

## Inferring the flood frequency distribution for an ungauged basin using a spatially distributed rainfall-runoff model

G. Moretti<sup>1</sup> and A. Montanari<sup>2</sup>

<sup>1</sup>Independent Engineer, Stuttgart, Germany

<sup>2</sup>Faculty of Engineering, University of Bologna, Bologna, Italy

Received: 28 November 2007 – Published in Hydrol. Earth Syst. Sci. Discuss.: 2 January 2008

Revised: 20 June 2008 – Accepted: 22 July 2008 – Published: 26 August 2008

**Abstract.** The estimation of the peak river flow for ungauged river sections is a topical issue in applied hydrology. Spatially distributed rainfall-runoff models can be a useful tool to this end, since they are potentially able to simulate the river flow at any location of the watershed drainage network. However, it is not fully clear to what extent these models can provide reliable simulations over a wide range of spatial scales. This issue is investigated here by applying a spatially distributed, continuous simulation rainfall-runoff model to infer the flood frequency distribution of the Riarbero River. This is an ungauged mountain creek located in northern Italy, whose drainage area is 17 km<sup>2</sup>. The hydrological model is first calibrated by using a 1-year record of hourly meteorological data and river flows observed at the outlet of the 1294 km<sup>2</sup> wide Secchia River basin, of which the Riarbero is a tributary. The model is then validated by performing a 100-year long simulation of synthetic river flow data, which allowed us to compare the simulated and observed flood frequency distributions at the Secchia River outlet and the internal cross river section of Cavola Bridge, where the basin area is 337 km<sup>2</sup>. Finally, another simulation of hourly river flows was performed by referring to the outlet of the Riarbero River, therefore allowing us to estimate the related flood frequency distribution. The results were validated by using estimates of peak river flow obtained by applying hydrological similarity principles and a regional method. The results show that the flood flow estimated through the application of the distributed model is consistent with the estimate provided by the regional procedure as well as the behaviors of the river banks. Conversely, the method based on hydrological similarity delivers an estimate that seems to be not as

reliable. The analysis highlights interesting perspectives for the application of spatially distributed models to ungauged catchments.

### 1 Introduction

Estimation of the peak discharge for an assigned probability of exceedance (the design flood; NERC, 1975) is frequently carried out in applied hydrology to design flood control measures. In the case of ungauged or scarcely gauged catchments this is not an easy task (Jothityangkoon and Sivapalan, 2003). In fact, the limited availability of historical peak flow data over a sufficiently long observation period induces a relevant uncertainty in the flood frequency analysis. A possible solution to overcome this problem is the application of hydrological simulation studies in order to generate long synthetic river flow series. These may allow one to perform the flood frequency analysis on an extended data base so that uncertainty is in principle reduced. If this method is applied, to verify the representativity of the synthetic data is a relevant issue.

Several applications of rainfall-runoff models in the framework of continuous simulations to estimate the flood frequency distribution have been proposed by the scientific literature (Naden et al., 1996; Lamb, 1999; Cameron et al., 1999, 2000a,b; Blazkova and Beven, 1997, 2002). The simulation approach is based on the use of rainfall-runoff models of various complexity for transforming a precipitation records in river discharge. The aim is to generate a synthetic series of peak river flows that are used for inferring the flood frequency distribution at a given site. This approach presents many advantages (Moretti and Montanari, 2004). Spatially distributed rainfall-runoff models, that are potentially able to



Correspondence to: A. Montanari  
(alberto.montanari@unibo.it)

produce synthetic river flows at any locations of the river network, can be a valuable tool in the context of the simulation approach. However, it is still not fully clear to what extent these models are able to simulate the river flow over a wide range of spatial scales (Brath et al., 2004).

This issue is investigated here by applying a continuous simulation, spatially distributed rainfall-runoff model to the 1294 km<sup>2</sup> gauged Secchia River basin, located in northern Italy, for which discharge data are available at the outlet and some internal cross sections. The model has been parameterised by using hourly rainfall over the basin and discharge records at the basin outlet. Then, the model has been validated by simulating a 100-year long sequence of hourly river flows and then comparing observed and simulated frequency distributions of annual peak flows. The above comparison was carried out for both the basin outlet and the internal cross section of Cavola Bridge, where the drainage area is 337 km<sup>2</sup>. This site was treated as ungauged. Finally, the model has been used to estimate the peak flow for a given probability of exceedance (or return period) for the ungauged Riarbero River, a tributary of the Secchia River. The result has been compared with estimates of the peak river flow for the Riarbero River obtained by applying a classical procedure based on hydrological similarity principles and a regional method for flood frequency estimation.

The model herein used is AFFDEF (Moretti and Montanari, 2007) which has been shown to be parameter parsimonious and not computer intensive. Therefore it is suited for scarcely gauged basins and for performing long term continuous simulations. The results of the flood frequency estimation for the Riarbero River are promising and indicate AFFDEF as a valuable tool for inferring the hydrological regime of ungauged catchments.

## 2 Study region and data

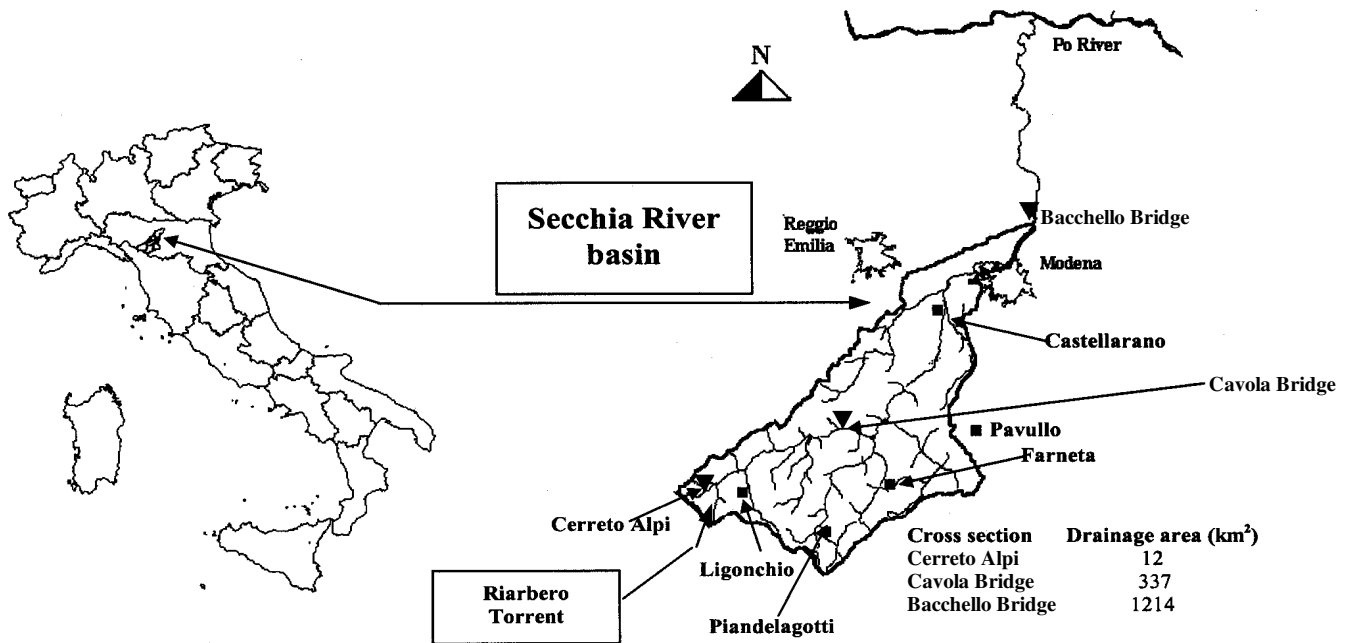
The simulation approach herein proposed was applied in order to infer the shape of the flood frequency distribution of the Riarbero River, which is a right tributary to the Secchia River, that is located in northern Italy (see Fig. 1). The Secchia River flows northwards in the Apennine Mountains and is a right tributary to the Po River. The contributing area at the Bacchello Bridge river cross section, which is located in the vicinity of the basin outlet, about 62 km upstream the confluence in the Po River, is 1294 km<sup>2</sup>. The maximum altitude is the Mount Cusna, which is 2121 m above sea level (a.s.l.) high, while the main stream length up to Bacchello Bridge is about 98 km. The mean annual rainfall depth over the basin ranges between 700 and more than 2000 mm/year. Intense meteoric events can occur in each season, even though the most critical precipitations typically occur in Autumn. The peak flow observed at Bacchello Bridge in the period 1923–1981 is 823 m<sup>3</sup> s<sup>-1</sup> (20 April 1960).

