

## A guide to good practice in modeling semantics for authors and referees

Keith Beven<sup>1</sup> and Peter Young<sup>1</sup>

Received 3 April 2013; revised 24 June 2013; accepted 25 June 2013; published 26 August 2013.

[1] This opinion piece makes some suggestions about guidelines for modeling semantics that can be referred to by authors and referees. We discuss descriptions of model structures, different forms of simulation and prediction, descriptions of different sources of uncertainty in modeling practice, the language of model validation, and concepts of predictability and fitness-for-purpose. While not expecting universal agreement on these suggestions, given the loose usage of words in the literature, we hope that the discussion of the issues involved will at least give pause for thought and encourage good practice in model development and applications.

**Citation:** Beven, K., and P. Young (2013), A guide to good practice in modeling semantics for authors and referees, *Water Resour. Res.*, 49, 5092–5098, doi:10.1002/wrcr.20393.

### 1. Introduction

[2] There are many examples of bad practice in the use of words in the hydrological modeling literature, and we have found it rather surprising that referees have not always picked up on this. Thus, we would respectfully like to suggest that it would be rather useful to have some guidelines for good practice in modeling semantics that can be referred to by authors and referees, with the aim of encouraging good practice in model development and applications. Here we discuss some of the issues involved, drawing on practice in other areas. Section 2 discusses descriptions of model structures; section 3 discusses the different forms of simulation and prediction; section 4 discusses descriptions of different sources of uncertainty in modeling practice; section 5 discusses the language of model validation, and section 6 concepts of predictability and fit-for-purpose.

### 2. Model Structures

[3] There are many different classifications of model structures in the hydrological literature dating back to at least *Clarke* [1973]. A variety of categories such as lumped or distributed, deterministic or stochastic, inductive or deductive, and black-box or process-based are commonly used, while particular models transcend such crisp boundaries by being described as semidistributed, gray-box, or “using stochastic ensembles of deterministic model runs.” One term that referees should actively discourage is “physically based.” This has become corrupted in its use (for example, in being applied to models such as SWAT) [e.g., *Gassman et al.*, 2010]. That is not to say that models

cannot have components that aim to represent different hydrological processes, or that models cannot be interpreted in physically meaningful ways. Such alternative descriptions are more acceptable and correct. Even simple transfer function models derived from observations can have a mechanistic interpretation in terms of meaningful time constants or gains [e.g., *Young*, 2003, 2013], but there are (at least as yet) no models that are based on correct physical principles in hydrology. Models based on the Darcy-Richards equation and Manning equation might claim to have a physical foundation, but the empirical “physics” on which they are based is clearly not universally applicable [see, for example, *Beven and Germann*, 2013; *Beven*, 2013]. In principle, of course, we can have deductive physically based models derived from principles of conservation of mass, energy, and momentum (as in the Representative Elementary Watershed “REW” framework of *Reggiani et al.* [2000, 2001]) and simplifications of the Navier-Stokes equations [e.g., *Neuman*, 1977]. In both cases, except for some special cases, the auxiliary and boundary conditions required to have a working model at useful scales will reintroduce an inductive element.

[4] If the difficulty of applying physical principles in hydrology is embraced, then it seems that we could simplify approaches to the classification of models to simply distinguish between models that are deductive from those that are inductive, possibly with an additional qualifier describing whether a model calculates spatially distributed values of variables. Deductive model structures are defined prior to applying the available observational data, while inductive model structures are normally identified on the basis of the available observational data.

[5] Of course, the deductive-inductive dichotomy is rather unsatisfactory [*Young*, 2011, 2013], and these approaches are certainly not mutually exclusive. The best aspects of both approaches can be exploited to produce a systematic approach to hydrological model development. For example, there are deductive models that are based on

<sup>1</sup>Lancaster Environment Centre, Lancaster University, Lancaster, UK.

Corresponding author: K. Beven, Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK. (K.Beven@lancaster.ac.uk)

principles originally derived by induction (Darcy-Richards, Manning), and the calibration (estimation) of parameters will introduce an inductive element (and associated uncertainty) to even the most deductive model [Beven, 2012; Young, 2011, 2013]. Referees should ensure, however, that the deductive and inductive elements of a modeling study are clearly differentiated.

