



## Estimating the uncertainty of hydrological forecasts: A statistical approach

Alberto Montanari<sup>1</sup> and Giovanna Grossi<sup>2</sup>

Received 5 February 2008; revised 3 September 2008; accepted 22 September 2008; published 16 December 2008.

[1] A method for quantifying the uncertainty of hydrological forecasts is proposed. This approach requires the identification and calibration of a statistical model for the forecast error. Accordingly, the probability distribution of the error itself is inferred through a multiple regression, depending on selected explanatory variables. These may include the current forecast issued by the hydrological model, the past forecast error, and the past rainfall. The final goal is to indirectly relate the forecast error to the sources of uncertainty in the forecasting procedure, through a probabilistic link with the explaining variables identified above. Statistical testing for the proposed approach is discussed in detail. An extensive application to a synthetic database is presented, along with a first real-world implementation that refers to a real-time flood forecasting system that is currently under development. The results indicate that the uncertainty estimates represent well the statistics of the actual forecast errors for the examined events.

**Citation:** Montanari, A., and G. Grossi (2008), Estimating the uncertainty of hydrological forecasts: A statistical approach, *Water Resour. Res.*, 44, W00B08, doi:10.1029/2008WR006897.

### 1. Introduction

[2] The development and improvement of real-time flood forecasting systems is a very topical subject today. In fact, the effectiveness of early warning systems for the sake of reducing the damages and casualties induced by floods is widely recognized and new avenues of research are continuously open by the recent technical advances in computer sciences, hydrological monitoring and modeling. Among the most crucial scientific interests the assessment of the forecast uncertainty plays a not negligible role. In general, a forecast does not eliminate uncertainty about a future event, but only reduces it. Therefore it is extremely important to quantify the residual uncertainty, in order to make the forecast useful for practical purposes [Krzysztofowicz, 2001]. A timely communication to end users requires though that the estimate of the uncertainty is performed quickly.

[3] The need for an effective quantification of the forecast uncertainty has been recently stressed by Krzysztofowicz [2001, 2002], who remarked that the prevailing format of operational hydrological forecasts is still deterministic. Information neglecting uncertainty may induce a misleading illusion of certainty in the end user or decision maker. In fact, when such a forecast turned out to be wrong, the consequences would probably be worse with respect to a situation where no forecast was available.

[4] Uncertainty assessment is the subject of an intense research activity in hydrology [Beven, 2006a, 2006b; Montanari, 2007; Mantovan and Todini, 2006]. Basically, the techniques so far proposed for tackling the problem can

be separated into statistical and non statistical approaches. While many researchers are convinced that all the hydrological forecasts should be expressed in terms of probabilities [Krzysztofowicz, 2001], others are convinced that ergodicity and stationarity, which are the basis for statistical inference, cannot be suitable working hypotheses in hydrology. The main reason is that the underlying physical processes are highly heterogeneous in space and time and therefore might be inherently non stationary. From there comes the motivation for using non statistical approaches to assess uncertainty in hydrology, like, for instance, the generalized likelihood uncertainty estimation (GLUE) [Beven, 2006a].

[5] No matter which method is applied, uncertainty assessment in hydrological forecasting is still considered a relevant practical problem. A valuable contribution was recently provided by Krzysztofowicz [2001, 2002] who developed the Bayesian forecasting system (BFS), a statistically based method for assessing the uncertainty of flood forecasts. This technique presents many advantages. The most relevant one is the capability to take explicitly into account the uncertainty of the precipitation forecast, if this latter is expressed in the form of a probability distribution of future rainfall.

[6] The present study aims to propose a statistical method for quantifying the total uncertainty in hydrological forecasting, which is based on the use of a meta-Gaussian statistical model to infer the probability distribution of the forecast error depending on selected explanatory variables. The meta-Gaussian multivariate probability distribution was introduced in hydrology by Kelly and Krzysztofowicz [1997]. It is obtained by fitting with a multivariate Gaussian distribution random variables with arbitrary marginal distributions by embedding the normal quantile transform (NQT) of each variate into the Gaussian law. For more details see section 2.

<sup>1</sup>Faculty of Engineering, University of Bologna, Bologna, Italy.

<sup>2</sup>Faculty of Engineering, University of Brescia, Brescia, Italy.

[7] The proposed method does not attempt to separately estimate the contribution of each individual source of uncertainty. It is operationally simple and fast and relies on mild assumptions that are frequently met in practical applications. In view of such features, this technique might provide useful perspectives for estimating the uncertainty of hydrological forecasts in real time within a robust framework.

## 2. Estimating the Probability Distribution of the Forecast Error

[8] In order to estimate the uncertainty of hydrological forecasts, it is assumed here that the forecast error is a stationary and ergodic stochastic process, denoted with the symbol  $E(t)$ . It is suggested that its statistical properties are inferred by analyzing a past realization  $e_{obs}(t) = Q_{obs}(t) - Q_{pred}(t)$  that it is assumed to be available, where  $Q_{obs}(t)$  and  $Q_{pred}(t)$  are true and forecasted river flows, respectively. Therefore, this method implies that the hydrological model is preliminarily applied in order to predict past observations by emulating an operational forecasting framework. In this way, the past realization of the forecast error,  $e_{obs}(t)$ , can be obtained.

[9] In order to derive a probabilistic model for  $E(t)$ , the main statistical behavior of the forecast error have to be taken into account. They can be summarized in the following two points. (1)  $E(t)$  is characterized by marginal statistics that change in time. Typically, the greater the predicted hydrological variable, the greater the forecast error. (2)  $E(t)$  is frequently affected by a strong persistence. However, such persistence does not mean that the error can be easily predicted. Generally, a significant forecast error is an indicator of the presence of relevant uncertainty in the predicting procedure at the forecast time and thus it is likely to be followed by high errors as well. However, the realizations  $e_{obs}(t)$  of hydrological forecast errors are usually characterized by infrequent and random changes of sign. These induce the presence of a sizable uncertainty in the prediction of the next forecast error.

