International IAHS-PUB Workshop 6-8 July, 2004 in Lugano, Italy
Uncertainty Analysis in Environmental Modeling

A generalised statistically-based approach for uncertainty assessment in rainfall-runoff modelling

Alberto Montanari

Faculty of Engineering, University of Bologna (alberto.montanari@unibo.it)
http://www.costruzioni-idrauliche.ing.unibo.it/people/alberto/indexeng.html

1. Introduction

This contribution presents a statistically based approach for uncertainty assessment in rainfall-runoff (R-R) modelling. The aim of the technique is to infer the probability distribution of the model error conditioned to a number of selected conditioning variables. These may include the current and previously simulated river flows as well as internal state variables of the model. The purpose is to indirectly relate the model error to the sources of uncertainty, through the conditioning variables.

The probability distribution of the model errors is derived in the Gaussian space, through a meta-Gaussian approach. The normal quantile transform is applied in order to make the series of the model errors and conditioning variables Gaussian.

2. Context of application and data requirements

The technique can be applied to any rainfall-runoff model, when this latter is run in simulation or in forecasting.

Uncertainty is estimated in a global form and it is not possible to separate the contribution of each source of approximation.

The proposed approach is calibrated by inferring the statistical properties of the model simulation errors. Therefore a series of observed versus simulated river flows is needed, in order to be able to obtain a record of simulation errors.

3. Description of the technique

Let us denote with the symbol R(t) a vector of explanatory variables at time t. Without loosing any generality, let us suppose that the rainfall-runoff model is run in real time in order to obtain a one-step-ahead forecasting of future river flows and that the vector of the explanatory variables is given by

R(t) = [S(t), E(t-1)],

where S(t) is the river flow simulated by the rainfall-runoff model at time t and E(t-1) is the model error at the previous time step (which is known since the model is run in real time).

The series S(t) and E(t) are in general non Gaussian. Let us denote with $P(E \le E(t))$ and $P(S \le S(t))$ the arbitrarily specified marginal cumulative probability distributions of E(t) and S(t), respectively, which are assumed to be strictly increasing and continuous. First, a standard normal quantile transform (NQT) is applied in order to make $P(E \le E(t))$ and $P(S \le S(t))$ Gaussian. The NQT is fully described by Kelly and Krzysztofowicz [1997].

The composition of the inverse Q-1 of the standard normal distribution and the marginal probability distribution $P(S \le S(t))$ defines the NQT of the original variate S(t), which will be referred to as NQT_S,

$$NS(t) = Q^{-1}\left[P(S \le S(t))\right]$$

in which N indicates that the variables are referred to the normalized space.

The cumulative probability $P(S \le S(t))$ of each observation of S(t) is approximated here with the corresponding sample frequency F(S(t)), which is estimated using the Weibull plotting position.

Therefore, the NQTS involves the following steps: (1) for each S(t) the cumulative frequency F(S(t)) is computed; (2) for each F(S(t)) the standard normal quantile NS(t) is computed and it is associated with the corresponding S(t). Thus, a discrete mapping of the NQT_S, is obtained. In order to be able to apply the inverse of the NQT_S, that is, NQT_S^{-1} for any $NS(t) \in \Re$, linear interpolation is used to connect the points of the discrete mapping previously obtained. The region beyond the minimum and maximum available NS(t) values is covered by linear extrapolation.

Once the normalized series NS(t) and NE(t) are derived by applying the respective transformations NQT_S and NQT_E, let us make the following hypotheses: (a) NS(t) and NE(t) are stationary and ergodic; and (b) the cross dependence between NS(t) and NE(t) is governed by the normal linear equation

$$NE(t) = C_1 NS(t) + C_2 NE(t-1) + N\varepsilon(t)$$

where C_1 and C_2 are constant coefficients and $N\varepsilon(t)$ is an outcome of the stochastic process $N\Theta(t)$ which is stochastically independent of NS(t) and $N\varepsilon(t-1)$ and is normally distributed with mean zero. Consequently, the conditional mean and variance of the R-R model error in the normalized space are

$$\mu\left(NE(t) \mid NS(t), NE(t-1)\right) = C_1 NS(t) + C_2 NE(t-1)$$

$$\sigma^2\left(NE(t) \mid NS(t), NE(t-1)\right) = 1 - C_1^2 - C_2^2$$

In applications, assumption (b) is subjected to testing.

