

## Uncertainty analysis of the rainfall runoff model LisFlood within the Generalized Likelihood Uncertainty Estimation (GLUE)

FLORIAN PAPPENBERGER and KEITH BEVEN, *Lancaster University, Lancaster, GB*

AD DE ROO, JUTTA THIELEN and BEN GOUWELEEUW, *European Commission, Joint Research Center, Ispra, Italy*

### ABSTRACT

The uncertainty of the GIS based rainfall runoff model LisFlood has been investigated within the Generalized Likelihood Uncertainty Estimation (GLUE) framework. Multipliers for the saturated and unsaturated hydraulic conductivity, the porosity of the upper and lower soil layer, channel and overland flow roughness and the maximum percolation from upper to lower storages have been sampled within a Monte Carlo analysis from a uniform random distribution. With each parameter set the model has been computed with input for the 1995 flood event of the river Meuse situated in France, Belgium and The Netherlands. Eight gauging stations have been used for model evaluation by the Multicomponent Mapping ( $M^x$ ) method. All parameters demonstrate equifinality and no parameter set could be classified as behavioural for all the evaluation datasets. However, the results of the prediction of uncertainty percentiles on the flow are very satisfactory and encouraging. The model did further show the capability to predict the uncertainty for estimating the exceedence of threshold levels, which can be used in flood warning decision making and river basin management.

*Keywords:* LisFlood; rainfall runoff; GLUE; uncertainty; Meuse; flood.

### Introduction

According to the definitions of Abbott *et al.* [1,2], hydrological modelling has now reached the 5th model generation which is characterised by highly complex physically based models linked to Geographical Information Systems and Graphical User Interfaces. This phase has been driven by an incredible increase in computer power, the increased availability of spatial data of different types and the need for integrated river basin management. However, the linkage is still incomplete due to the missing 4th dimension (time) in most of the major commercial GIS software packages. Therefore, the development of a dynamically modelling language within a GIS framework like PCRaster [3] can be seen as a further important step within this model generation.

The development of such a modelling language allows complex models, like the rainfall runoff model LisFlood, to be implemented and facilitates the use of the increasing wealth of spatial data sets. Such spatially distributed models also have the potential to help answer questions from policy makers about the impact of spatial changes (e.g. land use change) or react to the increasing demand for information by the public for more accurate spatial predictions. This has become an especially important topic after the widespread floods in 1996, 2000 and 2002 that had severe impacts all over Europe. However, the information given is still very often deterministic and ignores all uncertainties. This is unrealistic. A proper assessment

of uncertainty is essential to establish trust in model predictions by stakeholders of different types [4–9] and to inform risk based decision making [10–13].

This is particularly the case for the LisFlood model which is being implemented as part of a prototype for a European Flood Forecasting System [14,15] and is used for actual predictions of pre-warning levels for national river authorities. In this paper the uncertainties of this model are evaluated within the Generalized Likelihood Uncertainty Estimation (GLUE) method of Beven and Binley [16] on the data of the 1995 flood event on the river Meuse.

### Description of the model LisFlood and Model region

LisFlood is a distributed, raster based model, which is embedded in a dynamic modelling language (PCRaster) especially designed to describe distributed modelling problems [15,17–20]. The model is driven by rainfall and temperature time series, which are interpolated over the whole domain by Thiessen polygons. Precipitation, interception, snowmelt, evapotranspiration, infiltration, percolation, groundwater flow, lateral flow and surface runoff are computed for each cell. Table 1 and Figure 1 give a brief overview of the key processes included in the model.

In this paper the model has been calibrated and evaluated with 1995 flood data on the river Meuse catchment situated in France, Belgium and The Netherlands (Figure 2). In the

Table 1 Description of key processes in the Rainfall Runoff model LisFlood.

Process	Method/Additional Information	Reference
Interception of rainfall	Takes account of seasonal changes in the Leaf Area Index	[44] For Forest: [45]
Evapotranspiration	Penman-Monteith method	[46,47] For Forest: [45]
Snowmelt	A degree-day method	[48]
Infiltration*	Smith-Parlange equation	[49]
Soil freezing	A degree-day method	[50]
Vertical transport of water	Richard's Equation	
Soil water retention and conductivity curves	van Genuchten's Equation	[51]
Water retention and conductivity curves from soil texture	Pedotransfer-functions from the HYPRES project	[52]
Percolation to the groundwater	Darcy Equation	[53]

\*The capillary drive value is based on topsoil texture. Saturated hydraulic conductivity values are based on topsoil texture and land use. A correction factor is applied to account for land use influences on infiltration.

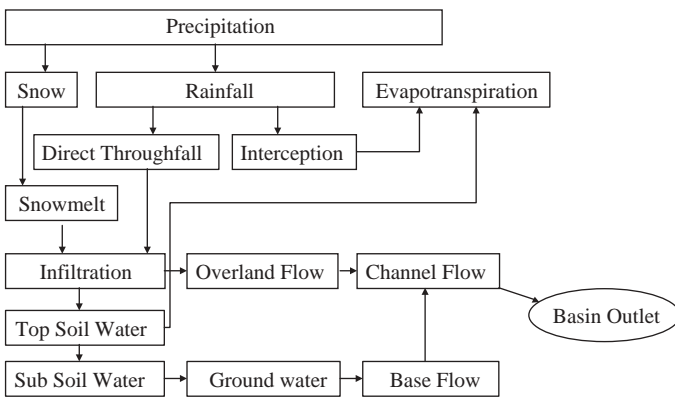


Figure 1 Flow Chart of LisFlood Model.

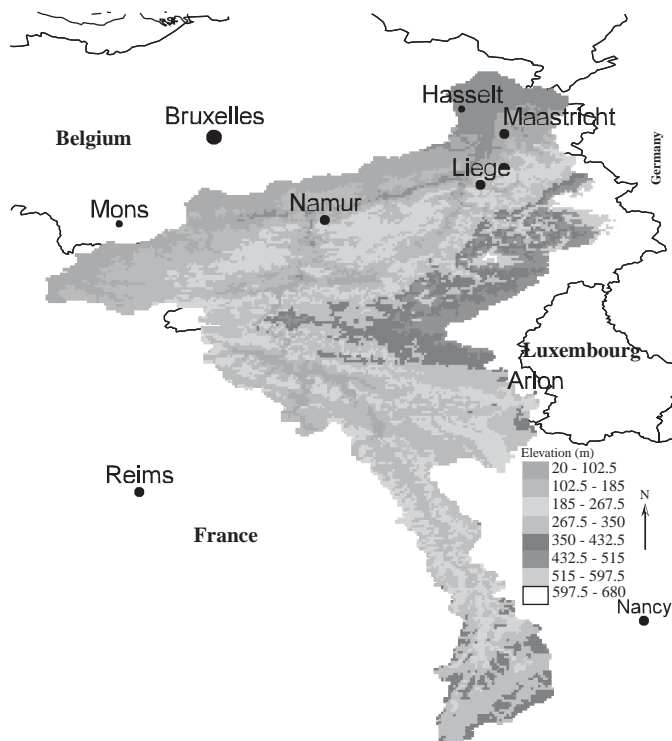


Figure 2 Topography and location of the Meuse catchment upstream of the Borgharen gauging station.

upper part of the catchment the number of tributaries is low, the river slope gradient is low and the river valley that may inundate at high water level is broad. This causes a slow

river response during high precipitation events. During periods of low precipitation discharge rates are relatively high, as the water storage capacity in this part is high. In the middle part, Meuse Ardennes, the river slope is relatively steep. The number of tributaries is high, while the infiltration capacity of the hilly tributary catchments is low. Water is discharged quickly, resulting in high discharge peaks and low water levels during dry periods. Inundated surface area is relatively small in this part. The lower part of the catchment is not part of this analysis which has Borgharen as the downstream boundary [21].