[10] The use of a meta-Gaussian model is then proposed to derive the time varying probability distribution of the forecast error. Basically, the probability distribution of  $E(t)$  is inferred on the basis of its dependence on  $M$  selected explanatory random variables. These are in charge of explaining the variability in time of the marginal statistics of  $E(t)$ . The statistical inference is performed in the Gaussian domain, by preliminarily transforming  $E(t)$  and the explanatory variables to the Gaussian probability distribution. The above transformation is operated through the Normal Quantile Transform (NQT).

[11] Let us refer to  $E(t)$  for explaining the NQT, which involves the following steps: (1) for the  $j$ th data  $e_{obs}(t_j)$  of the realization  $e_{obs}(t)$  the cumulative frequency  $F[e_{obs}(t_j)]$  is computed by using the Weibull plotting position [Stedinger et al., 1993], that is,  $F[e_{obs}(t_j)] = k_j/(n + 1)$ . Here  $k_j$  is the position occupied by  $e_{obs}(t_j)$  in the sample sorted in ascending order and  $n$  is the sample size of  $e_{obs}(t)$ . (2) For each  $F[e_{obs}(t_j)]$  the standard normal quantile  $Ne_{obs}(t_j)$  is computed and associated with the corresponding  $e_{obs}(t_j)$ . Thus, a discrete mapping from  $e_{obs}(t)$  to its transformed counterpart  $Ne_{obs}(t)$ , which gives the NQT, is obtained. In order to be able to apply the inverse of the NQT, that is  $NQT^{-1}$ , for any

value of the transformed forecast error, a linear interpolation is used to connect the points of the discrete mapping previously obtained. The region beyond the minimum and maximum of  $Ne_{obs}(t)$  is covered by linear extrapolation. For more details about the operational use of the NQT see Kelly and Krzysztofowicz [1997] and also Montanari and Brath [2004].

[12] In practice, the probabilistic model for  $E(t)$  is built as follows. First of all, it is assumed that positive and negative errors come from 2 different statistical populations  $E^{(+)}(t)$  and  $E^{(-)}(t)$ . Therefore, the probability model for  $E(t)$  is given by a mixture of two probability distributions, one for  $E^{(+)}(t)$  and one for  $E^{(-)}(t)$ . The mixture is composed such that the area of the probability distribution of  $E^{(+)}(t)$  is equal to the percentage,  $P^{(+)}$ , of positive errors over the total sample size of the available past realization  $e_{obs}(t)$  of the forecast error.

[13] The two realizations  $e^{(+)}_{obs}(t)$  and  $e^{(-)}_{obs}(t)$  are transformed through the NQT, therefore obtaining the normalized realizations  $Ne^{(+)}_{obs}(t)$  and  $Ne^{(-)}_{obs}(t)$ . Then,  $M$  explanatory variables,  $X^{(i)}(t)$  with  $i = 1, \dots, M$  (which should be readily available at the forecast time) are selected in order to explain the variability in time of the marginal statistics of  $E^{(+)}(t)$  and  $E^{(-)}(t)$ . The values of such explanatory variables for the realizations  $e^{(+)}_{obs}(t)$  and  $e^{(-)}_{obs}(t)$  above are estimated and then transformed by using the NQT, therefore obtaining the normalized explanatory variables  $NX^{(i)}_{obs}(t)$  with  $i = 1, \dots, M$ .

[14] In the Gaussian domain, it is assumed that the forecast error can be expressed as a linear combination of the selected explanatory variables. Let us focus on the positive error. The linear combination can be expressed through the following relationship:

$$Ne^{(+)}(t_j) = C_1^{(+)}NX^{(1)}(t_j) + C_2^{(+)}NX^{(2)}(t_j) + \dots + C_M^{(+)}NX^{(M)}(t_j) + \varepsilon^{(+)}(t_j), \quad (1)$$

where  $\varepsilon^{(+)}(t_j)$  is an outcome of a homoscedastic and Gaussian random variable. An analogous relationship holds for  $Ne^{(-)}(t)$ . It is assumed that positive and negative errors are conditioned by the same explanatory variables, but the fit of the linear regression (1) leads to a different set of coefficient values. Such coefficients are estimated by plugging in (1) the past realizations of transformed forecast error,  $Ne^{(+)}_{obs}(t)$ , and explanatory variables,  $NX^{(i)}_{obs}(t)$ , and then by identifying the coefficient values that lead to the best fit (for instance by minimizing the sum of the squares of  $\varepsilon^{(+)}(t_j)$ ).

[15] The goodness of the fit provided by (1) can be verified by drawing a normal probability plot and a residual plot for  $\varepsilon^{(+)}(t)$  as in the work by Montanari and Brath [2004]. In the case the goodness-of-fit test is not satisfied, a better result can be obtained by calibrating the regression on the basis of the data points corresponding to the higher river flows only. Goodness-of-fit checking for the applications reported here is shown in section 4.1.

[16] Once the linear regression (1) has been calibrated, for positive as well as for negative errors, the probability distribution of the transformed positive forecast error can be easily derived for potential real-time and real-world

applications. Such distribution is Gaussian and is expressed by the following relationship:

$$Ne^{(+)}(t_j) \sim G\left\{\mu\left[Ne^{(+)}(t_j)\right], \sigma\left[Ne^{(+)}(t_j)\right]\right\}, \quad (2)$$

where  $\sim$  means equality in probability distribution and  $G$  indicates the Gaussian distribution whose parameters are given by

$$\mu\left[Ne^{(+)}(t_j)\right] = C_1^{(+)}Nx^{(1)}(t_j) + C_2^{(+)}Nx^{(2)}(t_j) + \dots + C_M^{(+)}Nx^{(M)}(t_j) \quad (3)$$

$$\sigma\left[Ne^{(+)}(t_j)\right] = \sigma\left[\varepsilon^{(+)}(t_j)\right]. \quad (4)$$

Analogous relationships (from (2) to (4)) hold for the negative error. Therefore, the confidence bands (CBs) for the transformed forecast at an assigned significance level can be straightforwardly derived. In detail, the difference,  $Ne_{90\%}^{(+)}(t_j)$ , between the forecast and the upper CB, in the Gaussian domain, at the  $\alpha$  significance level is given by the  $1 - \alpha/(2 \cdot P^{(+)})$  quantile of the Gaussian distribution given by (2), (3) and (4). Given that  $P^{(+)}$  can be arbitrarily close to 0, in the technical computation one may obtain values greater than 1 of  $\alpha/(2 \cdot P^{(+)})$ . This means that the probability of getting a positive forecast error is small enough to make equal to 0 the width of the upper CB at the  $\alpha$  significance level.

