Evaluating uncertainty in integrated environmental models: A review of concepts and tools

L. Shawn Matott,[1,2] Justin E. Babendreier,[1] and S. Thomas Purucker[1]

Received 21 July 2008; revised 16 February 2009; accepted 7 April 2009; published 17 June 2009.

[1] This paper reviews concepts for evaluating integrated environmental models and discusses a list of relevant software-based tools. A simplified taxonomy for sources of uncertainty and a glossary of key terms with ''standard'' definitions are provided in the context of integrated approaches to environmental assessment. These constructs provide a reference point for cataloging 65 different model evaluation tools. Each tool is described briefly (in the auxiliary material) and is categorized for applicability across seven thematic model evaluation methods. Ratings for citation count and software availability are also provided, and a companion Web site containing download links for tool software is introduced. The paper concludes by reviewing strategies for tool interoperability and offers guidance for both practitioners and tool developers.

Citation: Matott, L. S., J. E. Babendreier, and S. T. Purucker (2009), Evaluating uncertainty in integrated environmental models: A review of concepts and tools, *Water Resour. Res.*, *45*, W06421, doi:10.1029/2008WR007301.

1. Introduction

[2] Integrated environmental models have emerged as useful tools supporting research, policy analysis, and decision making [e.g., *Argent*, 2004; *Babendreier et al.*, 2007; *Clark and Gelfand*, 2006; *Matthies et al.*, 2007; *Gerber*, 2007]. In this regard, model integration often utilizes an underlying framework, a set of consistent, interdependent, and compatible science components (i.e., models, data, and assessment methods) presented in a context of organizing principles, standards, infrastructure, and software [*U.S. Environmental Protection Agency* (*USEPA*), 2008]. Examples include Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) [*Lahlou et al.*, 1998], Community Hydrology Prediction System (CHPS) [*McEnery et al.*, 2005], Framework for Risk Analysis of Multimedia Environmental Systems (FRAMES) [*Babendreier and Castleton*, 2005], Modeling Environment for Total Risk (MENTOR) [*Georgopoulos and Lioy*, 2006], Modular Modeling System (MMS) [*Leavesley et al.*, 1996], Multiscale Integrated Models of Ecosystem Services (MIMES) [*Van Bers et al.*, 2007], Object Modeling System (OMS) [*Ahuja et al.*, 2005], and Open Modeling Interface (OpenMI) [*Gregersen et al.*, 2007].

[3] Besides facilitating model integration, many frameworks provide or leverage model-independent tools, additional software codes that are not intrinsic components of any particular modeling program. Ideally, model-independent tools should be easily applied to arbitrary models and the user should not have to write additional software. In practice, many model-independent tools require writing an interface program to connect to a given model.

[4] A variety of model-independent tools support model evaluation, the process of determining model usefulness and estimating the range or likelihood of various interesting outcomes. This paper focuses on model evaluation technologies and is motivated by the complexities of integrating process-based numerical models. Integrating and evaluating these types of models is challenging in many respects [*Beven*, 2007; *Jakeman and Letcher*, 2003; *Johnston et al.*, 2004; *Newbold*, 2002], but such activity can yield a solid foundation for environmental assessment.

[5] Although evaluating integrated numerical models was the primary incentive for this work, the material is generally applicable to a broader array of environmental models. Thus, the intended audience is both (1) system level ''frame workers'' seeking to incorporate model evaluation tools in an integrated modeling framework and (2) modelers who need evaluation tools for a standalone (i.e., nonframework) application. The two groups tend to approach tool integration and usage from different perspectives.

1.1. Motivation and Overview

[6] To be useful in a policy context, models must be evaluated using reproducible and robust procedures [e.g., *Beven*, 2007; *McGarity and Wagner*, 2003; *Nicholson et al.*, 2003; *Oreskes*, 2003; *Gaber et al.*, 2009; *Pascual*, 2005; *Schultz*, 2008; *USEPA*, 1992; *van der Slujis*, 2007]. Much literature on model evaluation has been published, ranging from discourse on the nature and meaning of uncertainty [e.g., *Beck et al.*, 1997; *Beven*, 2002; *Konikow and Bredehoeft*, 1992; *Montanari*, 2007; *Norton et al.*, 2006; *Rotmans and van Asselt*, 2001a, 2001b; *Trucano et al.*, 2006; *van Asselt and Rotmans*, 2002; *Walker et al.*, 2003] to specific applications of various methods [e.g., *Balakrishnan et al.*, 2005; *Barlund and Tattari*, 2001; *Gallagher and Doherty*, 2007; *Lindenschmidt et al.*, 2007; *Muleta and Nicklow*, 2005; *Schuol and Abbaspour*, 2006; *Yu et al.*, 2001]. Related to these efforts is the development

[1]Ecosystems Research Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Athens, Georgia, USA.

[2]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario, Canada.

Published in 2009 by the American Geophysical Union.

of quality assurance (QA) guidelines for environmental modeling [e.g., *Aumann*, 2007; *Jakeman et al.*, 2006; *Refsgaard and Henriksen*, 2004; *Refsgaard et al.*, 2005, 2007]. Although model evaluation is a major element of QA, the topic is typically presented as a generic portion of an overall modeling strategy that may or may not be implemented.

[7] The present work complements QA guidance by summarizing available tools for model evaluation; emphasizing approaches that characterize, quantify, or propagate uncertainty. To facilitate an organized presentation of tools, the paper first reviews sources and types of uncertainty and categorizes methods of model evaluation. Existing literature contains multiple uncertainty taxonomies and generally utilizes an inconsistent model evaluation terminology. These inconsistencies are partially addressed by introducing both a glossary of 'standard' definitions for key terms and a simple taxonomy for sources of uncertainty.

[8] Following the uncertainty review and method categorization is a tabulation of model evaluation tools; published algorithms or software codes that implement evaluation methods in a model-independent manner. The main manuscript includes a functionality matrix for identifying and comparing tool capabilities, and the auxiliary material contains a brief overview of each tool.[1] Such a compendium is inherently subjective, and deserving tools may have been left out. The functionality matrix has been translated into a publicly accessible Web site. An "in press" snapshot of the site is reproduced in the auxiliary material, and a continually updated site is also available (www.epa.gov/athens/research/modeling/modelevaluation/index.html). The Web site contains links for downloading full tool descriptions and associated software codes (if available).

[9] The paper also discusses the use of multiple independently developed tools for a given model evaluation exercise. Barriers to tool interoperability are reviewed along with strategies to overcome these barriers. The paper concludes with a commentary on tool selection and the future of integrated modeling.

2. Methods

[10] Several references provide excellent introductions to the topic of evaluating environmental models [e.g., *Beck*, 1987, 2002; *Burnham and Anderson*, 2002; *Cox and Baybutt*, 1981; *Cullen and Frey*, 1999; *Funtowicz and Ravetz*, 1990; *Hamby*, 1994; *Helton et al.*, 2006; *Morgan and Henrion*, 1990; *Saltelli et al.*, 2000, 2004; *Vose*, 2000]. In addition, aspects of model evaluation have been the focus of recent conferences and workshops [e.g., *Hanson and Hemez*, 2004; *von Krauss et al.*, 2004] (see also IAHS-PUB Workshop on Uncertainty Analysis in Hydrologic Modeling, Lancaster University, Lancaster, United Kingdom, 2004, www.es.lancs.ac.uk/hfdg/uncertainty_workshop/uncert_intro.htm, and TransAtlantic Uncertainty Colloquium, University of Georgia, Athens, 2005, www.modeling.uga.edu/tauc/index.html). Combining these information sources with an appraisal of the relevant peer-reviewed literature yielded a fairly comprehensive review.

2.1. Review of Model Evaluation Concepts

[11] In the context of regulation, model evaluation is motivated by a desire to minimize the possibility of making a "wrong" decision about a potentially adverse environmental outcome. Central to such activity is the need to characterize, quantify and propagate uncertainty, while recognizing that both quantitative and qualitative components are present [*Funtowicz and Ravetz*, 1990; *Refsgaard et al.*, 2007; *Walker et al.*, 2003]. The desire to be comprehensive has yielded a broad variety of model evaluation methods and packaged software tools.

[12] Numerous uncertainty taxonomies have been advocated [e.g., *Cullen and Frey*, 1999; *Funtowicz and Ravetz*, 1990; *Jager and King*, 2004; *Li and Wu*, 2006; *Refsgaard et al.*, 2007; *Regan et al.*, 2003; *Rotmans and van Asselt*, 2001b; *Trucano et al.*, 2006; *van der Sluijs et al.*, 2003; *Vose*, 2000; *Walker et al.*, 2003]. In some cases, different taxonomies assign different meanings to the same terms; in others, different taxonomies reflect alternative perspectives. This intrinsic vagueness is an example of "linguistic uncertainty" [*Regan et al.*, 2003] and can engender significant confusion among practitioners, stakeholders, and decision makers. To establish a point of reference for comparing tools, a glossary of common terms is included in the auxiliary material.

2.1.1. Types of Uncertainty in Integrated Environmental Modeling

[13] Uncertainty may be classified as reducible (i.e., stemming from erroneous knowledge or data) or irreducible (i.e., stemming from inherent variability). From a decision-making perspective, these have different ramifications and should be separated, to the extent possible, when evaluating models [*Cullen and Frey*, 1999; *Hoffman and Hammonds*, 1994; *Nauta*, 2000; *Sonich-Mullin*, 2001]. The total uncertainty of a given quantity may be characterized in one of four ways: purely irreducible, the quantity varies and the associated population has been completely sampled without error; partly reducible and partly irreducible, the quantity varies and the associated population has been partially sampled or sampled with error; purely reducible, the quantity does not vary but has been sampled with error; and certain, the quantity does not vary and has been sampled without error [*Cullen and Frey*, 1999].

[14] While model evaluation traditionally utilizes a probabilistic representation of uncertainty, alternative representations have also been considered [*Helton and Oberkampf*, 2004]. Types of reducible (i.e., empirical) input uncertainty, for example, have been explicated in a variety of ways, popularly subclassified by: random error; systematic error; input sampling error; model output sampling error; inherent randomness; correlation; and disagreement [*Cullen and Frey*, 1999; *Morgan and Henrion*, 1990].

2.1.2. Sources of Uncertainty in Integrated Environmental Modeling

[15] Numerous classification schemes for sources of uncertainty have been introduced, and it is not always possible to reconcile the differing taxonomies. Examples include *Linkov and Burmistrov* [2003] (parameter, model, and modeler uncertainty); *Beck* [1987] (initial system state, parameter, input, and output uncertainty); *Refsgaard et al.* [2007] (context, input, parameter, structural, and technical uncertainty); and *Morgan and Henrion* [1990] (statistical

[1] Auxiliary materials are available in the HTML. doi:10.1029/2008WR007301.

variation, subjective judgment, linguistic imprecision, variability, inherent randomness, disagreement, approximation). We propose a simplified taxonomy (described below), consisting of quantitative input and model uncertainty under an umbrella of qualitative modeler uncertainty.

