Uncertainty in Predictions of Floods and Hydraulic Transport

Keith BEVEN
Environmental Science/Lancaster Environment Centre, Lancaster University, UK, and Geocentrum, Uppsala University, Sweden
e-mail: k.beven@lancs.ac.uk

Abstract
This paper provides a review of work within the Generalised Likelihood Uncertainty Estimation (GLUE) methodology on estimating uncertainties in predicting flood frequency, flood inundation, and hydraulic transport of solutes in rivers and soils. The issue of prediction uncertainty as an input decision making is also discussed. It is concluded that in real applications it is unlikely that a fully objective approach to uncertainty estimation is possible. It is therefore important that the assumptions made are stated explicitly so that they can be agreed or disputed with the users of the resulting predictions. It is also important that the modelling process be considered as a learning process of constraining uncertainty by adding new information.

1. Uncertainty about uncertainty in flood and transport predictions

There is currently significant debate about how to estimate the uncertainties associated with environmental predictions. This discussion has been prompted by the more widespread availability of computer power, especially Beowulf-type parallel systems of cheap PCs, that has allowed the application of Monte Carlo methods of different types to a wider range of environmental models. Clearly, there are still some limitations. Fine grid scale 2D and 3D hydrologic and hydraulic models with very large numbers of elements, still cannot easily be run in Monte Carlo experiments without access to very large scale resources, but we can probably expect computer power to continue to increase more quickly than changes in modelling concepts for the foreseeable future, so that the uncertainty analysis will become feasible for more and more model applications.

This raises some interesting questions: in particular, is it possible to agree on an uncertainty estimation methodology and how should prediction uncertainties be used in decision making? A full discussion of these questions is the subject of a forthcoming book (Beven 2008) and only a brief outline can be given here. Arguments for the routine application of uncertainty analysis can be found in Pappenberger and Beven (2006).
With respect to the first question, the answer is – as yet – no; though some suggestions can be given about different methods for use in different circumstances (see the Risk and Uncertainty Decision Tree Wiki pages at www.floodrisknet.org.uk/methods/Introduction). There are many people who believe that statistics is the only way of estimating uncertainties associated with model predictions (see O’Hagan and Oakley 2004, Mantovan and Todini 2006) but it is clear that in many applications of environmental models there are sources of uncertainty that are not statistical (aleatory) in nature (see Beven 2005, 2006a) and the use of formal statistical assumptions might lead to misleading results even in near-ideal cases (Beven et al. 2007). In non-ideal cases (i.e. nearly all real applications), non-statistical (epistemic) uncertainties may dominate. Examples of epistemic uncertainties are bias and nonstationarity in input errors, model structural errors and commensurability errors (where a variable or parameter in a model is different to an equivalent quantity that can be measured in the field, see Beven 1989, 2002, 2006a, b, Freer et al. 1996, 2004).

It can be easily shown that in all real applications it is impossible to separate out different sources of aleatory and epistemic uncertainties unless very strong (and difficult to justify) assumptions are made (Beven 2006a). This leaves plenty of scope for uncertainty about how to estimate prediction uncertainties. In what follows we will consider only one methodology, the Generalised Likelihood Uncertainty Estimation (GLUE) first proposed by Beven and Binley (1992). This is a very flexible technique for model conditioning given some past observations of system responses (it includes both formal statistical and fuzzy methods as special cases) but one that has been criticised as being based on too many subjective assumptions. It is based on the equifinality thesis: the concept that in real applications there may be many different model structures and sets of parameter values for each model structure that produce acceptable or behavioural predictions of the system of interest (Beven 1993, 2006a). Equifinality can be visualised in plots of some evaluation measure, such as residual variance or Nash-Sutcliffe efficiency, against single parameter values (e.g. Figs. 1 and 2). Such plots represent projections of points on a response surface in the model space onto the single parameter axes. As such they cannot reveal all the complex parameter interactions within a model structure that lead to behavioural or non-behavioural performance; they can reveal that very often the best model performances are found across a wide range of individual parameter values.