[17] For instance, if  $P^{(+)} = 0.5$  and  $\alpha = 10\%$ ,  $Ne_{90\%}^{(+)}(t_j)$  is given by the well known relationship

$$Ne_{90\%}^{(+)}(t_j) = \mu\left[Ne^{(+)}(t_j)\right] + 1.96\sigma\left[Ne^{(+)}(t_j)\right]. \quad (5)$$

Finally, by applying back the NQT one obtains the CBs for the assigned significance level in the untransformed domain. It is important to put in evidence that the  $\alpha$  significance level corresponds to the  $1 - \alpha$  confidence level. This means that the identified CBs of the hydrological forecast are such that there is a probability of  $1 - \alpha$  for the true value of the hydrological variable to fall between them.

[18] The reason why positive and negative errors are treated separately is that a good fit was not achieved through the linear regression (1) when the errors were pooled together. In fact, in this case, it appears that the NQT is not effective in making the errors homoscedastic and therefore the assumption of linearity does not hold. The reason for this result is that the NQT is not efficient in assuring homoscedasticity if the mean of the model error is not significantly changing across the range of the error itself, as it often happens when dealing with hydrological models. By treating positive and negative errors separately the problem disappears and the assumptions of the linear regression are met. Finally, it is important to note that the only assumption made about the sign of the future forecast error is that it has a probability equal to  $P^{(+)}$  to be positive. Therefore, no inference is made on the sign of the forecast error on the basis of the explanatory variables.

### 3. Case Studies

[19] The test sites herein considered refer to two medium-sized watersheds located in north central Italy: the Secchia

River and the Toce River. The Secchia River was used as reference basin for performing an extensive test by using synthetic hydrological data while the Toce River was used for performing an operational experiment within a real-time flood forecasting system that is under development. The synthetic case study was developed in order to be able to test the proposed procedure against a very extended data set.

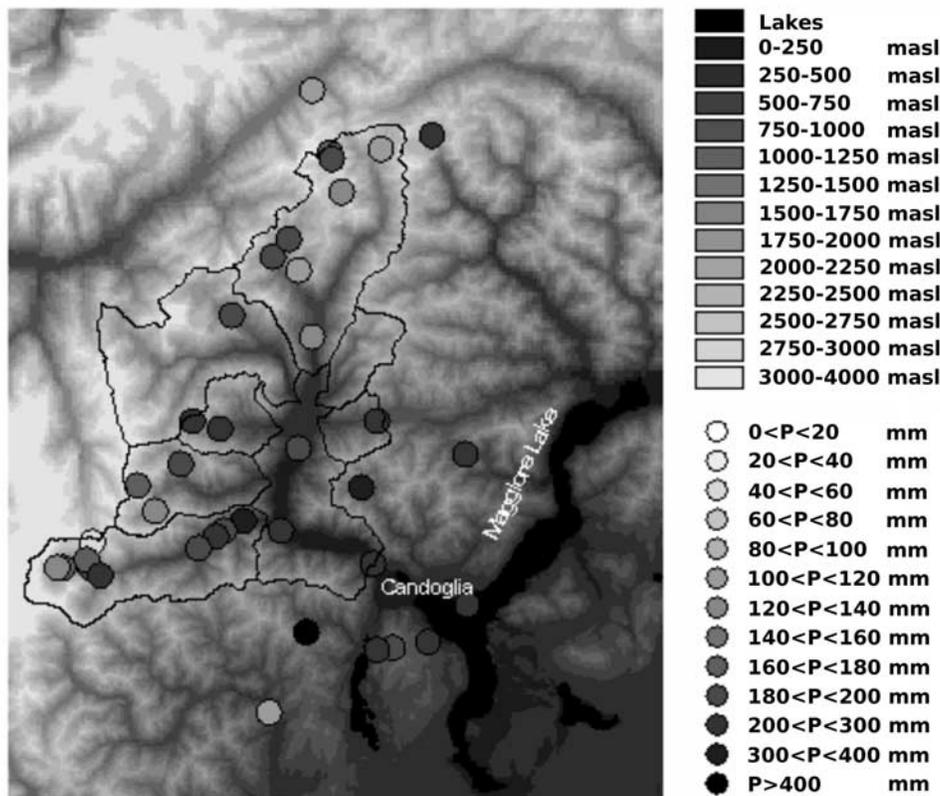
[20] The Secchia River flows northward across the Apennine Mountains and is a right tributary to the Po River. The contributing area is 1214 km<sup>2</sup> at the Bacchello Bridge river cross section that is located about 62 km upstream of the confluence in the Po River. The mainstream length up to Bacchello Bridge is about 98 km and the basin concentration time is about 15 h. The mean annual rainfall depth ranges between 700 and more than 2000 mm/a over the basin area, while snowmelt does not significantly contribute to the formation of the flood flows. The maximum peak discharge observed at Bacchello Bridge in the period 1923–1981 was 823 m<sup>3</sup>/s (20 April 1960).

[21] The Toce River is located in Northern Italy, west of Maggiore Lake (see Figure 1). It is a subbasin of the Ticino River watershed which has already been selected as the target area of several international projects (see for example *Bacchi and Ranzi* [2003]; *Bougeault et al.* [2001]). At the Candoglia cross section the drainage area is 1532 km<sup>2</sup> and 40% of it is higher than 2000 m, being the mean altitude equal to 1724 m. The mainstream is around 80 km long and the concentration time is about 9 h. Average annual precipitation is 1557 mm, ranging from about 1300 to 2381 mm. The runoff regime shows some differences with respect to the precipitation regime. In fact snow fall, snowmelt and icemelt play an important role in the time distribution of available water resources.