[16] The modeler is responsible for determining and assembling both an input vector (**X**) and the model (f(**X**)) operating on an input vector to simulate output. Different modelers may make different decisions about the form and content of **X** and f(**X**). Such modeler uncertainty can be measured by comparative study (contrasting results of multiple independent modelers) or by reduction via expert elicitation (having multiple modelers develop a single, consensus model or subset of models).

[17] Input uncertainty is associated with quantities in the input and output vectors; these can be subdivided into input data, response data, and model parameters. Input data refer to the forcing functions, sources and sinks, and initial and boundary conditions consumed by a model. Response data represent site-specific measurements and expert testimony that can be compared to model-simulated output. In general, the usefulness (i.e., information content) of a given set of input and response data will depend on the model structure and the degree to which model input influences model output. Model parameters are akin to input data, except parameters are typically carefully tailored (e.g., via parameter estimation) to suit a particular modeling effort.

[18] Unlike input and response data, quantifying parameter uncertainty requires both the model and data. Also, whereas input and response data can be irreducible, model parameters usually are treated as purely reducible (i.e., constants for a given site and whose values are uncertain). Systematic parameter variability or nonconstancy, for example, is an important indicator of model error that often is not explicitly addressed by modelers [*Beck*, 1987; *Kuczera et al.*, 2006; *Wagener et al.*, 2003]. Failure to recognize such variability confounds the associated uncertainty quantification by mixing together aspects of parameter, model and data uncertainty [*Kavetski et al.*, 2002, 2006a, 2006b]. The resulting bias may be viewed as an expression of hybrid uncertainty associated with an unknown probability of occurrence [*Vose*, 2000]. Depending on its significance, the bias may complicate efforts to compare or assimilate model parameter estimates generated for different models, sites, or scenarios.

[19] Model uncertainty reflects the inability of a model, even when provided with perfect (i.e., certain or purely irreducible) input, to generate output indistinguishable from corresponding real-world observations. Model uncertainty may be subdivided according to uncertainties in model structure (e.g., scientific hypotheses and governing equations), model resolution (e.g., spatiotemporal discretization, boundary specification and scale dependence), and model code (e.g., algorithms, numerical solvers). Another aspect of model uncertainty is the correspondence (or lack thereof) between model input-output and available data. For example, values assigned to areal or volumetric model inputs are often derived from point measurements. Likewise, areal or volumetric model outputs are often compared against point measured response data. Similar considerations apply to the temporal domain and are particularly relevant to steady state models.

[20] Various aspects of model uncertainty are amenable to quantitative methods. For example, sensitivity analysis can investigate the influence of model resolution and solver precision [e.g., *Ahlfeld and Hoque*, 2008; *Bedogni et al.*, 2005], and identifiability analysis can reveal systematic errors in model structure [e.g., *Beck*, 1987; *Wagener et al.*, 2003]. Propagation of model uncertainty is the subject of ongoing research. One approach introduces systematic or random model error (e.g., for numerical models) at model runtime [*Marin et al.*, 2003]. A similar tactic augments the model expression to include explicit error terms (e.g., f(**X**) + ε). Bayesian network modeling can accommodate analogous approaches [*Clark*, 2005; *Clark and Gelfand*, 2006]. Multimodel strategies, in juxtaposition, propagate model uncertainty by considering multiple candidate models of the system; each candidate is treated as a sample of the complete distribution of possible models.

2.1.3. Methods of Model Evaluation

[21] Seven subjective categories of methods for quantitative model evaluation were identified (see Table 1, which also includes method subclassifications): data analysis (DA), identifiability analysis (IA), parameter estimation (PE), uncertainty analysis (UA), sensitivity analysis (SA), multimodel analysis (MMA), and Bayesian networks (BN). The identified categories reflect common terminology used in the literature when describing the purpose of a given model evaluation tool or algorithm. Assigning tools to a particular method was occasionally difficult and necessarily subjective, as there is a certain degree of overlap among the various methods. For example, identifiability analysis contains elements of both parameter estimation and sensitivity analysis; multimodel analysis typically incorporates parameter estimation; and Bayesian networks are analogous to simultaneous parameter estimation and uncertainty analysis. Although not the focus of this paper, two essential evaluation methods are model verification and model validation, further defined in the glossary.

[22] Data analysis (DA) refers to analytical, statistical and graphical procedures for evaluating and summarizing input, response, or model output data. Capabilities typically include data screening and parameterization of distributions for input and response data. A key attribute of geospatial and time series data analysis [*Brockwell and Davis*, 1996; *de Smith et al.*, 2008] is the presence of scale-dependent correlation structures. When characterizing sampled populations, descriptions could be expected to cover biotic and abiotic entities.

[23] Identifiability analysis (IA) seeks to expose inadequacies in the data or suggest improvements in the model structure [*Jakeman and Hornberger*, 1993; *Ljung*, 1999; *Söderström and Stoica*, 1989; *Young et al.*, 1996]. In many cases, IA utilizes sequential parameter estimation or performance-based sensitivity analysis to identify deviations of the model from expected behavior [*Beck and Chen*, 2000; *Beck*, 1987; *Wagener et al.*, 2003]. Alternatively, IA can reveal model parameters that cannot be constrained adequately because of insufficient quantity or diversity of the response data.

[24] Parameter estimation (PE) quantifies uncertain model parameters on the basis of the use of model simulations and available response data. PE techniques yield either a single solution or multiple solutions (i.e., parameter sets).

Table 1. Quantitative Methods of Model Evaluation

Method	Purpose of the Method	Subclassifications
Data analysis (DA)	to evaluate or summarize input, response, or model output data	time series, population, geospatial
Identifiability analysis (IA)	to expose inadequacies in the data or suggest improvements in the model structure	temporal, behavioral, spatial
Parameter estimation (PE)	to quantify uncertain model parameters using model simulations and available response data	single solution, multiple solution
Uncertainty analysis (UA)	to quantify output uncertainty by propagating sources of uncertainty through the model	sampling methods, approximation methods
Sensitivity analysis (SA)	to determine which inputs are most significant	screening, local, global
Multimodel analysis (MMA)	to evaluate model uncertainty or generate ensemble predictions via consideration of multiple plausible models	quantitative, qualitative
Bayesian networks (BN)	to combine prior distributions of uncertainty with general knowledge and site-specific data to yield an updated (posterior) set of distributions	hierarchical Bayesian,Bayesian decision networks

Single-solution approaches view PE as an optimization problem whose unique solution is a single, "best fit" configuration of model parameters. A variety of search algorithms have been successfully applied to calibrate various environmental models, and these can be subdivided into local, global, and hybrid search techniques [e.g., *Bekele and Nicklow*, 2005; *Bell et al.*, 2002; *Duan et al.*, 1992; *Essaid et al.*, 2003; *Gan and Biftu*, 1996; *Giacobbo et al.*, 2002; *Matott and Rabideau*, 2008; *Mugunthan and Shoemaker*, 2006; *Raghavan*, 2003; *Solomatine et al.*, 1999; *Tolson and Shoemaker*, 2007]. Single-solution parameter estimation yields point estimates of parameter values and provides no information regarding the confidence associated with such estimates. To address this limitation, many PE codes [*Doherty*, 2004; *Matott*, 2005; *Poeter and Hill*, 1998] compute various assumption-laden postcalibration parameter statistics, including linear confidence intervals, parameter correlation coefficients, and parameter sensitivities. Some studies have estimated model parameter uncertainty by examining PE global search history [*Evers and Lerner*, 1998; *Seibert and McDonnell*, 2002; *van Griensven and Meixner*, 2006].

[25] Approaches for multiple-solution PE fall into two categories (importance sampling, and Markov chain Monte Carlo (MCMC) sampling) and yield full parameter distributions rather than simple point estimates. PE, via importance sampling, seeks to identify a family of "acceptable" (i.e., behavioral or plausible) model parameter configurations. Sampled model parameter configurations are divided into behavioral and nonbehavioral groups, according to an acceptance threshold for some objective function. This is referred to as rejection sampling because the nonbehavioral group is discarded (i.e., rejected) and model parameter distributions are estimated using a weighted or bias-corrected combination of the behavioral parameter sets. Within the environmental modeling community, Generalized Likelihood Uncertainty Engine (GLUE) [*Beven and Binley*, 1992] is arguably the most popular tool for PE via importance sampling.

[26] MCMC parameter estimation incorporates aspects of importance and rejection sampling into a mathematically robust procedure for evaluating conditional probability distributions. To apply the MCMC technique, model parameter distributions initially are specified independently of the response data (e.g., by assigning a uniform distribution using literature-derived parameter bounds). The sampler evolves these prior model parameter distributions into *posterior* distributions that are updated and conditioned on the response data.

[27] Regardless of the PE approach, several additional considerations exist: selection of an objective function [e.g., *Burnham and Anderson*, 2002; *Gan et al.*, 1997; *Hill*, 1998; *Kavetski et al.*, 2002; *Nash and Sutcliffe*, 1970; *Yapo et al.*, 1998], incorporation of prior information [e.g., *Gupta et al.*, 1998; *Hill*, 1998; *Khadam and Kaluarachchi*, 2004; *Mertens et al.*, 2004; *Rankinen et al.*, 2006; *Seibert and McDonnell*, 2002], and regularization of parameters [e.g., *Doherty*, 2003; *Doherty and Skahill*, 2006; *Tonkin and Doherty*, 2005; *Vermeulen et al.*, 2005; *Vermeulen et al.*, 2006]. A recent trend in calibrating environmental models is considering multiple or competing objectives [e.g., *Gupta et al.*, 1998; *Vrugt et al.*, 2003] and numerous multiobjective optimization algorithms have been applied [e.g., *Deb et al.*, 2002; *Hogue et al.*, 2000; *Vrugt et al.*, 2003; *Vrugt and Robinson*, 2007; *Yapo et al.*, 1998].

[28] Uncertainty analysis (UA) methods propagate sources of uncertainty through the model to generate statistical moments or probability distributions for various model outputs. UA strategies include approximation and sampling methods. Approximation methods characterize model output uncertainty by propagating one or more moments (e.g., mean, variance, skewness, and kurtosis) of the various input distributions through the modeling system. Examples include error propagation equations [*Gertner*, 1987], point estimate methods [*Tsai and Franceschini*, 2005], and various reliability methods [*Hamed et al.*, 1996; *Portielje et al.*, 2000; *Skaggs and Barry*, 1996].