A mixture of hourly and daily rainfall fields are used to drive the model on a  $1 \text{ km} \times 1 \text{ km}$  grid of 21, 128  $\text{km}^2$  total size. The model has been initialised with the water balance version computing on a daily time step for one year and the flood period has been simulated over 68 days at an hourly timestep [14]. The flood event has been split into a calibration period of 27 days and an evaluation period of 41 days (see Figures 6 and 7).

The soil properties are taken from the European Soil data base and the saturated hydraulic conductivity as well as the soil water content at saturation is based on the HYPRES database. Land use and soil cover are based on the Corine database while temporal changes in leaf area index are derived from AVHRR satellite images [15,19,20].

Figure 3 and Table 2 shows a flow chart of the eight gauging stations which are used for model evaluation. Additional sub-catchment discharge data sets would have been available, but have been ruled out because the small catchment size was not adequately represented by the  $1 \text{ km}^2$  grid size of the modelling domain. Further analysis of the volume of runoff generated in each area in relation to the main outflow at Borgharen during the evaluation period indicates that the greater part of the runoff is generated within area 2 whereas all other regions contribute lower percentages. This, of course, neglects the timing of the contributions which influences the position of the peak and has therefore an impact on the result of the uncertainty analysis presented.

### Generalized Likelihood Uncertainty Estimation (GLUE)

Empirical results of studies in various fields of environmental modelling (for a summary see [22]) show that many different

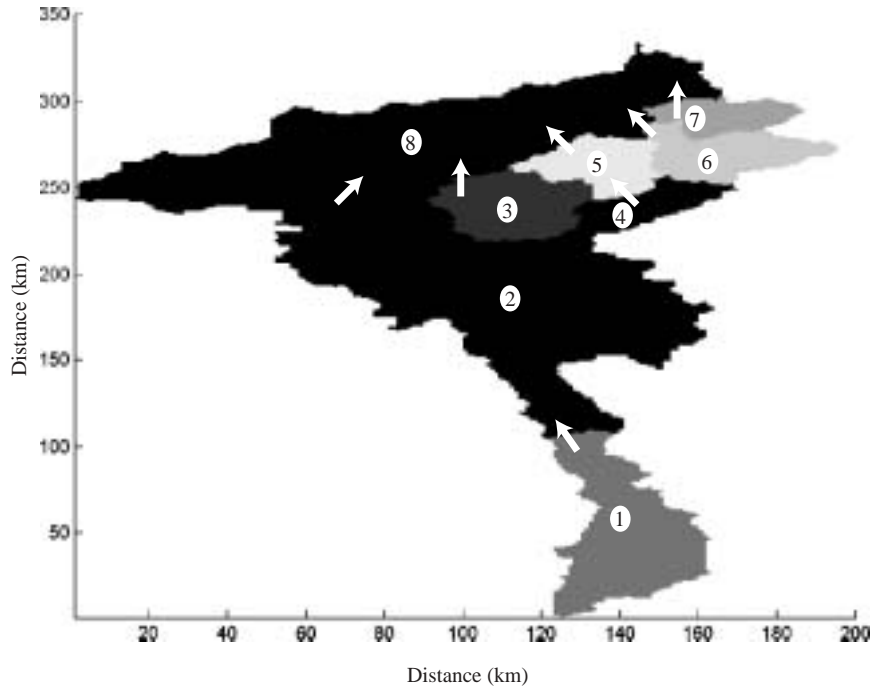


Figure 3 Flow diagram of the Meuse sub-catchment.

Table 2 Description of the gauging stations, which are available for model evaluation.

No of Upstream Area (Fig 3)	Station	Measured Interval
1	Saint Mihiel	Daily
2	Chooz	Daily
3	Lesse (Gendron)	Daily
4	Ourthe Nisramont	Daily
5	Ourthe (Tabreux)	Daily
6	Ambleve (Martinrive)	Daily
7	Vesdre (Chaufontaine)	Daily
8	Borgharen	Hourly

sets of parameters and model structures can fit evaluation data acceptably well and therefore, it may be impossible to decide between them as useful predictors of the system (the ‘equifinality’ concept of [23]). However, some realisations can be classified on the basis of the evaluation data as non-representative of the system or ‘non-behavioural’ and thus can be excluded from further analysis. GLUE has been criticised for the subjectivity of this distinction [24,25], but as Freer *et al.* [26] pointed out, subjectivity is unavoidable if no consistent error structure can be assumed. Within GLUE a likelihood weight is assigned to each of the behavioural parameter sets, according to how well that parameter set has performed during a calibration period.

The likelihood weights of the behavioural data set can be used to assess parameter sensitivity and to compute prediction limits on hydrographs [16,27]:

$$P(\hat{Z}_t < z) = \sum_{i=1}^{i=N} L[M(\theta_i) | \hat{Z}_{t,i} < z] \quad (1)$$

$N$ : Number of behavioural models

$P$ : Prediction quantile

$\theta_i$ : Model parameter set  $i$

$\hat{Z}_{t,i}$ : The value of the variable  $Z$  at time  $t$  simulated by the model  $M(\theta_i)$

$L$ : Likelihood measure

Furthermore it is possible to update the likelihood weights when new measurements arrive [22,27].

This method requires the modeller to specify a prior sampling distribution of the parameter space including any known interrelationships between the parameters. These specifications are not always straightforward, because the model parameters are effective parameters, which will be affected by all sources of error in the modelling system. The sampling strategy will inevitably influence the result and is governed also by the available computer power. It is a particular problem with parameters that can be allowed to vary spatially in a distributed model, since this will result in very high dimensional parameter spaces and the possibility of quite different spatial patterns giving similar predictions for good physical reasons.

In this particular application, to reduce the dimensionality of the problem it was decided to sample all parameters from uniform distributions and further, to fix the spatial relationships between the parameters and apply multipliers to ‘base’ values. This approach assumes that the basic spatial patterns have correct relative magnitudes. This can be questioned considering uncertainty and scaling issues [36], but would appear to be the most pragmatic way of reducing the dimensionality of the calibration problem for a distributed model of this type.