#### 3.1. Generation of Synthetic Hydrological Data for the Secchia River and Testing of a Forecasting Procedure

[22] The Secchia River case study was developed by first generating a synthetic 100-year-long sample of hourly rainfall, temperature and river flow data. The synthetic rainfall data were generated by using the generalized multivariate Neyman-Scott rectangular pulses model. For more details the interested reader is referred to *Cowpervtwait* [1995] and *Montanari* [2005]. The rainfall model was applied to the Secchia River basin by generating data for five locations where rain gauges are present. By using the method of moments, historical hourly rainfall data observed in the 2-year period 1972–1973 were used to calibrate the Neyman-Scott model parameters. The model provided a satisfactory fit to the selected local and spatial statistics of the observed rainfall field (including cross correlation between rain gauges).

[23] A hundred years of hourly rainfall data were subsequently generated for the five rain gauges. The depth duration frequency curves in each of them were well simulated; in particular, the percentage error in the simulation of the 12-h (a time span comparable to the concentration time of the basin) cumulated rainfall with return period of 100 years was always lower than 6% in all the rain gauges. The mean areal hourly rainfall over the basin was then estimated through a weighted average of the hourly rainfall data in the five locations. The weights  $w_j, j = 1, \dots, 5$  were estimated with the Thiessen polygons method.



**Figure 1.** The Toce watershed: the dots outline the location of rain gauges working during MAP-SOP IOP02 (intensive observation period 2, September 1999).

[24] The rainfall data were subsequently corrupted in order to emulate a real situation, where the observed records are always affected by uncertainty. In detail, the weights  $w_j$  used to compute the mean areal rainfall over the basin were perturbed by randomly generating, at each time step, each  $w_j$  accordingly to a uniform distribution in the range  $\pm 20\%$  of their actual value. Then, the  $w_j$  obtained at each time step are rescaled to make their cumulative sum equal to one.

[25] The rainfall data corruption introduced an uncertainty that can be quantified by the value equal to 0.76 of the coefficient of determination of the linear regression of corrupted versus uncorrupted mean areal rainfall depths. Finally, the rainfall generation allowed to obtain a 100-year record of lumped and uncertain hourly rainfall  $P_m(t)$  over the basin.

[26] A generation of a synthetic hourly temperature record was performed by referring to a location where historical data were available and was carried out by applying a fractionally differenced ARIMA model (FARIMA) [Montanari et al., 1997]. A mean areal value of the hourly temperature data was obtained by rescaling the synthetic observations to the mean altitude of the basin area, by adopting a mean vertical thermal gradient. Temperature data were not corrupted given that temperature uncertainty is not significantly effective in the formation of the Secchia River peak flows, in view of the negligible contribution of snowmelt.

[27] Finally, synthetic river flow data were obtained by using the previously generated synthetic rainfall and temperature records as input data to the lumped rainfall-runoff

model ADM [Franchini, 1996], requiring the estimation of 11 parameters. The model was calibrated by automatically optimizing the continuous simulation of historical hourly river flow data observed at Bacchello Bridge in the year 1972 and part of the year 1973. It was finally validated by simulating the fraction of the 1973 flows that was not used for the calibration. The resulting *Nash and Sutcliffe* [1970] efficiency was 0.81 for the validation period. The ADM model was subsequently applied to generate a 100-year-long record of hourly synthetic river flows of the Secchia River.

[28] The synthetic river flows were subsequently corrupted by multiplying each observation by a coefficient that assumed different values at each hourly time step. The time series of the coefficient values was obtained by generating outcomes from a uniform distribution in the range 0.8–1.2. The coefficient of determination of the linear regression of corrupted versus uncorrupted river flow data is equal to 0.86.

[29] The synthetic hydrometeorological data set was finally used to perform an extensive forecasting experiment. From the generated data set, some flood events were selected that lasted at least 200 h, during which the river flow was always higher than 30 m<sup>3</sup>/s and reached a peak higher than 200 m<sup>3</sup>/s. This selection led to picking up 25 floods, with peak flows ranging from about 200 to 672 m<sup>3</sup>/s.

[30] The above flood events were used to emulate an operational real-time flood forecasting, that was performed by using the HYMOD rainfall-runoff model, with an hourly computational time step. HYMOD is a lumped rainfall-

**Table 1.** Secchia River Synthetic Case Study Regression Coefficients and Explained Variance of the Regression Model<sup>a</sup>

Dependent Variables	Explanatory (Independent) Variables			Explained Variance (%)
	Past Forecast Error $E_p(t)$	Past 6-h Cumulative Rainfall $P_p(t)$	Forecasted River Flow $Q_{pred}(t + \Delta t)$	
Positive 1-h-ahead forecast error	0.953	0.105	-0.010	95
Negative 1-h-ahead forecast error	0.946	0.052	0.012	91
Positive 6-h-ahead forecast error	0.710	0.275	0.075	63
Negative 6-h-ahead forecast error	0.621	0.538	-0.096	71

<sup>a</sup>The transformed forecast error is expressed as a linear combination of the transformed explanatory variables.

runoff model that was introduced by Boyle [2000] and recently used by Wagener *et al.* [2001], Vrugt *et al.* [2003], and Montanari [2005]. Since HYMOD has only five parameters, it can be considered a modeling approach of reduced complexity with respect to ADM, the output of which is therefore assumed to reproduce the “real” response of the basin. HYMOD was then applied to forecast the river flow for each time step of the selected synthetic flood events, with lead time of 1 h and 6 h. The precipitation forecasting needed for issuing the 6-h lead time river flow forecasts was obtained by using the persistent model: the future hourly rainfall is assumed to be equal to the value observed at the forecast time.

[31] Therefore, for each of the synthetic flood events two forecasted time series of river flows and two series of forecast errors were obtained for the lead times of 1 h and 6 h, respectively.