[29] Sampling methods characterize model output distributions by propagating an intensive random sampling of each input distribution. Sampling methods for uncertainty analysis may be classified as Monte Carlo sampling (MCS), stratified sampling, importance sampling, or a combination [*Cox and Baybutt*, 1981; *Helton et al.*, 2006]. Sampled inputs are treated as statistically independent unless correlation is inherent in the assumed distributions (e.g., multivariate normal) or restricted pairing techniques [*Iman and Conover*, 1982] are utilized to handle covariance structures.

[30] Monte Carlo sampling (MCS) draws unbiased random samples from a prescribed distribution [*Vose*, 2000] such that statistical sampling error can be determined using standard theory [*Cullen and Frey*, 1999]. Given a desired level of sampling accuracy and precision, the required sample size can be determined unambiguously.

[31] Stratified sampling, e.g., Latin hybercube sampling (LHS) [*McKay et al.*, 1979], divides a given input distribution into intervals, then generates samples from each interval. For efficient sampling (i.e., fewer model runs), intervals typically are constructed so that each has an equal probability of occurrence. Two assumptions in obtaining reliable model output distributions with fewer model runs are the expectation that fewer model inputs drive output variance, and that model behavior is monotonic and linear [*Campolongo et al.*, 2000; *Helton and Davis*, 2003; *Morgan and Henrion*, 1990].

[32] For uncertainty analysis, importance sampling typically is used to generate intentionally biased samples, ensuring that particular kinds of behavior (e.g., rare but highly consequential events) are analyzed [*Helton et al.*, 2006]. Importance sampling is also useful when an unknown or difficult-to-sample input distribution is "enveloped" by an easily sampled alternative distribution [*Chen*, 2005]. In these cases, the alternative distribution is sampled and the actual distribution is inferred using importance weighting, in which sample weights reflect the probability that a given alternative distribution sample could have come from the actual distribution of interest. This feature makes importance sampling an appealing choice for parameter estimation.

[33] UA sampling schemes can be layered (i.e., nested, hierarchical, or *n* stage). The amount of layering defines the "order" or "dimension" of the resulting sampling procedure. Second-order schemes that separately analyze reducible and irreducible uncertainty are the most common [e.g., *Marin et al.*, 2003; *Wu and Tsang*, 2004]. In principle, any separation of uncertainties can be nested [*Suter*, 2006].

[34] Sampling approaches can be computationally expensive when applied to complex models because of long run times. Parallel processing can reduce the "wall time" (i.e., run time from a human perspective) of such analyses and is particularly suited to embarrassingly parallel sampling methods [*Babendreier and Castleton*, 2005]. Replacing complex models with cheaper surrogate expressions is another way to improve sampling efficiency. Surrogate approaches include the stochastic response surface method (SRSM) [*Isukapalli et al.*, 1998], high-dimensional model representation (HDMR) [*Rabitz and Aliş*, 1999; *Wang et al.*, 2005], Gaussian emulation machines (GEM)[*O'Hagan*, 2006], artificial neural networks (ANN) [*van der Merwe et al.*, 2007], and multivariate adaptive regression splines (MARS) [*Friedman*, 1991]. Generalized polynomial chaos approaches [e.g., *Xiu and Karniadakis*, 2002; *Xiu et al.*, 2002] and stochastic collocation [*Xiu and Hesthaven*, 2005] can also reduce the runs needed for uncertainty estimation. Although listing all available ANN tools is beyond the scope of the work, the Fast Artificial Neural Network library [*Nissen*, 2003] is provided as a representative free and open source implementation.

[35] Sensitivity analysis (SA) studies the degree to which model output is influenced by changes in model inputs or, more generally, the model itself. SA methods help to identify critical areas where knowledge or data are lacking. Ideally, such information leads to reduction of uncertainties in model output via model refinements or additional observations of the system under study, keying on sensitive inputs with large reducible uncertainties.

[36] SA methods are classified as screening, local, and global. Screening methods are efficient, simplistic techniques to rank the importance of inputs, generally without regard to possible interactions [e.g., *Campolongo and Saltelli*, 1997; *Campolongo et al.*, 2007; *Morris*, 2006; *Saltelli et al.*, 2004]. For highly parameterized models, screening methods are extremely useful because model output is often heavily influenced by just a few key inputs. Thus, screening can help guide the subsequent application of more rigorous methods. Screening methods are also employed to compare the influence of different types of uncertainty.

[37] Local (i.e., differential) SA methods utilize gradient (i.e., derivative) information to quantify sensitivity around a specifically configured input (i.e., a single point in multidimensional input space). Such methods are the core of many single-solution, local search PE methods, e.g., Gauss-Marquardt-Levenberg (GML) [*Levenberg*, 1944; *Marquardt*, 1963]. A key factor in local SA is the computation of derivatives; relevant approaches include finite difference perturbation methods and automatic differentiation [*Bischof et al.*, 1996; *Elizondo et al.*, 2002; *Griewank et al.*, 1996; *Yager*, 2004]. Adjoint SA performs reverse local SA to quantify sensitivity around a specifically configured output, e.g., via specialized automatic differentiation techniques [*Hill et al.*, 2004; *Petzold et al.*, 2006; *Sandu et al.*, 2005].

[38] Global SA methods evaluate input sensitivity or importance over the entire range of a model's input space, and can be categorized as variance decomposition, regression-based, correlation-based, parameter bounding, or some combination thereof. Correlation- and regression-based procedures involve graphical or statistical postprocessing of possibly rank-transformed samples collected using Monte Carlo or stratified sampling [*Hamby*, 1994; *Helton and Davis*, 2003; *Helton et al.*, 2006; *Hornberger and Spear*, 1981; *Manache and Melching*, 2008; *Spear and Hornberger*, 1980]. Higher-order variance decomposition procedures, on the other hand, require a more computationally demanding sampling in order to evaluate complex multidimensional integrals [*Chan et al.*, 2000]. In total effect variance decomposition procedures (e.g., the Fourier Amplitude Sensitivity Test (FAST) [*Cukier et al.*, 1978] and the *Sobol'* [2001] method), the total output variance is expressed as the sum of variances contributed by individual inputs (i.e., main effects) plus those of their interactions (i.e., joint effects). Inputs that are more sensitive contribute to a greater fraction of the total variance. Parameter bounding techniques [e.g., *Norton*, 1987, 1996] allow for inversion of an acceptable, bounded output set through a model to estimate input parameter bounds through set-to-set mapping. Originally for models with small numbers of parameters, recent advances have expanded its applicability [*Norton et al.*, 2005].

[39] Multimodel Analysis (MMA) typically evaluates a given site-specific problem statement when multiple plausible models can be developed by, for example, considering alternative processes, using alternative modeling codes, or by defining alternative boundary conditions [*Pachepsky et*

al., 2006]. The resulting consideration of multiple models is an important component of model evaluation [*Burnham and Anderson*, 2002].

[40] Quantitative MMA methods assign performance scores to each candidate model [e.g., *Burnham and Anderson*, 2002; *Ye et al.*, 2008]. The scores help to rank the best models or assign importance weights (e.g., for use in an ensemble forecasting). Qualitative MMA methods rely on expert elicitation, stakeholder involvement, and quality assurance/quality control procedures to assess the relative merits of alternative models [*Funtowicz and Ravetz*, 1990; *van der Slujis*, 2007].

[41] Bayesian networks (BN) are probabilistic graphical models that combine prior distributions of uncertainty with general knowledge (e.g., as one or more models) and site-specific data to yield an updated (posterior) set of distributions. In theory, Bayesian networks can simultaneously treat uncertain input and response data, reducible and irreducible model parameter distributions, and qualitative errors in model code, structure and resolution [*Clark*, 2005; *Clark and Gelfand*, 2006]. Two subclasses, hierarchical Bayesian networks (HBN) [e.g., *Borsuk et al.*, 2004; *Clark*, 2005; *Clark and Gelfand*, 2006; *Elsner and Jagger*, 2004; *Wikle et al.*, 1998; *Wikle*, 2003] and Bayesian decision networks (BDN) [e.g., *Ames*, 2002; *Neilson et al.*, 2002; *Sadoddin et al.*, 2005; *Varis*, 1997], were identified. Developing a BN involves: defining a directed acyclic graph that specifies a network of conditional probability dependencies; defining prior probability distributions for all graph nodes (i.e., sources of uncertainty); and defining a likelihood function and sampling strategy (e.g., MCMC) for inducing posterior distributions from prior distributions. Credal networks can supplement BNs by allowing variables to be associated with interval and set valued probabilities [*Cano et al.*, 1993; *Cozman*, 2000], leading to studies of robustness and incomplete knowledge of uncertainties.

2.2. Catalog of Model Evaluation Tools

[42] Sixty-five tools were identified and these are briefly summarized in the auxiliary material. Using the seven model evaluation methods discussed in section 2.1.3, the overall tool coverage consisted of: data analysis (5 tools), identifiability analysis (10 tools), parameter estimation (32 tools), uncertainty analysis (26 tools), sensitivity analysis (33 tools), multimodel analysis (6 tools), Bayesian networks (5 tools). Some are available for download as standalone executables, complete with user manual; some are provided as source code on request from a designated contact; and others are available only as published algorithm descriptions. In general, if an algorithm description was readily available as part of a given software package, only the software package(s) (i.e., not the algorithm descriptions) were considered to be tools. For this reason, common algorithms (e.g., LHS, FAST) are not listed as separate tools. Although free and open source tools are better fits for integrated modeling [*Jolma et al.*, 2008; *Voinov et al.* 2008], some popular proprietary tools are also included in the catalog.

[43] Table 2 is a functionality matrix for the tools evaluated, and Table 3 has a corresponding list of tool acronyms. Each row of the matrix corresponds to a unique model evaluation tool and the columns identify the capabilities and availability of a given tool. Shorthand notation used in the columns is expanded in the legend, and maps to the subclassifications described in section 2.1.3. Citation counts for each tool were generated using the SCOPUS database [*Burnham*, 2006]. The counts provide a rough gauge of the maturity and impact of a given tool, but do not distinguish between actual tool usage and its simple citation, e.g., as part of a literature review or as an alternative that was considered, but disregarded. Citation counts were not restricted to any particular set of publications. Thus, they serve as an indirect and imperfect measure of the state of practice in the environmental modeling community.

[44] A Web site (www.epa.gov/athens/research/modeling/modelevaluation/index.html), also presented in the auxiliary material, was developed as a companion to the tool catalog. The site augments Table 2 by including download links for identified tools. To request table additions (e.g., for new or overlooked tools) and modifications (e.g., changing a Web address or other corrections) please contact the authors.