[6] We also recognize that a distinction between deterministic and stochastic model structures can be useful, even if the deductive role of purely deterministic model runs would appear to be rather limited (see section 4). Given the nature of hydrological systems and the uncertainties in the modeling process, referees should expect that deterministic models are applied with a proper account taken of the relevant uncertainties. They should not, therefore, accept papers based on single deterministic model runs, unless a convincing argument is made as a special case. This should otherwise be considered bad practice.

### 3. Simulation, Prediction, Projection, Forecasting, and Hindcasting

[7] There is considerable confusion in the literature about the use of the terms simulation, prediction, projection, forecasting, and hindcasting; the usage in hydrological modeling in some cases conflicts with usage in other areas. It would be desirable for authors and referees to use of these terms with more rigor. The terms themselves are, with one exception, not that ambiguous (see Table 1). Simulation is the quantitative reproduction of the behavior of a system, given some defined inputs but without reference to any observed outputs. Forecasting is the quantitative reproduction of the behavior of a system ahead of time, but given observations of the inputs, state variables (where applicable), and outputs up to the present time (the forecasting origin). Hindcasting is the application of forecasting to past data as if the inputs and outputs were only available up to a certain point in time (most forecasting that is reported in the literature is, for obvious reasons, actually hindcasting). We should also add projection to the list of activities. Projection, or “what-if” simulation, is the simulation of the future behavior of a system given prior assumptions about future input data.

[8] That leaves “prediction” that *has* been used ambiguously in hydrology. In many other areas of modeling, prediction is used synonymously with forecasting, i.e., forecasting into the future with inputs and other observa-

tions known only up to the present forecasting origin. This is consistent with the Latin origin of the word (from *praedice*, to make a statement about the future). In hydrology, however, it has been used more generally for the case of any model output in time and in space, whether produced by simulation or forecasting and in particular, for the temporal or spatial outputs of simulation models used in model evaluation (for example, in “validation” or “verification” runs of a simulation model, see section 4). But this is not making a statement about the future, it is a simulated model output. To avoid this ambiguity we would thus suggest that the word prediction should generally be avoided (the terms in Table 1 cover all useful circumstances). We suspect, however, that it will continue to be used, with a variety of meanings but would suggest that referees should ensure that this and the other different terms are at least defined and used clearly by authors in ways consistent with Table 1.

[9] The situation is further confused by two different types of forecasting that have not always been distinguished in the hydrological literature. It is necessary to differentiate between “ex-ante” and “ex-post” forecasting. These are terms that are well known in the forecasting literature [Young, 2013]. Here ex-ante forecasts are real forecasts, only dependent on data up to the forecasting origin. Ex-post forecasting, on the other hand, are model outputs calculated *as if* future *measured* data (particularly the inputs, such as rain and evapotranspiration, which are particularly difficult to forecast) are also known *after* the forecasting origin. Note the difference here between ex-post forecasting and simulation. Both assume that the input data required are known, but ex-post forecasting also assumes that observations of the system response are also available up to the start of the forecasting interval. Table 1 summarizes these definitions.

[10] Ex-post forecasting (actually more strictly ex-post hindcasting since measured inputs cannot be available into the actual future) should never be portrayed either as real forecasting or as real simulation, since this is highly misleading: indeed, it is difficult to justify ex-post forecasting in any realistic application [e.g., Beven, 2009; Young, 2013], unless it is made quite clear that it is being used only as a check on the *stochastic simulation* abilities of the model. Again we suggest that referees should act to ensure good practice by requiring revision to papers that use “prediction” without clarification of its purpose, and in particular where it is used to mean ex-post forecasting. In rainfall-

**Table 1.** Forms of Modeling<sup>a</sup>

Term	Inputs	Known Outputs Used in Model Simulations or Forecasts	Model Simulated or Forecast Outputs
Simulation	$I(t), t = 1 : N$		$\hat{O}(t), t = 1 : N$
Ex-ante forecasting	$I(t), t = 1 : t_o$	$O(t), t = 1 : t_o$	$\hat{O}(t), t = t_o + 1 : t_o + k$
Ex-post forecasting	$I(t), t = 1 : t_o + k$	$O(t), t = 1 : t_o$	$\hat{O}(t), t = t_o + 1 : t_o + k$
Hindcasting	$I(t), t = 1 : t_f$	$O(t), t = 1 : t_f$	$\hat{O}(t), t = t_f + 1 : t_f + k$
Projection	$\hat{I}(t), t = t_o + 1 : N$		$\hat{O}(t), t = t_o + 1 : N$

