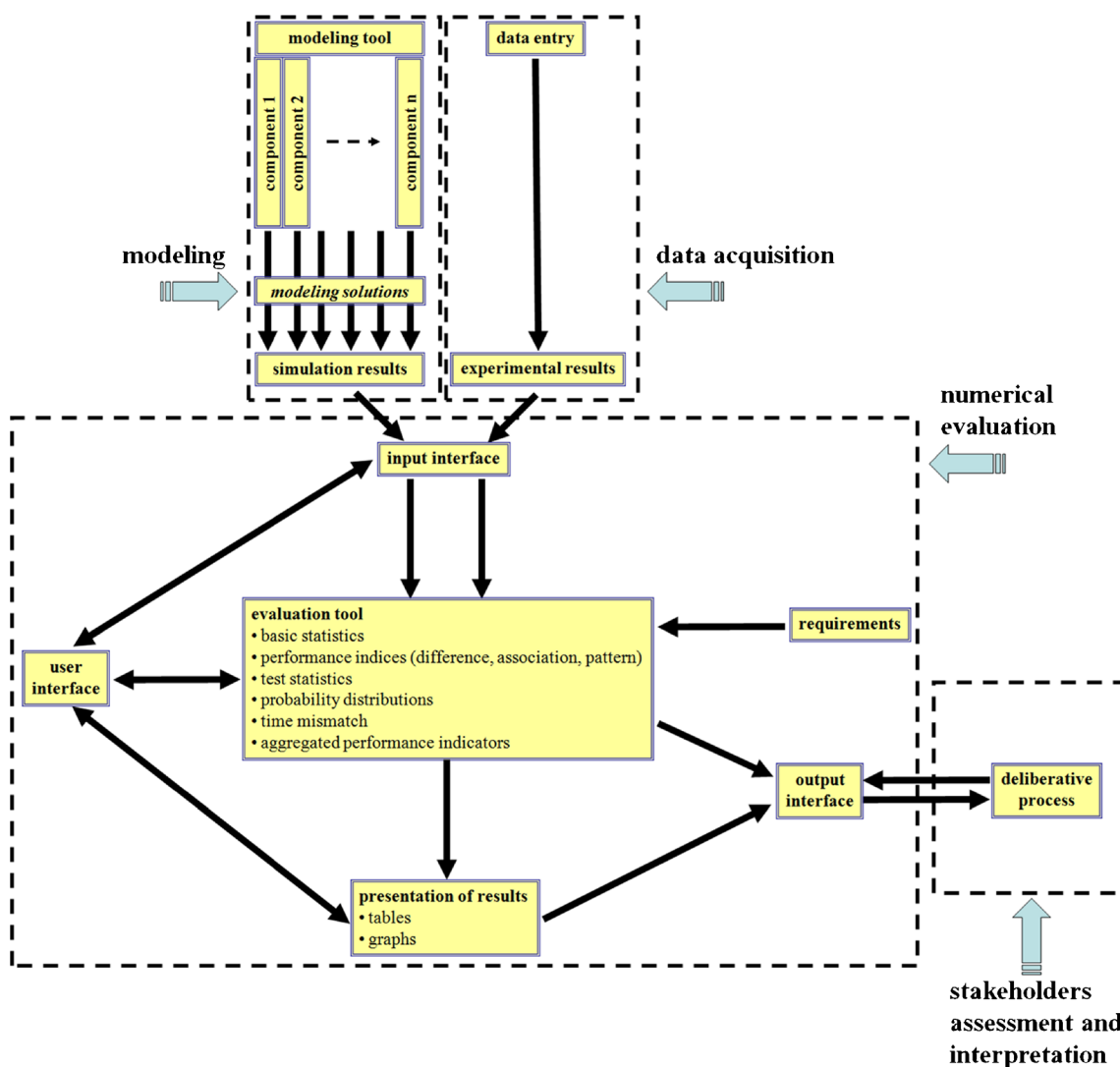
Fig. 5 Integrated system of modeling tools, data provider, and evaluation tools. The evaluation component communicates with both the modeling structure and the data provider via a suitable protocol, while allowing the user to interact via a dedicated graphical user interface. Adjustments in the modeling system or critical reviewing of data used to evaluate the model can be made afterward (via deliberative process), if the results are deemed to be not satisfactory. A new evaluation/interpretation cycle can be run any time that new versions of the modeling system are developed by modifying basic model components and generate new modeling solutions

can be run any time that new versions (solutions) of the modeling system are developed. Again, a well-designed, component-based evaluation system can be easily extended toward including further evaluation approaches to keep up with evolving methodologies, i.e., statistical or fuzzy-based (e.g., Carozzi et al. 2013; Fila et al. 2014).

The scheme of Fig. 5 closely resembles the coupling of freeware, Microsoft COM-based tool IRENE_DLL (Integrated Resources for Evaluating Numerical Estimates_Dynamic Link Library, Fila et al. 2003b; available for downloading through the web site http://www.sipeaa.it/tools) with the model for rice production WARM (Confalonieri et al. 2005; Acutis et al. 2006; Bellocchi et al. 2006), in which the double arrow at the level of stakeholder assessment and interpretation indicates that the deliberative part informs the selection of

metrics and module for evaluation, and the setting of thresholds and weights in the evaluation tool. The modular structure of IRENE_DLL allowed it to be integrated into the WARM application software, including a calibration tool for evaluation of objective functions (Acutis and Confalonieri 2006). In this way, evaluation runs can be automated and executed on either individual model components (e.g., Bregaglio et al. 2012; Donatelli et al. 2014) or the full model at any time that components are added or modified, using a wide range of integrated metrics, as also shown by Fila et al. (2006) with a tailored application for evaluation of pedotransfer functions.