3. Discussion

[45] The catalog presented in section 2.2 illustrates the wide variety of tools covering all aspects of model evaluation. Integrating these tools would facilitate routine and comprehensive model evaluations within the environmental modeling community. However, there are numerous barriers to achieving such integration. Many different programming languages, compilers, and development platforms have been utilized, leaving largely incompatible source codes. Furthermore, independent source codes are necessarily a combination of user and model interface code, algorithmic code, and execution management code (i.e., code which exercises arbitrary model executables and, if necessary, performs associated error handling). If the code is not carefully constructed, separating tool science from tool interface and execution management can be difficult. But, such separation is needed if integrated tools are to have a consistent look and feel as well as consistent execution.

[46] Another common integration barrier is the prevalence of different input-output (I/O) file formats utilized by the various standalone executable tools. I/O incompatibility complicates both tool integration (i.e., linking outputs of one tool to the inputs of another) and tool comparison (i.e., comparing different tools with similar functionality). Recent attempts at I/O "standardization" include adopting PEST I/O conventions [*Doherty*, 2004; *Gallagher and Doherty*, 2007; *Skahill et al.*, 2009]), using the XML (extensible markup language) format [*Gernaey et al.*, 2006], and several proposed APIs [*Banta et al.*, 2006; *Reichert*, 2006]). Whether or not one of these proposed standards can be universally accepted and adopted remains to be seen. Rather than developing yet another "locally arbitrary" I/O format, future tool developers may prefer to adopt at least one of the proposed "standards."

[47] From a modeler's perspective, the ability to select multiple standalone tools and have them interoperate via a standardized I/O scheme is all that may be desired or, indeed, required. For example, if DUE, DYNIA, GLUE and MMA all followed the same I/O conventions, modelers could conveniently exercise all seven of the model evaluation methods with a single syntax. On the other hand, for an integrated modeling frame worker, simply agreeing on standard I/O may not be the preferred end game for tool interoperability. One reason is that many frameworks em-

Table 2. Tools for Model Assessment[a]

	Assessment Method							Impact and Availability		
Tool Name	DA[b]	IA[c]	PE[d]	UA[e]	SA[f]	MM[g]	BN[h]	CIT[i]	AV[j]	DIS[k]
ACE		2						2	2, 4	1
ACUARS			4	3	1			2	1	3
AMALGAM			3					5	2	2
BaRE			4	3	1			75	1	3
BATEA			5	3	1			34	1	3
BFL		1						1	2–3	1
BFS							1	68	1	3
BMC			4	3	1			39	1	3
BMElib	1–3							54	2–4	1
BUGS							1	576	2–4	1
CANOPI				3				4	1	3
DAKOTA			1–3	1–2, 4	2			72	2–4	1
DBM		2						187	2–4	1
DDS, DDS-AU			2, 4	3	1			2	2, 4	1
DUE	1–3							7	2–4	1
DYNIA		1–2						51	2–4	1
EESA					1			4	2	1
FANN	NA	NA	NA	NA	NA	NA	NA	6	2–4	1
GEM			5	1	5			9	3–4	1
GLUE			4	3	1			539	3–4	1
HBC							1	0	2–4	1
HDMR	NA	NA	NA	NA	NA	NA	NA	44	2	1
IBUNE			5	3	1	2		3	1	3
JAGS							1	3	2–4	1
JUPITER			1		2			4	2–4	1
LH-OAT					1			10	2–4	1
MARS	NA	NA	NA	NA	NA	NA	NA	814	3–4	1
MCAT		1–2			1,3			9	2–4	1
MCMC-SRSM			5	3	1			5	2–4	2
mGLUE			4	3	1			6	1	3
MICA			5	3	1			5	2, 4	2
MICUT			1, 3					4	2, 4	2
MLBMA, BMA						2		24/300	2, 3	1
MMA						2		11	2–4	1
MOCOM			2					159	2	2
MOGSA					3			63	2	2
MOSCEM			5	3	1			69	2–4	1
NLFIT			2, 5	3	1			135	2, 3	2
NSGA			2					814	2	1
NUSAP						1		37	NA	NA
OSTRICH			1–3		2	2		7	2–4	1
ParaSol			4	3	1			2	2–4	1
PEAS			1		2			2	2–4	2
PEST			1–3	3–4	2			197	2–4	1
PIMLI		1	5	3	1			24	1	3
PSO			2					2473	2–4	1
PyMC			5	3	1		1	0	2–4	1
R	1–3							2439	2–4	1
ReBEL		1						6	2–4	1
RIMME				3	3			4	2	1
SADA	1–2							2	3–4	1
SAMPLING/ANALYSIS					1			5	2	2
SARS-RT					3–4			11	1	3
SCE			2					533	2–4	1
SCEM			5	3	1			74	2–4	1
SIMLAB				1–2	1, 3–5			101	2–4	1
SODA		1	5	3	1			26	1	3
SOLO		2						29	2	2
SRSM	NA	NA	NA	NA	NA	NA	NA	32	2	1
SUFI,SUFI-2			4					16	1	3
SUNGLASSES				3				3	2–4	1
UCODE			1	4	2			148	2–4	1
UNCERT	1–3							20	2–4	1

Table 2. (continued)

Tool Name	Assessment Method							Impact and Availability		
	DA[b]	IA[c]	PE[d]	UA[e]	SA[f]	MM[g]	BN[h]	CIT[i]	AV[j]	DIS[k]
UNCSIM		1	2, 4–5	1–3	2			8	2–4	1
WebGUAC						1		17	NA	NA

[a]NA means not applicable; tool for surrogate-based modeling.
[b]DA, data analysis; 1, population data; 2, geospatial data; 3, time series data.
[c]IA, identifiability analysis; 1, temporal; 2, behavioral; 3, spatial.
[d]PE, parameter estimation; 1, local; 2, global; 3, hybrid; 4, importance sampling; 5, MCMC sampling.
[e]UA, uncertainty analysis; 1, Monte Carlo; 2, stratified sampling; 3, importance sampling; 4, approximate.
[f]SA, sensitivity analysis; 1, screening; 2, local; 3, correlation based; 4, regression based; 5, variance based.
[g]MMA, multimodel analysis; 1, qualitative; 2, quantitative.
[h]BN, Bayesian networks; 1, hierarchical Bayesian network; 2, Bayesian decision network.
[i]CIT, number of citations determined by a search of SCOPUS database.
[j]AV, available materials; 1, method description only; 2, source code; 3, manual; 4, executable.
[k]DIS, form of software distribution; 1, Web download; 2, on request; 3, software not available.

ploy database, dictionary, or object-oriented concepts to store and transfer data through the integrated modeling system. In this paradigm, frameworks do not rely on file-based I/O and associated formatting to transfer data; where file-based I/O is required (e.g., when interfacing with legacy models) special translation software must be constructed.

[48] Another drawback to purely I/O-based standardization, again from the frame worker's perspective, is that execution management issues are not addressed. Like many integrated modeling frameworks, standalone model evaluation tools tend to ''wrap'' around some underlying executable; where ''wrap'' is defined as having control over program inputs and execution, and responsibility for handling errors and exceptions. While model evaluation tools focus on wrapping themselves around models, frameworks go one level further and seek to wrap themselves around both the model and the evaluation tools. Successfully incorporating model evaluation tools within an integrated modeling framework thus requires modifying or kludging (i.e., tricking) the tool to defer to the enveloping framework for execution management.

[49] Because of execution management issues and a general lack of file-based I/O, attempts at framework standardization have focused on defining core ''interface level'' programming standards. For example, the calibration, optimization, and sensitivity and uncertainty (COSU) [*Babendreier*, 2003] API emerged from a recently convened international workshop on environmental modeling [*Nicholson et al.*, 2003]. Rather than specifying a file format for data exchange, COSU specifies a universal data type and subroutines for separately requesting and invoking a computational task. In addition to supporting parallel operations, this construct facilitates deference of execution management from evaluation tool to integrated framework (i.e., a tool requests a computational task and the framework invokes it). Alternatives to the COSU API are also available. For example, the OpenMI [*Gregersen et al.*, 2007] framework specifies efficient data exchange protocols that could be utilized to ensure tool interoperability.

4. Conclusions

[50] A total of 65 software tools were identified and categorized according to seven model evaluation methods: data analysis, identifiability analysis, parameter estimation, sensitivity analysis, uncertainty analysis, multimodel analysis, and Bayesian networks. A Web site (www.epa.gov/athens/research/modeling/modelevaluation/index.html) was developed to aid tool distribution. The identified tools are model-independent and can, in principle, be applied to evaluating any environmental model or modeling system. However, tool interoperability and comparison is complicated by the use of different coding languages, input-output formats, and approaches to execution management.

[51] The review portion of this work is intended to serve as a springboard for identifying and understanding relevant concepts, methods, and issues. The tool catalog and associated Web site are designed to facilitate selection and acquisition of necessary tools for comprehensive model evaluation. An extensive tool catalog has been compiled, and this may minimize redundant tool development in the future.

[52] The assembled list of tools contains a considerable amount of overlapping functionality. This redundancy confounds practitioners tasked with selecting the best tool for the job. Unfortunately, recommending specific tools from the list is beyond the scope of this work, since few of the tools were tested or exercised. In this context, it is more useful to prioritize the underlying methodologies, as opposed to actual technology implementations. Ideally, core model evaluation activities performed for decision support should include (1) data analysis to characterize any available input and response data, (2) sensitivity analysis to determine the most important set of inputs, and (3) uncertainty analysis to establish the range or likelihood of predicted outcomes. If sufficient response data is available, additional activities may also include identifiability analysis and parameter estimation. Parameter estimation approaches, in order of decreasing preference, are those that yield (1) full parameter distributions, (2) point estimates supported with relevant parameter statistics, and (3) point estimates only. Alternatively, and if sufficient expertise is available, Bayesian networks may be utilized in lieu of, or in addition to, parameter estimation and uncertainty analysis. Given adequate resources and knowledge base, several alternative models should be developed and subjected to multimodel analysis.