Table 3 shows the parameters that were sampled and the origin of the ‘base’ values (see also [28]). The sampling range has been adjusted according to an initial analysis and is also shown in this table. In this methodology the posterior likelihood value derived

Table 3 Parameters, which have been varied by a multiplier for the Monte Carlo analysis of the LisFlood model.

Parameter	Initial data source	Multiplier (uniform dist)
Saturated hydraulic conductivity (upper soil layer, Ks(S1))	HYPRES database	0.1–10
Saturated hydraulic conductivity (lower soil layer, Ks(S2))	HYPRES database	0.1–10
Porosity (upper soil layer, Theta(S1))	Soil cover/Corine data base	0.1–1.2
Porosity (lower soil layer, Theta(S2))	Soil cover/Corine data base	0.1–1.2
Saturated hydraulic conductivity (upper groundwater layer, Ks(G1))	HYPRES database	0.1–10
Saturated hydraulic conductivity (lower groundwater layer, Ks(G2))	HYPRES database	0.1–10
Manning Roughness Channel (NC)	Channel Order	0.1–5
Manning Roughness Overlandflow (NO)	Land use/Corine data base	0.1–10
Maximum Leaf Area Index (ML)	AVHRR satellite images	0.1–10
Maximum percolation from upper to lower response box (P)	Expert judgement	0.1–10

for each parameter set realisation will reflect all the interactions between the parameters in giving behavioural simulations. It is important to note that within the GLUE methodology, it is the parameter set that is evaluated rather than values of individual parameters.

The choice of an appropriate likelihood measure is therefore very important in the GLUE methodology. This will depend on the type of observational data available for a given application. It would be expected that the analysis is strongly influenced by the choice of the likelihood measure, though this is not necessarily the case as Freer *et al.* [29] have proven with different likelihoods in an application of the rainfall-runoff model TOP-MODEL. An interesting variation to this approach has been reported by Aronica *et al.* [30], Franks *et al.* [31], and Blazkova and Beven [32] who have successfully used fuzzy evaluation criteria.

The role and importance of the likelihood measure has been discussed in Pappenberger and Beven [33]. This manuscript also introduced a new methodology, Multicomponent Mapping, for the evaluation of hydrographs. The latter is used in this paper and maps the error distribution by subdividing the hydrograph ‘area’ into boxes. Unfortunately, as always in this type of real application, no information about the real distribution of the errors arising from different sources exists. Here, the expected final

prediction errors have been assumed to be heteroscedastic with larger error bands at high flows and lower error bounds at low flows. Similar assumptions have to be made about the error structure in time. For this exercise it has been arbitrarily assumed (due to the lack of data) that a timing error of  $\pm 5$  h for the hourly and  $\pm 1$  day for the daily observations still results in a model with perfect performance. With larger errors the performance of individual points decreases linearly with a zero performance value at  $\pm 10$  h ( $\pm 2$  days). The error structure in the magnitude was fixed in a way to reflect the heteroscedasticity and is based on experience of acceptable model errors (due to the lack of further information on input, rating curve, and model structural errors:

$$BoxSize \propto (0.1 * q_b / \text{mean}(o)) \quad (2)$$

$o$ : Vector of Observations

$q_b$ : Flow magnitude at centre of box

$BoxSize$ : Size of box in vertical direction

0.1: Error percentage assumed at mean flow

The reader is referred to Pappenberger and Beven [33] for a more detailed discussion of the Multicomponent Mapping approach.

These values for the allowable modelling error can be questioned and especially their large magnitude is of concern. However, Table 4 shows the values of the popular Nash-Sutcliffe criteria [34] when these specific error structures are assumed for the gauging station at Borgharen (hourly discharge). The first two columns show the change of the Nash-Sutcliffe criteria with an increase in flow magnitude (e.g. in the first row all measurements have been increased by 10%). Table 4 demonstrates the possibility of achieving a Nash-Sutcliffe of 0.65 with an increase of 50% – obviously mass conservation would be violated in this case. In the third and fourth row of this table the influence of a timing error is shown. Even when the original hydrograph is delayed by 20 hours, the Nash-Sutcliffe value still reaches 0.68, which is sometimes still regarded as a level for acceptable model results [26,35]. Such an experiment could be extended to other classical measures which have been proven not very discriminatory in some cases [36]. Furthermore, the change of the Nash-Sutcliffe values depends largely on the hydrograph used. However, this small exercise suggests that other ways of differentiating between behavioural and non-behavioural models within

Table 4 Nash-Sutcliffe values for various errors at the gauging station of Borgharen.

Magnitude Error (percentage increase at each measurement point)	Nash-Sutcliffe	Timing Error (steps)	Nash-Sutcliffe
+10	0.9745	1	0.9957
+20	0.9142	5	0.9947
+30	0.8354	10	0.8890
+40	0.7477	24	0.6093
+50	0.6566	48	-0.1783

the Generalized Likelihood Uncertainty Estimation framework might be valuable.

The value of the likelihood measure and the distinction between behavioural and non-behavioural models have been computed in two steps. Firstly, all simulated hydrographs, which fell outside of the specified bandwidth of acceptability have been considered as non-behavioural. Secondly, the mean value of all boxes which have been hit has been computed and forms the relative likelihood measure.

A Monte Carlo analysis naturally requires a large number of model runs and is therefore not necessarily suited for complex models. With the above described set-up one model run takes a minimum of 4 h on a 1.5 Ghz/512 RAM Linux machine and thus limits the number of runs significantly. For this river basin a total of 6,632 runs have been performed and will be analysed.

## Results and discussion

Figure 4 shows the results of the Monte Carlo runs for the Lesse (at Gendron) with the parameter multipliers on the  $x$  axis and the likelihood measure on the  $y$  axis. This picture is representative for all other gauging stations, which show very similar results. Each dot represents the evaluation of the simulation resulting from a particular set of randomly chosen parameter values and it can be seen that behavioural models range over the full extent of

the multipliers. Different patterns, however, might be expected for results from different catchments in different climatic regions (see [35] and uniqueness of place argument [37]).

Figure 5 enhances the one dimensional projections of the response surface shown in Figure 4 by plotting two parameters against each other and grey coding the  $M^x$  likelihood measure. Only a subset of all possible combinations is shown for demonstration. The mesh has been interpolated with a cell size of 0.05 in each dimension. All values have been averaged with each cell (various step sizes have been tried and none altered the findings significantly).

The plots show that there is apparently no strong two-dimensional covariation between parameters (only all the plots with Theta seem to be stretched, which is due to the restricted sampling range of this parameter. Furthermore, behavioural models can be achieved in various regions of the parameter space and again indicate equifinality. Unfortunately, it is very difficult to visualize and recognize patterns in higher dimensions (especially the full 11 dimensions of this application). It is not unexpected that models showing good fits to the observations should be scattered throughout the parameter space [26,37,38] when a high dimensional model (11 parameters in this case) is constrained by only limited amount of observational data.