The Secchia river basin is monitored by raingauges and hydrometers managed by the National Hydrographic Service of Italy. For the purposes of the present analysis, historical hourly river discharges at Bacchello Bridge for the year 1972 have been collected. In addition, annual peak flows are available at Bacchello Bridge and Cavola Bridge (55 and 16 values, respectively). The drainage area at Cavola Bridge is 337 km<sup>2</sup>. Annual peak discharges are at disposal as well for the period 1955–1960 (6 observations) at Cerreto Alpi, where the drainage area of the Secchia River is 12 km<sup>2</sup>. All the river flow measurements were collected by observing the river stage, which is then converted to river discharge by means of rating curves derived on the basis of flow velocity measures and field surveys of the river cross section geometry.

Hourly temperature data and hourly rainfall depths over the Secchia River basin for the years 1972 and 1973 have been gathered for five gauging stations located in the basin area or in the vicinity. Depth duration frequency curves for rainfall are also available for the five raingauges, for storm duration ranging from 1 to 24 h and return period ranging from 20 to 100 years. They were derived by fitting long records of annual maximum rainfall depths for different storm durations. The topography of the Secchia River basin is described by a Digital Elevation Model (DEM), whose resolution is 250×250 m. An extensive data base of soil type and soil use at local scale is at disposal, retrieved from recent surveys.

The Riarbero River originates from the slopes of Mount Ischia (1724 m a.s.l.) and Mount Scalocchi (1727 m a.s.l.) in the Apennine Mountains and flows in a narrow valley, joining the right bank of the Secchia River in the neighbourhood of Collagna, few kilometres downstream of Cerreto Alpi. The main stream length is 7 km, while the drainage area is 17 km<sup>2</sup>. The drainage basin has an average altitude of 1360 m a.s.l. and is mostly covered by high altitude meadows and woods. The Riarbero River is located within a national park and is immersed in a natural environment of significant value.

Estimation of the peak river flow for the Riarbero River is an important issue as there is the need to design erosion control works along the main stream. Being the Riarbero River ungauged, it is usual practice for professional hydrologists in Italy to estimate the required design flow on the basis of hydrological similarity principles. This technique is briefly described in Sect. 3.1 and provided an estimate of the peak river flow which may be not reliable, being extremely high (see Sect. 3.1). Therefore, the regional procedure described in Sect. 3.2 and the river flow simulation study described in Sect. 3.3 were implemented to derive further indications about the magnitude of the peak river flows.



**Fig. 1.** Ubication and map of the Secchia River basin. Black squares and triangles indicate the position of the raingauges and hydrometers, respectively.

### 3 Description of the analysis

The estimation of the peak river flow for the Riarbero River was first of all performed by applying a traditional procedure based on hydrological similarity principles. Subsequently, a regional flood frequency estimation method was applied and, finally, the simulation study was performed, which is structured in the following steps: a) calibration and validation of the spatially distributed rainfall-runoff model for the whole Secchia River basin by using the historical rainfall data and the river flow record referred to Bacchello Bridge. b) Calibration of a rainfall simulation model and generation of a 100-year long record of hourly rainfall data over the basin. These were routed through AFFDEF to obtain synthetic 100-year long records of hourly river flows at Bacchello Bridge and Cavola Bridge, in order to check the simulation reliability by comparing observed and simulated flood frequency distributions. c) Generation of a 100 year long record of hourly rainfall data and river flows for the Riarbero River basin, in order to infer the shape of the unknown flood frequency distribution.

#### 3.1 Estimation of the peak river flow for the Riarbero River through hydrological similarity

It was mentioned above that it is usual practice in Italy to apply hydrological similarity principles for estimating the peak river flow in ungauged catchments. In particular, the peak river flow,  $q(T)$ , per unit catchment area (specific

peak flow) and return period  $T$ , for a catchment of area  $A$ , is estimated through the following monomial relationship (Maione, 1995):

$$q(T) = q'(T) \left( \frac{A}{A'} \right)^{-2/3}, \quad (1)$$

where  $q'(T)$  is the specific peak flow for return period  $T$  in a catchment of area  $A'$ , which is hydrologically similar to the study catchment. The exponent  $-2/3$  in Eq. (1) was estimated by using historical flood data for many regions of Italy, that prevalently refer to small and medium size river basins where small scale processes play a significant role in the formation of the flood flows.

$q'(T)$  was estimated here by fitting the observed record of the Secchia River annual peak flow at Cerreto Alpi (see Sect. 2) with a type I extreme value probability distribution (Gumbel distribution). It should be noted that such record comprises only 6 observations and therefore the derived estimate for  $q'(T)$  may be affected by a relevant uncertainty, especially for high return periods. Moreover, by looking at the Cerreto Alpi flow record, one notes the presence of two very high values, around  $250 \text{ m}^3/\text{s}$  for a watershed area of  $12 \text{ km}^2$ . These values are unusual for the geographical context and therefore one may suspect that the rating curve is unreliable when extrapolated in the high flow domain.

Applying Eq. (1) as described above in order to estimate the peak flow for the Riarbero River, for a return period of 20 years, one obtains a value of  $295 \text{ m}^3/\text{s}$ , that would be the

design flow suggested by the hydrological similarity procedure. This is a very high flow value, which is likely to be affected by a relevant uncertainty since it implies an extension of the wetted perimeter of the river cross section that appears to be not consistent with the vegetation cover along the main stream.