Since IRENE_DLL was developed, the component-oriented paradigm has evolved, specifying new requirements in order to increase software quality, reusability, extensibility,

and transparency for components providing solutions in the biophysical domain (Donatelli and Rizzoli 2007). A .NET (http://www.microsoft.com/net) redesign was performed (Criscuolo et al. 2007; Simulation Output Evaluator, through http://agsys.cra-cin.it/tools/default.aspx) to provide third parties with the capability of extending methodologies without recompiling the component. This ensures greater transparency and ease of maintenance, also providing functionalities such as the test of input data versus their definition prior to computing any simple or integrated evaluation metric. Making it in agreement with the modern developments in software engineering, the component for model evaluation better serves as a convenient means to support collaborative model testing among the network of scientists involved in creating component-oriented models in the agroecological domain (Donatelli et al. 2012; Bergez et al. 2013).

3.5 Summary

Easy to maintain and reusable code is of paramount importance in model development. Component-based programing is an affordable way to effectively reduce the development cost (or recover it in the long run). In this respect, it is essential that model evaluation becomes an integral part of the overall model development and application process. Hence, we would argue that a great emphasis should be put on evaluation plans within scientific projects in which model applications cover a variety of time and space scales. Matching these scales and ensuring consistency in the overall modeling flow are not a trivial process and may be difficult to automate without access to model environments preventing from hard coding to couple simulation and evaluation tools. This calls for the need to develop evaluation techniques at the same pace with which the models themselves are created and improved (by developers) and applied (by users), while also model outputs are exploited (by beneficiaries).

4 Deliberation processes in model evaluation: the examples of MACSUR and MODEXTREME

Whether evaluation is a scheduled action in modeling, little work is published in the open literature (e.g., conference proceedings and journals) describing the evaluation experience accumulated by modeling teams (including interactions with the stakeholders). Failing to disseminate the evaluation experience may result in the repetition of the same mistakes in future modeling projects. Learning from the past experience of others is an excellent and cost-effective educational tool. The return on such an investment can easily be realized by preventing the failures of modeling projects and thus avoiding

wrong simulation-based decisions. This section deals with the kind of deliberation on model evaluation proposed by discussion forums established within two international projects.

4.1 Modelling European Agriculture with Climate Change for Food Security—a FACCE JPI Knowledge Hub

The MACSUR (http://www.macsur.eu, 2002–2015) knowledge hub (as well as parallel programs such as AgMIP, http://www.agmip.org, or other initiatives of the FACCE JPI, http://www.jpifacce.org) holds potential to help advance good modeling practice in relation with model evaluation (including access to appropriate software tools), an activity which is frequently neglected in the context of time-limited projects. In MACSUR CropM-LiveM (crop-livestock modeling) cross-cutting activities, a questionnaire-based survey (through http://limesurvey.macsur.eu) on fuzzy-logic-based multimetric indicators for model evaluations helped understanding of the multifaceted knowledge and experience required and the substantial challenges posed by the deliberative process. A composite indicator (MQI_m: Model Quality Indicator for multisite assessment), elaborated by a limited group of specialists, was first revised by a broader representative group of modelers and then assessed via questionnaire survey of all project partners (scientists and end users, including trade modelers), the results of which were presented in an international conference (Bellocchi et al. 2014a). The indicator aggregates the three components of model quality—agreement with actual data, complexity, and stability—represented in Fig. 2 and described by Bellocchi et al. (2014b). Seven questions were asked about indicator characteristics, which were answered by 16 respondents. The responses received (Fig. 6) reflect a general consensus on the key terms of the original proposal, although caution is advised on how metrics were formulated (question 3). In particular, some remarks and considerations suggested that other factors than purely climatic ones (such as soil conditions) may play a role in the concept of robustness and the construction of its metric (after Confalonieri et al. 2010). For a first attempt, this was encouraging, even if the relatively small number of answers received (~20 % of partner teams) indicates that there is still a difficulty to transfer the brainstorming items into a formal model evaluation process based on fuzzy-logic-based rules.

4.2 MODelling vegetation response to EXTREMe Events—European Community's Seventh Framework Programme

The EU-FP7 project MODEXTREME (http:///www.modextreme.org, 2013–2016) is an example of science-stakeholder dialogue where a platform of diverse stakeholders (from local actors to institutional parties) is established to

