

# The role of observation uncertainty in the calibration of hydrologic rainfall-runoff models

T. Ghizzoni<sup>1,2</sup>, F. Giannoni<sup>3</sup>, G. Roth<sup>1,2</sup>, and R. Rudari<sup>1,2</sup>

<sup>1</sup>CIMA, Centro di ricerca Interuniversitario in Monitoraggio Ambientale, Universities of Genoa and Basilicata, Italy

<sup>2</sup>DIST, Dipartimento di Informatica, Sistemistica e Telematica, University of Genoa, Italy

<sup>3</sup>ARPAL, Agenzia Regionale per l'Ambiente Ligure, Genoa, Italy

Received: 5 March 2007 – Revised: 21 May 2007 – Accepted: 3 June 2007 – Published: 28 June 2007

**Abstract.** Hydrologic rainfall-runoff models are usually calibrated with reference to a limited number of recorded flood events, for which rainfall and runoff measurements are available. In this framework, model's parameters consistency depends on the number of both events and hydrograph points used for calibration, and on measurements reliability. Recently, to make users aware of application limits, major attention has been devoted to the estimation of uncertainty in hydrologic modelling. Here a simple numerical experiment is proposed, that allows the analysis of uncertainty in hydrologic rainfall-runoff modelling associated to both quantity and quality of available data.

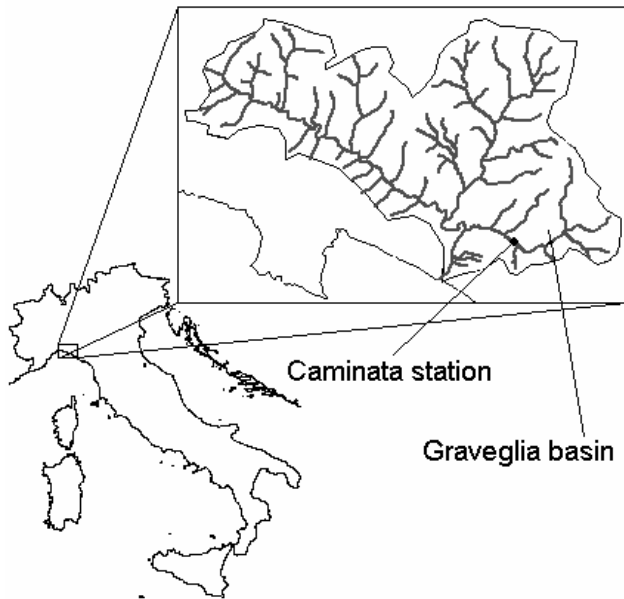
A distributed rainfall-runoff model based on geomorphologic concepts has been used. The experiment involves the analysis of an ensemble of model runs, and its overall set up holds if the model is to be applied in different catchments and climates, or even if a different hydrologic model is used. With reference to a set of 100 synthetic rainfall events characterized by a given rainfall volume, the effect of uncertainty in parameters calibration is studied. An artificial truth – perfect observation – is created by using the model in a known configuration. An external source of uncertainty is introduced by assuming realistic, i.e. uncertain, discharge observations to calibrate the model. The range of parameters' values able to “reproduce” the observation is studied. Finally, the model uncertainty is evaluated and discussed. The experiment gives useful indications about the number of both events and data points needed for a careful and stable calibration of a specific model, applied in a given climate and catchment. Moreover, an insight on the expected and maximum error in flood peak discharge simulations is given: errors ranging up to 40% are to be expected if parameters are calibrated on insufficient data sets.

*Correspondence to:* T. Ghizzoni  
(tatiana.ghizzoni@cima.unige.it)