### 3.2 Estimation of the peak river flow for the Riarbero River through regionalisation of peak flows

In view of the result obtained by applying hydrological similarity, the peak river flow for the Riarbero River was also estimated by applying a regional procedure. The Secchia River basin is included in an extended area for which a regional flood frequency estimation method was developed by Franchini and Galeati (1996) and Brath et al. (2001). In detail, for a given site the peak flow  $Q(T)$  for return period  $T$  can be estimated according to the relationship

$$Q(T) = \mu \cdot x(T) \quad (2)$$

where  $\mu$  is the index flood, which corresponds to the mean value of the annual maximum peak flow, and  $x(T)$  is a  $T$ -year return period dimensionless quantile, which is assumed to be invariant over the homogenous geographical regions that are preliminarily identified for the study area. When no observations of the annual maximum flood are available for the site of interest, computation of the index flood  $\mu$  can be carried out through the relationship

$$\mu = 2.797 \times 10^{-5} A^{1.235} \mu_{H,1}^{3.513} L^{-0.720} \quad (3)$$

where  $A$  and  $L$  are the catchment area and main stream length, respectively, and  $\mu_{H,1}$  is the mean value of the annual maximum rainfall depth for storm duration of 1 h for which local values are proposed by Brath et al. (2001).  $x(T)$  can be estimated as the  $T$ -year quantile of the Two Component Extreme Value Distribution (TCEV, Castellarin, 2000), whose parameters for the study region are provided by Franchini and Galeati (1996).

By applying the relationship (2) and (3) to the Riarbero River one obtains an estimate of 123 m<sup>3</sup>/s for the 20-year return period peak flow. Therefore, the regional estimate is considerably lower than the value derived through the hydrological similarity (see Sect. 3.1) which, once again, appears to be affected by a significant uncertainty.

### 3.3 Estimation of the peak river flow for the Riarbero River through hydrological simulation

#### 3.3.1 The rainfall-runoff model

The rainfall-runoff model applied here is AFFDEF (Moretti and Montanari, 2007). Basically, it was developed in order to satisfy the following general requirements: (a) the model should be reliable in making predictions for ungauged or scarcely gauged catchments or where little information about

the contributing area is available; (b) the model should allow a spatially distributed description of the geomorphological characteristics of the catchment in order to generate river flow data at any cross section of the river network; (c) the model should have some physical basis in order to constrain the range of some parameters by means of in situ measurements or physical reasoning and in order to decrease parameter uncertainty; (d) the model should be computationally inexpensive in such a way that long simulation runs could be performed at short time steps in a reasonably limited time, even for medium size basins.

Therefore, AFFDEF was conceived in order to be applicable to a wide spectrum of real world case studies, even when only a limited data base of hydrometeorological and geomorphological records is available. It is a continuous simulation and conceptual approach, which is primarily in charge of providing a sufficiently reliable reproduction of the peak flows. Some of the hydrological processes involved in the rainfall-runoff transformation have been schematised by using conceptual schemes. These need to be parameterised on the basis of some historical hydrometeorological records and therefore the need for observed data is not eliminated. However extensive analyses (Brath et al., 2001) proved the efficiency and robustness of the model when applied to data limited catchments, especially in comparison with lumped approaches. Such efficiency is believed to be due to the capability of the model to take advantage from the spatially distributed description of the basin topography, soil type and use.

The catchment hydrologic response is determined by the composition of the two processes of hillslope runoff and channel propagation along the river network. The model discretises the basin in square cells coinciding with the pixels of the DEM. The river network is automatically extracted from the DEM itself by applying the D-8 method (Band, 1986; Tarboton, 1997), which allows one to estimate the flow paths and the contributing area to each cell. In detail, the network determination is carried out by first assigning to each DEM cell a maximum slope pointer and then processing each cell in order to organise the river network. Digital pits are filled in a preprocessing step, before extracting the channel network from the DEM of the catchment. Each cell receives water from its upslope neighbours and discharges to its downslope neighbour. For cells of flow convergence, the upstream in-flow hydrograph is taken as the sum of the outflow hydrographs of the neighbouring upslope cells.

Distinction between hillslope rill and network channel is based on the concept of constant critical support area (Montgomery and Foufoula-Georgiou, 1993). Accordingly, rill flow is assumed to occur in each cell where the upstream drainage area does not exceed the value of the critical support area  $A_0$ , while channel flow occurs otherwise.

The interaction between soil, vegetation and atmosphere is modelled by applying a conceptual approach. The model firstly computes the local gross rainfall  $P_l [t, (i, j)]$ , for each

DEM cell of coordinates  $(i, j)$ , by interpolating the observations referred to each rain gauge through an inverse distance approach. Then, for each cell a first rate of rainfall depth is accumulated in a local reservoir (interception reservoir) which simulates the interception operated by the vegetation. The capacity of such interception reservoir is equal to  $C_{int} S(i, j)$ , being  $C_{int}$  a parameter, constant in space and time, and  $S(i, j)$  the local storativity. The latter is computed depending on soil type and use accordingly to the Curve Number method (CN method, Soil Conservation Service, 1987; Chow et al., 1988).

Once the interception reservoir is full of water, the exceeding rainfall reaches the ground. Then, surface and sub surface flows are computed accordingly to a modified CN approach that is able to simulate the redistribution of the soil water content during interstorm periods. In detail, it is assumed that a linear reservoir (infiltration reservoir), which collects the infiltrated water, is located in correspondence of each DEM cell at the soil level. The local surface runoff and the infiltration are computed according to the relationship

$$\frac{P_n[t, (i, j)]}{P[t, (i, j)]} = \frac{F[t, (i, j)]}{H \cdot S(i, j)} \quad (4)$$

where  $P[t, (i, j)]$  is the rainfall intensity that reaches the ground at time  $t$ ,  $P_n[t, (i, j)]$  is the intensity of surface runoff,  $F[t, (i, j)]$  is the water content at time  $t$  of the infiltration reservoir located in correspondence of the cell  $(i, j)$ , and  $H \cdot S(i, j)$  is the capacity of the infiltration reservoir itself, computed by multiplying the calibration parameter  $H$  by the soil storativity previously introduced. The quantity  $I[t, (i, j)] = P[t, (i, j)] - P_n[t, (i, j)]$  represents the intensity of the infiltrated water. The outflow  $W[t, (i, j)]$  from the infiltration reservoir to the sub surface river network, which is assumed to coincide with the surface one, is given by the linear relationship

$$W[t, (i, j)] = \frac{F[t, (i, j)]}{H_S} \quad (5)$$

where  $H_S$  is a calibration parameter.  $H$  and  $H_S$  are assumed to be constant with respect to both space and time.