The GLUE methodology is essentially very simple in concept. A large number of runs of a model are made using randomly generated sets of effective parameter values (chosen from defined prior distributions if that information is available, otherwise from uniform prior distributions). The outputs from each model run are compared with the observational data, taking account of observational error where appropriate (see, e.g. Beven 2006a, Freer et al. 2004). Those models providing acceptable or behavioural results are retained for use in prediction; those that do not are rejected. Each behavioural model is assigned a likelihood weight dependent on performance (zero for non-behavioural models) that is used to weight the predictions of that model in a formal cumulative distribution of predictions over the whole set of behavioural models. Different model structures as well as different parameter sets can be included in this process if the same methods of evaluation and likelihood assignment can be applied.
Different types of evaluation (analogous to multi-criteria calibration) are easily combined in this methodology, using either Bayes equation or some other chosen combination method (e.g. fuzzy union/intersection). Demonstration software for GLUE can be found at http://www.es.lancs.ac.uk/hfdg/freeware/hfdg_freeware_glue.htm.

The use of GLUE will be demonstrated in applications to flood frequency estimation, flood inundation predictions for risk mapping, and hydraulic transport predictions. A final section considers the use of prediction uncertainties in decision making, with a focus on inundation predictions.

![Fig. 1](image)

**Fig. 1.** Dotty plots of a coefficient of determination in a pesticide transport model fitted to observed atrazine concentrations in a large undisturbed soil column. Each dot represents one run of the model with different randomly chosen parameter values. The four parameters are: (top) an effective pore water velocity, a dispersion coefficient (the ranges for which were previously determined by fitting bromide concentration data assumed to be a near conservative tracer on the same column), (bottom) a retardation coefficient and a degradation coefficient. The best models of the realisations simulated by uniform sampling in the model space are at the top of each plot. The error bars shown on the bottom plots are ±2 standard errors on the parameters estimated by nonlinear regression (after Zhang et al. 2006).

### 2. Uncertainty in flood frequency estimation

Flood frequency estimation is often considered as a statistical problem. Given a sequence of historical flood events, a statistical distribution is fitted using either annual maximum or peaks over threshold data so that an estimate of the flood peak for any
given return period can be estimated. If done properly, this can also yield a statistical estimation of the uncertainty in the predicted peak discharges, the uncertainty that increases rapidly for return periods longer than the length of the historical series (and that therefore might be important in decisions based on 100 year return period events or longer).

Fig. 2. Dotty plots for channel and flood plain roughness coefficients in an application of the 1D HEC-RAS model for the flood of 1997 on the River Morava, Czech Republic. Each dot represents one run of the model with randomly chosen roughness coefficients, assumed constant for the whole reach. The combined likelihood reflects model performance in reproducing both observed inundation extent and the downstream hydrograph (after Pappenberger et al. 2005a).

This is a nice example where different sources of uncertainty and statistical assumptions might affect the result significantly. There is little agreement in the literature on what distribution should be chosen, and different distributions fitting the data more or less equally well might result in quite different predictions at higher return periods (this is an epistemic uncertainty analogous to model structural error). It is (usually) necessary to assume that the historical flood data are correct, even though it is known that out-of-bank flows are notoriously difficult to estimate accurately (again
a form of epistemic uncertainty). It is necessary to assume that the historical data are samples from a stationary distribution even where a catchment is known to have undergone significant land use change and been subject to longer time scale climate variability (again a form of epistemic uncertainty). Fitting a statistical distribution also necessarily deals with any hydrological and hydraulic process changes in different flood events (e.g., extent of surface and subsurface runoff contributing areas, transition to overbank flow) implicitly (again a form of epistemic uncertainty – though there have been rare examples of trying to fit mixed distributions for different runoff generation mechanisms, where the distributional assumptions then apply to each mechanism). Finally, it is not often realised that the parameter values for the fitted distribution (and consequent predictions and uncertainties) depend strongly on very specific assumptions about the statistical nature of the residuals that may or may not be valid – they are rarely checked.