Fig. 6 Questionnaire survey of MACSUR partners about fuzzy-logic-based indicator for model evaluation. The Model Quality Indicator for multisite assessment (MQI_m) integrates different components of model quality such as the agreement between model outputs and actual data, elements of model complexity, and an assessment of the stability of model performance over a variety of conditions. The results show that most respondents agree on the settings (metrics, modules, thresholds, weights) configured by specialists
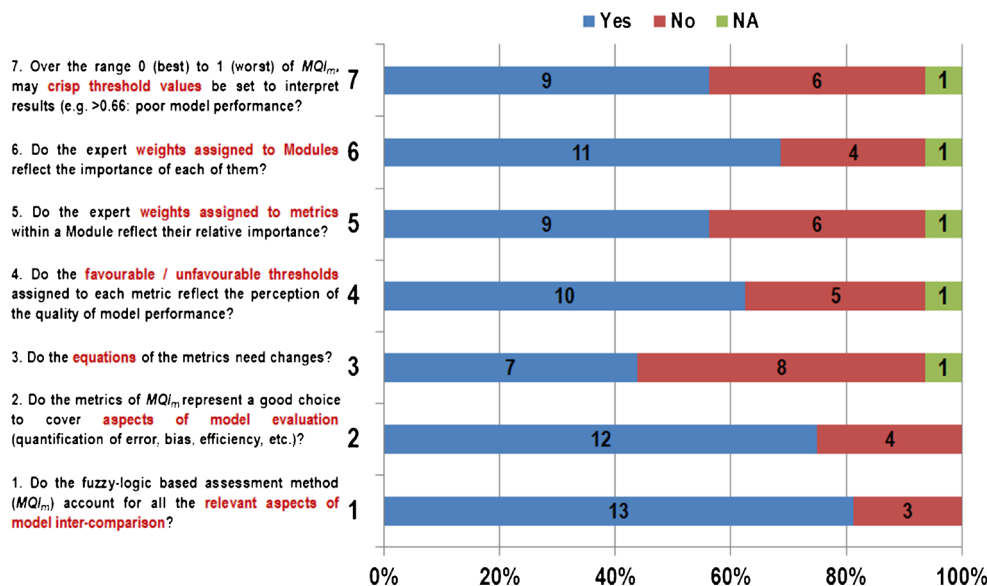
evaluate agroecological models. The aim is to represent the impact of extreme weather events on agricultural production. The strategy is to develop and implement modeling solutions and couple them with numerical evaluation tools (as described in Sect. 3.4) using the capabilities of the platform BioMA (http://mars.jrc.ec.europa.eu/mars/About-us/AGRI4CAST/ Models-Software-Tools/Biophysical-Model-Application-BioMA). The mass of stakeholder engagement reveals four clusters (after Spitzeck and Hansen 2010), as presented in Fig. 7.

The cluster "dialogue and issues advisory" demonstrates a high diversity of stakeholders with low power; i.e., broad types of stakeholders within operational and managerial scope (farmers, providers of agricultural services, field research agronomists) are identified locally, mainly by non-European project partners (Brazilian Corporation of Agricultural Research, Argentinian National Agricultural Technology Institute, University of Pretoria in South Africa, Chinese Academy of Agricultural Sciences). The cluster "issues of collaboration" is characterized by a partner (Food and Agriculture

Organization of the United Nations) with considerable power, regarded as a stakeholder for the clear understanding of specific issues (food security) beyond the scope of the project and within a limited scope (local communities). In the cluster "strategic collaboration," stakeholders are a limited group of institutional actors (at the level of European Commission), regarded as partners for their direct involvement in research actions via survey techniques, meetings with representatives, and exchange of datasets (http://modextreme.org/event/dgagri2014). The power is high because the Joint Research Centre and the directorate for agriculture have the control to transfer scientific advances from the project into knowledge suitable for policy implementation in Europe (e.g., in-season crop monitoring and forecasts, integrated assessments in agriculture, and price regulation of agricultural commodities). The final cluster "strategic advisory and innovation" leads to institutional diversity, still at the level of European Commission. In contrast with the strategic collaboration predominant in the prior cluster, this cluster

Fig. 7 Clusters of stakeholder engagement. There are four clusters of stakeholder groups (A, B, C, and D). The scope increases from advisory and collaborative issues (A, B, and D) to a strategic level of interaction (C). Different levels of power include situations where no evidence of stakeholder influence is provided (A) to a stage where stakeholders have a say in decision making (C). Stakeholder diversity diminishes with increasing scope and power

INRA SCIENCE & IMPACT

Springer
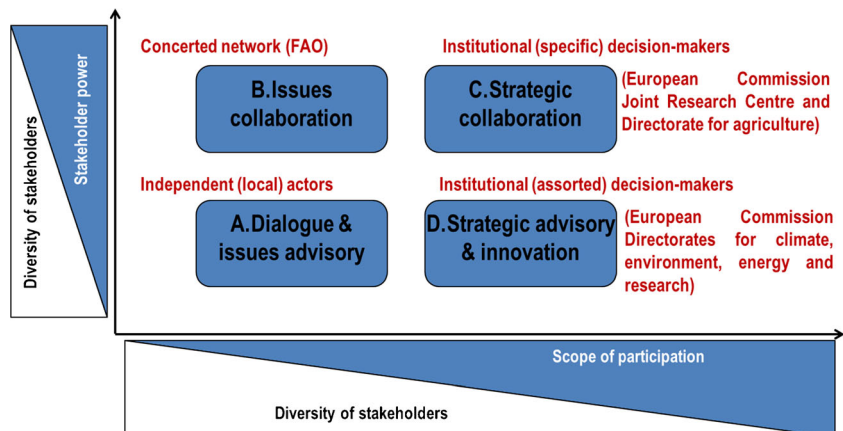
advises the dissemination strategy broadly (large scope extending to climate, environment, energy, and research), with less power for implementation.

4.3 Summary

The experiences in MACSUR and MODEXTREME demonstrate that deliberative engagement processes can be implemented within research projects and can be used to guide model evaluation. The kind of deliberation on this topic is not exhaustive, but the two projects are a good initial step to support evaluation of agroecological models with deliberation. The peculiarities of these forums, mostly characterized by asynchrony in written exchanges, absence of face-to-face interaction, and anonymity, suggest to reconsider the ways in which stakeholder may participate in the evaluation of models. In particular, the analysis of messages as well as interviews with participants indicates the need for an improvement of the rules which structure the participatory approach. The perception gap of agroecological models (and their use) between different actors can hinder expression. The arguments and skills used by participants in discussions should therefore be reconsidered.