### 3.2. A Real-Time Flood Forecasting System for the Toce River Basin

[32] The Ticino-Toce (TT) area was one of the target areas of the international research project RAPHAEL [Bacchi and Ranzi, 2003]. Four extreme rainfall and flood events that occurred in the previous decade in the target area were selected as case studies to demonstrate the potentials of coupling meteorological and hydrological models for forecasting flood events in mountain areas: (1) from 22 to 25 September 1993, the TT1 event, (2) from 11 to 14 October 1993, the TT2 event, (3) from 3 to 6 November 1994, the TT3 event, (4) from 27 to 30 June 1997, the TT4 event. Of course no real-time operational forecast was issued for these four (past) events, but a potential real-time framework was defined and used to simulate the storm and flood events, either coupling high-resolution limited area meteorological models to hydrological flood models or using rainfall and surface air temperature ground observations. The BOLAM hydrostatic model [Buzzi and Foschini, 2000] was coupled to a conceptual distributed hydrological model, called DIMOSOP, to simulate and predict the flood hydrographs at the basin outlet for the selected events [Ranzi *et al.*, 2003]. Actually none of the parameters of the hydrological model was finally set through a trial-and-error calibration procedure, since each parameter (including channel roughness and the soil hydraulic properties) was estimated on the basis of field surveys, laboratory analysis or literature reports [Bacchi *et al.*, 2002].

[33] The same research goal was actually addressed also by the international Mesoscale Alpine Project (MAP), mainly focused on the intense rainfall field forecasting in the alpine area. The MAP scientific committee organized a

special observing period in autumn 1999 (MAP-SOP'99 [Bougeault *et al.*, 2001]). Again the Toce River was selected as target area [Ranzi *et al.*, 2003]. The most intense storm and flood event (IOP02) from 19 to 21 September 1999 recorded during MAP-SOP'99 was then selected as a case study for a numerical experiment [Grossi and Kouwen, 2004].

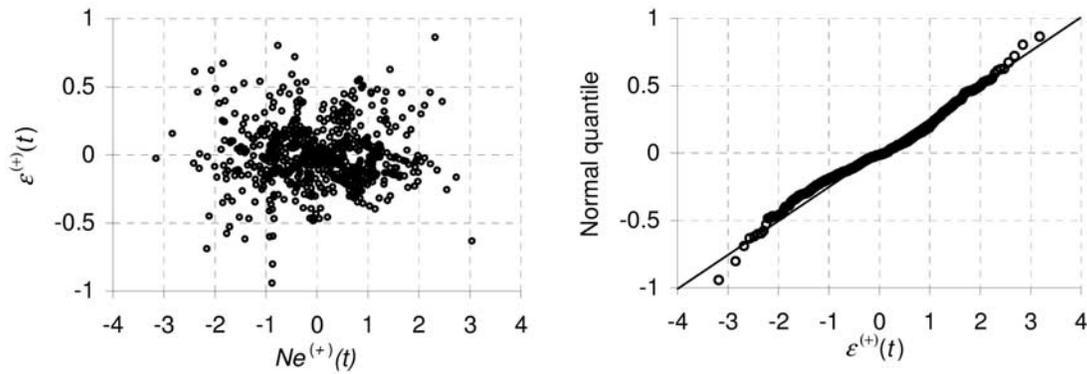
[34] An intense storm event occurred the year after (2000) again in the Toce area. Namely, in the period between 13 and 16 October 2000, heavy and persistent rainfall, locally exceeding 700 mm, caused extensive floods in several Alpine watersheds of the western Italian Alps. A detailed description of the meteorological evolution of the storm event is reported by Bacchi *et al.* [2002]. As for the IOP02 event, for the 2000 event (TT5 event) BOLAM run operationally and the 36 h ahead forecasts issued after 1200 UTC of every day were used to force the DIMOSOP hydrological model and forecast the flood hydrograph.

[35] Therefore, by summing up previous experiences, six major flood events that occurred in the Toce River basin in the last two decades (see Table 3) were simulated either by coupling meteorological and hydrological models or by forcing the hydrological model with ground observations of precipitation and surface air temperature. For two of the events (IOP02 and TT5) the hydrometeorological forecasting chain was run operationally. On average the percentage errors in the flood volumes simulated using the rain gauge data was -15%, while the timing of the flood model was accurate since observed peaks at 7 stream gauges along the main tributaries of the Toce River and at the basin outlet were anticipated by only 0.5 h. Using the predicted meteorological fields runoff volumes were underestimated, on average, by 10%. It is important to note that events TT2 and TT4 turned out to be difficult to simulate by the rainfall-runoff model, as a consequence of the relevant volume of water that was retained in those cases by the reservoirs displaced along the Toce-Ticino system. In view of the relevant features of the Toce case study from an operational point of view, the Toce River was selected as a real-world example of application of the statistical approach proposed here to estimate the forecast uncertainty.

## 4. Results

### 4.1. Extensive Test of the Uncertainty Assessment Procedure by Using Synthetic Data

[36] The first step for the application of the proposed uncertainty assessment procedure is the identification of the explanatory variables that, once transformed, are plugged in the linear regression (1). These variables should be readily



**Figure 2.** (left) Residual plot and (right) normal probability plot for the residuals of the linear regression (1) applied to the positive 1-h-ahead forecast errors of the Secchia River synthetic case study.

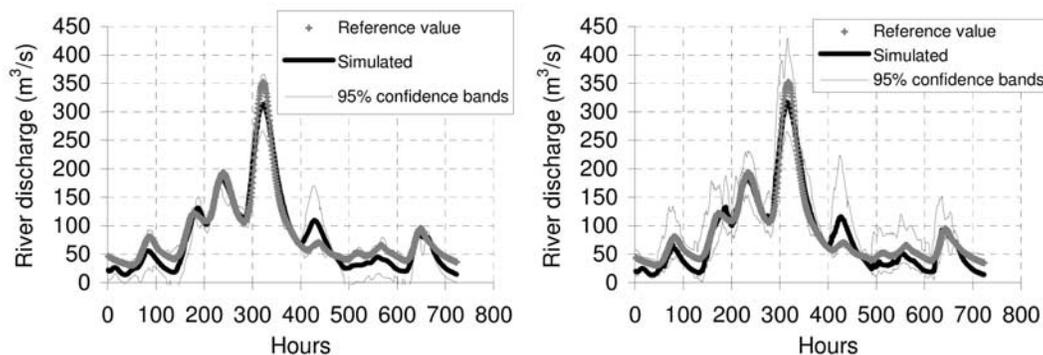
available at the forecast time  $t$ . By performing extensive tests on the synthetic data, the following explanatory variables were identified: (1) the forecasted river flow,  $Q_{pred}(t + \Delta t)$ , (2) the average of the absolute value of the past four 1-h-ahead forecast errors,  $E_p(t)$ , and (3) the cumulative rainfall in the 6 hours preceding the forecast time,  $P_p(t)$ .