Fig. 3. Flood frequency predictions for the Dolni Kralovice sub-catchment of the Zelivka River catchment, Czech Republic, using continuous 10000 year Topmodel simulations driven by a stochastic rainfall model using behavioural parameter sets after conditioning on observed flood peaks, flow duration curves and maximum snow water equivalents, combined within GLUE using a fuzzy rules method. Grey lines represent frequencies predicted by different parameter sets, dashed lines the 5 and 95% likelihood weighted prediction bounds derived from these simulations, circles represent frequencies estimated from observed annual maxima, dotted lines represent statistical estimates based on observed annual maxima assuming a Wakeby distribution (after Blazkova and Beven 2004).
This is clearly not only a statistical problem – even if we have been happy to use statistical fitting for convenience in the past – there are too many epistemic uncertainties. There is an alternative approach which can reflect the nonlinearities in the hydrological and hydraulic responses more directly and, as a result, might be more useful in predicting the effects of future change. This is the continuous simulation rainfall-runoff modelling approach, first used by Beven (1986, 1987) and more recently by Cameron et al. 1999, 2000, Lamb 1999, Blazkova et al. 2002, 2004, Lamb and Kay 2004, Cameron 2006). The papers by Cameron and Blazkova have applied this approach within the GLUE methodology (Figs. 3 and 4).

Fig. 4. River Wye catchment, Wales: Cumulative distributions of the 100 year return period flood peak estimated within the GLUE methodology using 1000 year continuous simulations with different behavioural parameter sets in the rainfall-runoff model Topmodel driven by a stochastic rainfall model for different climate change scenarios. Scenario Z (current conditions); Scenario A1 (2020s); Scenario B1 (2050s); Scenario C1 (2080s) (after Cameron et al. 2000).

3. Uncertainty in flood inundation predictions
Flood inundation predictions are required for a variety of purposes including flood risk mapping and real-time forecasting. Many different 1D and 2D models are available for
making these predictions which are often presented without any attempt to assess the
associated uncertainties. There are again, however, a number of different sources of
epistemic (non-statistical) uncertainties in real applications of such models. These in-
clude the uncertainty in the estimation of the upstream hydrograph, the estimation of
effective roughness coefficients for different sections of channel and flood plain
(which might be quite different from estimates derived at single points on the chan-
nel), the representation of the flood plain geometry, the effects of infrastructure on the
flood plain, and the implementation of the numerical algorithm (where different algo-
rithms generally involve more or less numerical dispersion). In real-time predictions
there might also be issues of embankment failures, blockages of culverts and bridges,
predictions of wind and surge effects in tidal situations, etc. that might also have a
significant effect on flood levels. Again, it is difficult to allow that these types of un-
certainties are easily handled by a purely statistical approach. 1D and 2D flood inun-
dation models have been applied in the GLUE methodology by Romanowicz et al.
Pappenberger et al. 2004, 2005a, b, 2006a, b (see Figs. 5 and 6).