The hourly intensity of potential evapotranspiration  $E_P[t, (i, j)]$  is computed at local scale by applying the radiation method (Doorenbos et al., 1984). When some water is stored in the interception reservoir, the effective evapotranspiration  $E[t, (i, j)]$  is assumed to be equal to  $E_P[t, (i, j)]$  and is subtracted from the water content of the interception reservoir itself. When this latter is empty, or is emptied while subtracting the evapotranspiration rate, the remaining part of  $E_P[t, (i, j)]$  is subtracted from the water content of the infiltration reservoir. In this case, it is assumed that  $E[t, (i, j)]$  is varying linearly from 0 when  $F[t, (i, j)] = 0$ , to  $E_P[t, (i, j)]$  when  $F[t, (i, j)] = H \cdot S(i, j)$ . Evapotranspiration is the only source of water losses in the model, which primarily depends on the capacity of the interception reservoir and hence on the parameter  $C_{int}$ . Therefore a first

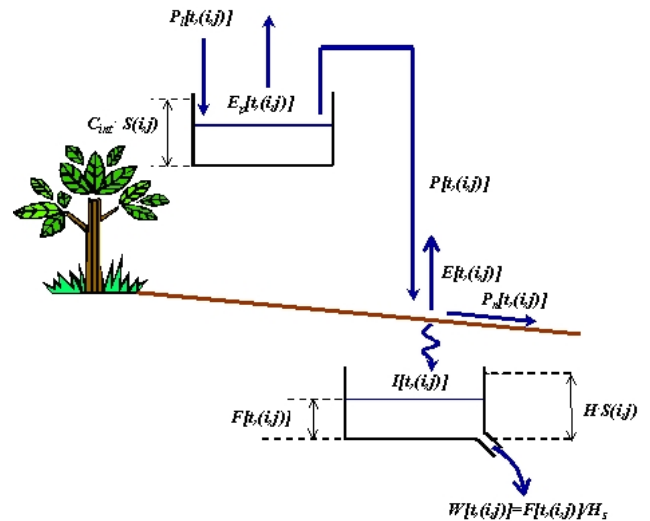


Fig. 2. Schematisation operated by the AFFDEF model of the interaction among soil, vegetation and atmosphere.

estimation of  $C_{int}$  can be obtained by comparing observed and simulated runoff coefficients. Figure 2 shows the scheme of the interaction among soil, vegetation and atmosphere operated by the model.

The continuity equation applied to the infiltration reservoir can be written as:

$$I[t, (i, j)] - W[t, (i, j)] = \frac{dF[t, (i, j)]}{dt} \quad (6)$$

By combining Eqs. (4), (5) and (6) and taking the effective evapotranspiration into account, the mass balance equation for the infiltration reservoir can be derived and written in the form

$$\frac{dF[t, (i, j)]}{dt} = -\frac{F[t, (i, j)]}{H_S} - E[t, (i, j)] + P[t, (i, j)] \left\{ 1 - \frac{F[t, (i, j)]}{H \cdot S(i, j)} \right\} \quad (7)$$

Surface and sub surface flows are propagated towards the basin outlet by applying the variable parameters Muskingum-Cunge model. Extensive details can be found in Cunge (1969) and Orlandini et al. (1999) for surface and sub surface propagation, respectively. For the surface flow, the kinematic celerity is computed by considering rectangular river cross section with fixed width/height ratio. The latter parameter and the channel roughness can assume different values along the river network and on the hillslopes. In particular, the channel roughness in the river network is allowed to vary from a minimum to a maximum value depending on the contributing area. For the subsurface flows, the kinematic celerity is instead computed as a function of the saturated hydraulic conductivity of the soil.

It is interesting to note that the model describes in a simplified manner the dynamics of the sub surface flows. In

**Table 1.** Rainfall-runoff model parameters and their values. These latter were partly optimised by calibration (C) and partly estimated by in situ measurements or physical reasoning (E).

Strickler roughness for the hillslopes	$k_{sv}$ [ $\text{m}^{1/3} \text{s}^{-1}$ ]	C	0.906
Channel width/height ratio for the hillslopes	$w_v$ [-]	C	500 000
Maximum and minimum Strickler roughness for the channel network	$k_{sr}^0, k_{sr}^1$ [ $\text{m}^{1/3} \text{s}^{-1}$ ]	E	10-22
Channel width/height ratio for the channel network	$w_r$ [-]	E	20
Constant critical source area	$A_0$ [ $\text{km}^2$ ]	E	1
Saturated hydraulic conductivity	$K_i$ [ $\text{m s}^{-1}$ ]	C	0.05
Width of the rectangular cross section of the sub surface water flow	$B_p$ [m]	C	0.5
Bottom discharge parameter for the infiltration reservoir capacity	$H_S$ [s]	C	5 000 000
Multiplying parameter for the infiltration reservoir capacity	$H$ [-]	C	0.1092
Multiplying parameter for the interception reservoir capacity	$C_{\text{int}}$ [-]	C	0.1

particular, it does not distinguish between near surface and deep water flow, and assumes that the calibration parameters  $H$  and  $H_S$  are constant with respect to both space and time. This simplified description has been used in order to reduce the number of model parameters and, consequently, the amount of historical data required for model calibration. On the other hand, one may expect a significant approximation in the simulation of the low river discharges, especially when referring to highly permeable basins.

Moreover, the formation of the surface runoff is modelled accordingly to a scheme that is very similar to the one adopted by the CN method, which is considered by many authors as an infiltration excess approach (Beven, 2000). Therefore one may expect that the proposed model is better suited for basins characterised by low permeability and prevalently impervious hillslope, where the surface runoff is more likely to be given by excess of infiltration instead of excess of saturation.

### 3.3.2 Application of the rainfall-runoff model to the Secchia River basin

AFFDEF has been calibrated for the Secchia River basin by using historical rainfall data and the river flow record observed at Bacchello Bridge. Most of the model parameters have a well defined physical meaning and were estimated on the basis of in situ surveys. However, it was necessary to optimise some of them by means of a trial and error procedure through a manual calibration which has been performed by comparing observed and simulated river flows for the year 1972.

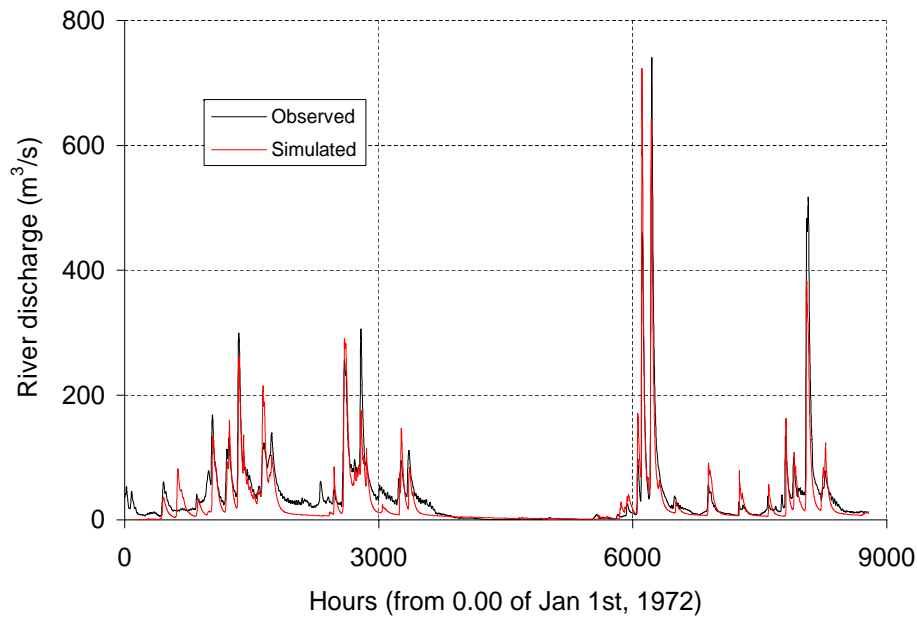
Table 1 reports a list of the model parameters and indicates which ones were estimated by in situ measurements or physical reasoning and which ones were instead derived by man-

ual calibration. Figure 3 shows the observed and simulated hydrographs for the year 1972 at Bacchello Bridge. A dispersion diagram of observed versus simulated 1972 hourly flows is reported in Fig. 4. The efficiency and explained variance of the simulation are 0.81 and 0.83, respectively.