As already mentioned, all the other gauging stations show very similar behaviour. However the total number of behavioural models is different for each station (Table 5). For example Ambleve

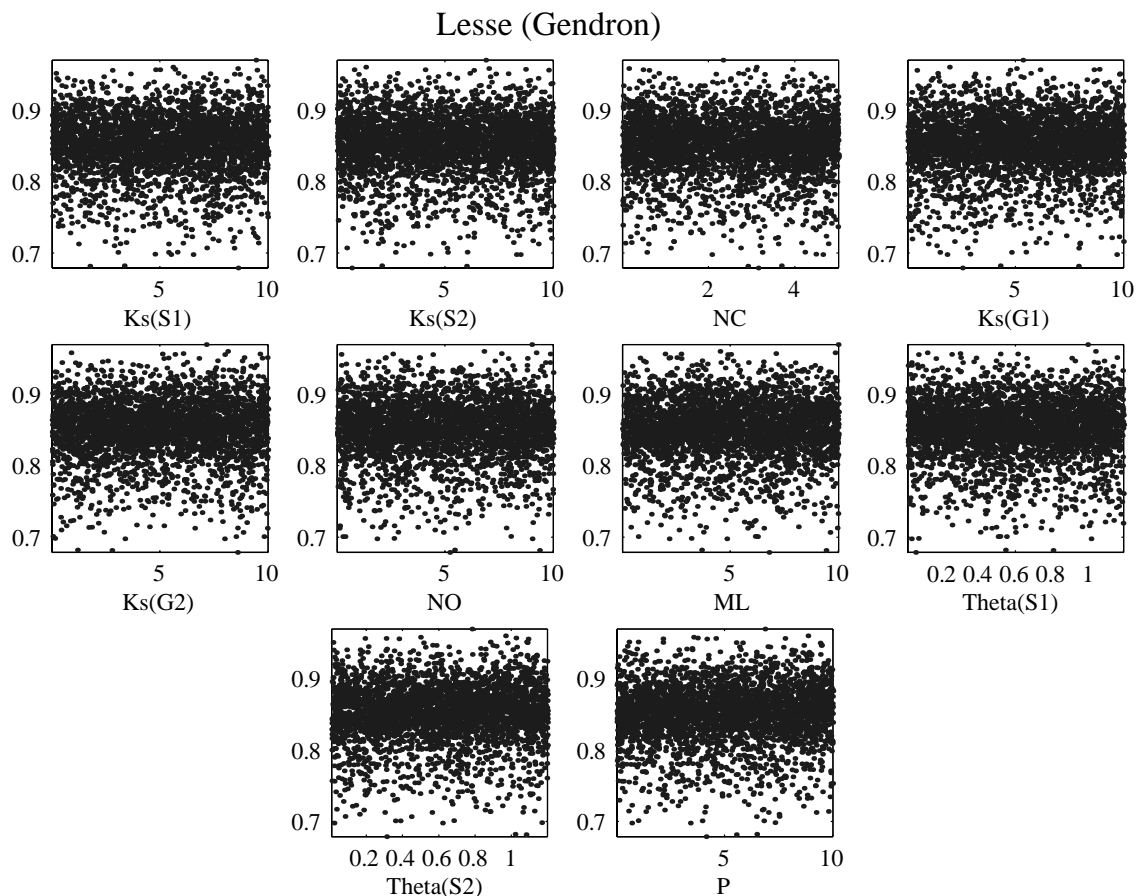


Figure 4 Dotty Plots of the 10 parameter values against the  $M^x$  model performance. Each dot represents a parameter set. For a description of the abbreviations, please see Table 3.

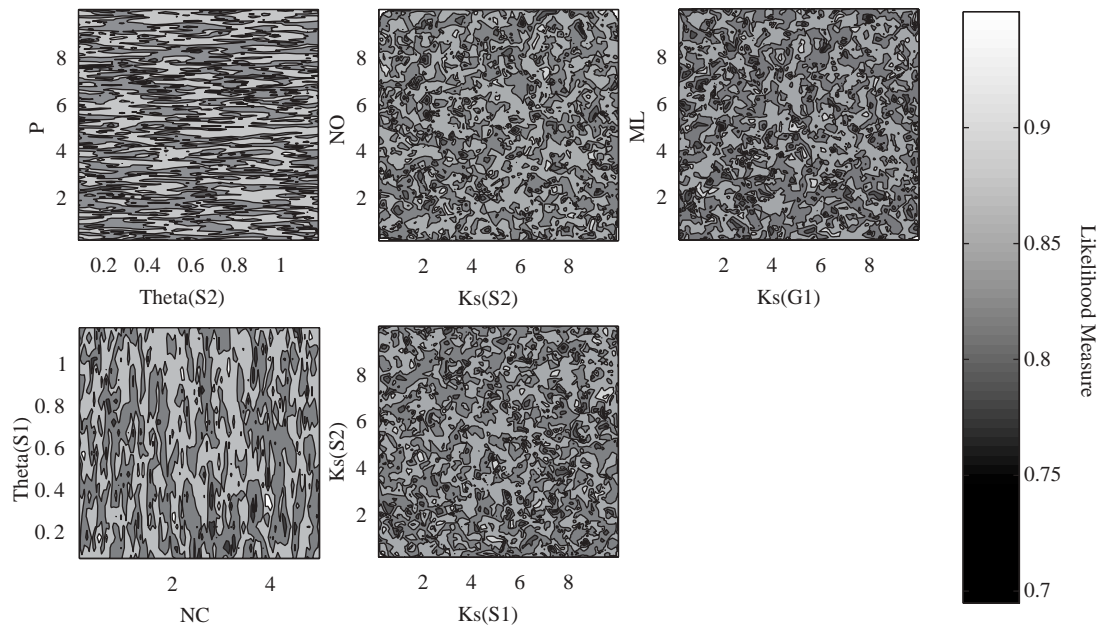


Figure 5 Two dimensional dotted plots. Model parameters are plotted on the y and x axis. The grey scale shows the likelihood measure. The model results have been grouped in classes of 0.05 in each dimension. An average value has been computed for each class.

Table 5 Number of behavioural models at each gauging station.

Station	Ourthe Nisramont	Vesdre (Chaufontaine)	Ourthe (Tabreux)	Ambleve (Martinrive)	Lesse (Gendron)	Saint Mihiel	Chooz	Borgharen
Nr. of behavioural parameter sets	287	566	2813	2729	3260	3713	685	35

Table 6 Comparison between catchments of number of behavioural models (No. of behavioural models in both catchments/No. of behavioural models in each catchment).