Fig. 5. The 5 and 95% Inundation Quantiles for an example reach of the River Morava, Czech
Republic, determined within the GLUE methodology using the HEC-RAS model after
conditioning on both observed inundation and downstream hydrograph data (see also Fig. 2)
(after Pappenberger et al. 2005a).
Fig. 6. Flood hazard map for part of the flood plain of the Alzette River, Luxembourg, conditioned on inundation extent derived from EnviSat ASAR images for a flood event in January 2003. 5% and 95% quantiles determined within the GLUE methodology using the 1D HEC-RAS model are shown (after Pappenberger et al. 2006a).
The experience of using such distributed hydraulic models within GLUE has been interesting. In particular, it rapidly becomes clear that equifinality is an issue in the application of such models when used with global or individual reach scale evaluation measures (e.g. Fig. 2 above). It also becomes clear that there are some reaches or cross-sections where it is very difficult for any of the models tried to reproduce the historical flood data. This might be a problem of the observational data itself (in one case in the application of Pappenberger et al. 2005a, this was obvious as the recorded water levels on the two banks of the large Morava river were 10 m different at one cross-section but in many cases this might not be so obvious). It might also be a result of any of the other sources of epistemic uncertainty noted above. In such cases, all the models tried could be rejected as non-behavioural. This could provoke a review of the modelling strategy and data – it is perhaps more likely, however, to result in the neglect of extreme errors as “outliers” or the use of global evaluation measures where the effects of local failures are not so obvious (e.g. Romanowicz et al. 1998, Romanowicz and Beven 2003, Bates et al. 2004, Hunter et al. 2005). Rejecting all the models is not a good result in presenting a report to a decision maker, of course. One strategy for retrieving such a situation is discussed in Section 5 below.

4. Uncertainty in hydraulic transport predictions

Similar issues arise in distributed hydraulic solute transport predictions. Such predictions are often based on an implementation of the advection-dispersion equation (ADE), either assuming a steady discharge in the channel or linked to the velocities predicted by a hydraulic flow model. The ADE can be justified theoretically on the basis of shear dispersion due to a logarithmic vertical velocity profile, once a solute is well mixed with the flow (e.g. Rutherford 1994). For steady flow conditions it predicts a concentration distribution for an impulse input that is Gaussian in space at any particular point in time, and slightly skewed in time at any particular cross-section downstream of the mixing length. These characteristics are often quite different from observations of real tracer concentration curves which very often are skewed in both time and space and have very long tails. It seems that in real rivers shear mixing is often dominated by “dead zone” mixing. The result is that a simple transfer function model might provide much more accurate predictions than the ADE, which simply has the wrong process assumptions (see, for example, the review of Young and Wallis 1993). The ADE can be modified to include dead zone effects (e.g. Bencala and Walters 1983) at the expense of adding additional parameters that will need to be fitted in the same way as the roughness coefficients in flood inundation models above, and which will be subject to similar equifinality. Dispersion model calibration within the GLUE framework has been considered by Hankin and Beven 1998a, b, Hankin et al. 2001, 2002, Kettle and Beven 2002, Kettle et al. 2002 (e.g. Fig. 7), and for the case of solute transport in soils by Zhang et al. 2005 (see Figs. 1 and 8).

5. Uncertain predictions as an input to decision making

There are two main reasons for using models in hydrology and hydraulics. The first is to show that we understand how a system is working (although even if a model does
Fig. 7. 2D predictions of tracer transport in a short reach of the River Severn at Leighton, UK, based on velocity fields produced by Telemac-2D. A. Concentration patterns at time 600 s after end of tracer input for 8 different runs of the model using randomly chosen parameter values. B. Pattern of uncertainty in concentration predictions, relative to mean concentration field after fuzzy conditioning on point tracer concentration observations (after Hankin et al. 2001, copyright John Wiley and Sons Limited. Reproduced with permission).
Fig. 8. A comparison of prediction bounds from GLUE and CXTFIT (which uses a nonlinear regression method) after fitting parameters to atrazine pesticide breakthrough curves from four large undisturbed soil columns (see also Fig. 1). The model is the same in both cases. Dots are observed data (after Zhang et al. 2006).

give good predictions of the available observations it might still not be doing so for the right reason Beven 2001, Kirchner 2004). The second is to provide predictions to inform a decision. But if, as argued above, predictions of hydrological and hydraulic models are intrinsically uncertain, how should estimates of that uncertainty be presented to decision makers and used in the decision making process, especially when there are many ways of estimating the uncertainty? This has caused some recent debate in the hydrological literature (Beven 2006, Hall et al. 2007, Mantovan et al. 2007). Uncertainty appears at first sight to making the decision making more difficult, but this is not necessarily the case. Decision makers always make decisions under uncertainty, whether they do so formally or informally. Most will already have a healthy scepticism about any predictions provided to them by the modeller, whether the predictions are presented as a single deterministic outcome or as an uncertain (probabilistic or possibilistic) range.