## 1 Introduction

In recent years, major interest has been devoted to environmental monitoring and modelling. Thanks to a growing understanding of physical processes, it is now possible to obtain realistic catchment-scale descriptions of the dynamics of the involved components. These improvements led many modellers to study new methods to calibrate and validate models. Despite these advancements, hydrologic modelling is still uncertain. To discriminate different sources of uncertainty in complex non-linear systems, as those that describe hydrological processes, is very hard (Beven, 2004a, b). A different way to improve hydrological model reliability is to advance calibration methods. As a consequence, a number of calibration methods have been proposed: for example, Bayesian (Kuczera, 1994), multi-objective (Sorooshian et al., 1993; Gupta et al., 1998), the GLUE method (Beven and Binley, 1992) and Kalman filter based (Bras and Rodriguez-Iturbe, 1985). Nowadays, Markov Chain Monte Carlo algorithms have achieved a growing importance, e.g. Kuczera and Parent (1998) have used Metropolis and Hastings algorithms (Metropolis et al., 1953; Hastings, 1970) to evaluate parameters' uncertainty. Nevertheless, calibration is performed by using a limited number of observations. As a consequence, it is important not only to apply calibration best procedures, but also to estimate the uncertainty associated with calibration results. This approach gives a double advantage: users are conscious of not having a “perfect” tool and, by knowing its strength and weak points, they can use it at its best.

The purpose of this work is to delineate an experiment able to improve our confidence in hydrologic models calibration based on standard hydrologic observations. Since calibration reliability increases with the number of events and the number of hydrograph points – usually peaks – available to describe each event, the aim of the present work is to explore external uncertainty associated to model calibration in terms of the number of both available events and hydrograph



**Fig. 1.** Location of the Graveglia basin and of the Caminata station, Northern Italy.

points. This experiment proposes an alternative reading of a more classical sensitivity analysis: in fact, to obtain information about the reliability of a model, an analysis of its potential accuracy is set up. While this approach provides the best performance the model can give for specified climate and catchment, the overall procedure here proposed can be taken as a guide for model calibration. Independently from the hydrologic model used, concepts introduced in this work can be seen as useful in delineating an optimal calibration strategy.

The morphology of the Graveglia basin (Fig. 1), in Northern Italy, closed at Caminata station, with drained area  $\cong 40 \text{ km}^2$  is taken as laboratory for the experiment. Its elevation is modelled through a  $40 \times 40 \text{ m}$  grid Digital Elevation Model. To highlight the role of surface propagation processes, and their relations to observational uncertainty, the catchment is considered impervious, i.e. infiltration processes are not modelled in this experiment.

## 2 Experimental set up

Hydrologic modelling is performed using the DRiFt – Discharge River Forecast – model: a linear, semi-distributed event model based on a geomorphologic approach (Giannoni et al., 2000, 2003, 2005; Boni et al., 2007). This model is focused on the efficient description of the drainage system in its essential parts: hillslopes and the channel network are addressed with two kinematic scales, which determine the base of the geomorphologic response of the basin.

In the DRiFt model, the geomorphologic module is coupled with a simple distributed representation of soil infiltration properties, while the rainfall event is schematized with its variability in time and space. The model presents four parameters: two for drainage network recognition and two kinematic parameters. The runoff volume is routed to the outlet with a Time Variant Unit Hydrograph (TVUH) technique, which, although representing the basin as a linear system, takes into account the spatial and temporal variability of the runoff generation. While in standard IUH convolution procedures the whole elementary response of the basin, multiplied for an average rainfall excess intensity, is considered and then summed, here, at each temporal step, only those cells where an actual runoff is produced are used to compute the UH. This results in that each time step UH is different from the other. This approach allows the representation of the runoff generation dispersion due to both rainfall and soil characteristic heterogeneity, if any, producing a more realistic simulation of basin behaviour.

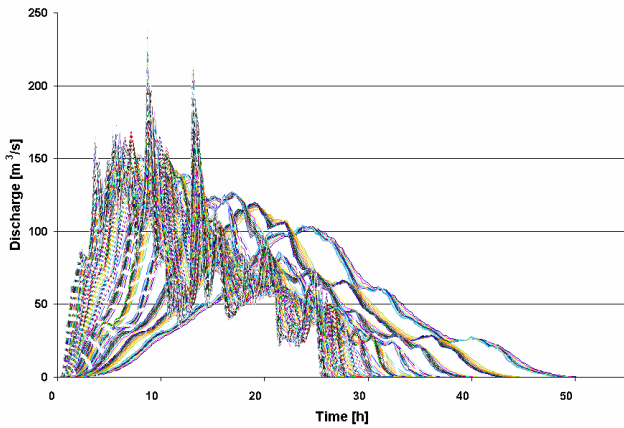
To simulate rainfall events, the RainFARM model (Rainfall Filtered Autoregressive Model, Rebora et al., 2006) has been used. RainFARM is a method for stochastic rainfall downscaling. It is based on a non-linear transformation of a linearly-correlated Gaussian random field; it conserves the total amount of precipitation cumulated on a chosen spatial and temporal box, and it takes into account anisotropy between space and time, if any. It conserves the spatial and temporal correlations of observed meteorological rainfall fields, and, to take into account orography effects, it conserves also the spatial position of large rainfall structures. RainFARM model uses LAM rainfall output and, assuming that rainfall's spectrum has a power-law form, it estimates its spectral slopes. Using RainFARM, 100 realistic and independent events, with fixed rainfall volume  $- 7 \times 10^8 \text{ m}^3$  within a 24 h overall event duration – have been generated, and used as input for hydrologic modelling.