	Ourthe Nisramont	Vesdre (Chaufontaine)	Ourthe (Tabreux)	Ambleve (Martinrive)	Lesse (Gendron)	Saint Mihiel	Chooz	Borgharen
Ourthe Nisramont	100%	58%	97%	76%	93%	51%	10%	3%
Vesdre (Chaufontaine)	–	100%	62%	61%	60%	41%	11%	11%
Ourthe (Tabreux)	–	–	100%	61%	91%	75%	68%	20%
Ambleve (Martinrive)	–	–	–	100%	69%	71%	67%	31%
Lesse (Gendron)	–	–	–	–	100%	75%	68%	46%
Saint Mihiel	–	–	–	–	–	100%	79%	63%
Chooz	–	–	–	–	–	–	100%	<1%
Borgharen	–	–	–	–	–	–	–	100%

Martinrive (Figure 3, Catchment Nr. 6) has 2729 behavioural models compared to the nearby sub-catchment Ourthe Nisramont (Figure 3, Catchment Nr. 4) with 287 behavioural simulations. Both catchments share very similar characteristics in the inputs (e.g. geology, soil type, landuse and rainfall) and the main difference is in the catchment size with Ambleve Martinrive being about 1/3 larger. A closer examination of the results shows that the majority of simulations overpredict the flow at Ambleve Martinrive and thus this explains the worse performance. Therefore, either the measurement has to be considered in error or the 1 km grid square LisFlood representation of the catchment, in conjunction with the input maps, neglects certain underlying processes.

This problem is similar at the main outflow at Borgharen and is induced by the bad model performance at Chooz (Figure 3,

Catchment Nr. 2) from which 45% of the discharge volume is contributed. Moreover the overlap of behavioural parameter sets between Chooz and Borgharen is minimal.

Table 6 compares the sub-catchments to each other by dividing the number of common behavioural parameter sets through the number of maximum possible common behavioural parameter sets. It shows that some catchments are in very good agreement whereas others have lower numbers of behavioural models, particularly at Borgharen. This chart could indicate a probable problem in the process representation of the model structure or errors in the input and output data. It is very difficult to separate these different sources of error (see discussion of Beven and Young [39]). However, a lower number could also mean that the data contain a higher information content which allows greater

rejection and thus can be seen as very positive. If Figure 3 is used and the number of common parameter sets along the flow path is evaluated a rapid decrease can be seen. Unfortunately, no universal parameter set (Table 7) survives this rigorous evaluation procedure, which suggests that some of the basic assumptions of the model (e.g. the distribution of the rainfall) should be reconsidered in future applications.

An individual calibration for each sub-catchment seems to be inadequate from a physical modelling point of view, because only multipliers on definite physical characteristics have been applied. However, the uniqueness of place argument by Beven [37] suggests that different parameter sets might be required in individual catchment areas due to the aggregation of different small scale behaviours. If this argument is extended to make the case that effective parameters have to compensate for different errors even

Table 7 Number of behavioural models along the flow path.

Stations	Behaviourals
Saint Mihiel/Chooz	540
Saint Mihiel/Chooz/Lesse (Gendron)	387
Ourthe Nisramont/Ourthe (Tabreux)	278
Ourthe Nisramont/Ourthe (Tabreux)/ Ambleve (Martinrive)	122
Ourthe Nisramont/Ourthe (Tabreux)/ Ambleve (Martinrive)/Vesdre (Chaufontaine)	57
All	0

in very similar catchments, processed in the same way and at similar scales, then an individual sub-catchment calibration is appropriate. The question remains as to how far the calibration of our grid should be disaggregated in order to aggregate to a reasonable result. Furthermore, an increase of equifinality and non-identifiability would be unavoidable with an increase of freedom of calibration [40]. Nevertheless, predictions can be made for each sub-catchment hydrograph considering only the behavioural models at this point. Figures 6 and 7 display the flow hydrographs with the 5 and 95% percentiles. The black vertical line marks the end of the calibration period and the beginning of the evaluation time steps. In general the uncertainty bounds bracket the observations very well and describe the flow regime in the calibration period. Even in the prediction period, with a quite different range of discharges, the observations are bracketed satisfactorily. As expected the discharge prediction quantiles cover a very large range especially at the higher peaks because the calibration period consisted of much smaller peaks. Special attention has to be given to the hydrographs at Borgharen, which are important for prediction of flood inundation [41]. Although only a small sample of hydrographs was shown to be behavioural, the prediction bounds for the calibration as well as the evaluation period proved to be reasonable. However, it can be questioned if this relative small sample, limited by the available computer power, is representative enough for these predictions.

Of further interest is the plot at Saint Mihiel where the second evaluation peak occurs later than the prediction bounds. This

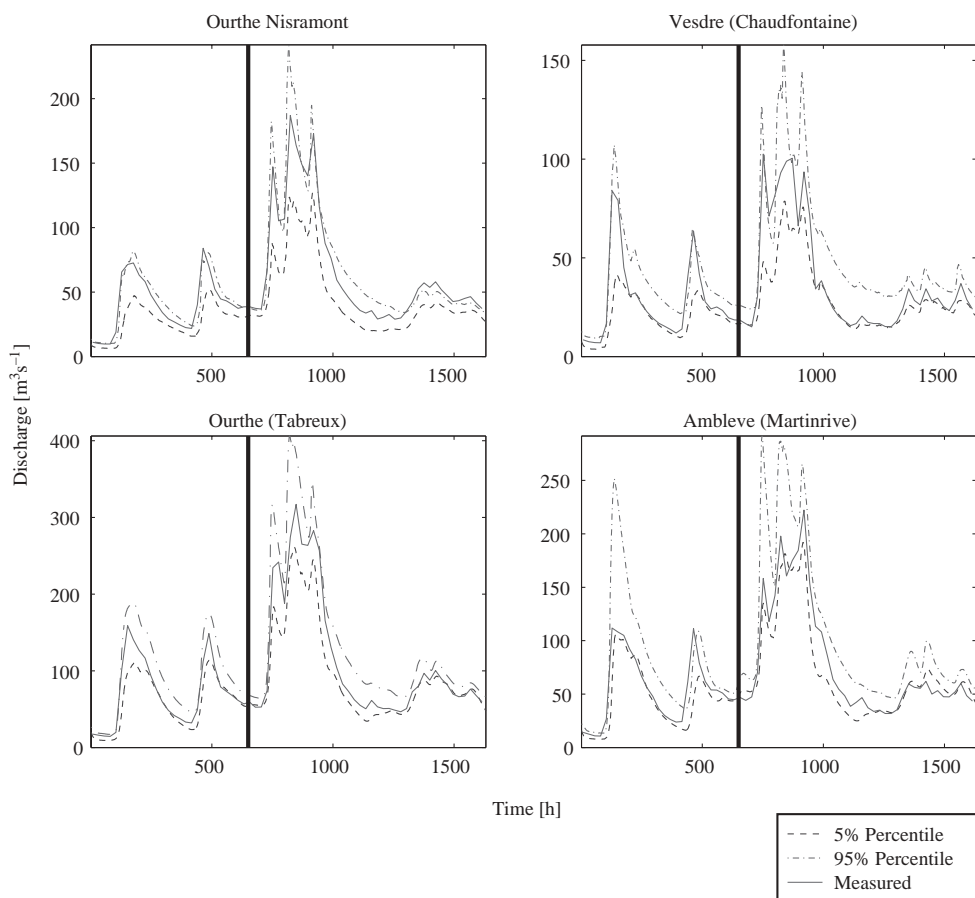


Figure 6 Uncertainty bounds for calibration (before vertical line) and evaluation period with observed rainfall.