5 Conclusion

This review has covered the issues of model evaluation in agroecology. Model evaluation is a multifaceted complex process that is strongly influenced by the nature of the models as well as the conditions where they are applied. There is an increasing interest in the use of biophysical models to analyze agroecological systems, quantify outcomes, and drive decision making. Modeling applications have increased in the last decades, and the concept of model-based simulation of complex systems sounds attractive to support problem solving. However, problems exist when systematic and generalized evidence based on abstract knowledge is used by modelers, leaving potential model beneficiaries with less influence on decisions. The participatory and deliberative feature suggests that the beneficiaries of model outputs may voice their complaints and desires to the model providers, discuss with each other and with the model providers, and, to some extent, influence and take responsibility for model content. A transition from model evaluation as academic research toward model evaluation as a participative, deliberative, and dialogue-based exercise (illustrated with two examples from international projects) is therefore desirable to raise the bar of model credibility and thus legitimate the use of agroecological models in decision making. Currently, the software technology to assist participatory approaches for model evaluation exists. The major limitation remains the difficulty to establish disciplined approached, effective management, and

well-educated personnel within the time limitation and budgetary constraints of research projects. However, the continuing interest in the use of agroecological models to set ground for decisions offers opportunities to look at model evaluation with a fresh angle of vision and to question about opening new ways to see the principles of deliberative processes and software model development to converge.

References

Aaslyng JM, Lund JB, Ehler N, Rosenqvist E (2003) IntelliGrow: a greenhouse component-based climate control system. Environ Model Softw 18:657–666. doi:10.1016/S1364-8152(03)00052-5

Acutis M, Confalonieri R (2006) Optimization algorithms for calibrating cropping systems simulation models. A case study with simplex-derived methods integrated in the WARM simulation environment. Italian Journal of Agrometeorology 11:26–34

Acutis M, Confalonieri R, Genovese G, Donatelli M, Rodolfi M, Mariani L, Bellocchi G, Trevisiol P, Gusberti D, Sacco D (2006) WARM: a new model for rice simulation. In: Fotyma M., Kaminska B. (eds) Proceedings of the 9th European Society for Agronomy Congress, 6–9 September, Warsaw, pp 259–260

Alexandrov GA, Ames D, Bellocchi G, Bruen M, Crout N, Erechtchoukova M, Hildebrandt A, Hoffman F, Jackisch C, Khaiter P, Mannina G, Matsunaga T, Purucker ST, Rivington M, Samaniego L (2011) Technical assessment and evaluation of environmental models and software. Environ Model Softw 26:328–336. doi:10.1016/j.envsoft.2010.08.004

Argent RM (2004a) An overview of model integration for environmental applications—components, frameworks and semantics. Environ Model Softw 19:219–234. doi:10.1016/S1364-8152(03)00150-6

Argent RM (2004b) Concepts, methods and applications in environmental model integration. Environ Model Softw 19:217. doi:10.1016/S1364-8152(03)00149-X

Argent RM (2005) A case study of environmental modelling and simulation using transplantable components. Environ Model Softw 20:1514–1523. doi:10.1016/j.envsoft.2004.08.016

Argent RM, Voinov A, Maxwell T, Cuddy SM, Rahman JM, Seaton S, Vertessy RA, Braddock RD (2006) Comparing modelling frameworks—a workshop approach. Environ Model Softw 21:895–910. doi:10.1016/j.envsoft.2005.05.004

Asseng S, Ewert F, Rosenzweig C, Jones JW, Hatfield JL, Ruane A, Boote KJ, Thorburn P, Rötter RP, Cammarano D, Brisson N, Basso B, Martre P, Aggarwal PK, Angulo C, Bertuzzi P, Biernath C, Doltra J, Gayler S, Goldberg R, Grant R, Heng L, Hooker JE, Hunt LA, Ingwersen J, Izaurralde RC, Kersebaum KC, Müller C, Naresh Kumar S, Nendel C, O'Leary G, Olesen JE, Osborne TM, Palosuo T, Priesack E, Ripoche D, Semenov MA, Shcherbak I, Steduto P, Stöckle CO, Stratonovitch P, Streck T, Supit I, Travasso M, Tao F, Waha K, Wallach D, White JW, Wolf J (2013) Uncertainties in simulating wheat yields under climate change. Nature Clim Change 3:827–832. doi:10.1038/nclimate1916

Bair ES (1994) Model (in)validation—a view from courtroom. Ground Water 32:530–531. doi:10.1111/j.1745-6584.1994.tb00886.x

Balci O (1997) Principles of simulation model validation, verification, and testing. Trans Soc Comput Simul Int 14:3–12

Balci O, Ormsby WF (2002) Expanding our horizons in verification, validation, and accreditation research and practice. In: Yücesan E, Chen C-H, Snowdon JL, Charnes JM (eds) Proceedings of 2002 Winter Simulation Conference, 8–11 December, San Diego, pp 653–663. doi:10.1109/WSC.2002.1172944

Bassu S, Brisson N, Durand JL, Boote K, Lizaso J, Jones JW, Rosenzweig C, Ruane AC, Adam M, Baron C, Basso B, Biernath C, Boogaard H, Conijn S, Corbeels M, Deryng D, De Sanctis G, Gayler S, Grassini P, Hatfield J, Hoek S, Izaurralde C, Jongschaap R, Kemanian AR, Kersebaum KC, Kim SH, Kumar NS, Makowski D, Müller C, Nendel C, Priesack E, Pravia MV, Sau F, Shcherbak I, Tao F, Teixeira E, Timlin D, Waha K (2014) How do various maize crop models vary in their responses to climate change factors? Global Change Biol 20:2301–2320. doi:10.1111/gcb.12520