[53] Looking to the future of integrated environmental modeling, it is worth noting that high level languages (e.g., MatLAB, Octave, Python, and R) are becoming increasingly

Table 3. List of Tool Acronyms

Tool Name	Acronym Description
ACE	alternating conditional expectation
ACUARS	automatic calibration and uncertainty assessment using response surfaces
AMALGAM	a multialgorithm genetically adaptive multiobjective method
BaRE	Bayesian recursive estimation
BATEA	Bayesian total error analysis
BFL	Bayesian filtering library
BFS	Bayesian forecasting system
BMC	Bayesian Monte Carlo
BMElib	Bayesian maximum entropy library
BUGS	Bayesian inference using Gibbs sampling
CANOPI	confidence analysis of physical inputs
DAKOTA	design analysis kit for optimization and terascale applications
DBM	data-based mechanistic modeling
DDS, DDS-AU	dynamically dimensioned search, DDS for approximation of uncertainty
DUE	data uncertainty engine
DYNIA	dynamic identifiability analysis
EESA	elementary effects sensitivity analysis
FANN	fast artificial neural network library
GEM	Gaussian emulation machine
GLUE	generalized likelihood uncertainty engine
HBC	hierarchical Bayesian compiler
HDMR	high-dimensional model representation
IBUNE	integrated Bayesian uncertainty estimator
JAGS	just another Gibbs sampler
JUPITER	joint universal parameter identification and evaluation of reliability
LH-OAT	Latin hypercube sampling one factor at a time
MARS	multivariate adaptive regression splines
MCAT	Monte Carlo analysis toolbox
MCMC-SRSM	Markov chain Monte Carlo stochastic response surface method
mGLUE	modified GLUE
MICA	model-independent Markov chain Monte Carlo analysis
MICUT	model-independent calibration and uncertainty analysis toolbox
MLBMA, BMA	maximum likelihood Bayesian model averaging
MMA	multimodel analysis
MOCOM	multiobjective complex evolution
MOGSA	multiobjective generalized sensitivity analysis
MOSCEM	multiobjective shuffled complex evolution Metropolis
NLFIT	Bayesian nonlinear regression suite
NSGA	nondominated sorting genetic algorithm
NUSAP	numeral, unit, spread, assessment, and pedigree
OSTRICH	optimization software toolkit for research in computational heuristics
ParaSol	parameter solutions
PEAS	parameter estimation accuracy software
PEST	parameter estimation toolkit
PIMLI	parameter identification method localization of information
PSO	particle swarm optimization
PyMC	Python-based Markov chain Monte Carlo library
R	package for statistical computing
ReBEL	recursive Bayesian estimation library
RIMME	random-search inverse methodology for model evaluation
SADA	spatial analysis and decision assistance
SAMPLING/ANALYSIS	screening level sampling and sensitivity analysis tool
SARS-RT	sensitivity analysis based on regional splits and regression trees
SCE	shuffled complex evolution
SCEM	shuffled complex evolution Metropolis
SIMLAB	simulation laboratory for UA/SA
SODA	simultaneous optimization and data assimilation method
SOLO	self-organizing linear output map
SRSM	stochastic response surface method
SUFI, SUFI-2	sequential uncertainty-fitting algorithm
SUNGLASSES	sources of uncertainty global assessment using split samples
UCODE	universal code for inverse modeling
UNCERT	uncertainty analysis, geostatistics and visualization toolkit
UNCSIM	uncertainty simulator
WebGUAC	Web site guidance for uncertainty assessment and communication

popular delivery mechanisms for emerging tools (e.g., see tool summaries in the auxiliary material) and environmental models (e.g., TimML [*Bakker*, 2006] and qtcm [*Lin*, 2008]). Frameworks that can easily accommodate such languages will have an edge over those that only support lower-level languages (e.g., FORTRAN, C/C++, and Java).

[54] In the short term, a lack of consistent standards for dealing with I/O and execution management is likely to be an ongoing problem for both tool development and model integration. Efforts to overcome these barriers have resulted in a plethora of institutional frameworks, each offering a unique approach. Evidenced by this survey, model evaluation tool developers have provided a wide array of useful, but highly repetitive, noninteroperable model evaluation technologies. On some level, tool developers are themselves frame workers, except that they prescribe I/O and execution management schemes to facilitate model evaluation rather than model integration. Thus, the irony in design of both model evaluation tools and integrated modeling systems is that everyone wants to define the ‘‘standard’’ and be the integrative framework. This presents natural, yet unresolved, conflict for a community that would benefit from sharing models and model evaluation tools.

[55] Ultimately, we anticipate robust community support for only a small number of de facto frameworks within different regulatory and application modeling domains. Ongoing multi-institutional efforts will then establish consistent standards across these frameworks (such as those discussed by *USEPA* [2008]). Such advancements will send a clear message to developers: tools that adhere to interoperability standards will have broader support, greater usage, and more impact. In this way, standards and frameworks will encourage enhanced tool interoperability and facilitate a much more comprehensive model evaluation paradigm.

[56] **Acknowledgments.** Comments from Fran Rauschenberg, Robert Swank, and five anonymous reviewers greatly improved the manuscript. Although this work was reviewed by the EPA and was approved for publication, it may not necessarily reflect official agency policy. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation, or favoring by the United States government or any agency thereof.

References

Ahlfeld, D. P., and Y. Hoque (2008), Impact of simulation model solver performance on ground water management problems, *Ground Water*, *46*(5), 716–726, doi:10.1111/j.1745-6584.2008.00454.x.

Ahuja, L. R., J. C. Ascough, and O. David (2005), Developing natural resource models using the object modeling system: Feasibility and challenges, *Adv. Geosci.*, *4*, 29–36.

Ames, D. (2002), Bayesian decision networks for watershed management, Ph.D. thesis, Civ. and Environ. Eng. Dep., Utah State Univ., Logan.

Argent, R. M. (Ed.) (2004), Concepts, methods and applications in environmental model integration, *Environ. Modell. Software*, *19*(3), 217–335, doi:10.1016/S1364-8152(03)00149-X.

Aumann, C. A. (2007), A methodology for developing simulation models of complex systems, *Ecol. Modell.*, *202*, 385–396, doi:10.1016/j.ecolmodel.2006.11.005.

Babendreier, J. (2003), Calibration, Optimization, and Sensitivity and Uncertainty Algorithms Application Programming Interface (COSU-API), in *Proceedings of the International Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling, Rep. NUREG/CP-0187*, edited by T. J. Nicholson et al., pp. A1–A54, Off. of Nucl. Regul. Res., U. S. Nucl. Regul. Comm., Washington, D. C.

Babendreier, J. E., and K. J. Castleton (2005), Investigating uncertainty and sensitivity in integrated, multimedia environmental models: Tools for FRAMES-3MRA, *Environ. Modell. Software*, *20*(8), 1043–1055, doi:10.1016/j.envsoft.2004.09.013.

Babendreier, J., et al. (2007), Managing multimedia pollution for a multimedia world, *Environ. Manage. Mag.*, *10*(12), 6–11.

Bakker, M. (2006), Analytic element modeling of embedded multiaquifer domains, *Ground Water*, *44*(1), 81–85.

Balakrishnan, S., A. Roy, M. G. Ierapetritou, G. P. Flach, and P. G. Georgopoulos (2005), A comparative assessment of efficient uncertainty analysis techniques for environmental fate and transport models: Application to the FACT model, *J. Hydrol.*, *307*(1–4), 204–218, doi:10.1016/j.jhydrol.2004.10.010.

Banta, E. R., E. P. Poeter, J. E. Doherty, and M. C. Hill (Eds.) (2006), JUPITER: Joint Universal Parameter Identification and Evaluation of Reliability—An application programming interface (API) for model analysis, *U.S. Geol. Surv. Tech. Methods*, *TM 6-E1*.

Barlund, I., and S. Tattari (2001), Ranking of parameters on the basis of their contribution to model uncertainty, *Ecol. Modell.*, *142*, 11–23, doi:10.1016/S0304-3800(01)00246-0.

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, *23*(8), 1393–1442, doi:10.1029/WR023i008p01393.

Beck, M. B. (Ed.) (2002), *Environmental Foresight and Models: A Manifesto*, 473 pp., Elsevier Sci., Oxford, U. K.

Beck, M. B., and J. Chen (2000), Assuring the quality of models designed for predictive tasks, in *Sensitivity Analysis*, edited by A. Saltelli et al., pp. 401–420, John Wiley, Chichester, U. K.

Beck, M. B., J. R. Ravetz, L. A. Mulkey, and T. O. Barnwell (1997), On the problem of model validation for predictive exposure assessments, *Stochastic Hydrol. Hydraul.*, *11*, 229–254, doi:10.1007/BF02427917.

Bedogni, M., V. Gabusi, G. Finzi, E. Minguzzi, and G. Pirovano (2005), Sensitivity analysis of ozone long term simulations to grid resolution, *Int. J. Environ. Pollut.*, *24*, 51–63, doi:10.1504/IJEP.2005.007384.

Bekele, E. G., and J. W. Nicklow (2005), Automatic calibration of a semi-distributed hydrologic model using particle swarm optimization, *Eos Trans AGU*, *86*(52), Fall Meet. Suppl., Abstract H54B-07.

Bell, L. S., P. J. Binning, G. Kuczera, and P. M. Kau (2002), Rigorous uncertainty assessment in contaminant transport inverse modeling: A case study of fluoride diffusion through clay liners, *J. Contam. Hydrol.*, *57*(1–2), 1–20, doi:10.1016/S0169-7722(01)00221-2.

Beven, K. (2002), Towards a coherent philosophy for modelling the environment, *Proc. R. Soc. London, Ser. A*, *458*(2026), 2465–2484, doi:10.1098/rspa.2002.0986.

Beven, K. (2007), Towards integrated environmental models of everywhere: Uncertainty, data and modelling as a learning process, *Hydrol. Earth Syst. Sci.*, *11*(1), 460–467.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Bischof, C., P. Khademi, A. Mauer, and A. Carle (1996), Adifor 2.0: Automatic differentiation of Fortran 77 programs, *IEEE Comput. Sci. Eng.*, *3*(3), 18–32, doi:10.1109/99.537089.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow (2004), A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis, *Ecol. Modell.*, *173*, 219–239, doi:10.1016/j.ecolmodel.2003.08.020.

Brockwell, P. J., and R. A. Davis (1996), *Introduction to Time Series and Forecasting*, Springer, New York.

Burnham, J. F. (2006), Scopus database: A review, *Biomed. Digital Libr.*, *3*(1), doi:10.1186/1742-5581-3-1.

Burnham, K. P., and D. R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed., 488 pp., Springer, New York.

Campolongo, F., and A. Saltelli (1997), Sensitivity analysis of an environmental model: An application of different analysis methods, *Reliab. Eng. Syst. Saf.*, *57*(1), 49–69, doi:10.1016/S0951-8320(97)00021-5.

Campolongo, F., A. Saltelli, T. Sorensen, and S. Tarantola (2000), Hitchhiker's guide to sensitivity analysis, in *Sensitivity Analysis*, edited by A. Saltelli et al., pp. 15–47, John Wiley, Chichester, U. K.

Campolongo, F., J. Cariboni, and A. Saltelli (2007), An effective screening design for sensitivity analysis of large models, *Environ. Modell. Software*, *22*(10), 1509–1518, doi:10.1016/j.envsoft.2006.10.004.