<sup>a</sup> $I$  and  $O$  represent vectors of known input and output values, respectively;  $\hat{I}$  represents a vector of assumed input values;  $\hat{O}$  represents a vector of simulated or forecast values;  $t_o$  represents now (or the start time of a simulation run);  $t_f$  represents a point in the past used as the start time for a hindcast treated as an ex-ante forecast;  $k$  represents a forecasting horizon or lead time; and  $N$  represents a period of simulation (note that discrete time is assumed here, so that  $t$  is the sampling index).

flow modeling, for instance, an ex-post forecast that uses future observed rainfall to forecast flow over the forecast horizon is clearly misleading on its own, since a true ex-ante forecast would require a forecast of future rainfall beyond the forecasting origin, based *only* on measurements of rainfall (and any other related variables required, such as potential evapotranspiration) available at the forecasting origin.

#### 4. Describing Uncertainties

[11] There has been significant discussion in the literature about what methods of uncertainty analysis to use. Because many of the sources of uncertainty are epistemic in nature, the debate about uncertainty estimation techniques will not be resolved soon (see recent discussions in *Beven* [2006, 2008, 2012], *Montanari* [2007], *Beven et al.* [2008], *Clark et al.* [2011, 2012], *Beven et al.* [2012], *McMillan et al.* [2012], *Beven and Smith* [2013], and *Young* [2013]). What is important is that some consideration of uncertainty is incorporated into any modeling study, and as noted above, referees should ensure that this is the case.

[12] There is also an extensive literature on the classification of different types of uncertainty [*Hoffman and Hammonds*, 1994; *Helton and Burmaster*, 1996; *Walker et al.*, 2003; *Van der Sluijs et al.*, 2005; *Refsgaard et al.*, 2006, 2007, 2013; *Warmink et al.*, 2010; *Spiegelhalter and Riesch*, 2011]. There are multiple classifications of varying degrees of complexity. Operationally, however, we would suggest that sources of uncertainty can be simply differentiated as either aleatory or epistemic (see examples in Table 2). Aleatory uncertainties (from the Latin *alea* meaning a die, or game of dice) are concerned with apparent random variability and can be treated directly in probabilistic terms. In this definition, aleatory uncertainties can be treated as statistical variables, potentially of complex structure [e.g., *Koutsyiannis*, 2010]. It is sometimes suggested that aleatory uncertainties are irreducible, but in the working definition we are suggesting here that will not necessarily be the case.

[13] Epistemic uncertainties (from the Greek *πιστημη*, for knowledge or science), on the other hand, arise from lack of knowledge and understanding and, it is often suggested, could be reduced in principle by having more or better measurements; or by new science. Epistemic

**Table 2.** Examples of Aleatory and Epistemic Uncertainties in Hydrological Modeling

Source of Uncertainty	Aleatory Component	Epistemic Component
Rainfall observations	Gauge errors, after allowing for consistent bias associated with height, wind speed, shield design etc. Radar reflectivity residuals, after corrections for type of reflector, drop size distribution, attenuation, bright band, and other anomalies	Neglect of, or incorrect corrections for, gauge errors and radar estimates. Errors associated with lack of knowledge of spatial heterogeneity
Rainfall interpolations	Residuals for any storm given choice of interpolation method	Rain gauge network or choice of interpolation method might not be appropriate for all storms, unobserved cells, nonstationary spatial covariance characteristics/choice of fractal dimension
Evapotranspiration estimates	Random measurement errors in meteorological variables	Biases in meteorological variables relative to effective values required to estimate catchment average evapotranspiration Choice of assumptions in process representations The choice of functions in simulation or forecasting equation. Neglect of local boundary condition effects over a heterogeneous surface Neglect of wet/dry surface effects
Discharge observations	Fluctuations in water level observations Observation error in direct discharge measurements for rating curve definition Residuals from rating curve	Poor methodology and operator error in direct discharge measurement for rating curve definition Unrecorded nonstationary changes in cross section from vegetation growth and sediment transport Inappropriate choice of rating curve, particularly in extrapolating beyond the range of available discharge observations Errors due to unmeasured and distributed flow effects
Internal state variables (soil moisture, water tables, etc.)	Point observation errors	Commensurability errors of predicted variables with respect to observed values arising from inappropriate theory and scale effects
Remote sensing data	Random error in correction of sensor values to fields of digital numbers (sensor drift, atmospheric corrections, etc.) Random error in converting digital numbers to hydrologically relevant variables	Inappropriate correction algorithms or assumptions about parameters Inappropriate conversion algorithms in obtaining hydrologically relevant variables