Bellocchi G, Confalonieri R, Donatelli M (2006) Crop modelling and validation: integration of IRENE_DLL in the WARM environment. Italian Journal of Agrometeorology 11:35–39

Bellocchi G, Rivington M, Donatelli M, Matthews KB (2010) Validation of biophysical models: issues and methodologies. A review. Agron Sustain Dev 30:109–130. doi:10.1051/agro/2009001

Bellocchi G, Rivington M, Acutis M (2014) Deliberative processes for comprehensive evaluation of agro-ecological models. FACCE MACSUR Mid-term Scientific Conference, "Achievements, Activities, Advancement," 01–04 April, Sassari. http://ocs.macsur.eu/index.php/Hub/Mid-term/paper/view/193. Accessed 06 November 2014

Bellocchi G, Rivington M, Acutis M (2014) Protocol for model evaluation. FACCE MACSUR Reports 2(1): D-C1.3. http://ojs.macsur.eu/index.php/Reports/article/view/D-L2.2. Accessed 06 November 2014

Bennett ND, Croke BFW, Guariso G, Guillaume JHA, Hamilton SH, Jakeman AJ, Marsili-Libelli S, Newham LTH, Norton JP, Perrin C, Pierce SA, Robson B, Seppelt R, Voinov AA, Fath BA, Andreassian V (2013) Characterising performance of environmental models. Environ Model Softw 40:1–20. doi:10.1016/j.envsoft.2012.09.011

Bergez J-E, Debaeke P, Deumier J-M, Lacroix B, Leenhardt D, Leroy P, Wallach D (2001) MODERATO: an object-oriented decision tool for designing maize irrigation schedules. Ecol Model 137:43–60. doi:10.1016/S0304-3800(00)00431-2

Bergez J-E, Chabrier P, Gary C, Jeuffroy MH, Makowski D, Quesnel G, Ramat E, Raynal H, Rousse N, Wallach D, Debaeke P, Durand P, Duru M, Dury J, Faverdin P, Gascuel-Odoux C, Garcia F (2013) An open platform to build, evaluate and simulate integrated models of farming and agro-ecosystems. Environ Model Softw 39:39–49. doi:10.1016/j.envsoft.2012.03.011

Bernstein PA, Bergstraesser T, Carlson J, Pal S, Sanders P, Shutt D (1999) Microsoft repository version 2 and the open information model. Inf Syst 24:71–98

Boe J (2007) Changement global et cycle hydrologique: une étude de régionalisation sur la France. PhD thesis, University Paul Sabatier(in French)

Booch G, Rumbaugh J, Jacobson I (1999) The unified modeling language user guide. Addison-Wesley, Reading

Bredehoeft JD, Konikow LF (1993) Ground water models: validate or invalidate. Ground Water 3:178–179. doi:10.1111/j.1745-6584.2012.00951.x

Bregaglio S, Donatelli M, Confalonieri R, De Mascellis R, Acutis M (2012) Comparing modelling solutions at submodel level: a case on soil temperature simulation. In: Seppelt R, Voinov AA, Lange S, Bankamp D (eds) International Environmental Modelling and Software Society (iEMSs), 2012 International Congress on Environmental Modelling and Software, Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig. http://www.iemss.org/sites/iemss2012//proceedings/D3_1_0851_Bregaglio_et_al.pdf. Accessed 06 November 2014

Caminiti JE (2004) Catchment modelling—a resource manager's perspective. Environ Model Softw 19:991–997. doi:10.1016/j.envsoft.2003.11.002

Carberry PS, Hochman Z, McCown RL, Dalgliesh NP, Foale MA, Poulton PL, Hargreaves JNG, Hargreaves DMG, Cawthray S, Hillcoat N, Robertson MJ (2002) The FARMSCAPE approach to decision support: farmers', advisers', researchers' monitoring, simulation, communication and performance evaluation. Agr Syst 74:141–177. doi:10.1016/S0308-521X(02)00025-2

Carozzi M, Bregaglio S, Scaglia B, Bernardoni E, Acutis M, Confalonieri R (2013) The development of a methodology using fuzzy logic to assess the performance of cropping systems based on a case study of maize in the Po Valley. Soil Use Manage 29:576–585. doi:10.1111/sum.12066

Chopin P, Blazy J-M, Dore T (2014) Indicators for the assessment of the sustainability level of agricultural landscapes. In: Pepó P, Csajbók J (eds) Proceedings of the 13th Congress of the European Society for Agronomy, 25–29 August, Debrecen, pp 149–150. http://www.esa2014.hu/doc/esa2014_proceedings.pdf

IPCC (Intergovernmental Panel on Climate Change) (2013) IPCC 5th Assessment Report "Climate Change 2013: the Physical Science Basis. University Press, Cambridge. http://www.ipcc.ch/report/ar5/wg1/#.Uk7O1xBvCVq. Accessed 06 November 2014

Colomb B, Carof M, Aveline A, Bergez J-E (2013) Stockless organic farming: strengths and weaknesses evidenced by a multicriteria sustainability assessment model. Agron Sustain Dev 33:593–608. doi:10.1007/s13593-012-0126-5

Confalonieri R, Acutis M, Donatelli M, Bellocchi G, Mariani L, Boschetti M, Stroppiana D, Bocchi S, Vidotto F, Sacco D, Grignani C, Ferrero A, Genovese G (2005) WARM: a scientific group on rice modelling. Italian Journal of Agrometeorology 2:54–60