In fact, what the decision maker is really interested in is not the uncertainty of a prediction but the risk of a potential outcome and its potential impact or consequences for the decision. There is no doubt that taking proper account of such risks can affect the decision made (see for example Todini 2004). This is therefore a reason why uncertainty estimation should be part of any and every modelling exercise. The debate arises because the risk, in formal risk-based decision making theory, is normally defined as the product of probability * consequences (very often economic conse-
quences). This would appear to be a major argument for the use of probabilistic assessments of uncertainties. Thinking more deeply, however, if the assumptions required for such a formal probabilistic assessment cannot easily be justified in the face of the type of epistemic uncertainties described above, then perhaps other approaches might be useful (see for example the evidential reasoning approach of Wang et al. 2006, and the Info-Gap approach of Ben-Haim 2006). Use of these alternative decision making techniques is described in more detail in those references and Beven (2008).

6. Uncertain futures

Many decisions are, of course, concerned with the impacts of future change. Particular current issues are the impacts of land management practices and future climate change. Assessing such changes necessarily depends on assumptions both about the changes to boundary conditions and changes in model structure or effective values of parameters. The type of conditioning against data, such as that of Figs. 1 and 2, can only be carried out for current (or historical) conditions. There is an implication, however, that if the complex interactions between parameters and boundary conditions that lead to behavioural models for current conditions, then similarly the complex interactions will be involved in making predictions of future conditions. Such interactions are local in the model space, not easily described by either a single point or global covariance matrices.

In some cases where only the boundary conditions are assumed to change (e.g., in assessing the impacts of climate change on flood frequency in Cameron et al. 2000, Cameron 2006), then the behavioural parameter sets for current conditions can be used in predicting future impacts. Where it is expected that future change will result in changed parameter values, then it might not be possible to change single parameter values independently of others to form new “behavioural” parameter sets. What is needed is to “drift” the (complex) cloud of behavioural parameters sets through the model space to where they might best represent the new conditions. There does not as yet seem to be an easy way of doing this, but the task can be set up as a learning process. We know the starting point (the best estimate of a set of behavioural models under current conditions). As time evolves and more data is gathered we can start to study whether any resultant drift is apparent in the evolving parameter sets.

A similar approach can be taken to the ungauged site problem. Initial estimates of parameter sets can be conditioned on either quantitative or qualitative observations to gradually refine the representations of the site: in particular, parameter sets that are clearly inconsistent with the observations can be rejected (this might sometimes be all the models tried, e.g. Choi and Beven 2007). This type of learning process will become increasingly necessary as models of everywhere are implemented (Beven 2007).

This brief review of uncertainty estimation for flood inundation and transport calculations based on the GLUE methodology can only serve as an illustrated introduction to a complex subject area where the answers you get are dependent on the assumptions you make. Since it is difficult to be at all sure about the real nature of different sources of uncertainty in real applications then many different sets of assump-
lations could be argued for. It is unlikely in real applications that a fully objective approach to uncertainty estimation is possible. It is therefore important that the assumptions made are stated explicitly so that they can be agreed or disputed with the users of the resulting predictions. In the present state of the science, this requirement of making assumptions quite open to encourage a thoughtful approach to uncertainty estimation is probably more important than the differences between different methods. The topic will be discussed in much more detail in Beven (2008).

References


Cameron, D., 2006, An application of the UKCIP02 climate change scenarios to flood estimation by continuous simulation for a gauged catchment in the northeast of Scotland, UK (with uncertainty), *J. Hydrol.* **328**, 212-226.


Pappenberger, F., P. Matgen, K.J. Beven, J.-B. Henry, L. Pfister and P. de Fraipont, 2006a, Influence of uncertain boundary conditions and model structure on flood