uncertainties may not be treated easily in probabilistic terms, but it will not usually be clear what other uncertainty framework could be used usefully to represent them. Epistemic errors are often treated in terms of probabilities *as if* they were aleatory, but the probabilities are much more likely to be incomplete, imprecise, or nonstationary in nature. Consequently, any representation will require subjectively chosen assumptions (actually, by definition, since if we had the knowledge to describe them properly, they would no longer be epistemically uncertain, see the discussion of Rougier and Beven [2012]). Note that some epistemic errors may lead to disinformation in learning about the hydrological response of a catchment, with consequences for model calibration, estimation, and evaluation. Examples are where rainfalls are recorded, and there is no observed discharge response in a humid catchment, or where apparent runoff coefficients are greater than 1 [e.g., Beven and Westerberg, 2011; Beven et al., 2011; Young, 2013].

[14] Thinking a little more deeply about uncertainty introduces still more complications. For instance, we might have a source of uncertainty that appears to be aleatory and therefore, appropriately represented in probabilistic terms, but it might not be clear as to whether we are using the correct probabilistic model (at base, a distribution or joint distribution, together with any consistent bias or correlation structure). Techniques have been developed (e.g., the meta-Gaussian transform, or copula methods) to try to convert apparently complex series of errors into simple distributional forms for which the theory is well developed (in particular, multivariate Gaussian forms) [e.g., Montanari and Koutsoyiannis, 2012]. But there might then be epistemic uncertainty about the choice of distributional form or transform (e.g., the choice of a particular distribution or copula). We do not know a priori what the correct form is for a given type of error arising in a particular type of application, although there are often forms accepted as “standard practice” (e.g., the distributions commonly used in flood frequency analyses). Thus, deeper thought will often reveal that sources of uncertainty are more epistemic than aleatory (see Table 2), even if the choice of treating them *as if* aleatory may be justifiable on the basis of checking the associated assumptions about the structure of the errors in both calibration and forecasting or simulation.

[15] Note that ex-ante adaptive forecasting is one area where it is most appropriate to treat all sources of uncertainty *as if* they are aleatory in nature. Useful serial correlation in the residual error may well be captured by an aleatory model, such as the heteroscedastic autoregressive, moving average (ARMA) process used in Young [2013], thus enhancing the stochastic description of the data and leading to improvement in ex-ante forecasting ability. The adaptive nature of the forecasting will also allow some account to be taken of epistemic errors as they occur.

[16] To summarize, while referees should ensure that all modeling studies are associated with some form of uncertainty estimation, it has often been difficult in the past to assess how authors have differentiated between aleatory and epistemic errors. Perhaps it is sufficient at this stage to ensure that authors make a clear statement about how aleatory and epistemic errors have been treated, with a clear statement about the assumptions made, and their evaluation.

## 5. The Language of Model Evaluation

[17] Another vexed issue is the language used for model evaluation, confirmation, validation, and verification that has generated so much discussion in the literature. Refsgaard and Henriksen [2004] provide a useful discussion of this topic in relation to hydrological modeling. As pointed out by Oreskes et al. [1994], Rykiel [1996], and others, this is partly because there are multiple uses of these words that range from the colloquial to the technical. There are also quite different uses in different domains of the environmental sciences. In meteorological modeling it is standard practice to present information on different model verification statistics in the evaluation of weather forecasts [e.g., Pappenberger et al., 2008; Cloke and Pappenberger, 2008]. In other domains, model verification is normally only used for the evidence that a computer code is consistent with the definition of the model concepts that it purports to represent, without any requirement that the resulting outputs will be consistent with observations in a particular application.