Confalonieri R, Acutis M, Bellocchi G, Donatelli M (2009) Multi-metric evaluation of the models WARM, CropSyst, and WOFOST for rice. Ecol Model 220:1395–1410. doi:10.1016/j.ecolmodel.2009.02.017

Confalonieri R, Bregaglio S, Acutis M (2010) A proposal of an indicator for quantifying model robustness based on the relationship between variability of errors and of explored conditions. Ecol Model 221:960–964. doi:10.1016/j.ecolmodel.2009.12.003

Creighton J (1983) The use of values: public participation in the planning process. In: Daneke GA, Garcia MW, Priscoli JD (eds) Public involvement and social impact assessment. Westview Press, Boulder, pp 143–160

Criscuolo L, Donatelli M, Bellocchi G, Acutis M (2007) Component and software application for model output evaluation. In: Donatelli M, Hatfield J, Rizzoli AE (eds) Farming Systems Design 2007, Int. Symposium on Methodologies on Integrated Analysis on Farm Production Systems, September 10–12, Catania, Vol 2, pp 211–212

Dietz T, Stern PC, Rycroft RW (1989) Definition of conflict and the legitimation of resources: the case of environmental risk. Sociol Forum 4:47–69

Donatelli M, Rizzoli AE (2007) A design for framework-independent model components of biophysical systems. In: Donatelli M, Hatfield J, Rizzoli AE (eds) Farming Systems Design 2007, International Symposium on Methodologies on Integrated Analysis on Farm Production Systems, September 10–12, Catania, Vol 2, pp 208–209

Donatelli M, Omicini A, Fila G, Monti C (2004) Targeting reusability and replaceability of simulation models for agricultural systems. In: Jacobsen SE., Jensen CR, Porter JR (eds) Proceedings of the 8th European Society for Agronomy Congress, 11–15 July, Copenhagen, pp 237–238

Donatelli M, Carlini L, Bellocchi G, Colauzzi M (2005) CLIMA: a component-based weather generator. In: Zerger A, Argent RN

(eds) MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 12–15 December, Melbourne, pp 627–633

Donatelli M, Cerrani I, Fanchini D, Fumagalli D., Rizzoli AE (2012) Enhancing model reuse via component-centered modeling frameworks: the vision and example realizations. In: Seppelt R, Voinov AA, Lange S, Bankamp D (eds.) International Environmental Modelling and Software Society (iEMSs), 2012 International Congress on Environmental Modelling and Software, Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig. http://www.iemss.org/sites/iemss2012//proceedings/D3_1_0847_Donatelli_et_al.pdf. Accessed 06 November 2014

Donatelli M, Bregaglio S, Confalonieri R, De Mascellis R, Acutis M (2014) A generic framework for evaluating hybrid models by reuse and composition—a case study on soil temperature simulation. Env Modell Softw. doi:http://dx.doi.org/10.1016/j.envsoft.2014.04.011

Dryzek J (2000) Deliberative democracy and beyond: liberals, critics, contestations (Oxford Political Theory). Oxford University Press, New York

Ethier GJ, Livingston NJ (2004) On the need to incorporate sensitivity to CO_2 transfer conductance into the Farquhar-von Caemmerer–Berry leaf photosynthesis model. Plant Cell Environ 27:137–153

Fearon JD (1998) Deliberation as discussion. In: Elster J (ed) Deliberative democracy. Cambridge University Press, Cambridge, pp 44–68

Fila G, Bellocchi G, Acutis M, Donatelli M (2003a) IRENE: a software to evaluate model performance. Eur J Agron 18:369–372. doi:10.1016/S1161-0301(02)00129-6

Fila G, Bellocchi G, Donatelli M, Acutis M (2003b) IRENE_DLL: a class library for evaluating numerical estimates. Agron J 95:1330–1333

Fila G, Donatelli M, Bellocchi G (2006) PTFIndicator: an IRENE_DLL-based application to evaluate estimates from pedotransfer functions by integrated indices. Environ Model Softw 21:107–110. doi:10.1016/j.envsoft.2005.01.001

Fila G, Di Lena B, Gardiman M, Storchi P, Tomasi D, Silvestroni O, Pitacco A (2014) Calibration and validation of grapevine budburst models using growth-room experiments as data source. Agr Forest Meteorol 160:69–79. doi:10.1016/j.agrformet.2012.03.003

French S, Geldermann J (2005) The varied contexts of environmental decision problems and their implications for decision support. Environ Sci Policy 8:378–391. doi:10.1016/j.envsci.2005.04.008

Girot P, Ehrhart C, Oglethorpe J (2012) Integrating community and ecosystem-based approaches in climate change adaptation responses. Ecosystems & Livelihood Adaptation Network Report. http://www.careclimatechange.org/files/adaptation/ELAN_IntegratedApproach_150412.pdf. Accessed 06 November 2014

Glasow PA, Pace DK (1999) SIMVAL'99: making VV&A effective and affordable workshop. The Simulation Validation Workshop 1999, January 26–29, Laurel

Gliessman SR (2007) Agroecology: the ecology of sustainable food systems. CRC Press, Boca Raton

Hamilton MA (1991) Model validation: an annotated bibliography. Communications Stat-Theor M 20:2207–2266

Hutchins MG, Urama K, Penning E, Icke J, Dilks C, Bakken T, Perrin C, Saloranta T, Candela L, Kamari J (2006) The BMW model evaluation tool: a guidance document. Archiv für Hydrologie: Large Rivers Supplement 17:23–48

Jakeman AJ, Letcher RA, Norton JP (2006) Ten iterative steps in development and evaluation of environmental models. Environ Model Softw 21:602–614. doi:10.1016/j.envsoft.2006.01.004