Cano, J., M. Delgado, and S. Moral (1993), An axiomatic framework for propagating uncertainty in directed acyclic networks, *Int. J. Approx. Reasoning*, *8*, 253–280, doi:10.1016/0888-613X(93)90026-A.

Chan, K., S. Tarantola, A. Saltelli, and I. M. Sobol' (2000), Variance-based methods, in *Sensitivity Analysis*, edited by A. Saltelli et al., pp. 167–197, John Wiley, Chichester, U. K.

Chen, Y. (2005), Another look at rejection sampling through importance sampling, *Stat. Probab. Lett.*, *72*(4), 277–283, doi:10.1016/j.spl.2005.01.002.

Clark, J. S. (2005), Why environmental scientists are becoming Bayesians, *Ecol. Lett.*, *8*, 2–14, doi:10.1111/j.1461-0248.2004.00702.x.

Clark, J. S., and A. E. Gelfand (2006), A future for models and data in environmental science, *Trends Ecol. Evol.*, *21*(7), 375–380, doi:10.1016/j.tree.2006.03.016.

Cox, D. C., and P. Baybutt (1981), Methods for uncertainty analysis: A comparative survey, *Risk Anal.*, *1*(4), 251–258, doi:10.1111/j.1539-6924.1981.tb01425.x.

Cozman, F. G. (2000), Credal networks, *Artif. Intell.*, *120*, 199–233, doi:10.1016/S0004-3702(00)00029-1.

Cukier, R. I., H. B. Levine, and K. E. Shuler (1978), Nonlinear sensitivity analysis of multiparameter model systems, *J. Comput. Phys.*, *26*(1), 1–42, doi:10.1016/0021-9991(78)90097-9.

Cullen, A. C., and H. C. Frey (1999), *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Plenum, New York.

Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002), A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.*, *6*(2), 182–197, doi:10.1109/4235.996017.

de Smith, M. J., M. F. Goodchild, and P. A. Longley (2008), *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, 2nd ed., Troubador, Leicester, U. K.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177, doi:10.1111/j.1745-6584.2003.tb02580.x.

Doherty, J. (2004), *PEST: Model-Independent Parameter Estimation, User Manual*, 5th ed., Watermark Numer. Comput., Brisbane, Queensland, Australia.

Doherty, J., and B. E. Skahill (2006), An advanced regularization methodology for use in watershed model calibration, *J. Hydrol.*, *327*(3–4), 564–577, doi:10.1016/j.jhydrol.2005.11.058.

Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, doi:10.1029/91WR02985.

Elizondo, D., B. Cappelaere, and C. Faure (2002), Automatic versus manual model differentiation to compute sensitivities and solve non-linear inverse problems, *Comput. Geosci.*, *28*(3), 309–326, doi:10.1016/S0098-3004(01)00048-6.

Elsner, J. B., and T. H. Jagger (2004), A hierarchical Bayesian approach to seasonal hurricane modeling, *J. Clim.*, *17*, 2813–2827, doi:10.1175/1520-0442(2004)017 < 2813:AHBATS > 2.0.CO;2.

Essaid, H. I., I. M. Cozzarelli, R. P. Eganhouse, W. N. Herkelrath, B. A. Bekins, and G. N. Delin (2003), Inverse modeling of BTEX dissolution and biodegradation at the Bemidji, MN crude-oil spill site, *J. Contam. Hydrol.*, *67*(1–4), 269–299, doi:10.1016/S0169-7722(03)00034-2.

Evers, S., and D. N. Lerner (1998), How certain is our estimate of a wellhead protection zone?, *Ground Water*, *36*(1), 49–57, doi:10.1111/j.1745-6584.1998.tb01064.x.

Friedman, J. H. (1991), Multivariate adaptive regression splines, *Ann. Stat.*, *19*(1), 1–67, doi:10.1214/aos/1176347963.

Funtowicz, S. O., and J. R. Ravetz (1990), *Uncertainty and Quality in Science for Policy*, Kluwer Acad., Dordrecht, Netherlands.

Gaber, N., P. Pascual, N. Stiber, and E. Sunderland (2009), Guidance on the development, evaluation, and application of environmental models, *Rep. EPA/100/K-09/003*, U.S. Environ. Prot. Agency, Washington, D. C. (Available at www.epa.gov/crem/library/cred_guidance_0309.pdf)

Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty analysis for a watershed model, *Environ. Modell. Software*, *22*(7), 1000–1020, doi:10.1016/j.envsoft.2006.06.007.

Gan, T. Y., and G. F. Biftu (1996), Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, *Water Resour. Res.*, *32*(12), 3513–3524, doi:10.1029/95WR02195.

Gan, T. Y., E. M. Dlamini, and G. F. Biftu (1997), Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling, *J. Hydrol.*, *192*(1–4), 81–103, doi:10.1016/S0022-1694(96)03114-9.

Georgopoulos, P. G., and P. J. Lioy (2006), From a theoretical framework of human exposure and dose assessment to computational system implementation: The modeling Environment for Total Risk Studies (MENTOR), *J. Toxicol. Environ. Health, Part B*, *9*(6), 457–483.

Gerber, N. (Ed.) (2007), *US EPA CREM Workshop on Integrated Modeling for Integrated Environmental Decision Making*, U.S. Environ. Prot. Agency, Research Triangle Park, N. C. (Available at www.epa.gov/crem/crem_integmodelwkshp.html)

Gernaey, K. V., C. Rosen, D. J. Batstone, and J. Alex (2006), Efficient modelling necessitates standards for model documentation and exchange, *Water Sci. Technol.*, *53*(1), 277–285, doi:10.2166/wst.2006.030.

Gertner, G. (1987), Approximating precision in simulation projections: An efficient alternative to Monte Carlo methods, *For. Sci.*, *33*, 230–239.

Giacobbo, F., M. Marseguerra, and E. Zio (2002), Solving the inverse problem of parameter estimation by genetic algorithms: The case of a groundwater contaminant transport model, *Ann. Nucl. Energy*, *29*(8), 967–981, doi:10.1016/S0306-4549(01)00084-6.

Gregersen, J. B., P. J. A. Gijsbers, and S. J. P. Westen (2007), OpenMI: Open modelling interface, *J. Hydroinf.*, *9*(3), 175–191, doi:10.2166/hydro.2007.023.

Griewank, A., D. Juedes, and J. Utke (1996), Algorithm 755: ADOL-C: A package for the automatic differentiation of algorithms written in C/C++, *Trans. Math. Software*, *22*(2), 131–167, doi:10.1145/229473.229474.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–764, doi:10.1029/97WR03495.

Hamby, D. M. (1994), A review of techniques for parameter sensitivity analysis of environmental models, *Environ. Monit. Assess.*, *32*(2), 135–154, doi:10.1007/BF00547132.

Hamed, M. M., P. B. Bedient, and C. N. Dawson (1996), Probabilistic modeling of aquifer heterogeneity using reliability methods, *Adv. Water Resour.*, *19*(5), 277–295, doi:10.1016/0309-1708(96)00004-8.

Hanson, K. M. and F. M. Hemez (Eds.) (2004), *Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output (SAMO 2004)*, Los Alamos Natl. Lab., Santa Fe, N. M. (Available at http://library.lanl.gov/ccw/samo2004/)

Helton, J. C., and F. J. Davis (2003), Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliab. Eng. Syst. Saf.*, *81*(1), 23–69, doi:10.1016/S0951-8320(03)00058-9.

Helton, J. C., and W. L. Oberkampf (Eds.) (2004), Alternative representations of epistemic uncertainty, *Reliab. Eng. Syst. Saf.*, *85*(1–3), 1–369.

Helton, J. C., J. D. Johnson, C. J. Sallaberry, and C. B. Storlie (2006), Survey of sampling-based methods for uncertainty and sensitivity analysis, *Reliab. Eng. Syst. Saf.*, *91*(10–11), 1175–1209, doi:10.1016/j.ress.2005.11.017.

Hill, C., V. Bugnion, M. Follows, and J. Marshall (2004), Evaluating carbon sequestration efficiency in an ocean circulation model by adjoint sensitivity analysis, *J. Geophys. Res.*, *109*, C11005, doi:10.1029/2002JC001598.

Hill, M. C. (1998), Methods and guidelines for effective model calibration, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *98-4005*.

Hoffman, F. O., and J. S. Hammonds (1994), Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability, *Risk Anal.*, *14*(5), 707–712, doi:10.1111/j.1539-6924.1994.tb00281.x.

Hogue, T. S., S. Sorooshian, H. Gupta, A. Holz, and D. Braatz (2000), A multistep automatic calibration scheme for river forecasting models, *J. Hydrometeorol.*, *1*(6), 524–542, doi:10.1175/1525-7541(2000)001 < 0524:AMACSF > 2.0.CO;2.

Hornberger, G. M., and R. C. Spear (1981), Approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, *12*(1), 7–18.

Iman, R. L., and W. J. Conover (1982), A distribution-free approach to inducing rank correlation among input variables, *Comm. Stat. Simul. Comput.*, *11*(3), 311–334, doi:10.1080/03610918208812265.

Isukapalli, S. S., A. Roy, and P. G. Georgopoulos (1998), Stochastic response surface methods (SRSMs) for uncertainty propagation: Application to environmental and biological systems, *Risk Anal.*, *18*(3), 351–363, doi:10.1111/j.1539-6924.1998.tb01301.x.

Jager, H. I., and A. W. King (2004), Spatial uncertainty and ecological models, *Ecosystems*, *7*, 841–847, doi:10.1007/s10021-004-0025-y.

Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, *29*(8), 2637–2649, doi:10.1029/93WR00877.

Jakeman, A. J., and R. A. Letcher (2003), Integrated assessment and modelling: Features, principles and examples for catchment management, *Environ. Modell. Software*, *18*(6), 491–501, doi:10.1016/S1364-8152(03)00024-0.

Jakeman, A. J., R. A. Letcher, and J. P. Norton (2006), Ten iterative steps in development and evaluation of environmental models, *Environ. Modell. Software*, *21*(5), 602–614, doi:10.1016/j.envsoft.2006.01.004.

Johnston, J. M., J. H. Novak, and S. R. Kraemer (2004), Multimedia integrated modeling for environmental protection: Introduction to a collaborative framework, *Environ. Modell. Assess.*, *63*(1), 253–263.

Jolma, A., D. P. Ames, N. Horning, H. Mitasova, M. Neteler, A. Racicot, and T. Sutton (2008), Free and open source geospatial tools for environmental modelling and management, in *Environmental Modelling, Software and Decision Support: State of the Art and New Perspectives*, vol. e, edited by A. J. Jakeman et al., pp. 163–180, Elsevier, Amsterdam.