Figure 5 shows a detail of the observed and simulated hydrographs for two high flow events occurred in the year 1972 in two different seasons, precisely in the months of February (event 1, left panel) and September (event 2, right panel). Event 1 was a moderate one, while event 2 was an extreme flood with a return period of about 100 years. This latter was originated by a long rainstorm that lasted about one week and provoked a series of floods. The flood volume was simulated by the model with an underestimation of 15% for event 1 and 5% for event 2. The peak river flow was underestimated by about 15% for both events, although one may note the the first peak was overestimated in both the events. The time to peak was well simulated in event 1, while it was anticipated by 3 h in event 2. A detailed analysis of the sensitivity of the model performances was carried out, showing that the error in the representation of the spatial distribution of rainfall can well explain the magnitude of the model error depicted by the above statistics.

### 3.3.3 Generation of a long sequence of river flow data for the Secchia River

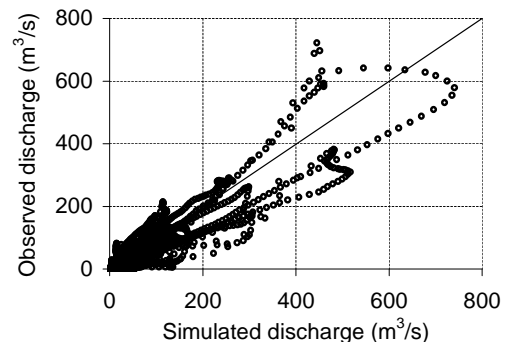
The goodness of the fit provided by the model has been validated by simulating a 100-year hourly sequence of river flows at the Secchia outlet and comparing the observed and simulated frequency distributions of the annual peak flows. The same comparison has been performed for the Secchia River at Cavola Bridge, where 16 observed annual peak flows are at disposal.



**Fig. 3.** Observed versus simulated hourly hydrograph at the Bacchello Bridge river cross section; year 1972.

For performing the hydrological simulation, a 100-year multisite hourly rainfall record was firstly generated, by using the multivariate Neyman-Scott rectangular pulses model, which represents the total rainfall intensity at time  $t$  as the sum of the intensities given by a random sequence of rain cells active at time  $t$ . Extensive details can be found in Cowpertwait (1996). The model is a generalisation of the well known single site Neyman-Scott rectangular pulses model (Rodriguez-Iturbe et al., 1987; Burlando, 1989) and is capable of generating a synthetic multisite record referred to the five raingauges displaced over the basin and considered within this study. The rainfall model has been parameterised by using the method of moments for fitting selected statistics of the hourly rainfall records observed in the years 1972 and 1973 in the five raingauges. Estimated parameters have been adjusted through a trial and error procedure in order to match the simulated mean areal Depth Duration Frequency (DDF) curve for rainfall over the basin, estimated for a return period of 50 years, with the corresponding DDF curve computed by fitting extreme rainfall data collected for storm duration ranging from 1 to 24 h in the above five raingauges. The synthetic rainfall record for the Secchia River basin will be referred to with the symbol  $P_l^*(t, s)$ , where the subscript  $l$  stands for gross rainfall, the subscript  $*$  indicates that the record is synthetic,  $t$  is time and  $s=1, \dots, 5$  indicates the raingauge.

Then, a 100-year record of hourly temperature data was generated by using a stochastic model, namely, a fractionally differenced ARIMA model (FARIMA). This kind of model has been shown in many applications to be able to well fit

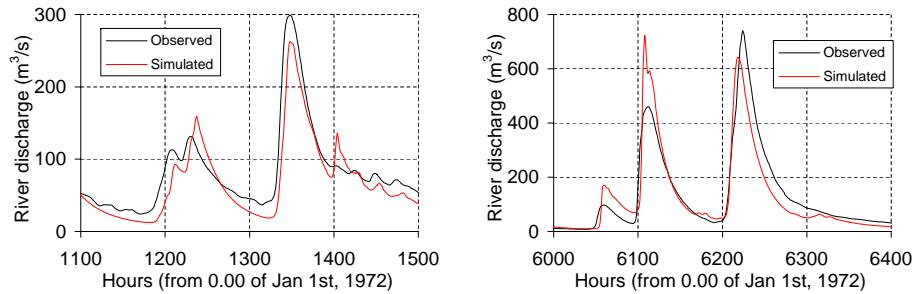


**Fig. 4.** Dispersion diagram of observed versus simulated hourly discharges at the Bacchello Bridge river cross section; year 1972.

the autocorrelation structure of temperature series which, for increasing lag, is very often affected by a slow decay that may suggest the presence of long term persistence. FARIMA models, which are characterised by a high flexibility in their autocorrelation structure, are capable of fitting long term persistence by means of the fractional differencing operator. More details on FARIMA models and the simulation procedure herein applied for the temperature data can be found in Montanari (2003) and Montanari et al. (1997).

The synthetic rainfall and temperature data have been subsequently routed through the calibrated rainfall-runoff model, therefore obtaining a 100-year long sequence of river flows referred to Bacchello Bridge, Cavola Bridge and Cerreto Alpi. A first check of the reliability of the obtained





**Fig. 5.** Observed versus simulated hourly hydrograph at the Bacchello Bridge river cross section for two high flow events occurred in the year 1972.

simulation was performed by referring to the Bacchello Bridge River flows. In detail, the long term runoff coefficient,  $C$ , for the simulated record was computed by obtaining  $C=0.61$ . The  $C$  value estimated by the National Hydrographic Service of Italy by using long records of daily river flows observed at Bacchello Bridge is equal to 0.48, while in the year 1972 the occurrence of severe floods induced a value  $C=0.75$ . If one considers that the hydrological model was calibrated by using data observed in 1972, it can be concluded that a satisfactory agreement is obtained. Moreover, for each annual maximum flood we computed the event runoff coefficient, therefore obtaining an average value of 0.94 with a standard deviation of 0.22. These statistics are reasonable, even if one may note that the computation of the event runoff coefficient is affected by a significant uncertainty, which is mainly induced by the separation of the event water from the pre-event base flow.

Furthermore, 100 values of annual peak flow for Bacchello Bridge, Cavola Bridge and Cerreto Alpi were extracted and compared with the actual observations. The comparison between observed and simulated sample frequency distributions of the annual peak flows at Bacchello Bridge and Cavola Bridge is shown in Fig. 6. It can be seen that the model operates a slight overestimation of the peaks. In detail, the mean value of the annual peak flow is overestimated of about 20% at both the river cross sections. However this is not detrimental for the purpose of the present analysis, since it leads to a precautionary estimation of the design flood. This approximation might be due to the uncertainty in the generation of synthetic data. In particular it might be ascribed to the limited number of raingauges available, which prevents an accurate representation of the spatial variability of rainfall. However, the frequency plots show a satisfactory fit of the sample frequency distributions.