[18] The loose usage of the language of model evaluation was first criticized by Oreskes et al. [1994], following the a posteriori evaluation of hydrogeological simulation models published in Konikow and Bredehoeft [1992] and Anderson and Wössner [1992]. They pointed out that the word verification comes from a Latin root (*verus*) meaning truth. But models are not truthful representations of reality, they are necessarily approximations, so verification is not an appropriate term to use in respect of models. Rykiel [1996], following Oreskes et al. suggested that the only possibility of having a logical truth in environmental modeling would be in the implementation of the model concepts as computer code. He suggested that verification should therefore be restricted to this context, and not used when referring to how a model might represent the actual response of the system of interest. As such, verification should be qualified as conditional verification since any such exercise for a nonlinear model will be conditional on the range of tests to which the model has been submitted. We know from experience that even widely distributed model codes can crash with certain combinations of parameter values and boundary conditions.

[19] Model validation is quite complex in its origins. The word validation is derived from the Latin *validus* (meaning strong or effective) and only came to signify being supported by evidence or authority in the seventeenth century. It is now widely used in environmental modeling to mean the successful testing of model output variables against some criteria of performance [see, for example, Klemesš, 1986; Refsgaard and Knudsen, 1996; Refsgaard and Henriksen, 2004]. There have, for example, been many papers that have declared success in validating the SWAT model [e.g., Gassman et al., 2010]. But as Mitro [2001] points out, the statement that a model has been validated needs some qualification as to how that validation has been carried out. A decision maker might be faced by predictions from a number of different models, all of which have been validated in some sense. The degree of belief in each of the models should then depend on an assessment of the validation process, so the relevant information needs to be communicated. Young [2001] introduces the term “conditional validation,” where the model is validated on data other than that used in its estimation (sometimes termed “predictive

validation”). A model that is conditionally valid is one that has not yet been falsified by tests against observational data (see later for a discussion of predictability and falsification). Models that are considered to be conditionally valid in this sense clearly have immediate practical utility in simulating within the range of the calibration and evaluation data, while allowing for their updating in the light of future research and development or change in catchment characteristics.

[20] In this sense, conditional validation represents good practice, even in its weakest form when a model that has been calibrated for one data set is evaluated against a different data set of the same variables, at the same site as used in the calibration (the classical “split-record test”). Stronger forms of conditional validation are also possible by using additional variables or sites that have not been used in calibration [Klemeš, 1986; Refsgaard, 1997; Refsgaard and Knudsen, 1996]. In general, validation will be both criterion and data set dependent [see, for example, Choi and Beven, 2007]. Thus, predictive validation can only ever be conditional [Young, 2001], dependent on the conditions of the evaluation process. Mitro [2001] therefore suggests that the term validation should also be avoided by simply stating explicitly the conditions and results of the evaluations carried out. Referees should consider this to be good practice. If the word validation is used, then, as pointed out previously, it should always be associated with the word conditional, and the conditions of the model evaluation should be stated clearly (this is consistent with the suggestions of Refsgaard and Henriksen [2004]). The implication is that further testing in the future might reveal model (or data) deficiencies that up to now have not been apparent. Suggestions for the clear use of terms in this context are suggested in Table 3.

## 6. Conditional Validation, Falsification, and Fit-For-Purpose

[21] A suggestion that a model has satisfied some conditions in evaluation suggests that the model is (to some level of belief) fit-for-purpose. But it is important that the conditional nature of the validation should be made explicit so that it can be evaluated by potential users of the model output. The limitations of conditional validation have been discussed by Kumar [2011] in relation to different types of