Keating BA, Gaydon D, Huth NI, Probert ME, Verburg K, Smith CJ, Bond W (2002) Use of modelling to explore the water balance of dryland farming systems in the Murray-darling basin, Australia. Eur J Agron 18:159–169. doi:10.1016/S1161-0301(02)00102-8

Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, Holzworth D, Huth NI, Hargreaves JNG, Meinke H, Hochman Z, McLean G, Verburg K, Snow V, Dimes JP, Silburn M, Wang E,

Brown S, Bristow KL, Asseng S, Chapman S, McCown RL, Freebairn DM (2003) An overview of APSIM, a model designed for farming systems simulation. Eur J Agron 18:267–288. doi:10.1016/S1161-0301(02)00108-9

Kolkman MJ, van der Veen A (2006) Without a common mental model a DSS makes no sense (a new approach to frame analysis using mental models). In: Voinov A, Jakeman AJ, Rizzoli AE (eds) Proceedings of the 3rd Biennial Meeting of the International Environmental Modelling and Software Society (iEMSs), July 9–13, Burlington. http://www.iemss.org/iemss2006/papers/s10/140_Kolkman_1.pdf. Accessed 11 June 2014

Konikow LF, Bredehoeft JD (1992) Ground water models cannot be validated. Adv Water Resour 15:75–83. doi:10.1016/0309-1708(92)90033-X

Landry M, Oral M (1993) In search of a valid view of model validation for operations research. Eur J Oper Res 66:161–167. doi:10.1016/0377-2217(93)90310-J

Laniak GF, Olchin G, Goodall J, Voinov A, Hill M, Glynn P, Whelan G, Geller G, Quinn N, Blind M, Peckham S, Reaney S, Gaber N, Kennedy R, Hughes A (2013) Integrated environmental modeling: a vision and roadmap for the future. Environ Model Softw 39:3–23. doi:10.1016/j.envsoft.2012.09.006

Li T, Hasegawa T, Yin X, Zhu Y, Boote K, Adam M, Bregaglio S, Buis S, Confalonieri R, Fumoto T, Gaydon D, Marcaida III M, Nakagawa H, Oriol P, Ruane AC, Ruget F, Singh B, Singh U, Tang L, Tao F, Wilkens P, Yoshida H, Zhang Z, Bouman B (2014) Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Global Change Biol, in press. doi:10.1111/gcb.12758

Liu S, Sheppard A, Kriticos D, Cood D (2011) Incorporating uncertainty and social values in managing invasive alien species: a deliberative multi-criteria evaluation approach. Biol Invasions 13:2323–2337. doi:10.1007/s10530-011-0045-4

Lizaso JI (2014) Improving crop models: incorporating new processes, new approaches, and better calibrations. In: Pepó P, Csajbók J (eds) Proceedings of the 13th Congress of the European Society for Agronomy, 25–29 August, Debrecen, pp 5–10. http://www.esa2014.hu/doc/esa2014_proceedings.pdf. Accessed 06 November 2014

Maiorano A, Cerrani I, Fumagalli D, Donatelli M (2014) New biological model to manage the impact of climate warming on maize corn borers. Agron Sustain Dev 34:609–621. doi:10.1007/s13593-013-0185-2

Matthews R (2006) The People and Landscape Model (PALM): towards full integration of human decision-making and biophysical simulation models. Ecol Model 4:329–343. doi:10.1016/j.ecolmodel.2005.10.032

Matthews KB, Buchan K, Sibbald AR, Craw S (2006) Combining deliberative and computer-based methods for multi-objective land-use planning. Agr Syst 87:18–37. doi:10.1016/j.agsy.2004.11.002

Matthews KB, Rivington M, Buchan K, Miller DG (2008a) Communicating climate change consequences for land use, Technical Report on Science Engagement, Grant No 42/07 2007-08. Macaulay Institute, Aberdeen

Matthews KB, Schwarz G, Buchan K, Rivington M, Miller D (2008b) Wither agricultural DSS? Comput Electron Agr 61:149–159. doi:10.1016/j.compag.2007.11.001

Matthews KB, Rivington M, Blackstock K, McCrum G, Buchan K, Miller DG (2011) Raising the bar?—the challenges of evaluating the outcomes of environmental modelling and software. Environ Model Softw 26:247–257. doi:10.1016/j.envsoft.2010.03.031

Matthews KB, Miller DG, Warden-Johnson D (2013) Supporting agricultural policy—the role of scientists and analysts in managing political risk. In: Piantadosi J, Anderssen RS, Boland J (eds) MODSIM 2013 International Congress on Modelling and Simulation, 1–6 December, Adelaide, pp 2152–2158

McCown RL (2002a) Changing systems for supporting farmers' decisions: problems, paradigms, and prospects. Agr Syst 74:179–220. doi:10.1016/S0308-521X(02)00026-4

McCown RL (2002b) Locating agricultural decision support systems in the troubled past and socio-technical complexity of models for management. Agr Syst 74:11–25. doi:10.1016/S0308-521X(02)00020-3

McCown RL, Hochman Z, Carberry PS (2005) In search of effective simulation-based intervention in farm management. In: Zerger A, Argent RM (eds) MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 12–15 December, Melbourne, pp 232–238

Meinke H, Baethgen WE, Carberry PS, Donatelli M, Hammer GL, Selvaraju R, Stöckle CO (2001) Increasing profits and reducing risks in crop production using participatory systems simulation approaches. Agr Syst 70:493–513. doi:10.1016/S0308-521X(01)00057-9