[37] The reason for selecting  $Q_{pred}(t + \Delta t)$  is that the forecast error is likely to increase for increasing river flows and therefore it is expected to depend on the forecasted discharge as well. Therefore the explaining power of  $Q_{pred}(t + \Delta t)$  is expected to be significant. The second variable,  $E_p(t)$ , was selected by considering the significant persistence that usually affects the time series of forecast errors, while the third one,  $P_p(t)$ , was selected because the uncertainty is generally much more significant when predicting the raising limb of the hydrograph.

[38] Among the 25 flood events extracted from the synthetic data set, generated by the ADM run, 5 were picked up in order to calibrate the linear regression (1). Calibration was performed by using the data points corresponding to river flows greater than  $60 \text{ m}^3/\text{s}$ , therefore obtaining a sample of 683 forecast errors. The percentage of positive errors resulted equal to 79% and 81% for the 1-h and 6-h lead times, respectively, which means that HYMOD operated a general underestimation of the river discharge in a potential real-time prediction. Table 1 shows the calibrated

regression coefficients and the explained variance of the regression models for positive and negative errors. Some valuable remarks can be drawn.

[39] First, the presence of a significant dependence of the forecast error on  $E_p(t)$  is clear, while the dependence on  $Q_{pred}(t + \Delta t)$  and  $P_p(t)$  is less evident. This was an expected result and shows the great amount of information that is conveyed by the previous forecast error. The dependence on the past forecast error decreases for increasing lead time of the forecast, as it is expected, because the effect of the current forecast error vanishes when moving far apart in the future. Second, the explaining power of  $P_p(t)$  increases for the 6-h-ahead lead time. This result is explained by the use of the naïve persistent model for predicting the future rainfall (see section 3.1), which implies that an error in rainfall is affecting the forecast error to a greater extent and therefore the rainfall error is prevailing over the rainfall-runoff model structural error. Finally, it can be seen that the coefficients of the regression model are not much different for positive and negative errors. Of course, the shape of the two NQTs is instead well different, also because the transformed data are characterized by the same probability distribution (the standard Gaussian distribution), while the untransformed observations have opposite sign. Therefore the different statistics of positive and negative errors are mainly accounted for by the NQT in the present case.



**Figure 3.** Secchia River synthetic case study. Example of computation of the 95% confidence bands of the forecast with (left) 1-h and (right) 6-h lead time. The reference river flow values are the synthetic data that were treated as observed.

**Table 2.** Secchia River Synthetic Case Study Percentage of Observations Lying Outside the Forecast Confidence Bands for Different Confidence Levels

	95% Confidence Bands (%)	98% Confidence Bands (%)	99% Confidence Bands (%)
Positive error 1-h lead time	2.01	1.61	1.40
Negative error 1-h lead time	5.30	3.68	3.34
Positive error 6-h lead time	1.32	1.04	0.84
Negative error 6-h lead time	9.63	5.37	2.64

[40] Figure 2 shows the residual plot and the residual normal probability chart for the linear regression (1) applied to the positive 1-h-ahead forecast errors. The residual plot shows the progress of the residuals versus the linear combination of the predictors, where each predictor is weighted through the respective regression coefficient. In fact, the variable  $Ne^{(+)}(t)$  displayed on the abscissa is given by the linear regression (1). The residual plot allows to assess the goodness of the fit through just one picture [Cook and Weisberg, 1994].

[41] It can be noticed that there is a good agreement with the hypothesis of homoscedastic and Gaussian residuals. A similar result was obtained for the regressions related to the negative forecast errors and the 6-h lead time (again for both positive and negative errors). It is worth noting that a fit to the regression model was attempted also on the basis of the whole sample of  $E(t)$ , without separating between positive and negative errors. As it was expected, the residuals of the regression turned out to be non Gaussian and heteroscedastic, therefore providing a further confirmation that splitting the error series into positive and negative errors is a necessary step.

[42] Then, the procedure was applied for assessing the 1-h-ahead and 6-h-ahead forecast uncertainty for the other 20 selected events in validation mode. A total of 6654 emulations of a real-time forecast were performed for both lead times. First of all, it is worth noting that the regression model given by (1) exhibited in validation an explained variance of 84% and 88% for positive and negative forecast errors, respectively, at 1-h lead time. For the 6-h lead time the explained variance was 58% and 59%, respectively. Figure 3 shows an example of observed and forecasted hydrographs, along with the related 95% CBs, for both the 1-h and 6-h lead times: the distance between the CBs increases for increasing lead time, as expected.

[43] The extensive forecasting test performed with synthetic data allows us to perform a systematic testing of the reliability of the identified CBs. Laio and Tamea [2007] provided a comprehensive overview of verification tools for probabilistic forecasts of continuous hydrological variables. In particular, they proposed a statistically based technique for testing the accurateness of probabilistic forecasts like the one derived here. However, such method requires the assessment of the whole probability distribution of the forecast in the untransformed domain, and not just the CBs. Therefore such test, in the context of the technique that is proposed here, would be computationally expensive.

[44] To avoid the problem, it is suggested that the reliability of the CB is verified by simply computing the percentage  $PI$  of observations that fall between the CB in

validation mode. Ideally, in absence of any nonstationarity and sampling variability  $PI$  should be equal to 95%. Computation has been done separately for positive and negative errors. The disadvantage of this test is that any observation located outside the CBs is treated as a failure of the uncertainty assessment method, regardless of its distance from them. Therefore, we also indicate the  $PI$  values for the CBs at the 98% and 99% confidence level.

[45] Table 2 shows the computed  $PI$  values. It can be seen that the points falling outside the confidence bands are slightly more frequent than expected, summing to a total of 7.31% and 10.95% for the 1-h and 6-h lead times, respectively, to be compared with an expected theoretical value of about 5%. This effect was induced by using calibration events where the forecast error was slightly more contained with respect to the validation data set. As a matter of fact, the calibration sample has a significant effect on the performances of the uncertainty assessment and therefore it is advisable to pay particular attention to its selection. However, Table 2 shows that a significant part of the outside observations are located between the 95% and 99% confidence bands and therefore the reliability of the uncertainty assessment can be considered satisfactory.