We believe the result obtained for the Cavola Bridge cross section is particularly significant. In fact, given that the Secchia River was treated as ungauged at Cavola Bridge, and in view of the significant reduction of the contributing area from the calibration to the validations sites (1294 and 337 km<sup>2</sup>, respectively), the outcome therein obtained provides a signifi-

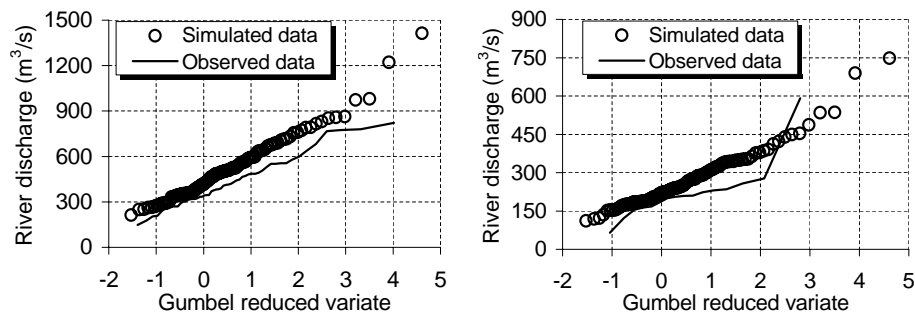
cant support to the capability of AFFDEF to effectively simulate the peak river flow regime in internal river cross sections. This result provides some support to the application of the simulation procedure for estimating the design flood of the Riarbero River.

However, a different outcome is obtained for the simulation at Cerreto Alpi. In fact, here the highest annual peak flow simulated by the rainfall-runoff model is 106.1 m<sup>3</sup>/s, while the mean value of the 6 annual maxima observed from 1955 to 1960 is 131.6 m<sup>3</sup>/s. It is clear that in this case the river flows simulated by the hydrological model are considerably lower than the observations. Therefore one may suspect that the reliability of the hydrological model is not confirmed for small spatial scales. One reason for this possible unreliability could be that rainfall is not well simulated by the stochastic model at local scale (see the correction at the rainfall simulation introduced in Sect. 3.3.4). On the other hand, it was already mentioned in Sect. 3.1 that the peak discharges observed at Cerreto alpi are unusually high for the geographical context. This unusual behaviour was the motivation for using a spatially distributed rainfall-runoff model for better inspecting the spatial distribution of the Secchia River flows. A refined reliability analysis for the output of the hydrological model with respect to the situation depicted by the observations collected at Cerreto Alpi is carried out in Sect. 3.3.4 by also taking into account the results obtained with the regional peak flow estimation.

### 3.3.4 Application of the simulation procedure to the Riarbero River and uncertainty analysis

Given the small size of the Riarbero River, a key point of the simulation study is the generation of the rainfall record. In fact, a good representation of rainfall at the spatial scale of the whole Secchia basin does not assure that rainfall is well reproduced at local scale as well. In particular, when the focus of the analysis is restricted to the Riarbero River, one may note that the nearest of the five raingauges used for calibrating the rainfall model is located in Ligonchio and therefore may not provide a good representation of the rainfall regime





**Fig. 6.** Sample frequency distribution of the annual peak flows at the Bacchello Bridge (left) and Cavola Bridge (right) river cross sections.

over the Riarbero watershed. In fact, Ligonchio is located at a much lower altitude (about 930 m a.s.l., against an average altitude of the Riarbero River basin of 1360 m a.s.l.).

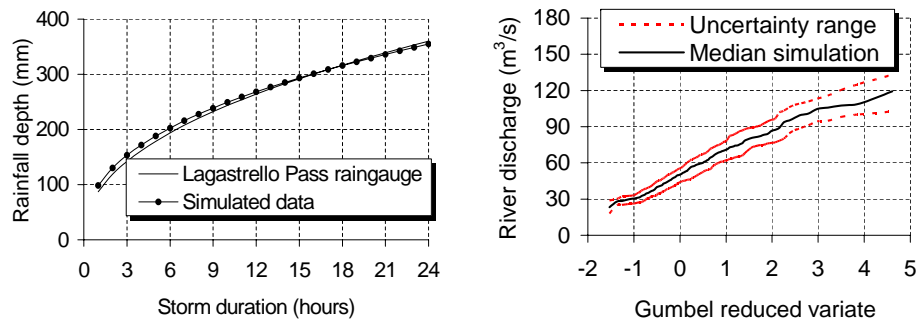
In order to provide a better fit of the local climatic forcing, the parameters of the rainfall model were adjusted through a trial and error procedure, by matching the simulated DDF curve referred to the Riarbero watershed with the DDF curve of the Lagastrello Pass raingauge. This latter is located outside the Secchia River basin at an altitude of about 1200 m a.s.l. and is the highest rainfall measuring station placed in the vicinity of the Riarbero watershed. Figure 7 shows a comparison between simulated and observed DDF curves at the Lagastrello Pass for a return period of 50 years after the above adjustment. 100 years of hourly rainfall were subsequently generated. This record will be referred to with the symbol  $P_l^*(t, R)$  where  $R$  stands for the location of the Riarbero watershed.

In order to account for the presence of uncertainty in the simulation procedure, the generation of synthetic river flows for the Riarbero River was done by allowing some of the parameters of the rainfall-runoff model to vary within a feasible range, therefore applying the Generalise Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992). Accordingly to this procedure, we rejected the concept of an optimal parameter set for AFFDEF, given that several parameter combinations may provide equally likely simulations in the presence of uncertainty (Cameron et al., 2000a).

It is important to note that GLUE does not provide a statistically based uncertainty estimation. By allowing the parameter set of the rainfall-runoff model to change, GLUE implicitly accounts for the presence of uncertainty in the modelling process, therefore evaluating the approximation in the results that might be induced by model misspecification. This type of uncertainty assessment is subjective (for an extensive discussion see Montanari, 2005, 2007, and Beven, 2006). Therefore it is necessary to make clear the underlying assumptions so that GLUE can provide the user with an indication of the variability of the response depending on the rainfall-runoff model uncertainty and sensitivity.

In detail, the GLUE simulation was performed as follows. 3000 parameter sets for AFFDEF were generated by allowing the parameters  $H$ ,  $H_S$ ,  $k_{sv}$  and  $w_v$  to vary uniformly in the range  $\pm 20\%$  of their optimal values calibrated for the Secchia River and shown in Table 1. These parameters were selected as they are the most effective on the AFFDEF response (Moretti and Montanari, 2007). For each of the sets, the hourly synthetic rainfall data set  $P_l^*(t, s)$ ,  $s=1, \dots, 5$ , previously generated for the Secchia River basin was routed through AFFDEF therefore obtaining a 100-year long simulation of hourly flows at Bacchello Bridge. The annual peak flows were extracted and a sample frequency distribution was derived by plotting them against the related Gumbel reduced variate (GRV), therefore obtaining a diagram like Fig. 6 (left) for each parameter set. Then, the simulated peak flow corresponding to the GRVs of the observed record was estimated. Given that the GRVs of observed and simulated records do not coincide, linear interpolation was used to identify the simulated river flows corresponding to the GRVs of the observed data. Then, the Nash efficiency of the simulated peak flows was computed by comparing them with the observed value corresponding to the same GRV. Those parameter sets that did not provide a minimum efficiency of 0.50 were rejected as non behavioural. At the end of this simulation procedure for the Secchia River, 565 sets were retained as behavioural. The efficiency  $E_j$ ,  $j=1, \dots, 565$ , of each behavioural simulation were rescaled, in order to make  $\sum_{j=1}^{565} E_j = 1$ , and associated to the related parent set.