Although the proposed experiment is a synthetic one, it is strictly linked with reality: in fact the Graveglia basin morphology is considered, and the rain maps reproduce the characteristic of intense rainfall events.

### 2.1 Synthetic hydrographs simulation

The aim of this work is to evaluate hydrologic model uncertainty related to parameter sets calibration. For the specific hydrologic model used, parameters subject to calibration on the basis of rainfall and discharge observations are those related to the dynamic of water movement, i.e. hillslope and channel velocities. As a consequence, an ensemble of hydrographs has been first produced with reference to all the available rainfall events for a wide range of parameter sets.

The investigated parameter sets range includes channel velocities from 0.2 to 4 m/s ( $V_c$ ) and hillslope velocities from 0.02 to 0.44 m/s ( $V_h$ ). Within those boundaries, 12 channel velocities and 18 hillslope velocities have been investigated,



**Fig. 2.** Hydrographs ensemble obtained from a single rainfall event with reference to the investigated parameter sets (216 channel and hillslope velocities couples).

for a total of 216 hydrographs for each rainfall event. The hydrographs ensemble for a selected rainfall event is shown in Fig. 2.

Calibration reliability increases with the number of events and hydrograph points – i.e., flow peaks – made available to describe each event. In the present work, the uncertainty associated with the calibration is therefore described as a function of the number of both events and hydrograph points made available for calibration. Since the results of this analysis strictly depend on the type – e.g., peaks – and quality – e.g., reported in terms of measurement accuracy – of the available hydrograph points, an objective sampling procedure has been implemented as follows.

## 2.2 Sampling procedure

A common use of hydrologic modelling is to evaluate peak flow discharge, in order to identify flooding risks. As a consequence, the proposed sampling procedure firstly identifies peaks. For a given event, one could easily recognize the first peak, i.e. the overall discharge maximum. More difficult is the objective identification of secondary peaks, often confused by automated recognition processes with minor discharge fluctuation. To disregard peaks generated by minor discharge fluctuations, a smoothing procedure, based on a moving average technique, is applied as

$$Q_s(i) = \sum_{j=0}^N \left[ Q \left( i - \frac{N}{2} + j \right) \times f(j) \right] \quad (1)$$

where  $Q_s(\cdot)$  is the smoothed discharge,  $N$  the number of filters' coefficients, and  $f(j)$  the Gaussian filters' coefficients. The filter  $f$ , by means of 11 coefficients, weights the hydrograph points' values in function of their distance from the analyzed one, assuming a Gaussian distribution.

Once the hydrograph is smoothed, maximum points are first identified, by calculating the first order derivative, and then sorted in decreasing order. This procedure, with different number of filter's coefficients, has been applied iteratively to the entire hydrograph set to identify, and disregard, peaks generated by minor discharge fluctuations.

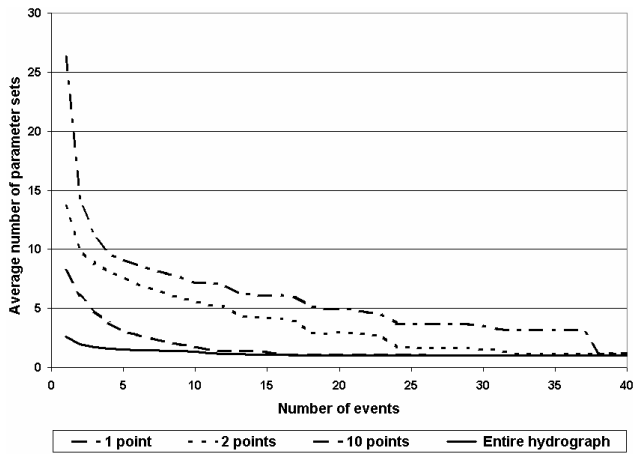
## 3 Hydrologic model sensitivity to uncertainty in discharge measurements