#### 4.2. Real-World Application to the Toce River Flood Forecasting System

[46] The limited number of flood events observed for the Toce River introduces a relevant uncertainty in the calibration of the linear regression given by (1). In fact, the experiment carried out with the synthetic data proved the importance of using an extended calibration data set to prevent estimation errors induced by local variability of the model performances. In order to limit the overparameterization problem, only two explanatory variables were considered for the case study of the Toce River, that is, the forecasted river flow and the average of the absolute value of the past four 1-h-ahead forecast errors. These were found to outperform the current and past climatic data (namely, rainfall and temperature) in terms of explained variance of the forecast error.

[47] For each of the 6 flood events mentioned above and synthetically described in Table 3, either real-time flood forecasting experiments or hindcast experiments were performed by using a unique meteorological prediction provided by the limited area model at the beginning of the event, for a lead time varying from 1 to about 96 h. In the case the uncertainty in the hydrometeorological coupled forecast had to be evaluated, only one prediction would be

**Table 3.** Toce River Case Study Summary of Flood Events<sup>a</sup>

Event	Date of Occurrence	Number of the Issued Forecasts	Peak Flow (m <sup>3</sup> /s)
TT1	21–25 Sep 1993	57	2535.1
TT2	11–15 Oct 1993	118	940.7
TT3	3–7 Nov 1994	120	978.1
TT4	26 Jun to 1 Jul 1997	150	927.9
MAP-SOP'99	19–22 Sep 1999	47	1495.2
TT5	12–17 Oct 2000	112	2681.0

<sup>a</sup>Data were analyzed during the developing phase of the real-time flood forecasting system.

**Table 4.** Toce River Case Study Regression Coefficients and Explained Variance of the Regression Model<sup>a</sup>

Dependent Variables	Explanatory (Independent) Variables		
	Past Forecast Error $E_p(t)$	Forecasted River Flow $Q_{pred}(t + \Delta t)$	Explained Variance (%)
Positive 1-h-ahead forecast error	0.637	0.187	59
Negative 1-h-ahead forecast error	0.499	0.185	92

<sup>a</sup>The transformed forecast error is expressed as a linear combination of the transformed explanatory variables.

available per each lead time and event. It follows that a reliable calibration of the regression given by (1) for a given lead time would not be possible because of the small size of the sample.

[48] To overcome this problem, the uncertainty assessment methodology was tested by focusing on the lead time of 1-h only. In fact, given that the rainfall-runoff model runs at hourly time step, the 1-h ahead prediction can always be obtained by using the observed rainfall data available at the forecast time. In this way, many predictions can be issued per each event with a lead time of 1 h and the data set of forecast errors becomes extended enough to allow a reliable calibration. The outcome of this experiment can provide a first indication of the applicability of the proposed method to the Toce River case study.

[49] Since for the TT2 and TT4 events the shape of the flood hydrograph was affected by the retention operated by the reservoirs, the calibration of the uncertainty assessment method was performed by using the event TT1, with a river flow threshold of 80 m<sup>3</sup>/s, while events TT3, IOP02 and TT5 were used for the validation. The calibration procedure provided a good fit in term of normality and homoscedasticity of the regression residuals. The regression coefficients and the explained variance of the regression models for positive and negative errors are reported in Table 4. The percentage of positive errors resulted equal to 87%. Figure 4 shows the results of the validation operated on event TT3 and IOP02. Analogous results were obtained for event TT5.

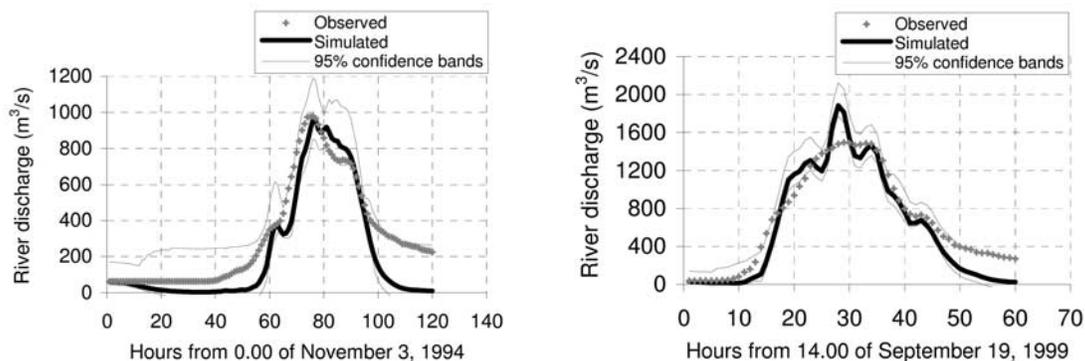
[50] Figure 4 shows that the CBs were estimated quite satisfactorily, even if only one flood event was used for

calibrating the uncertainty assessment method. One can note that the uncertainty resulted underestimated for the flow values around the peak of the event IOP02. This outcome is due to the underestimation of a few relevant negative errors close to the peak of the hydrograph, as a consequence of the limited percentage of negative errors in the calibration event. Once again, it is confirmed the importance of an accurate selection of the calibration database.

## 5. Concluding Remarks

[51] A statistical approach for assessing the total uncertainty of hydrological forecasts is proposed, which is based on the use of a probabilistic model of the forecast error. This is built by using a meta-Gaussian approach to infer the probability distribution of the forecast error on the basis of its dependence on selected explanatory variables. The procedure herein proposed was first tested by using an extensive synthetic data set. Subsequently, a second test was performed by using flood events observed on the Toce River basin, that were forecasted by a flood forecasting system. The synthetic case study is motivated by the need for an extended database in order to be able to perform an extensive test of the reliability of the method. The real-world case study is instead performed in order to prove that the method works well even with real data, which are not as well behaved as synthetic records.