model use. He differentiates the use of a model to reduce the uncertainty in estimating a future trajectory of the catchment response, and the use of a model to explore potentially new trajectories for the catchment response as a result of changing boundary conditions or the complexity of evolving catchment dynamics. The first can be subject to conditional validation (for example, in terms of reduction of uncertainty relative to the “climatology” of the historical responses) but does not guarantee successful simulation or forecasts in future if there are changes in the system. The examples of postaudit analysis of groundwater models cited earlier are good examples of this. The second cannot be subject to validation until the future evolves along a particular new trajectory. Since this might involve other modes of response and process interactions to the historical response, a different model might be required that cannot easily be tested in the ways outlined above. Conditional validation in the sense of Young [2001] allows for these possibilities by stressing the need for continued recursive updating of model parameters and data assimilation under the assumption that the parameters might change over time. Statistically significant changes in parameters will then reveal changing dynamic behavior caused by changes in the system and the need to re-evaluate the model.

[22] It is quite possible, of course, that a model structure or parameter set may fail such a continuing evaluation. Model falsification or rejection is important because in progressing the science, we normally learn more from falsification than from (conditional) model validation. Falsification implies some improvement is required, either in the data that is being used to drive and evaluate the model or in the model structure itself. However, testing models as hypotheses in this manner is difficult in hydrology and other areas of environmental science because of the inherent epistemic nature of many sources of uncertainty, as discussed earlier [see Beven et al., 2012; Clark et al., 2011, 2012]. Since all models are approximations, a close enough look at the model outputs will always reveal some deficiencies. Thus, defining fit-for-purpose and testing models as hypotheses in some posterior analysis must also be conditional. In hydrology and as far as we can tell, in many other areas of environmental modeling, the conditional nature of model validation and what constitutes being fit-for-purpose have not been given

**Table 3.** Suggestions for the Use of Terms in Hydrological Model Evaluation

Term	Suggested Usage	Comments	Synonyms
Verification	Only for checks that a model computer code properly reproduces the equations it purports to solve. Will generally be only conditional verification dependent on the range of tests carried out	Sometimes used in context of comparing model outputs and observables (as in forecast verification); this should be avoided in favor of conditional validation or model evaluation. Conditions of model verification should be clearly stated	
Validation	Conditional evaluation that a model computer code reproduces some observed behavior of the system being studied to an acceptable degree	Should always be used in the form conditional validation with an explanation of the conditions of the model evaluation	Conditional model evaluation
Falsification	An assessment that a model computer code does not reproduce some observed behavior of the system to an acceptable degree	Should always allow for potential errors in observed behavior so as to avoid Type I error of rejecting a model as a result only of errors in the data	Model rejection To fail a hypothesis test
Fit-for-purpose	Conditional evaluation that a computer model code is suitable for use in a particular type of application	Should always state conditions of how fitness-for-purpose has been assessed	To have (conditional) predictability

sufficient consideration, despite the publication of various papers on this topic [see, e.g., *Beven*, 2001; *Young*, 2001; *Kirchner*, 2006; *Clark et al.*, 2011; *Beven et al.*, 2012]. Authors and referees should, however, at least recognize the nature of these issues in the presentation of model results and posterior analysis and ensure that the conditions of accepting a model as useful are clearly set out.

## 7. Conclusion

[23] In this opinion piece, we have sought to provide some guidance to authors and referees about the proper use of words in describing model structures, in the use of terms describing model outputs, in the description of different types of uncertainties, and in the use of language in model evaluation and falsification. We have (mostly) limited this opinion paper to comments about good practice in modeling semantics rather than good practice in modeling. However, we would hope that if referees insist on more rigor in the technical semantics, the resulting papers will be both clearer and result in better practice in model applications.

[24] **Acknowledgments.** Jens-Christian Refsgaard, David Huard, and a third anonymous referee are thanked for their comments, which helped improve the final version of this manuscript.

