The Real-time Stochastic Flow Forecast

Lucie BŘEZKOVÁ, Miloš STARÝ and Petr DOLEŽAL

Brno University of Technology, Faculty of Civil Engineering, Brno, Czech Republic

Abstract: In the Czech Republic, deterministic flow forecasts with the lead time of 48 hours, calculated by rainfall-runoff models for basins of a size of several hundreds to thousands square kilometers, are nowadays a common part of the operational hydrological service. The Czech Hydrometeorological Institute (CHMI) issues daily the discharge forecast for more than one hundred river profiles. However, the causal rainfall is a random process more than a deterministic one, therefore the deterministic discharge forecast based on one precipitation prediction is a significant simplification of the reality. Since important decisions must be done during the floods, it is necessary to take into account the indeterminity of the input meteorological data and to express the uncertainty of the resulting discharge forecast. In the paper, a solution of this problem is proposed. The time series of the input precipitation prediction data have been generated repeatedly (by the Monte Carlo method) and, subsequently, the set of discharge forecasts based on the repeated hydrological model simulations has been obtained and statistically evaluated. The resulting output can be, for example, the range of predicted peak discharges, the peak discharge exceeding curve or the outflow volume exceeding curve. The properties of the proposed generator have been tested with acceptable results on several flood events which occurred over the last years in the upper part of the Dyje catchment (Podhradí closing profile). The rainfall-runoff model HYDROG, which has been in operation in CHMI since 2003, was used for hydrological simulation.

Keywords: discharge forecast; flow; hydrological model; Monte Carlo; operative hydrological forecast; rainfall-runoff model; stochastic

Czech Hydrometeorological Institute (CHMI) – the national hydrological service – issues daily deterministic hydrological forecasts with lead time of 48 h for more than one hundred watergauge stations. The discharge forecasts for the basins of size several hundreds to thousands square kilometers are calculated by rainfall-runoff models. The basic input for hydrological model is a rainfall and temperature prediction, which is issued according to the synoptic situation by meteorologists. Often the output of the numerical weather prediction model ALADIN (NWP ALADIN, ALADIN International Team 1997), operated by CHMI, is a decisive information. The error of the precipitation forecast is usually the main factor that influences the accuracy of the final flow forecast – at least when we speak about the forecasting of summer floods caused by heavy precipitation. The importance of accurate precipitation forecast is described for example in (Rabuffetti *et al.* 2008).

The deterministic discharge forecast based on one rainfall scenario is a great simplification of the real situation – the indeterminations, which influence the whole calculation process, are not expressed in the final discharge forecast. When significant decisions are to be made according to an actual discharge forecast (e.g. evacuation of inhabitants in the flood-threatened area), it is necessary to estimate the hazard factor.

The stochastic discharge forecast is nowadays a very actual topic. The transformation of meteorological uncertainties to the hydrological uncer-

Supported by the Czech Science Foundation, Grant No. 103/07/1620.

tainty is solved by many (local and international) projects (e.g. COST 731 – Concerted Research Action 731 – Propagation of uncertainty in advanced meteo-hydrological forecast systems). The EFAS (European Flood Alert System, see Ramos *et al.* 2007), which issues discharge forecasts for main European rivers, uses 50 meteorological scenarios of the global NWP (numerical weather prediction) model ECMWF for calculation of the stochastic discharge forecast. The set of scenarios such that each of them is calculated with a little bit different initial conditions is called an ensemble. The problem is that the space grid of the ECMWF model is about 80 km at mid latitudes, which is too coarse for the conditions of the Czech Republic; hence, ECMWF is capable of predicting only "large-scale" events. On the other hand, the meteorological ensemble (a set of several model runs which are calculated with a little bit different initial conditions) from local NWP models caanot be calculated operationally due to the long calculation time (one run of the ALADIN model takes about 4–5 h). It is also possible to estimate the precipitation forecast uncertainty by an analogy method (where the set of meteorological forecasts is derived from similar synoptic situations in the past), although the actual NWP models usually give better results (Diomede *et al.* 2006).