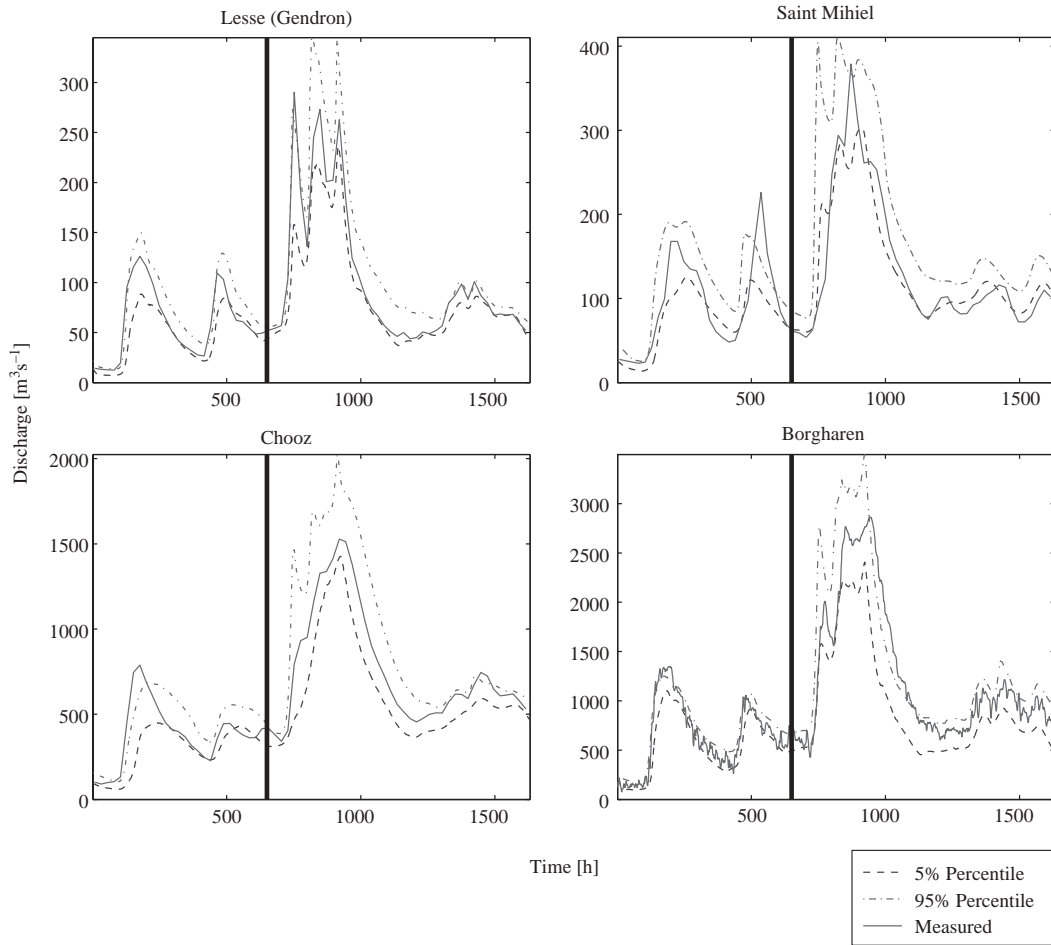


Figure 7 Uncertainty bounds for calibration (before vertical line) and evaluation period with observed rainfall.

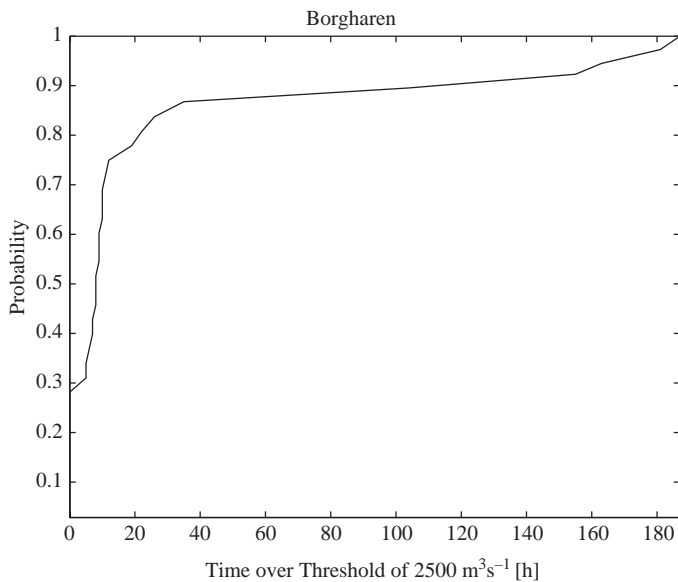


Figure 8 Cumulative Distribution Function for time over a threshold of  $2500 \text{ m}^3 \text{ s}^{-1}$  at Borgharen.

suggests an intrinsic error in the model structure or the input and the data for this station should be reexamined. However, this would have not changed any of the findings and was therefore left as an example.

Another method of evaluating predictions is to concentrate on exceedance levels [21,42], which is seen as a valuable way of

providing additional useful information to users for operational flood warning purposes. For illustration purposes a threshold level, which could reflect a warning level for the national river authorities, has been set arbitrarily to  $2500 \text{ m}^3 \text{ s}^{-1}$  for the main outflow at Borgharen. A cumulative distribution function (Figure 8) is estimated, for the duration of exceedance versus the cumulative relative likelihood, which is similar to the percentile evaluation at each time step and is for example a useful estimator for flood damage [43]. Calculated in this way, the exceedance duration estimates take proper account of the prediction uncertainties. As expected the curve jumps at 0 hours of exceedance and indicates a 28% possibility of the flow staying below the threshold value. However, this graph also shows a possibility to exceed this threshold for more than 200 hours which might be important for defence structures.

**Conclusions**

The parameter uncertainties of the raster based rainfall runoff model LisFlood have been evaluated within the Generalized Likelihood Uncertainty Estimation [16]. The calibration and evaluation data set consisted of the 1995 flood event on the river Meuse, which is situated in France, Belgium and The Netherlands.

Ten parameters (see Table 2) which control surface as well as subsurface processes have been sampled from independent uniform random distributions and the results compared to eight sets of evaluation hydrographs with the Multicomponent Mapping method [33]. The parameters are multipliers of effective values, which are based on estimates of physical properties or maps. None of the multipliers showed good identifiability and behavioural models could be found over the full range of the sampled parameter sets. Therefore, an uncertainty analysis should be compulsory for every model exercise even when the changes in the modelling domain are considered as controllable or minor.

In this study it was also not possible to establish clear two-dimensional relationships between the parameters of behavioural models.