Then, AFFDEF was run for the Riarbero River by using in turn each of the behavioral parameter sets, by utilising as input the 100-year synthetic local climatic forcing  $P_l^*(t, R)$  and the synthetic temperature series generated as described in Sect. 3.3.3. Therefore, 565 peak flows were obtained for each year of the simulation period. A probability distribution was then constructed for the peak flow of each year as follows: the 565 annual maxima for year  $t$  were ranked in ascending order. The rescaled efficiency  $E_j$  of the parent parameter set was associated to each value. Finally, an uncertainty range and a median simulation for the annual maximum flow of year  $t$  were identified by selecting the river flow



**Fig. 7.** Comparison of the 50-year return period DDF curves derived by simulated rainfall over the Riarbero catchment and rainfall extremes observed at Lagastrello Pass (left) and median simulation with uncertainty range for the frequency distribution of the annual peak flows for the Riarbero River (right).

values corresponding to a cumulative rescaled efficiency of 0.025, 0.975 and 0.5, respectively. The frequency distributions for the median simulation and uncertainty range for the whole simulation period are shown in Fig. 7. By fitting a type 1 extreme value probability distribution to the median simulation one obtains an estimate of about  $102 \text{ m}^3/\text{s}$  for the 20-year return period peak flow.

One should note that this type of GLUE approach does not account explicitly for the presence of uncertainty in the rainfall and temperature data. Input uncertainty could be explicitly evaluated by routing through AFFDEF many synthetic rainfall and temperature series for different parameter sets of the respective generating models. However, we believe such procedure, which is extremely computer intensive, is not strictly necessary as any contribution to the formation of the global uncertainty is implicitly accounted for by GLUE. As a matter of fact, it is the threshold value of the Nash efficiency, which is selected to identify the behavioural simulations, that determines the width of the uncertainty range. Therefore, the user should select such threshold value accordingly to expert knowledge, in order to effectively identify what is a reasonable performance of a behavioural model. In the presence of limited uncertainty the user should select a high threshold efficiency, while he should be much more tolerant when there is the feeling that uncertainty is relevant. In the context of the present study, in consideration of the limited availability of historical data, it was deemed that an efficiency of 0.5 could be a reasonable threshold value, that allows GLUE to implicitly account for any source of uncertainty.

### 3.4 Is the hydrological simulation satisfactory for the right reason?

In recent times, hydrologists are becoming more concerned about the physical fundament of hydrological models. In fact, it is deemed increasingly important to check that the model is well working for the right reasons (Kirchner, 2006). This is not an obvious task as in real world applications the

interaction among many types of uncertainties within a hydrological simulation may induce compensation among errors. For instance, it is well known that parameters may compensate for an erroneous input or a wrong model structure. In such a case, if the user only looks at the goodness of the fit of observed data he may erroneously conclude that the model is good. However, if the model is not correctly schematising the underlying physical process its performances might abruptly decrease if it is applied to ungauged/scarcely gauged basins. In fact, the limited data availability might not allow the user to tune the parameter values to compensate for the inherent uncertainties. Indeed, such abrupt decrease of performance in non ideal situations is frequently experienced in hydrology. Getting the right answer for the right reasons is compelling in order to improve our ability to understand, schematise and effectively model hydrological processes. The concern about the possibility that a hydrological model might work well for the wrong reason should arise any time that multiple sources of uncertainties are potentially interacting.

Indeed, many types of uncertainty are involved in the case study considered herein. The uncertainty in the observed and synthetic rainfall and temperature data may interact with the uncertainty in the AFFDEF structure and all the others uncertainties related to the AFFDEF input data (for instance, digital elevation model and CN values). These uncertainties could be compensated by the AFFDEF parameters, therefore leading to a typical situation where the results presented here could be satisfactory for the wrong reason.

It is well known that to evaluate the goodness of the schematisation operated by a mathematical model of complex hydrological processes is not an easy task. For the case of multistep simulations like the one we performed here it is advisable to check the goodness of the results not only at the end of the simulation itself, but also during the intermediate steps. Other important indications may be derived by using independent sources of information to check the model accuracy (orthogonal observations, Winsemius et al., 2006) at the intermediate and final steps of the simulation.

For the case of the simulation performed here, we believe all the available information was used to check the consistency of the approach at different steps. First, the accuracy of AFFDEF in performing continuous time simulation of hourly flows at the Secchia River basin outlet was checked in calibration mode. Second, the capability of the Neyman-Scott model to reproduce the statistics of the observed hourly precipitations in the five raingauges was assessed again in calibration mode. Third, an independent information, that is, the depth duration frequency curves for rainfall, was used to improve the fit provided by the rainfall simulation. Furthermore, another independent information, that is, the annual peak flows observed at the Secchia outlet and at the intermediate section of Cavola Bridge, allowed us to check the capability of AFFDEF to reproduce the frequency behaviours of peak flows at different spatial scales in validation mode. Moreover, a soft information was used to check the reliability of the final estimate of the peak flow for the Riarbero River. In fact, the obtained peak flow was critically analysed in view of the status of the vegetation along the main stream. Finally, the AFFDEF parameter values were carefully checked in view of their consistency with the physical processes they represent and the values they assumed in previous applications.

We believe that all the above considerations provide substantial support to the credibility of the overall simulation procedure.

#### 4 Conclusions

The design flood of  $102 \text{ m}^3/\text{s}$  obtained through hydrological simulation for the Riarbero River is close to the estimate of  $123 \text{ m}^3/\text{s}$  obtained with the regional method for flood frequency analysis. On the other hand, it is much smaller than the value of  $295 \text{ m}^3/\text{s}$  provided by the hydrological similarity procedure involving the Riarbero River at its outlet and the Secchia River basin at the Cerreto Alpi cross section. In view of the robustness of the model simulations proved at the cross sections of Bacchello Bridge and Cavola Bridge, and in view that the estimate provided by the hydrological similarity procedure seems to be too high when compared with the type of vegetation cover along the Riarbero River banks, the authors believe that the estimate obtained by hydrological simulation may provide a useful indication. However, it has also to be taken into account that the reliability of the rainfall-runoff model at medium scale (the contributing areas at Bacchello Bridge and Cavola Bridge are  $1294$  and  $337 \text{ km}^2$ , respectively) does not imply that such reliability is preserved also at the very small scale of the Riarbero River watershed. In the latter case, local conditions of climate, river cross section geometry, soil type and vegetation cover may exert significant forcing on the river flow magnitude, thus inducing a significant uncertainty in the peak flow estimate.

However, it should be considered that the required design flood is needed for the purpose of sizing river engineering works aimed at controlling river bed erosion control along the Riarbero River, whose watershed is inhabited. Therefore, a failure of erosion control works would not imply loss of lives but only a economical loss resulting from a possible need to re-structure the erosion control works. Furthermore, one should note that sizing the erosion control works on the basis of the design flood derived from hydrological similarity would imply the construction of massive engineering structures. This option would lead to loosing part of the great environmental value of the Riarbero watershed. On the light of these considerations we believe that it is not unreasonable to utilise the outcome of the hydrological simulation for the design of the required erosion control works.