Kavetski, D., S. W. Franks, and G. Kuczera (2002), Confronting input uncertainty in environmental modeling, in *Calibration of Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 49–68, AGU, Washington, D. C.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, *42*, W03407, doi:10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, *42*, W03408, doi:10.1029/2005WR004376.

Khadam, I. M., and J. J. Kaluarachchi (2004), Use of soft information to describe the relative uncertainty of calibration data in hydrologic models, *Water Resour. Res.*, *40*, W11505, doi:10.1029/2003WR002939.

Konikow, L. F., and J. D. Bredehoeft (1992), Ground-water models cannot be validated, *Adv. Water Resour.*, *15*(1), 75–83, doi:10.1016/0309-1708(92)90033-X.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*(1–2), 161–177, doi:10.1016/j.jhydrol.2006.05.010.

Lahlou, M., L. Shoemaker, S. Choundhury, R. Elmer, A. Hu, H. Manguerra, and A. Parker (1998), Better Assessment Science Integrating Point and Nonpoint Sources—BASINS, *Rep. EPA-823-B-98-006*, U.S. Environ. Prot. Agency, Washington, D. C.

Leavesley, G. H., S. L. Markstrom, M. S. Brewer, and R. J. Viger (1996), The modular modeling system (MMS)—The physical process modeling component of a database-centered decision support system for water and power management, *Water Air Soil Pollut.*, *90*(1–2), 303–311, doi:10.1007/BF00619290.

Levenberg, K. (1944), A method for the solution of certain problems in least squares, *Q. Appl. Math.*, *2*, 164–168.

Li, H., and J. Wu (2006), Uncertainty analysis in ecological studies: An overview, in *Scaling and Uncertainty Analysis in Ecology*, edited by J. Wu et al., pp. 43–46, Springer, Dordrecht, Netherlands.

Lin, J. (2008), qtcm 0.1.2: A Python implementation of the Neelin-Zeng quasi-equilibrium tropical circulation model, *Geosci. Model Dev. Discuss.*, *1*, 315–344.

Lindenschmidt, K.-E., K. Fleischbein, and M. Baborowski (2007), Structural uncertainty in a river water quality modelling system, *Ecol. Modell.*, *204*, 289–300, doi:10.1016/j.ecolmodel.2007.01.004.

Linkov, I., and D. Burmistrov (2003), Model uncertainty and choices made by modelers: Lessons learned from the International Atomic Energy Agency model intercomparisons, *Risk Anal.*, *23*(6), 1297–1308, doi:10.1111/j.0272-4332.2003.00402.x.

Ljung, L. (1999), *System Identification: Theory for the User*, 2nd ed., Prentice Hall, Englewood Cliffs, N. J.

Manache, G., and C. S. Melching (2008), Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters, *Environ. Modell. Software*, *23*(5), 549–562.

Marin, C. M., V. Guvanasen, and Z. A. Saleem (2003), The 3MRA risk assessment framework—A flexible approach for performing multimedia, multipathway, and multireceptor risk assessments under uncertainty, *Hum. Ecol. Risk Assess.*, *9*(7), 1655–1677, doi:10.1080/714044790.

Marquardt, D. (1963), An algorithm for least-squares estimation of nonlinear parameters, *SIAM J. Appl. Math.*, *11*, 431–441, doi:10.1137/0111030.

Matott, L. S. (2005), Ostrich: An optimization software tool, documentation and user's guide, version 1.6, Dep. of Civ., Struct., and Environ. Eng., Univ. at Buffalo, State Univ. of N. Y., Buffalo. (Available at www.groundwater.buffalo.edu)

Matott, L. S., and A. J. Rabideau (2008), Calibration of subsurface batch and reactive-transport models involving complex biogeochemical processes, *Adv. Water Resour.*, *31*(2), 269–286, doi:10.1016/j.advwatres.2007.08.005.

Matthies, M., C. Giupponi, and B. Ostendorf (Eds.) (2007), Environmental decision support systems, *Environ. Modell. Software*, *22*(2), 123–278.

McEnery, J., J. Ingram, Q. Duan, T. Adams, and L. Anderson (2005), NOAA's advanced hydrologic prediction service, *Bull. Am. Meteorol. Soc.*, *86*(3), 375–385, doi:10.1175/BAMS-86-3-375.

McGarity, T., and W. Wagner (2003), Legal aspects of the regulatory use of environmental modeling, *Environ. Law Rep.*, *33*, 10,751–10,774.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*(2), 239–245, doi:10.2307/1268522.

Mertens, J., H. Madsen, L. Feyen, D. Jacques, and J. Feyen (2004), Including prior information in the estimation of effective soil parameters in unsaturated zone modelling, *J. Hydrol.*, *294*(4), 251–269, doi:10.1016/j.jhydrol.2004.02.011.

Montanari, A. (2007), What do we mean by "uncertainty"? The need for a consistent wording about uncertainty in hydrology, *HPToday*, *21*(6), 841–845.

Morgan, M. G., and M. Henrion (1990), *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge Univ. Press, New York.

Morris, M. D. (2006), Input screening: Finding the important model inputs on a budget, *Reliab. Eng. Syst. Saf.*, *91*(10–11), 1252–1256, doi:10.1016/j.ress.2005.11.022.

Mugunthan, P., and C. A. Shoemaker (2006), Assessing the impacts of parameter uncertainty for computationally expensive groundwater models, *Water Resour. Res.*, *42*, W10428, doi:10.1029/2005WR004640.

Muleta, M. K., and J. W. Nicklow (2005), Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, *J. Hydrol.*, *306*(1–4), 127–145, doi:10.1016/j.jhydrol.2004.09.005.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part 1—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.

Nauta, M. J. (2000), Separation of uncertainty and variability in quantitative microbial risk assessment models, *Int. J. Food Microbiol.*, *57*(1–2), 9–18, doi:10.1016/S0168-1605(00)00225-7.

Neilson, B. T., D. P. Ames, and D. K. Stevens (2002), Application of Bayesian decision networks to total maximum daily load analysis, in *Proceedings of the 2002 Conference on Total Maximum Daily Load (TMDL) Environmental Regulations*, edited by A. Saleh, pp. 221–231, Am. Soc. of Agric. and Biol. Eng., St. Joseph, Mich.

Newbold, S. (2002), Integrated modeling for watershed management: Multiple objectives and spatial effects, *J. Am. Water Resour. Assoc.*, *38*(2), 341–352, doi:10.1111/j.1752-1688.2002.tb04321.x.

Nicholson, T. J., J. E. Babendreier, P. D. Meyer, S. Mohanty, B. B. Hicks, and G. H. Leavesley (2003), Proceedings of the International Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling, Rockville, MD, *Rep. EPA/600/R-04/117*, U.S. Environ. Prot. Agency, Washington, D. C.

Nissen, S. (2003), Implementation of a fast artificial neural network library (FANN), 92 pp., Dep. of Comput. Sci., Univ. of Copenhagen, Copenhagen. (Available at http://iweb.dl.sourceforge.net/sourceforge/fann/fann_doc_complete_1.0.pdf)

Norton, J. P. (1987), Identification and application of bounded-parameter models, *Automatica*, *23*, 497–507, doi:10.1016/0005-1098(87)90079-3.

Norton, J. P. (1996), Roles for deterministic bounding in environmental modelling, *Ecol. Modell.*, *86*, 157–161, doi:10.1016/0304-3800(95)00045-3.

Norton, J. P., F. H. S. Chiew, G. C. Dandy, and H. R. Maier (2005), A parameter-bounding approach to sensitivity assessment of large simulation models, in *MODSIM 2005 International Congress on Modelling and Simulation: Modelling and Simulation Society of Australia and New Zealand, December 2005*, edited by A. Zerger and R. M. Argent, pp. 2519–2525, Modell. and Simul. Soc. of Aust. and N. Z., Canberra.

Norton, J. P., J. D. Brown, and J. Mysiak (2006), To what extent, and how, might uncertainty be defined?, *Integr. Assess. J.*, *6*(1), 83–88.

O'Hagan, A. (2006), Bayesian analysis of computer code outputs: A tutorial, *Reliab. Eng. Syst. Saf.*, *91*(10–11), 1290–1300, doi:10.1016/j.ress.2005.11.025.

Oreskes, N. (2003), The role of quantitative models in science, in *The Role of Models in Ecosystem Science*, edited by C. D. Canham et al., pp. 13–31, Princeton Univ. Press, Princeton, N. J.

Pachepsky, Y. A., A. K. Guber, M. T. van Genuchten, T. J. Nicholson, R. E. Cady, J. Simunek, and M. G. Schaap (2006), Model abstraction techniques for soil water flow and transport, *Rep. NUREG/CR-6884*, 175 pp., Nucl. Regul. Comm., Washington, D. C.

Pascual, P. (2005), Wresting environmental decisions from an uncertain world, *Environ. Law Rep.*, *35*, 10,539–10,549.

Petzold, L., S. Li, Y. Cao, and R. Serban (2006), Sensitivity analysis of differential-algebraic equations and partial differential equations, *Comput. Chem. Eng.*, *30*(10–12), 1553–1559, doi:10.1016/j.compchemeng.2006.05.015.

Poeter, E. P., and M. C. Hill (1998), Documentation of UCODE: A computer code for universal inverse modeling, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *98-4080*.

Portielje, R., T. Hvitved-Jacobsen, and K. Schaarup-Jensen (2000), Risk analysis using stochastic reliability methods applied to two cases of deterministic water quality models, *Water Res.*, *34*(1), 153–170, doi:10.1016/S0043-1354(99)00131-1.

Rabitz, H., and Ö. Aliş (1999), General foundations of high-dimensional model representations, *J. Math. Chem.*, *25*(2–3), 197–233, doi:10.1023/A:1019188517934.

Raghavan, V. (2003), A comparison between gradient-based and heuristic algorithms for inverse modeling of two-dimensional groundwater flow, M.S. thesis, Univ. at Buffalo, State Univ. of N. Y., Buffalo.

Rankinen, K., T. Karvonen, and D. Butterfield (2006), An application of the GLUE methodology for estimating the parameters of the INCA-N model, *Sci. Total Environ.*, *365*(1–3), 123–139, doi:10.1016/j.scitotenv.2006.02.034.

Refsgaard, J. C., and H. J. Henriksen (2004), Modelling guidelines—Terminology and guiding principles, *Adv. Water Resour.*, *27*(1), 71–82, doi:10.1016/j.advwatres.2003.08.006.

Refsgaard, J. C., H. J. Henriksen, B. Harrar, H. Scholten, and A. Kassahun (2005), Quality assurance in model based water management—Review of existing practice and outline of new approaches, *Environ. Modell. Software*, *20*(10), 1201–1215, doi:10.1016/j.envsoft.2004.07.006.