As discussed above, calibration results depend on the number of available events and hydrograph points. Moreover, they depend also on observations accuracy, as discharge and rainfall data are far from “perfect”. For example, discharge estimates are always obtained from depth measurements by means of a rating curve (for this study area rating curve parameters are reported by the Italian National Technical Service, 1996). This involves two sources of uncertainty: the first related to the depth measurement, the second to the rating curve consistency. For this reason, to analyse calibration reliability in a “real” situation, uncertainty associated with discharge observations has been explicitly taken into account.

Once a hydrograph – and the corresponding parameter set – is assumed as the observed one, all the 21 600 members that constitute the hydrographs ensemble have been analyzed to identify congruent hydrographs, and corresponding parameter sets. A hydrograph is considered congruent with the observed one if respective discharge points are included in a specific range, whose width mimics the uncertainty associated to discharge measures. While there is neither reference nor an adequate theoretical explanation on how to select specific time and water level – i.e., discharge – range, hereinafter the time window has been selected to mimic the most common format for discharge records – i.e., one data point for each hour. Less a-priori knowledge is available for choosing the discharge range since it depends on different aspects that include, e.g., the instrument type – with its own accuracy – and the rating curve consistency.

If just one point, the peak, is available for calibration, congruent hydrographs present peak discharge and time to peak values that are in the vicinity of the observed values as follows:

1.  $Q_{\text{real}} \rightarrow$  rating curve:  $Q=72.51 \times (H-1.05)^{\frac{3}{2}} \rightarrow H_{\text{real}}$
2. Uncertainty assignment:  $\rightarrow H_{\text{max}}=H_{\text{real}}+15 \text{ cm}$   
 $\rightarrow H_{\text{min}}=H_{\text{real}}-15 \text{ cm}$
3.  $H_{\text{max}} \rightarrow$  rating curve:  $Q=72.51 \times (H-1.05)^{\frac{3}{2}} \rightarrow Q_{\text{max}}$
4.  $H_{\text{min}} \rightarrow$  rating curve:  $Q=72.51 \times (H-1.05)^{\frac{3}{2}} \rightarrow Q_{\text{min}}$
5.  $T_{\text{max}}=T_{\text{real}}+30 \text{ min}$
6.  $T_{\text{min}}=T_{\text{real}}-30 \text{ min}$



**Fig. 3.** Average number of parameter sets able to produce hydrographs that are congruent with the observed one as a function of the number of events and data points used for calibration.

where:  $Q_{real}$ =peak discharge of the observed hydrograph;  
 $T_{real}$ =time of peak discharge for the observed hydrograph.

As a consequence, congruent hydrographs must have a peak in:

$$T_{min} \div T_{max} \quad \text{and} \quad Q_{min} < Q(T_{real}) < Q_{max}$$

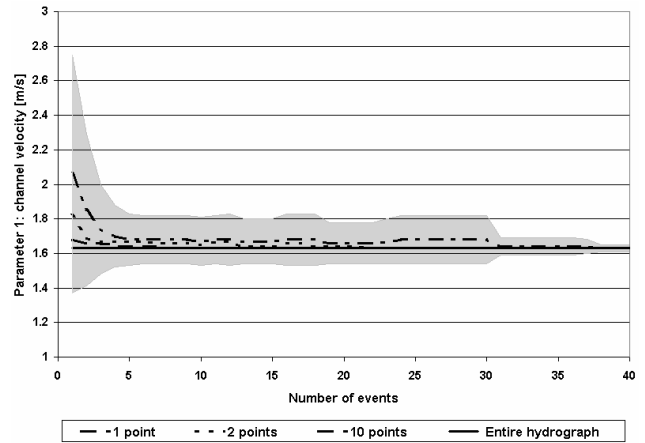
If more points are available for calibration, congruent hydrographs should present all of them in the vicinity of the corresponding observed values.

In the following, the hydrograph obtained with channel velocity  $V_c=1.63$  m/s and hillslope velocity  $V_h=0.15$  m/s has been considered as the truth, i.e. the “observed one”.