## References

- Anderson, M. P., and W. W. Woessner (1992), The role of the post-audit in model validation, *Adv. Water Resour.*, 15, 167–174.
- Beven, K. J. (2001), Dalton Medal Lecture: How far can we go in distributed hydrological modelling?, *Hydrol. Earth Syst. Sci.*, 5(1), 1–12.
- Beven, K. J. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36.
- Beven, K. J. (2008), On doing better hydrological science, *Hydrol. Processes*, 22, 3549–3553.
- Beven, K. J. (2009), Comment on “Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?” by Jasper A. Vrugt, Cajo J. F. ter Braak, Hoshin V. Gupta and Bruce A. Robinson, *Stochastic Environ. Res. Risk Assess.*, 23, 1059–1060, doi:10.1007/s00477-008-0283-x.
- Beven, K. J. (2012), Causal models as multiple working hypotheses about environmental processes, *C. R. Geosci. Acad. Sci. Paris*, 344, 77–88, doi:10.1016/j.crte.2012.01.005.
- Beven, K. J. (2013), “Here we have a system in which liquid water is moving; let’s just get at the physics of it” (Penman 1965), *Hydrol. Res.*, in press.
- Beven, K. J., and P. F. Germann (2013), Macropores and water flow in soils revisited, *Water Resour. Res.*, 49, 3071–3092, doi:10.1002/wrcr.20156.
- Beven, K. J., and P. J. Smith (2013), Concepts of information content and likelihood in parameter calibration for hydrological simulation models, *ASCE J. Hydrol. Eng.*, in press.
- Beven, K. J., and I. Westerberg (2011), On red herrings and real herrings: Disinformation and information in hydrological inference, *Hydrol. Processes*, 25, 1676–1680, doi:10.1002/hyp.7963.
- Beven, K. J., P. J. Smith, and J. Freer (2008), So just why would a modeller choose to be incoherent?, *J. Hydrol.*, 354, 15–32.
- Beven, K., P. J. Smith, and A. Wood (2011), On the colour and spin of epistemic error (and what we might do about it), *Hydrol. Earth Syst. Sci.*, 15, 3123–3133, doi:10.5194/hess-15-3123-2011.
- Beven, K., P. Smith, I. Westerberg, and J. Freer (2012), Comment on “Pursuing the method of multiple working hypotheses for hydrological modeling” by P. Clark et al., *Water Resour. Res.*, 48, W11801, doi:10.1029/2012WR012282.
- Choi, H. T., and K. J. Beven (2007) Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in distributed rainfall-runoff modelling within GLUE framework, *J. Hydrol.*, 332(3–4), 316–336.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, doi:10.1029/2010WR009827.
- Clark, M., D. Kavetski, and F. Fenicia (2012), Reply to comment by K. Beven et al. “Pursuing the method of multiple working hypotheses for hydrological modeling,” *Water Resour. Res.*, 48, W11802, doi:10.1029/2012WR012547.
- Clarke, R. T. (1973), A review of some mathematical models used in hydrology, with observations on their calibration and use, *J. Hydrol.*, 19, 1–20.
- Cloke, H. L., and F. Pappenberger (2008), Evaluating forecasts of extreme events for hydrological applications: An approach for screening unfamiliar performance measures, *Meteorol. Appl.*, 15, 181–190.
- Gassman, P. W., J. G. Arnold, R. Srinivasan, and M. Reyes (2010), The worldwide use of the SWAT model: Technological driver, networking impacts, and simulation trends, *Trans. ASABE*, 50(4), 1211–1250.
- Helton, J. C., and D. E. Burmaster (1996), Guest editorial: Treatment of aleatory and epistemic uncertainty in performance assessments for complex systems, *Reliab. Eng. Syst. Saf.*, 54, 91–94.
- 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, 707–712.
- Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362.
- Klemeš, V. (1986), Operational testing of hydrologic simulation models, *J. Hydrol.*, 31, 13–24.
- Konikow, L. F., and J. D. Bredehoeft (1992), Groundwater models cannot be validated, *Adv. Water Resour.*, 15, 75–83.
- Koutsoyiannis, D. (2010), HESS opinions “A random walk on water,” *Hydrol. Earth Syst. Sci.*, 14, 585–601.
- Kumar, P. (2011), Typology of hydrologic predictability, *Water Resour. Res.*, 47, W00H05, doi:10.1029/2010WR009769.
- McMillan, H., T. Krueger, and J. Freer (2012), Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality, *Hydrol. Processes*, 26(26), 4078–4111.
- Mitro, M. G. (2001), Ecological model testing: Verification, validation, or neither?, *Bull. Ecol. Soc. Am.*, 82, 235–237.
- Montanari, A. (2007), What do we mean by ‘uncertainty’? The need for a consistent wording about uncertainty assessment in hydrology, *Hydrol. Processes*, 21, 841–845.
- Montanari, A., and D. Koutsoyiannis (2012), A blueprint for process-based modeling of uncertain hydrological systems, *Water Resour. Res.*, 48, W09555, doi:10.1029/2011WR011412.
- Neuman, S. P. (1977), Theoretical derivation of Darcy’s law, *Acta Mech.*, 25, 153–170.
- Oreskes N., K. Schrader-Frechette, and K. Belitz (1994), Verification, validation and confirmation of numerical models in the earth sciences, *Science*, 263, 641–646.
- Pappenberger, F., K. Scipal, and R. Buizza (2008), Hydrological aspects of meteorological verification, *Atmos. Sci. Lett.*, 9, 43–52.
- Refsgaard, J. C. (1997), Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.*, 198, 69–97.
- Refsgaard, J. C., and H. J. Henriksen (2004), Modelling guidelines—Terminology and guiding principles, *Adv. Water Resour.*, 27, 71–82.
- Refsgaard, J. C., and J. Knudsen (1996), Operational validation and inter-comparison of different types of hydrological models, *Water Resour. Res.*, 32, 2189–2202.
- Refsgaard, J. C., J. P. Van der Sluijs, J. Brown, and P. Van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, 29(11), 1586–1597.
- Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg, and P. A. Vanrolleghem (2007), Uncertainty in the environmental modelling process—A framework and guidance, *Environ. Modell. Software*, 22, 1543–1556.
- Refsgaard, J.-C., K. Arnbjerg-Nielsen, M. Drews, K. Halsnæs, E. Jeppesen, H. Madsen, A. Markandya, J. E. Olesen, J. R. Porter, and J. H. Christensen (2013), The role of uncertainty in climate change adaptation strategies—A Danish water management example, *Mitigation Adaptation Strategies Global Change*, 18, 337–359.
- Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (2000), Conservation equations governing hillslope responses: Exploring the physical basis of water balance, *Water Resour. Res.*, 36, 1845–1863.
- Reggiani, P., M. Sivapalan, S. M. Hassanizadeh, and W. G. Gray (2001), Coupled equations for mass and momentum balance in a stream network: Theoretical derivation and computational experiments, *Proc. R. Soc. A.*, 457, 157.
- Rougier, J., and K. J. Beven (2012), Epistemic uncertainty, in *Risk and Uncertainty Assessment for Natural Hazards*, edited by J. Rougier, S. Sparks, and L. Hill, pp. 40–63, Cambridge Univ. Press, Cambridge, U. K.
- Rykiel E. J. (1996), Testing ecological models: The meaning of validation, *Ecol. Modell.*, 90, 229–244.