Remembering that the random variable E(t) is the model error, one can compute the 95% CI for the simulated river flow s_t by

$$s_t^+ = s_t + \mathrm{NQT}_E^{-1}\left[\mu(NE(t) \mid NS(t)) + 1.96\sigma(NE(t) \mid NS(t))\right]$$

$$s_t^+ = s_t + \mathrm{NQT}_E^{-1}\left[\mu(NE(t) \mid NS(t)) - 1.96\sigma(NE(t) \mid NS(t))\right]$$

4. Goodness of fit checking

The meta-Gaussian approach described above presents some relevant statistical properties which are well summarized by Kelly and Krzysztofowicz [1997]. In particular, whereas the regression of NE(t) on the transformed explanatory variables is Gaussian, linear and with constant dispersion, the regression of E(t) on the explanatory variables in the natural space may be non Gaussian, non linear and with varying dispersion. This property can provide a first means for verifying the goodness of the fit provided by the meta-Gaussian model, which consists in verifying the hypotheses that condition the validity of the linear regression model of NE(t). The residuals of such a model are given by

$$N\varepsilon(t) = Ne(t) - C_1\, NS(t) - C_2 NE(t-1)$$

and should be Gaussian, homoscedastic, with mean 0 and variance $1-C_1^2-C_2^2$. Gaussianity can be verified by applying the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D) and Probability Plot Correlation Coefficient (C-C) tests [D'Agostino and Stephens, 1986; Stedinger et al., 1993] and by drawing the normal probability chart. For checking the linearity and homoscedasticity assumptions, the graphical approach proposed by Cook and Weisberg [1994] and used, in a similar context as here, by Krzysztofowicz and Kelly [2000] may be applied, which consists of drawing the residual plots of Ne_t. Under the target model, the residual plots will have no systematic features. If they are curved, then the regression function is non linear. If the plots are fan shaped or otherwise show systematic changes in variation across the plot, then the variance of $N\varepsilon(t)$ is not constant.

In practical applications the goodness of fit checking may suggest that some of the basic assumptions of the meta-Gaussian model are not satisfied. The lack of fit can be resolved by preliminarily transforming the model error accordingly to the techniquesuggested by Montanari and Brath (2004).

5. Some relevant features and limitations

The meta-Gaussian model proposed above does not impose any restriction on the marginal dependence structure of NE(t) and NS(t) and therefore it is not influenced by the presence of serial correlation. The reliability of the meta-Gaussian approach might be affected by the hypothesis of stationarity and ergodicity for NE(t) and NS(t), which is necessarily introduced in order to be able to estimate the statistical properties of such random variables on the basis of a finite and perhaps limited sample. In order to limit the resulting approximation, it is advisable to fit the meta-Gaussian model on an extended data set, which should include a wide variety of hydrological scenarios. [1]

The presence of non stationarity in the model errors can be qualitatively checked by carrying out a validation of the meta-Gaussian model results, that allows one to check their reliability in real world applications. It is significant to point out that non stationarity might be due to R-R model inadequacy or unreliability of input and output data. Therefore in these circumstances the lack of fit could be resolved by changing the R-R modeling approach or by reconsidering the hydrological data set. However, this is not always the case since non stationarity might be originated by several other reasons, for instance a variation of river banks management or a change in the watershed land-use. The meta-Gaussian approach does not allow to identify the reasons for the presence of non stationarity and therefore it is not a means for verifying the suitability of a given R-R model. It gives an estimation of the uncertainty associated to the simulations provided by an assigned model and parameter set. Its lack of fit, which might be due to the presence of non stationarity and therefore even to model inadequacy, only indicates that the R-R uncertainty cannot be reliably estimated.

It is worth remarking that different model structures affect the analysis only through their simulation errors. Unsatisfactory R-R models are just expected to be characterized by wider CIs but do not compromise the applicability of the proposed uncertainty estimation technique, unless non stationarity is induced in the model errors.

Particular care has to be taken when trying to apply the meta-Gaussian approach outside the range of st covered by the (st,et) couples used in the calibration phase. Like any extrapolation, this procedure may result in unsatisfactory approximations if non stationarity is present in the dependence structure and probability distribution of E(t) and S(t).

Finally, it has to be considered that the meta-Gaussian approach leads to estimating the uncertainty in an aggregated solution. Input and output uncertainty, as well as parameter uncertainty, are accounted for implicitly, without separating their individual contribution. Therefore the meta-Gaussian approach reliability is expected to be conditioned by changes in the different uncertainty sources. Variations in the reliability of input and output variables, for instance those due to modification of the gauging methods, are likely to affect the R-R model performances, as well as using synthetic input data instead of observed ones, as it is often done when estimating design variables. In all these cases the effects of non stationarity in the uncertainty sources should be carefully evaluated.

6. Application

The application presented here refers to the case of uncertainty assessment in the simulation of river flow data (no forecasting). Therefore E(t) is regressed on S(t) only. The dependence of E(t) on E(t-1) is not exploited since in simulation the previous model errors are not known. This application is fully described in Montanari and Brath (2004).