Some sub-catchments have a significantly lower number of behavioural models than compared to others, although no strong differences in the underlying maps or inputs can be identified. It could be that the evaluation data allow a stronger constraint and thus lead to higher identifiability of the parameter sets. However, this difference could be also interpreted as a model failure, which may be either due to the large grid scale of 1 km<sup>2</sup> or processes which are neglected within the rainfall runoff model. It supports the argument of the uniqueness of place of Beven [37]. Even when all the processes and maps are derived in the same way for catchments with similar characteristics, effective parameters may have to compensate for deficiencies in the input data and model structures and therefore a determination of prior distributions is very difficult, if not impossible. Such an argument may justify the local calibration of individual sub-catchments, although here only global multipliers have been applied to the underlying physically based model parameters.

From this argument it is not surprising that this study fails to establish a global parameter set, which is behavioural for all sub-catchments. However, this total rejection of all parameter sets should not be seen as a deficit of the modelling process, rather as a chance to improve results or structures in future (for a similar result see [26]).

Certainly when the model has, as in this case, produced encouraging results in the prediction of uncertainty percentiles, the results might still be useful in decision making during flood periods. The model brackets the observations in the calibration as well as the evaluation events reasonably well and can be used for flow forecasting. The model also shows the capability to evaluate uncertainty in exceeding certain thresholds, which could be a warning level for overtopping of dykes or as linkage to cost functions [43], although this work clearly shows that this should not be done in a deterministic way.

The latter has to be investigated for different magnitudes to clarify the dependency of the uncertainty percentiles on magnitude, length and quality of the time series or other inputs such as rainfall forecasts. It is also of interest how scaling affects the results and thus it might be beneficial to evaluate the model on subcatchments with smaller grid sizes or on larger grid squares for larger regions of Europe.

## Acknowledgements

This work was funded by the EU Framework 5 Proposal EVG1-CT-1999-00011 European Flood Forecasting System (<http://effs.wdelft.nl/>) and NERC Grant NER/L/S/2001/0068. The authors wish to thank Thorsten Wagener for his insightful comments.

## Notations

$\theta_i$  = Model parameter set  $i$

*Box Size* = Size of box in vertical direction

$L$  = Likelihood measure

$N$  = Number of behavioural models

$o$  = Vector of Observations

$P$  = Prediction quantile

$q_b$  = Flow magnitude at centre of box

$\hat{Z}_{t,i}$  = The value of the variable  $Z$  at time  $t$  simulated by the model  $M(\theta_i)$

(see also Table 3)

## References

1. ABBOTT, M.B., BABOVIC, V.M. and CUNGE, J.A. (2001). "Towards the Hydraulics of the Hydroinformatics Era," *Journal of Hydraulic Research*, 39(4), 339–349.
2. HORRITT, M.S. and BATES, P.D. (2001). "Effects of Spatial Resolution on a Raster Based Model of Flood Flow," *Journal of Hydrology*, 253(1–4), 239–249.
3. N.N., *PCRaster Environmental Software*, 2003, <http://www.pcraster.nl/>
4. HUNT, J. and SHACKLEY, S. (1999). "Reconceiving Science and Policy: Academic, Fiducial and Bureaucratic Knowledge," *Minerva*, 37, 141–164.
5. ZEHR, S.C. (2000). "Public Representations of Scientific Uncertainty about Global Climate Change," *Public Understanding of Science*, 9(2), 85–103.
6. ORESKES, N., SHRADERFRETCHETTE, K. and BELITZ, K. (1994). "Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences," *Science*, 263(5147), 641–646.
7. SHACKLEY, S. (1998). "Introduction to Special Section on the use of Models in Appraisal and Policy-Making," *Impact Assessment and Project Appraisal*, 16(2), 81–89.
8. SHACKLEY, S. and DARIER, D. (1998). "Seduction of the Sirens: Global Climate Change and Modelling," *Science and Public Policy*, 25(5), 313–325.
9. HOFFMANN-RIEM, H. and WYNNE, B. (2002). "In Risk Assessment, One has to Admit Ignorance," *Nature*, 416, 123.
10. BRADSHAW, G.A. and BORCHERS, J.G. (2000). "Uncertainty as Information: Narrowing the Science-Policy Gap," *Conservation Ecology*, 4(1), Art. No. 7.
11. VAN ASSELT, M. and ROTMANS, J. (2002). "Uncertainty in Integrated Assessment Modelling - from Positivism to Pluralism," *Climatic Change*, 54(1–2), 75–105.