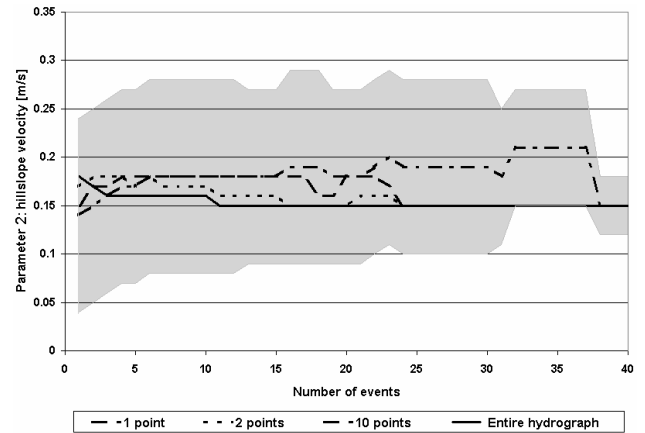
Figure 3 shows the average number of parameter sets able to produce hydrographs congruent with the observed one, as a function of the number of events available for calibration. The four curves show how this number changes with an increase of the number of hydrograph points available for calibration. It is possible to observe that either an increase in the number of hydrograph points or in the number of events implies an increase in hydrologic model reliability. For example, if only one hydrograph point, e.g. the peak, is available for calibration, 38 events are needed to identify the correct parameter set. On the other hand, only two events are needed if the entire hydrograph is available for calibration. This is however difficult with simple models targeted to simulate at best a certain range of discharges.

Figures 4 and 5 show results for the identification of parameter set values. Hillslope and channel velocities values are reported as functions of the number of events and hydrograph points available for calibration – the grey area highlights the variability, i.e. standard deviation, associated with the curve obtained with one hydrograph point.

One could observe that, having during calibration observed peak discharges for four rainfall events, it is possible to estimate the first parameter – channel velocity – with



**Fig. 4.** Average channel flow velocity as a function of the number of events and data points used for calibration. Shaded area represents the variability, in term of one standard deviation, associated with the one hydrograph point calibration.



**Fig. 5.** Average hillslope flow velocity as a function of the number of events and data points used for calibration. Shaded area represents the variability, in term of one standard deviation, associated with the one hydrograph point calibration.

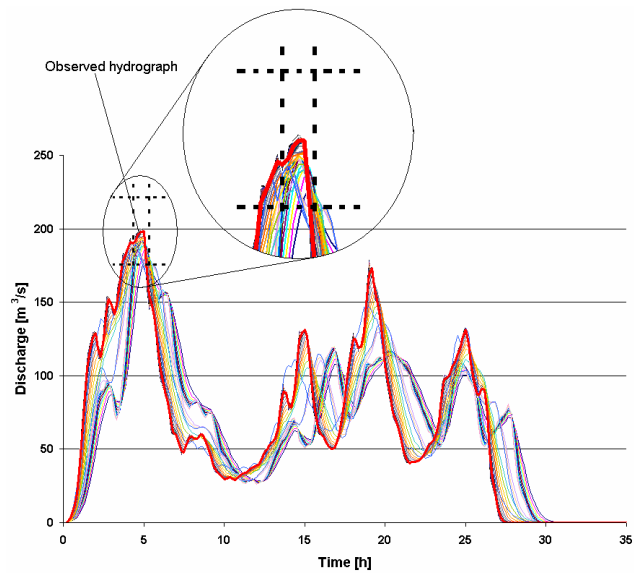
a 0.2 m/s uncertainty. On the other hand, it is very hard to reliably estimate the second parameter – hillslope velocity. Even if a greater number of events will be made available for calibration, the uncertainty associated with the estimation of this parameter is large and almost constant with respect to the number of available events.