*Acknowledgements.* The authors would like to thank the Editor Ralf Merz, Francesco Laio, Murugesu Sivapalan and an anonymous referee for providing valuable suggestions and assistance in revising the present paper. The work presented here has been carried out in the framework of the activity of the Working Group at the University of Bologna of the Prediction in Ungauged Basins (PUB) initiative of the International Association of Hydrological Sciences. The study has been partially supported by the Italian Government through its national grant to the program on “Advanced techniques for estimating the magnitude and forecasting extreme hydrological events, with uncertainty analysis”.

Edited by: R. Merz

#### References

- Band, L. E.: Topographic partition of watersheds with digital elevation models, *Water Resour. Res.*, 22, 15–24, 1986.
- Beven, K. J.: *Rainfall-Runoff modelling*, Wiley, United Kingdom, 2000.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36, 2006.
- Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Proc.*, 6, 279–298, 1992.
- Blazkova, S. and Beven, K. J.: Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOP-MODEL, *J. Hydrol.*, 195, 256–278, 1997.
- Blazkova, S. and Beven, K. J.: Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), *Water Resour. Res.*, 38, 1–14, 2002.
- Brath, A., Montanari, A., and Toth, E.: Comparing the calibration requirements and the simulation performances of lumped and distributed hydrological models: an Italian case study, AGU Spring Meeting, Boston, United States, 29 May–2 June 2001, H31D-04, 2001.
- Brath, A., Castellarin, A., Franchini, M., and Galeati, G.: Estimating the index flood using indirect methods, *Hydrol. Sci. J.*, 46, 399–418, 2001.
- Brath, A., Montanari, A., and Toth, E.: Analysis of the effects of different scenarios of historical data availability on the cali-

- bration of a spatially-distributed hydrological model, *J. Hydrol.*, 291, 232–253, 2004.
- Burlando, P.: Stochastic models for the predictions and simulations of rainfall in time, Ph.D. thesis, Politecnico di Milano, Italy, 183 pp, 1989.
- Cameron, D., Beven, K. J., Tawn, J., Blazkova, S., and Naden, P.: Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty), *J. Hydrol.*, 219, 169–187, 1999.
- Cameron, D., Beven, K. J., Tawn, J., and Naden, P.: Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation), *Hydrol. Earth Syst. Sci.*, 4, 23–34, 2000, <http://www.hydrol-earth-syst-sci.net/4/23/2000/>.
- Cameron, D., Beven, K. J., and Naden, P.: Flood frequency estimation by continuous simulation under climate change (with uncertainty), *Hydrol. Earth Syst. Sci.*, 4, 393–405, 2000b, <http://www.hydrol-earth-syst-sci.net/4/393/2000/>.
- Castellarin, A.: Criteri di affinità idrologica e tecniche di stima del fattore di scala nell'analisi regionale di frequenza delle piene, Ph.D. thesis, University of Bologna, Italy, 130 pp., 2000 (in Italian).
- Chow, V. T., Maidment, R. M., and Mays, L. W.: *Applied Hydrology*, McGrawHill, United Kingdom, 1988.
- Cowpertwait, P. S. P.: A generalized spatial-temporal model of rainfall based on a clustered point process, *Proc. of the R. Soc. of London, Series A*, 450, 163–175, 1996.
- Cunge, J. A.: On the subject of a flood propagation computation method (Muskingum method), *J. Hydraul. Res.*, 7, 205–230, 1969.
- Doorenbos, J., Pruitt, W. O., Aboukhaled, A., Damagnez, J., Dastane, N. G., van der Berg C., Rijtema, P. E., Ashford, O. M., and Frere, M.: Guidelines for predicting crop water requirements, *FAO Irrig. Drainage Pap.*, Rome, Italy, 1984.
- Franchini, M. and Galeati, G.: Analisi regionale dei massimi annuali delle portate al colmo per la regione Romagna-Marche (in Italian), *L'energia elettrica*, 73, 200–212, 1996.
- Jothityangkoon, C. and Sivapalan, M.: Towards estimation of extreme floods: Examination of the roles of runoff process changes and floodplains, *J. Hydrol.*, 281, 206–229, 2003.
- Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, doi:10.1029/2005WR004362, 2006.
- Lamb, R. L.: Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation, *Water Resour. Res.*, 35, 3103–3114, 1999.
- Maione, U.: *Le piene fluviali*, La Goliardica Pavese, Italy, 1995.
- Montanari, A.: Long range dependence in hydrology, in: *Theory and application of long range dependence*, Birkhauser, Boston, United States, 2003.
- Montanari, A.: Large sample behaviors of the Generalised Likelihood Uncertainty Estimator in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 41, W08406, doi:10.1029/2004WR003826, 2005.
- Montanari, A.: What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology, *Hydrol. Proc.*, 21, 841–845, 2007.
- Montanari, A., Rosso, R., and Taqqu, M. S.: Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation, *Water Resour. Res.*, 33, 1035–1044, 1997.
- Montgomery, D. R. and Foufoula-Georgiou, E.: Channel Network Source Representation Using Digital Elevation Models, *Water Resour. Res.*, 29, 3925–3934, 1993.
- Moretti, G. and Montanari, A.: Estimation of the peak river flow for an ungauged mountain creek using a distributed rainfall-runoff model, in: *Recent advances in peak river flow modelling, prediction and real-time forecasting – Assessment of the impacts of land-use and climate changes*, BIOS, Italy, 2004.
- Moretti, G. and Montanari, A.: Affdef: A spatially distributed grid based rainfall-runoff model for continuous time simulations of river discharge, *Environ. Mod. & Soft.*, 22, 823–836, 2007.
- Naden, P. S., Crooks, S. M., and Broadhurst, P.: Impact of climate and land use change on the flood response of large catchments, in: *Proceedings of the 31st MAFF Conference of River and Coastal Engineers*, Keele, United Kingdom, 3–5 July 1996, 16 pp, 1996.
- NERC: *Natural Environment Research Council Flood Studies Report*, Whitefriars Press., United Kingdom, 1975.
- Orlandini, S., Perotti, A., Sfondrini, G., and Bianchi, A.: On the storm flow response of upland Alpine catchments, *Hydrol. Proc.*, 13, 549–562, 1999.
- Rodriguez-Iturbe, I., Cox, D. R., and Isham, V.: Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London, Series A*, 410, 269–288, 1987.
- Soil Conservation Service: *National Engineering Handbook, Section 4, Hydrology*, US Department of Agriculture, United States, 1972.
- Tarboton, D. G.: A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resour. Res.*, 33, 309–319, 1997.
- Winsemius, H. C., Savenije, H. H. G., Gerrits, A. M. J., Zapreeva, E. A., and Klees, R.: Comparison of two model approaches in the Zambezi river basin with regard to model reliability and identifiability, *Hydrol. Earth Syst. Sci.*, 10, 339–352, 2006, <http://www.hydrol-earth-syst-sci.net/10/339/2006/>.