12. WHEATER, H. (1796). "Progress in and Prospects for Fluvial Flood Modelling," *Philosophical Transactions: Mathematical, Physical & Engineering Sciences*, 360, 1409–1431.
13. RICCI, P.F., RICE, D., ZIAGOS, J. and COX, L.A. (2003). "Precaution, Uncertainty and Causation in Environmental Decisions," *Environment International*, 29, 1–19.
14. DE ROO, A., GOUWEELEEUW, B., THIELEN, J., BARTHOLMES, J., BONGIOANNINI-CERLINI, P., TODINI, E., BATES, P., HORRITT, M., HUNTER, N., BEVEN, K., PAPPENBERGER, F., HEISE, E., RIVIN, G., HILLS, M., HOLLINGSWORTH, A., HOLST, B., KWADIJK, J., REGGIANI, P., VAN DIJK, M., SATTLER, K. and SPROKKEREEF, E. (2003). "Development of a European Flood Forecasting System," *International Journal of River Basin Management*, 1, 49–59.
15. DE ROO, A.P.J., ODIJK, M., SCHMUCK, G., KOSTER, E. and LUCIEER, A. (2001). "Assessing the Effects of Land use Changes on Floods in the Meuse and Oder Catchment," *Physics and Chemistry of the Earth (B)*, 26, 593–599.
16. BEVEN, K. and BINLEY, A. (1995). "The Future of Distributed Models: Model Calibration and Uncertainty Prediction," *Hydrological Processes*, 6, 279–298.
17. VAN DEURSEN, W.P.A. (1995). "Geographical Information Systems and Dynamic Models: Development and Application of a Prototype Spatial Modelling Language," *Netherlands Geographic Studies*, 190.
18. WESSELING, C.G., KRASSENBERG, D.J., BURROUGH, P.A. and VAN DEURSEN, W.P.A. (1996). "Integrated Dynamic Environmental Models in GIS: the Development of a Dynamic Modelling Language," *Transactions in GIS*, 1, 40–48.
19. DE ROO, A.P.J. (2000). LisFlood: A Rainfall-Runoff Model for Large River Basins to Assess the Influence of Land use Changes on Flood Risk, <http://natural-hazards.jrc.it/documents/floods/1998-docs/1998-papers/98Giessen.pdf>
20. DE ROO, A.P.J., WESSELING, C.G. and VAN DEURSEN, W.P.A. (2000). "Physically Based River Basin Modelling within a GIS: the LISFLOOD Model," *Hydrological Processes*, 14, 1981–1992.
21. GOUWEELEEUW, B., THIELEN, J., DE ROO, A. and BUIZZA, R. "Flood Forecasting using Probabilistic Weather Predictions," *Hydrology and Earth System Science*, accepted.
22. BEVEN, K.J., FREER, J., HANKIN, B. and SCHULZ, K. (2000). The use of generalised likelihood measures for uncertainty estimation in high order models of environmental systems, in P.C.Y. Young (ed.), *Nonlinear and Nonstationary Signal Processing*, Cambridge University Press, Cambridge, pp. 115–151.
23. BEVEN, K.J. (1993). "Prophecy, Reality and Uncertainty in Distributed Hydrological Modelling," *Advances in Water Resources*, 16, 41–51.
24. MELCHING, C. (1995). Reliability estimation, in V.P. Singh (ed.), *Computer Models of Watershed Hydrology*, Water Resources Publication, pp. 69–118.
25. THIEMANN, M., TROSSET, M., GUPTA, H. and SOROOSHIAN, S. (2001). "Bayesian Recursive Parameter Estimation for Hydrologic Models," *Water Resources Research*, 37(10), 2521–2535.
26. FREER, J., BEVEN, K.J. and PETERS, N. (2003). Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure, in Q.Y. Duan, H. Gupta, S. Sorooshian, A. Rousseau and R. Turcotte (eds.), *Calibration of Watershed Models*, American Geophysical Union: Washington, pp. 69–88.
27. ROMANOWICZ, R., BEVEN, K. and TWAN, J.A. (1994). Evaluation of predictive uncertainty in nonlinear hydraulic models using a Bayesian approach, in K.F. Turkman, (ed.), *Statistics for the Environment 2, Water Related Issues*, Wiley & Sons, New York.
28. PAPPENBERGER, F., DE ROO, A., BATES, P., BEVEN, K.J., GOUWEELEEUW, B., HUNTER, N. and THIELEN, J. (2003). Assessment of parameter estimates derived from GIS datasets for the first chosen application-Deliverable Report 5.1 – Contract no.: EVG1-CT-1999-00011. 2003.
29. FREER, J., BEVEN, K. and AMBROISE, B. (1996). "Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach," *Water Resource Research*, 32(7), 2161–2173.
30. Aronica, G., HANKIN, B. and BEVEN, K.J. (1998). "Uncertainty and Equifinality in Calibrating Distributed Roughness Coefficients in a Flood Propagation Model with Limited Data," *Advances in Water Resources*, 22(4), 349.
31. Franks, S.W., GINESTE, P., BEVEN, K.J. and MEROT, P. (1998). "On Constraining the Predictions of a Distributed Model: The Incorporation of Fuzzy Estimates of Saturated Areas into the Calibration Process," *Water Resources Research*, 34(4), 787–797.
32. BLAZKOVA, S. and BEVEN, K. (2002). "Flood Frequency Estimation by Continuous Simulation for a Catchment Treated as Ungauged (with Uncertainty)," *Water Resources Research*, 38(8), Art. No. 1139.
33. PAPPENBERGER, F. and BEVEN, K. (2004). "Functional Classification and Evaluation of Hydrographs based on Multicomponent Mapping," *International Journal of River Basin Management*, this issue.
34. NASH, J.E. and SUTCLIFFE, J.V. (1970). "River Flow Forecasting Through Conceptual Models, Part I – A Discussion of Principles," *Journal of Hydrology*, 10, 282–290.
35. MO, X., BEVEN, K. and PAPPENBERGER, F., "Water Budget Estimation and Prediction Uncertainty with the Distributed Model LisFlood Over the Lushi Basin, China", *Journal of Hydrology*, submitted.
36. GUPTA, H.V., SOROOSHIAN, S. and YAPO, P.O. (1998). "Toward Improved Calibration of Hydrologic Models: Multiple and Noncommensurable Measures of Information," *Water Resources Research*, 34(4), 751–763.
37. BEVEN, K.J. (2000). "Uniqueness of Place and Process Representations in Hydrological Modelling," *Hydrology and Earth System Sciences*, 4(2), 203.
38. DUAN, Q., SOROOSHIAN, S. and GUPTA, V. (1992). "Effective and Efficient Global Optimisation for Conceptual

- Rainfall-Runoff Models,” *Water Resources Research*, 28, 1015–1031.
39. BEVEN, K.J. and YOUNG, P.C. (2003). “Comment on Bayesian Recursive Parameter Estimation for Hydrologic Models by M. Thiemann, M. Trosset, H. Gupta and S. Sorooshian”, *Water Resource Research*, 9(5), p. doi: 10.1029/2001WR001183.
  40. SCHULZ, K. and BEVEN, K. (2002). Data supported robust parameterizations in land surfaces – atmosphere flux predictions: towards a Top-down approach. Hydrological Processes.
  41. HUNTER, N.M., BATES, P.D., HORRITT, M.S., DE ROO, A.P.G. and WERNER, M.G.F., “Utility of Different Data Types for Conditioning Flood Inundation Models within a GLUE Framework”, *Hydrology and Earth System Science*, submitted.
  42. KRZYSZTOFOWICZ, R. (2002). “Probabilistic Flood Forecast: Bounds and Approximations,” *Journal of Hydrology*, 268(1–4), 41–55.
  43. DUTTA, D., HERATH, S. and MUSIAKE, K. (2003). “A Mathematical Model for Flood Loss Estimation,” *Journal of Hydrology*, 277(1–2), 24–49.
  44. VON HOYNINGEN-HUENE, J. (1981). Die Interzeption des Niederschlags in landwirtschaftlichen Pflanzenbeständen. Verband für Wasserwirtschaft und Kulturbau, DVWK: Braunschweig, Germany. p. 63.
  45. SHUTTLEWORTH, W.S. and CALDER, I.R. (1979). “Has the Priestley-Taylor Equation any Relevance to Forest Evaporation?” *Journal of Applied Meteorology*, 18, 639–646.
  46. SUPIT, I., HOOIJER, A.A. and VAN DIEPEN, C.A. (1994). System description of the WOFOST 6.0 crop simulation model implemented in CGMS. Volume 1: Theory and Algorithms, European Commission EUR 15956 EN.
  47. VAN DER GOOT, E. (1997). Technical description of interpolation and processing of meteorological data in CGMS. Joint Research Centre, Space Applications Institute: Ispra, Italy.
  48. BAUMGARTNER, M.F., MARTINET, J., RANGO, A. and ROBERTS, R. (1994). Snowmelt runoff model (SRM) user’s manual. Univeristy of Bern. Dept. of Geography: Bern, Switzerland.
  49. SMITH, R.E. and PARLANGE, J.Y. (1978). “A Parameter-Efficient Hydrologic Infiltration Model,” *Water Resources Research*, 14, 533–538.
  50. MOLNAU, M. and BISSEL, (1983). A continuous frozen ground index for flood forecasting. 51st Annual Meeting of the Western Snow Conference.
  51. VAN GENUCHTEN, M. (1980). “A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils,” *Soil Science Society of America Journal*, 44, 892–898.
  52. WÖTEN, J.H.M., LILLY, A., NEMES, A. and LE BAS, C. (1998). Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning. Winand Staring Centre: Wageningen. p. 106.
  53. DARCY, H. (1856). Les Fontaines Publiques de la ville de Dijon. Paris: Dalmont.