Mijatović D, Van Oudenhoven F, Eyzaguirre P, Hodgkin T (2013) The role of agricultural biodiversity in strengthening resilience to climate change: towards an analytical framework. Int J Agr Sustain 11:95–107. doi:10.1080/14735903.2012.691221

Olesen HR, Chang JC (2010) Consolidating tools for model evaluation. Int J Environ Pollut 40:175–183

Oreskes N (1998) Evaluation (not validation) of quantitative models. Environ Health Persp 106:1453–1460

Papajorgji P, Beck HW, Braga JL (2004) An architecture for developing service-oriented and component-based environmental models. Ecol Model 179:61–76. doi:10.1016/j.ecolmodel.2004.05.013

Petts J (2001) Evaluating the effectiveness of deliberative processes: waste management case-studies. J Environ Plan Manage 44:207–226. doi:10.1080/09640560120033713

Raupach MR, Finnigan JJ (1988) Single layer models of evaporation from plant canopies are incorrect, but useful, whereas multilayer models are correct, but useless: discussion. Aust J Plant Physiol 15:705–716

Rauschmayer F, Wittmer H (2006) Evaluating deliberative and analytical methods for the resolution of environmental conflicts. Land Use Policy 23:108–122. doi:10.1016/j.landusepol.2004.08.011

Reynolds JF, Acock B (1997) Modularity and genericness in plant and ecosystems models. Ecol Model 94:7–16. doi:10.1016/S0304-3800(96)01924-2

Reynolds JF, Ford ED (1999) Multi-criteria assessment of ecological process models. Ecology 80:538–553

Richter K, Atzberger C, Hank TB, Mauser W (2012) Derivation of biophysical variables from Earth observation data: validation and statistical measures. J Appl Remote Sens 6:063557. doi:10.1117/1.JRS.6.063557

Ritter A, Muñoz-Carpena R (2013) Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. J Hydrol 480:33–45. doi:10.1016/j.jhydrol.2012.12.004

Rizzoli AE, Davis JR, Abel DJ (1998) A model management system for model integration and re-use. Decis Support Syst 4:127–144

Rotmans J, van Asselt MBA (2001) Uncertainty in integrated assessment modelling: a labyrinthic path. Integr Assess 2:43–55

Rykiel EJ Jr (1996) Testing ecological models: the meaning of validation. Ecol Model 90:229–244

Sargent RG (2001) Verification, validation and accreditation of simulation models. In: Peters BA, Smith JS, Medeiros DJ, Rohrer MW (eds) Proceedings of 2001 Winter Simulation Conference, December 10–13, Arlington, pp 106–114

Schneider SH (1983) CO_2, climate and society: a brief overview. In: Chen RS, Boulding E, Schneider SH (eds) Social science research and climate change: an interdisciplinary appraisal. Reidel, Boston, pp 9–15

Spitzeck H, Hansen EG (2010) Stakeholder governance: how stakeholders influence corporate decision making corporate governance. Int J Bus soc 10:378–391. doi:10.1108/14720701011069623

Steduto P, Hsiao T, Raes D, Fereres E (2009) AquaCrop—the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agron J 101:426–437. doi:10.2134/agronj2008.0139s

Szypersky C, Gruntz D, Murer S (2002) Component software—beyond object-oriented programming, 2nd edn. Addison-Wesley, London

Tanner CB, Sinclair TR (1983) Efficient water use in crop production: research or re-search? In: Taylor HMJ, Sinclair TR (eds) Limitations to efficient water use in crop production. American Society of Agronomy, Madison, pp 1–27

Tedeschi LO (2006) Assessment of the adequacy of mathematical models. Agr Syst 89:225–247. doi:10.1016/j.agsy.2005.11.004

Tengö M, Brondizio ES, Elmqvist T, Malmer P, Spierenburg M (2014) Connecting diverse knowledge systems for enhanced ecosystem governance: the multiple evidence base approach. Ambio 43:579–591. doi:10.1007/s13280-014-0501-3

Timothy B (1997) An introduction to object-oriented programming, 2nd edn. Addison-Wesley, Reading

Van Ittersum MK (2003) Modelling cropping systems—highlights of the symposium and preface to the special issues. Eur J Agron 18:189–191. doi:10.1016/S1161-0301(02)00095-3

Van Ittersum MK (2006) Integrated assessment of agriculture and environmental policies: towards a computerised framework for the EU (SEAMLESS-IF). In: Voinov A (ed) Proceedings of the 3rd Biennial Meeting of the International Environmental Modelling and Software Society (iEMSS), July 9–13, Burlington. http://www.iemss.org/iemss2006/papers/s10/280_vanIttersum_1.pdf. Accessed 11 June 2014

Vanclay JK (1994) Modelling forest growth and yield. CAB International, Wallingford

Vanclay JK, Skovsgaard JP (1997) Evaluating forest growth models. Ecol Model 98:1–12. doi:10.1016/S0304-3800(96)01932-1

Vischel T, Pegram G, Sinclair S, Wagner W, Bartsch A (2007) Comparison of soil moisture fields estimated by catchment modelling and remote sensing: a case study in South Africa. Hydrol Earth Syst Sci Discuss 4:2273–2306. doi:10.5194/hess-12-751-2008

Voinov A, Bousquet F (2010) Modelling with stakeholders. Environ Model Softw 25:1268–1281

Woodbury PB, Beloin RM, Swaney DP, Gollands BE, Weinstein DA (2002) Using the ECLPSS software environment to build a spatially explicit component-based model of ozone effects on forest ecosystems. Ecol Model 150:211–238. doi:10.1016/S0304-3800(01)00465-3

INRA
SCIENCE & IMPACT

Springer