- Spiegelhalter, D. J., and H. Riesch (2011), Don't know, can't know: Embracing deeper uncertainties when analysing risks, *Philos. Trans. R. Soc. A*, 369(1956), 4730–4750.
- Van der Sluijs, J. P., M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey (2005), Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: The NUSAP system, *Risk Anal.*, 25(2), 481–492.
- Walker, W. E., P. Harremoës, J. Rotmans, J. P. Van der Sluijs, M. B. A. Van Asselt, P. Janssen, and M. P. Kraayer von Krauss (2003), Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support, *Integr. Assess.*, 4(1), 5–17.
- Warmink, J. J., J. A. E. B. Janssen, M. J. Booij, and M. S. Krol (2010) Identification and classification of uncertainties in the application of environmental models, *Environ. Modell. Software*, 25, 1518–1527.
- Young, P. C. (2001), Data-based mechanistic modelling and validation of rainfall-flow processes, in *Model Validation: Perspectives in Hydrological Science*, edited by M. G. Anderson and P. D. Bates, pp. 117–161, John Wiley, Chichester, U. K.
- Young, P. C. (2003), Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale, *Hydrol. Processes*, 17, 2195–2217.
- Young, P. C. (2011), *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*, Springer, Berlin.
- Young, P. C. (2013), Hypothetico-inductive data-based mechanistic modeling of hydrological systems, *Water Resour. Res.*, 49, 915–935, doi:10.1002/wrcr.20068.