#### 4 Effects of uncertainty on flood forecast

If only one point – the peak – for one event is available for calibration, results show that, on average, 26 parameter sets, with a standard deviation equal to 16, are able to generate an output congruent with the observed one. On this basis, one could be interested in a quantification of the consequences of

**Table 1.** Calibration effects in flood forecast.

$V_c$ [m/s]	$V_h$ [m/s]	success percentage	average percentage error	maximum percentage error
0.97	0.041	13%	21%	39%
0.97	0.050	20%	20%	38%
0.97	0.060	22%	19%	38%
0.97	0.072	23%	18%	36%
0.97	0.086	25%	17%	36%
0.97	0.103	29%	16%	35%
0.97	0.124	32%	16%	34%
0.97	0.149	34%	15%	33%
0.97	0.178	34%	15%	32%
0.97	0.214	33%	14%	32%
0.97	0.257	35%	14%	31%
0.97	0.308	37%	14%	31%
0.97	0.370	37%	14%	31%
0.97	0.444	37%	14%	31%
1.25	0.035	13%	18%	31%
1.25	0.041	20%	17%	29%
1.25	0.050	25%	15%	28%
1.63	0.020	10%	21%	36%
1.63	0.024	17%	18%	32%
1.63	0.029	25%	16%	28%
1.63	0.035	35%	13%	24%
1.63	0.041	53%	11%	22%
1.63	0.050	67%	9%	17%
1.63	0.060	83%	7%	14%
1.63	0.072	96%	5%	11%
1.63	0.086	99%	4%	7%
1.63	0.103	100%	2%	5%
1.63	0.124	100%	1%	2%
1.63	0.149	100%	0%	0%
1.63	0.178	100%	1%	1%
1.63	0.214	100%	1%	1%
1.63	0.257	100%	2%	1%
1.63	0.308	100%	2%	1%
1.63	0.370	100%	3%	1%
1.63	0.444	100%	3%	1%
2.12	0.020	22%	17%	35%



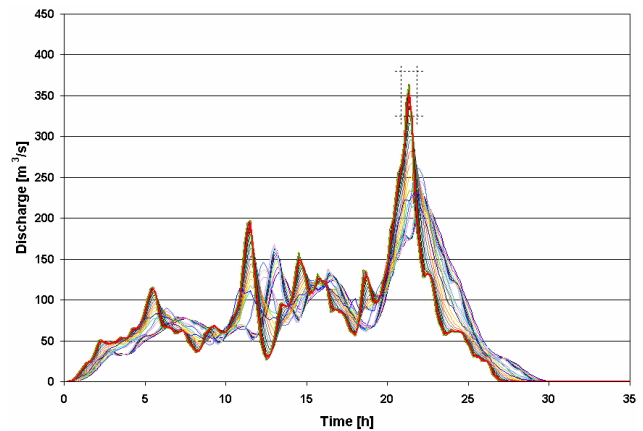
**Fig. 6.** Observed (red curve) and congruent hydrographs. The dashed box represents peak discharge and time to peak range used to identify congruent hydrographs. For this event, 36 hydrographs out of 216 are selected as congruent.

- the fifth column shows the maximum peak-discharge error in the remaining 99 events.

this uncertainty when those “congruent” parameter sets are used for the prediction of peak discharge. To answer this question, a simple experiment is proposed here.

First a rainfall event is selected, for which 36 parameter sets are able to identify a peak discharge congruent with the observed one (Fig. 6). Hydrograph peaks forecasted on the basis of these parameter sets are then investigated. The remaining 99 events have been used for the purpose, and mean and maximum error on peak discharge evaluation, together with the percentage of success using these parameter sets, identified. Results of this analysis are reported in Table 1. The red colour identifies the parameter set of the observed hydrograph. In detail:

- the first and the second columns show parameters values, channel and hillslope velocities, able to identify peak discharge for the rainfall event available for calibration;
- the third column shows the percentage of events, calculated on the remaining 99 events, in which peak discharge and time are correctly estimated, according to the assumed range;
- in the fourth column, the mean peak-discharge error is reported, as estimated on the remaining 99 events;



**Fig. 7.** Example of peak discharge forecasts obtained through the “36” parameter sets selected through calibration. The dashed box represents peak discharge and time to peak range used to identify congruent hydrographs. One could notice that many hydrographs are not congruent with the observed hydrograph (red curve), i.e. the peak falls outside of the dashed box.

On average, the peak discharge error is about 11%, while the max peak-discharge error is 39%. It can be noted that the couple  $V_c=1.63$  m/s– $V_h=0.149$  m/s provides a correct assessment of the peak-discharge only in 10% of the cases, with a mean error equal to 21% and a max error of 36%. Moreover, 9 velocity couples –  $V_c=1.93$  m/s;  $0.103 < V_h < 0.44$  m/s – are

able to correctly estimate peak-discharge for all the events: then, by considering the range previously assumed, a correct assessment of hillslope velocity is almost insignificant.