[52] In both case studies the uncertainty assessment was satisfactory, even though both applications highlighted the relevant role played by the calibration data set. Rainfall-runoff models are often characterized by a significant variability of their performances and therefore it is important to calibrate the uncertainty assessment method on the basis of an extended and representative data set, in order to keep track of a comprehensive sample of past events and different hydrological conditions. As a matter of fact, a sizable variability of the rainfall-runoff model performances may reduce the reliability of the uncertainty assessment, especially if a short data sample is used for calibrating the method. One should note that the above requirement for an extended data set might be difficult to meet, as extreme events are by definition rare. This is a frequent problem of statistical methods for uncertainty assessment in hydrology, which is often emphasized by scientists who prefer to use non statistical approaches. The aim of this work is not to



**Figure 4.** Toce River case study. The 95% confidence bands of the forecast with 1-h lead time are shown for (left) event TT3, which occurred on 3 November 1994, and (right) event IOP02, which occurred on 19 September 1999.

contribute to this specific debate but only to propose a tool and discuss its potential limitations.

[53] Other limitations of the proposed technique are the need to recalibrate the coefficients of the regression if the lead time of the forecast changes and the incapability to explicitly take into account the uncertainty in the precipitation forecast, which is accounted for implicitly. However, the technique presented here might be applied to the situation in which precipitation ensemble forecasting is available. In such a case, an ensemble confidence bands can be built for the river flow forecasting in the form of a mixture of the forecast probability distributions of each member of the ensemble. Finally, we would like to remark that the computational requirements and times of the proposed technique are extremely limited and therefore it can be successfully applied in real-time forecasting.

[54] **Acknowledgments.** We are thankful to Scott Tyler, Nick van de Giesen, Günter Blöschl, and Sjur Kolberg for providing very useful comments when reviewing our 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 simulating and forecasting extreme hydrological events, with uncertainty analysis.”

## References

- Bacchi, B., and R. Ranzi (2003), Hydrological and meteorological aspects of floods in the Alps: An overview, *Hydrol. Earth Syst. Sci.*, 7, 784–798.
- Bacchi, B., G. Grossi, R. Ranzi, and A. Buzzi (2002), On the use of coupled mesoscale meteorological and hydrological models for flood forecasting in midsize mountain catchments: Operational experience and verification, in *Flood Defence '2002*, B. Wu et al., pp. 965–972, Science, New York.
- Beven, K. J. (2006a), A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36.
- Beven, K. J. (2006b), On undermining the science?, *Hydrol. Processes*, 20, 3141–3146, doi:10.1002/hyp.6396.
- Bougeault, P., P. Binder, A. Buzzi, R. Dirks, R. Houze, J. Kuettner, R. B. Smith, R. Steinacker, and H. Volkert (2001), The MAP special observing period, *Bull. Am. Meteorol. Soc.*, 82, 433–462, doi:10.1175/1520-0477(2001)082<0433:TMSOP>2.3.CO;2.
- Boyle, D. P. (2000), Multicriteria calibration of hydrological models, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson.
- Buzzi, A., and L. Foschini (2000), Mesoscale meteorological features associated with heavy precipitation in the southern Alpine region, *Meteorol. Atmos. Phys.*, 72, 131–146.
- Cook, R. D., and S. Weisberg (1994), *An Introduction to Regression Graphics*, John Wiley, New York.
- Cowpertwait, P. S. P. (1995), A generalized spatial-temporal model of rainfall based on a clustered point process, *Proc. R. Soc. London, Ser. A*, 450, 163–175.
- Franchini, M. (1996), Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall runoff models, *Hydrol. Sci. J.*, 41, 21–39.
- Grossi, G., and N. Kouwen (2004), Intercomparison among hydrologic simulations coupled to meteorological predictions provided by different mesoscale meteorological models, paper presented at 29th National Conference on Hydraulics and Hydraulic Works, Univ. of Trento, Cosenza, Italy, 7–10 Sept. (Available at [http://dicata.ing.unibs.it/grossi/ggmk\\_trento2004.pdf](http://dicata.ing.unibs.it/grossi/ggmk_trento2004.pdf))
- Kelly, K. S., and R. Krzysztofowicz (1997), A bivariate meta-Gaussian density for use in hydrology, *Stochastic Hydrol. Hydraul.*, 11, 17–31, doi:10.1007/BF02428423.
- Krzysztofowicz, R. (2001), The case for probabilistic forecasting in hydrology, *J. Hydrol.*, 249, 2–9, doi:10.1016/S0022-1694(01)00420-6.
- Krzysztofowicz, R. (2002), Bayesian system for probabilistic river stage forecasting, *J. Hydrol.*, 268, 16–40, doi:10.1016/S0022-1694(02)00106-3.
- Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, 11, 1267–1277.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology, *J. Hydrol.*, 330, 368–381, doi:10.1016/j.jhydrol.2006.04.046.
- Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 41(8), W08406, doi:10.1029/2004WR003826.
- Montanari, A. (2007), What do we mean by ‘uncertainty’? The need for a consistent wording about uncertainty assessment in hydrology, *Hydrol. Processes*, 21, 841–845, doi:10.1002/hyp.6623.
- Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 40, W01106, doi:10.1029/2003WR002540.
- Montanari, A., R. Rosso, and M. S. Taqqu (1997), Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation, *Water Resour. Res.*, 33, 1035–1044, doi:10.1029/97WR00043.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models 1: A discussion of principles, *J. Hydrol.*, 10, 282–290, doi:10.1016/0022-1694(70)90255-6.
- Ranzi, R., B. Bacchi, and G. Grossi (2003), Runoff measurements and hydrological modelling for the estimation of rainfall volumes in an Alpine basin, *Q. J. R. Meteorol. Soc.*, 129, 653–672, doi:10.1256/qj.02.60.
- Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou (1993), Frequency analysis of extreme events, in *Handbook of Hydrology*, edited by D. R. Maidment, pp. 18.1–18.66, McGraw-Hill, New York.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39(8), 1201, doi:10.1029/2002WR001642.
- Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5, 13–26.

G. Grossi, Faculty of Engineering, University of Brescia, Via Branze 43, I-25123 Brescia, Italy.

A. Montanari, Faculty of Engineering, University of Bologna, Via del Risorgimento 2, I-40136 Bologna, Italy. (alberto.montanari@unibo.it)