The simulation of synthetic river flows is carried out for the Samoggia river basin, which is located in northern Italy and it is a left bank tributary to the Reno River (Figure 1). The total area of the basin, closed at the river cross section of Calcara, is 178 km^2. The topography of Samoggia River basin is described by a DEM whose resolution is 250 x 250 m (Figure 2). An extensive data base of soil texture, relative permeability and organic substance content has been derived from the soil map, at 1:250.000 scale and in digital format, provided by the local administration. Finally available meteorological data base is constituted by the hourly rainfall depths observed at the raingauges of Montepastore, Monteombraro and Montesanpietro during the four-year period 1994-1997 (Figure 2). For the same period, both hourly temperature data recorded at Monteombraro and historical data of hourly river discharges recorded at Calcara are also at disposal.

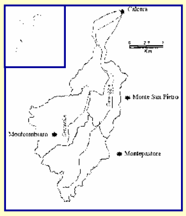

Figure 1. Location of the Samoggia River basin and the rainfall and river flow gauging stations.

Figure 2. Digital Elevation Model of the Samoggia River basin.

The rainfall-runoff model was calibrated by using data which refers to one flood event only, observed in 1996. The meta-Gaussian approach for uncertainty assessment was calibrated by using the river flow data collected in 1996-1997. Since the goodness of fit tests described in section 4 resulted not satisfied, a preliminary transformation of thre model error was applied (Montanari and Brath, 2004). Figure 3 shows the residual plot of NE(t) (after the transformation) on C_1NS(t).

The meta-Gaussian approach was validated by assessing the uncertainty of the simulated data which refers to the years 1994 and 1995. The computed confidence intervals were compared with the actual observations. Figure 4 shows the results of such comparison.

7. Comments

The proposed approach proved to be effective in the estimation of the uncertainty bands. His main advantages are:
1) It is statistically based and not subjective;
2) It is not computer intensive; its application is very simple;
3) It allows for investigating the sources of uncertainty through the analysis of the conditioning variables.
Its main disadvantages:
1) It needs to be calibrated by using some historical data;
2) It may be inffluenced by the presence of non-stationarity.

References

Beven, K.J., Prophesy, reality and uncertainty in distributed hydrological modeling, *Advances in Water Resources*, *16*, 41–51, 1993.
Beven, K.J., Towards an alternative blueprint for a physically based digitally simulated hydrologic response modeling system, *Hydrol. Proc.*, *16*, 189–206, 2002.
Beven, K.J., and A. Binley, The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Proc.*, *6*, 279–298, 1992.
Blaszkova, S., and K. Beven, Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), *Water Resour. Res.*, *38*, 1–14, 2002.
Cameron, D.S, K.J. Beven, J. Tawn, S. Blazkova, and P. Naden, Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty), *J. of Hydrol.*, *219*, 169–187, 1999.
Cook, R.D., and S. Weisberg, *An Introduction to Regression Graphics*, Wiley, New York, 1994.
Freer, J., K.J. Beven, and B. Ambroise, Bayesian uncertainty in runoff prediction and the value of data: an application of the GLUE approach, *Water Resour. Res.*, *32*, 2163–2173, 1996.
Kelly, K. S., and R. Krzysztofowicz, A bivariate meta-Gaussian density for use in hydrology, *Stochastic Hydrol. Hydraul.*, *11*, 17–31, 1997.
Kelly, K.S., and R. Krzysztofowicz, Precipitation uncertainty processor for probabilistic river stage forecasting, *Water Resour. Res.*, *36*, 2643–2653, 2000.
Krzysztofowicz, R., Bayesian system for probabilistic river stage forecasting, *J. Hydrol.*, *268*, 16–40, 2002.
Montanari, A., R. Rosso, and M.S. Taqqu, Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation, *Water Resour. Res.*, *33*, 1035–1044, 1997.
Montanari, A., Brath, A., 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. Water Resour. Res., 40, doi:10.1029/2003WR002540.
Montanari, A., 2004. Uncertainty assessment in rainfall-runoff modelling. A review. Proceedings of the workshop "Metodi statistici e matematici per l'analisi delle serie idrologiche", (Statistical and mathematical methods for the analysis of hydrological time series), Naples May 7 2004, edited by D. Piccolo e Lucio Ubertini, publication n. 2854 of the National Research Council of Italy, National Group for the Prevention of the Hygeological Disasters (available for download at http://www.costruzioni-idrauliche.ing.unibo.it/people/albert/indexeng.html).
Vrugt, J.A., H.V. Gupta, W. Bouten, and S. Sorooshian, A shuffled complex evolution Metropolis Algorithm for optimisation and uncertainty assessment of hydrological model parameters, *Water Resour. Res.*, *39*, 1201, doi:10.1029/2002WR001642, 2003.

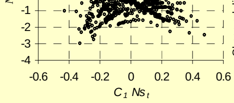

Figure 3

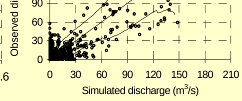

Figure 4