The statistical postprocessing of the predicted rainfall represents another option. The testing of a stochastic precipitation nowcast scheme (based on the "downscaling" of the precipitation forecast in such a way that it is statistically consistent with the recent radar-inferred precipitation fields) used for hydrological modelling is described in (Pierce *et al.* 2005) and shows that the statistical postprocessing of the NWP precipitation forecast is a way how to express the uncertainty of the flow prediction. The uncertainty of meteorological inputs can be expressed by repeated generation of meteorological scenarios with certain statistical properties (the Monte Carlo method). The set of corresponding river discharges obtained by repeated simulations (using a hydrological model) is then evaluated statistically. The peak discharge exceeding curve or the range of peak discharges are examples of the resulting stochastic discharge forecast. This method can be used for all meteorological quantities used as inputs of a hydrological model – rainfall, temperature and snow cover.

In this paper, a stochastic generator which produces the variants of input data scenarios based on several statistical parameters is described. The problem is solved from the point of view of a hydrological forecasting service. The attention is paid to the uncertainties of meteorological input data, because the routine experience confirms that these uncertainties usually create the main portion of the whole uncertainty of the rainfall-runoff simulation. The effect of the generator is tested in the conditions of the Dyje catchment. For hydrological simulation the HYDROG model (Starý & Tureček 2000) was used. This semi-distributed rainfall-runoff model is a software for simulation, operative prediction and operative control of water runoff from the catchment with reservoirs. It is routinely used by the regional offices Brno and Ostrava of the Czech Hydrometeorological Institute (Šálek *et al.* 2006) and by the Odra, Labe and Ohře Water Authorities.

METHOD

The uncertainties of discharge forecasts calculated by hydrological models

The process of creation of a discharge forecast proceeds under the conditions of uncertainty composed of:

- The choice of a hydrological model – e.g. the use of hydrological and hydraulic equations, parameterisations and concepts, which always simplify the reality. These simplifications express our ignorance of detailed natural conditions (e.g. soil properties). On the other side, these simplifications are necessary for the possibility of using a certain hydrological model in real time, when the quickness of the calculation process is one of the most important factors.
- Uncertainty of measured and predicted input data – e.g. the inaccuracy of measured and predicted meteorological data and also the simplification of the distributions of input quantities in space and time for the needs of a hydrological model (for example the prediction of the average rainfall sum for a certain part of the catchment). These types of uncertainties can be simulated by the proposed stochastic generator.

The uncertainty of the measured input data

The measured quantity can be considered as a random quantity with normal distribution. The meteorological quantities are usually not provided with an estimation of uncertainty. Their distribution is usually unknown (only one realization of the random quantity is available, the measurement is available in several points only, we do not know the spatial validity of the measured values etc.). In such a case usually the normal distribution is used. Since we usually have an insufficient set of measured data, its uncertainty can be expressed as the expanded standard uncertainty (± 3σ), estimated from the range of the data set. Sometimes the evaluation of historical data can help in estimating σ.

The uncertainty of the predicted input data

The predicted quantity can be obtained from various outputs of NWP models. The number of these outputs accessible to the operational hydrological service is very small and the accuracy of each NWP model is different (this is caused by different simplifications used in the models). Moreover, the forecast can be adjusted based on the experience of the meteorologist. Under these ambiguous conditions with a lot of independent effects influencing the resulting forecast, it is possible to assume the predicted quantity as a normally distributed one. Again, the uncertainty of the forecast can be expressed as the expanded standard uncertainty (± 3σ), estimated from the range of the set of forecasts or by the meteorologist.

Remark: The normal distribution can be replaced by another one, when such information is routinely available to the hydrological service. Then the proposed generator can be adapted to another kind of distribution – e.g. the skewed lognormal distribution.

The generator of a random vector

Let us assume the rainfall over the catchment measured by several raingauges. It is possible to consider the rainfall intensity as a random quantity with a normal distribution. Since the rainfall intensity varies in time, it is possible to speak about a random process. At every time step the measured values of the precipitation intensity create a random vector, the members are which correlated to each other. The uncertainty of rainfall can be mimicked by repeatedly generating this random vector.

Based on the assumptions mentioned above, the following properties of the generator are postulated (Starý 1985):

- The measured or predicted quantity (represented by a generated random vector) follows a normal distribution.
- The uncertainty of the generated quantity can be estimated as the expanded standard uncertainty (± 3σ).