Figure 7 shows the hydrographs produced through rainfall-runoff modelling for one of the 99 available events as obtained by using the 36 parameter sets estimated during the calibration phase. Consequences on peak-discharge forecasting are very strong: it would be possible to forecast a peak discharge value of 220 m<sup>3</sup>/s, much lower than the correct one (350 m<sup>3</sup>/s).

## 5 Conclusions

An experiment is described, which is able to improve our confidence in hydrologic models calibration. This is performed on the basis of standard hydrologic observations, and proposes an alternative reading of a more classical sensitivity analysis. The proposed approach can be taken as a guide for model calibration. In fact, its independence from the specific model used makes its application possible for a variety of hydrologic models applied for different catchments and climatic regimes. External uncertainty is associated to model calibration in terms of the number of both available events and hydrograph points.

Results give quantitative information about the number of events and hydrograph's points – usually peaks – required to improve calibration and, as a consequence, to increase model reliability, also in case of uncertain observations.

*Acknowledgements.* We gratefully acknowledge discussions with F. Siccardi, and the financial support of the Italian Civil Protection Department, through the “Proscenio” 2005–2008 program.

Edited by: P. Alpert, H. Saaroni, and E. Heifetz

Reviewed by: two anonymous referees

## References

- Beven, K. J.: Response to “Emergence of a new kind of relativism in environmental modelling: a commentary”, *Proc. Roy. Soc., Lond.*, A460, 2147–2151, 2004a.
- Beven, K. J.: Does an interagency meeting in Washington imply uncertainty?, *Hydrol. Processes*, 18, 1747–1750, 2004b.
- Beven, K. J. and Binley, A. M.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, 1992.
- Boni, G., Ferraris, L., Giannoni, F., Roth, G., and Rudari, R.: Flood probability analysis for un-gauged watersheds by means of a simple distributed hydrologic model, *Adv. Water Resour.*, doi:10.1016/j.advwatres.2006.08.009, in press, 2007.
- Bras, R. L. and Rodriguez-Iturbe I.: *Random Functions and Hydrology*, Addison Wesley, 1985, new printing, Dover Publications, New York, 559, 1994.
- Giannoni, F., Roth, G., and Rudari, R.: A Semi-Distributed Rainfall-Runoff Model Based on a Geomorphologic Approach, *Phys. Chem. Earth*, 25/7–8, 665–671, 2000.
- Giannoni, F., Roth, G., and Rudari, R.: Can the Behaviour of different basins be described by the same model's parameter set? A geomorphologic framework, *Phys. Chem. Earth*, 28/6–7, 289–295, 2003.
- Giannoni, F., Roth, G., and Rudari, R.: A procedure for drainage network identification from geomorphology and its application to the prediction of the hydrologic response, *Adv. Water Res.*, 28(6), 567–581, 2005.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J.: *Markov chain Monte Carlo in practice*, Chapman and Hall, London, 486, 1996.
- Gupta, H. V., Sorooshian, S., and Yapo, P.O.: Towards improved calibration of hydrologic models: multiple and incommensurable measures of information, *Water Resour. Res.*, 34, 751–763, 1998.
- Ibbitt, R. P. and O'Donnell, T.: Fitting methods for conceptual catchment models, *J. Hydr. Div., Am. Soc. Civ. Eng.*, 97(9), 1331–1342, 1971.
- Hastings, K.: *Monte Carlo Simulation Methods Using Markov Chains and Their Applications*, *Biometrika*, 57, 97–109, 1970.
- Kuczera, G.: *NLFIT A Bayesian nonlinear regression suite*, Users Manual, Department of Civil Engineering and Surveying, University of Newcastle, 1994.
- Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211, 69–85, 1998.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087–1092, 1953.
- National Technical Service of Italy: *Annali idrologici parte seconda: bacini con foce al litorale tirrenico dal Roja al Magra*, Istituto Poligrafico dello Stato, Roma, 56, 1996.
- Rebora, N., Ferraris, L., von Hardenberg, J., and Provenzale, A.: The RainFARM: Rainfall downscaling by a Filtered AutoRegressive Model, *J. Hydrometeorol.*, 7, 724–738, 2006.
- Sorooshian, S., Duan, Q., and Gupta, V. K.: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, *Water Resour. Res.*, 29, 1185–1194, 1993.