Refsgaard, J. C., J. P. van der Sluijs, A. L. Hojberg, and P. A. Vanrolleghem (2007), Uncertainty in the environmental modelling process—A framework and guidance, *Environ. Modell. Software*, *22*(11), 1543–1556, doi:10.1016/j.envsoft.2007.02.004.

Regan, H. M., H. R. Akcakaya, S. Ferson, K. V. Root, S. Carroll, and L. R. Ginzburg (2003), Treatments of uncertainty and variability in ecological risk assessment of single-species populations, *Hum. Ecol. Risk Assess.*, *9*(4), 889–906, doi:10.1080/713610015.

Reichert, P. (2006), A standard interface between simulation programs and systems analysis software, *Water Sci. Technol.*, *53*(1), 267–275, doi:10.2166/wst.2006.029.

Rotmans, J., and M. van Asselt (2001a), Uncertainty management in integrated assessment modeling: Towards a pluralistic approach, *Environ. Monit. Assess.*, *69*(2), 101–130, doi:10.1023/A:1010722120729.

Rotmans, J., and M. van Asselt (2001b), Uncertainty in integrated assessment modelling: A labrynthic path, *Integr. Assess. J.*, *2*, 43–55, doi:10.1023/A:1011588816469.

Sadoddin, A., R. A. Letcher, A. J. Jakeman, and L. T. H. Newham (2005), A Bayesian decision network approach for assessing the ecological impacts of salinity management, *Math. Comput. Simul.*, *69*(1–2), 162–176, doi:10.1016/j.matcom.2005.02.020.

Saltelli, A., K. Chan, and E. M. Scott (2000), *Sensitivity Analysis*, John Wiley, Chichester, U. K.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley, New York.

Sandu, A., D. N. Daescu, G. R. Carmichael, and T. Chai (2005), Adjoint sensitivity analysis of regional air quality models, *J. Comput. Phys.*, *204*(1), 222–252, doi:10.1016/j.jcp.2004.10.011.

Schultz, C. (2008), Responding to scientific uncertainty in U.S. forest policy, *Environ. Sci. Policy*, *11*, 253–271, doi:10.1016/j.envsci.2007.09.002.

Schuol, J., and K. C. Abbaspour (2006), Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa, *Adv. Geosci.*, *9*, 137–143.

Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.

Skaggs, T. H., and D. A. Barry (1996), Assessing uncertainty in subsurface solute transport: Efficient first-order reliability methods, *Environ. Software*, *11*(1–3), 179–184, doi:10.1016/S0266-9838(96)00025-1.

Skahill, B. E., J. S. Baggett, S. Frankenstein, and C. W. Downer (2009), More efficient PEST compatible model independent model calibration, *Environ. Model. Softw.*, *24*(4), 517–529, doi:10.1016/j.envsoft.2008.09.011.

Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6.

Söderström, T., and P. Stoica (1989), *System Identification*, Prentice-Hall Int., New York.

Solomatine, D. P., Y. B. Dibike, and N. Kukuric (1999), Automatic calibration of groundwater models using global optimization techniques, *Hydrol. Sci. J.*, *44*(6), 879–894.

Sonich-Mullin, C. (2001), Harmonizing the incorporation of uncertainty and variability into risk assessment: An international approach, *Hum. Ecol. Risk Assess.*, *7*(1), 7–13, doi:10.1080/20018091094187.

Spear, R. C., and G. M. Hornberger (1980), Eutrophication in Peel Inlet, II: Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, *14*(1), 43–49, doi:10.1016/0043-1354(80)90040-8.

Suter, G. W., III (Ed.) (2006), *Ecological Risk Assessment*, 2nd ed., CRC Press, Boca Raton, Fla.

Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, *43*, W01413, doi:10.1029/2005WR004723.

Tonkin, M. J., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, *41*, W10412, doi:10.1029/2005WR003995.

Trucano, T. G., L. P. Swiler, T. Igusa, W. L. Oberkampf, and M. Pilch (2006), Calibration, validation, and sensitivity analysis: What's what, *Reliab. Eng. Syst. Saf.*, *91*(10–11), 1331–1357, doi:10.1016/j.ress.2005.11.031.

Tsai, C. W., and S. Franceschini (2005), Evaluation of probabilistic point estimate methods in uncertainty analysis for environmental engineering applications, *J. Environ. Eng.*, *131*(3), 387–395, doi:10.1061/(ASCE)0733-9372(2005)131:3(387).

U.S. Environmental Protection Agency (USEPA) (1992), Guidelines for exposure assessment, *Rep. EPA/600/Z-92/001*, Off. of Res. and Dev., Washington, D. C.

U.S. Environmental Protection Agency (USEPA) (2008), Integrated modeling for integrated environmental decision making, *Rep. EPA 100/R-08/010*, Washington, D. C.

van Asselt, M., and J. Rotmans (2002), Uncertainty in integrated assessment modelling, *Clim. Change*, *54*(1–2), 75–105, doi:10.1023/A:1015783803445.

Van Bers, C., D. Petry, and C. Pahl-Wostl (Eds.) (2007), *Global Assessments: Bridging Scales and Linking to Policy*, *GWSP Issues Global Water Syst. Res.*, vol. 2, pp. 102–106, Global Water Syst. Proj., Bonn, Germany.

van der Merwe, R., T. K. Leen, Z. Lu, S. Frolov, and A. M. Baptista (2007), Fast neural network surrogates for very high dimensional physics-based models in computational oceanography, *Neural Networks*, *20*(4), 462–478, doi:10.1016/j.neunet.2007.04.023.

van der Slujis, J. P. (2007), Uncertainty and precaution in environmental management: insights from the UPEM conference, *Environ. Modell. Software*, *22*(5), 590–598, doi:10.1016/j.envsoft.2005.12.020.

van der Sluijs, J. P., et al. (2003), *RIVM/MNP Guidance for Uncertainty Assessment and Communication: Detailed Guidance*, pp. 2003–2163, Utrecht Univ., Utrecht, Netherlands.

van Griensven, A., and T. Meixner (2006), Methods to quantify and identify the sources of uncertainty for river basin water quality models, *Water Sci. Technol.*, *53*(1), 51–59, doi:10.2166/wst.2006.007.

Varis, O. (1997), Bayesian decision analysis for environmental and resource management, *Environ. Modell. Software*, *12*(2–3), 177–185, doi:10.1016/S1364-8152(97)00008-X.

Vermeulen, P. T. M., A. W. Heemink, and J. R. Valstar (2005), Inverse modeling of groundwater flow using model reduction, *Water Resour. Res.*, *41*, W06003, doi:10.1029/2004WR003698.

Vermeulen, P. T. M., C. B. M. te Stroet, and A. W. Heemink (2006), Model inversion of transient nonlinear groundwater flow models using model reduction, *Water Resour. Res.*, *42*, W09417, doi:10.1029/2005WR004536.

Voinov, A., R. R. Hood, J. D. Daues, H. Assaf, and R. Stewart (2008), Building a community modeling and information sharing culture, in *Environmental Modelling, Software and Decision Support: State of the Art and New Perspectives*, edited by A. J. Jakeman et al., pp. 345–365, Elsevier, Amsterdam.

von Krauss, M. K., M. Henze, and J. Ravetz (Eds.) (2004), *Proceedings of the International Symposium on Uncertainty and Precaution in Environmental Management (UPEM)*, Tech. Univ. of Denmark, Copenhagen. (Available at http://upem.er.dtu.dk/programme.htm)

Vose, D. (2000), *Risk Analysis: A Quantitative Guide*, 2nd ed., John Wiley, Chichester, U. K.

Vrugt, J., and B. Robinson (2007), Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(3), 708–711, doi:10.1073/pnas.0610471104.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Walker, W. E., P. Harremoes, J. Rotmans, J. P. van der Slujis, M. van Asselt, P. Janssen, and M. P. Krayer von Krauss (2003), Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support, *Integr. Assess.*, *4*(1), 5–17, doi:10.1076/iaij.4.1.5.16466.

Wang, S. W., S. Balakrishnan, and P. Georgopoulos (2005), Fast equivalent operational model of tropospheric alkane photochemistry, *AIChE J.*, *51*(4), 1297–1303, doi:10.1002/aic.10431.

Wikle, C. K. (2003), Hierarchical models in environmental science, *Int. Stat. Rev.*, *171*, 181–199.

Wikle, C. K., L. M. Berliner, and N. Cressie (1998), Hierarchical Bayesian space-time models, *Environ. Ecol. Stat.*, *5*, 117–154, doi:10.1023/A:1009662704779.

Wu, F.-C., and Y.-P. Tsang (2004), Second-order Monte Carlo uncertainty/variability analysis using correlated model parameters: Application to salmonid embryo survival risk assessment, *Ecol. Modell.*, *177*(3–4), 393–414, doi:10.1016/j.ecolmodel.2004.02.016.

Xiu, D., and J. S. Hesthaven (2005), High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.*, *27*(3), 1118–1139, doi:10.1137/040615201.

Xiu, D., and G. Karniadakis (2002), Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, *Comput. Methods Appl. Mech. Eng.*, *191*, 4927–4948, doi:10.1016/S0045-7825(02)00421-8.

Xiu, D., D. Lucor, C.-H. Su, and G. Karniadakis (2002), Stochastic modeling of flow-structure interactions using generalized polynomial chaos, *J. Fluids Eng.*, *124*, 51–59, doi:10.1115/1.1436089.

Yager, R. M. (2004), Effects of model sensitivity and nonlinearity on nonlinear regression of ground water flow, *Ground Water*, *42*(3), 390–400, doi:10.1111/j.1745-6584.2004.tb02687.x.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*(1–4), 83–97, doi:10.1016/S0022-1694(97)00107-8.

Ye, M., P. D. Meyer, and S. P. Neumann (2008), On model selection criteria in multimodel analysis, *Water Resour. Res.*, *44*, W03428, doi:10.1029/2008WR006803.

Young, P. C., S. Parkinson, and M. J. Lees (1996), Simplicity out of complexity in environmental modelling: Occam's razor revisited, *J. Appl. Stat.*, *23*, 165–210, doi:10.1080/02664769624206.

Yu, P.-S., T.-C. Yang, and S.-J. Chen (2001), Comparison of uncertainty analysis methods for a distributed rainfall-runoff model, *J. Hydrol.*, *244*(1–2), 43–59, doi:10.1016/S0022-1694(01)00328-6.

J. E. Babendreier and S. T. Purucker, Ecosystems Research Division, National Exposure Research Laboratory, Office of Research and Development, U.S. EPA, 960 College Station Road, Athens, GA 30605, USA. (babendreier.justin@epa.gov; purucker.tom@epa.gov)

L. S. Matott, Department of Civil and Environmental Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. (lsmatott@uwaterloo.ca)