The mathematical model was inspired by (Smith & Freeze 1979). Its modified version was used in (Starý 1983, 1985), where it was applied to the testing of reliability of silt cores of dams.

Mathematical description of the generator

Let us assume the plain area Ω divided to triangles with vertices $P_1, P_2, ..., P_n$. With the use of the autocorrelation theory (Smith & Freeze 1979), it is possible to write:

$$X_n = a_{1,n} \times X_1 + a_{2,n} \times X_2 + a_{3,n} \times X_3 + ... + a_{n-1,n} \times \times X_{n-1} + \varepsilon \qquad (1)$$

where:

X_i – values of a random variable in the vertices $P_1 ... P_n$

ε – normally distributed random variable

$a_{1,n} - a_{n-1,n}$ – autoregressive parameters expressing the degree of dependence of X_n on its neighbours X_1 to X_{n-1}

A typical location of the vertices $P_1, P_2, ..., P_n$ is depicted on Figure 1.

The Eq. (1) can be written for the total of n vertices as a set of n linear equations:

$$X = W \times X + \varepsilon \qquad (2)$$

where:

X – generated vector of random variables

ε – normally distributed random vector

W – matrix, which defines the autoregression (in this case, the linear interdependence) of X in the vertices $P_1, ..., P_n$; the autoregressive parameters $a_{i,j}$ are members of the matrix W

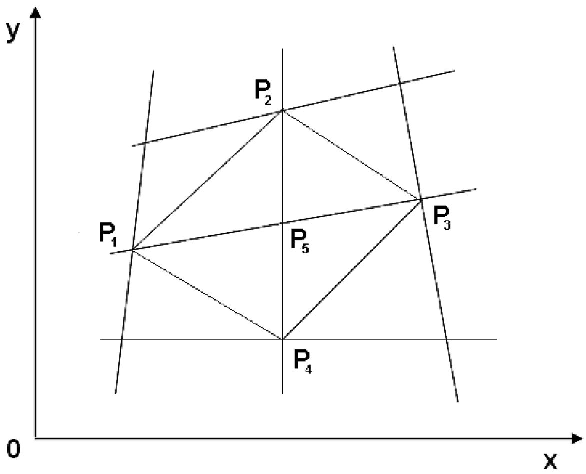

Figure 1. Example of a triangular division location of division of area Ω and location of vertices P in which the values of a random variable X are considered (n = 5)

It is possible to transform the Eq. (2) to the following (Starý 1985):

$$X = \mu_X + W \times X + \alpha \times \varepsilon \quad (3)$$

where:

μ_X – mean of X

α – matrix (derived from the relationship between the standard deviation σ_X and the correlation and covariance matrices of the vector X) which ensures the required value of the mean μ_X and the standard deviation σ_X of X

Then the Eq. (3) can be solved:

– Firstly, we need to know the matrix W, which can be calculated by the procedure described below. The other known inputs are σ_X and μ_X.
– The random vector ε is generated from the standard normal distribution N (0, 1) (in our case we were using the Delphi generator).
– With the use of Eq. (3) we obtain an autocorrelated vector X with the distribution $N(\mu_X, \sigma^2_X)$.

The generator creates the autocorrelated vector X for each time step. Since the input data for a hydrological model are correlated not only in space, but also in time, it is proper to correlate the following vector X with the previous one. For example, it is possible to keep the same normal random vector ε for all time steps or to change ε in time "slowly" (this must be tested during the operation in order to find the most appropiate method). The time autocorrelation of the vector X is then approximately modelled by the time autocorrelation of the input set of mean values μ_X.

In the following text, the methodology of calculation of the matrix W is described.

The Eq. (1) can be written for each raingauge station i in the form:

$$X_i = a_1 \times X_1 + a_2 \times X_2 + a_3 \times X_3 + \ldots + a_k \times X_k + \varepsilon_i \quad (4)$$

where:

k – number of neighbours surrounding the point i

$a_1 - a_k$ – autoregressive parameters, which express the dependence of X_i on X_1–X_k

If the set of data measured by all raingauges is known, the autoregressive parameters a_1 to a_k (members of the matrix W) in Eq. (4) can be solved with the use of the Yule-Walker equations (Eshel 2009):

$$\begin{pmatrix} 1 & r(1) & \ldots & r(k-2) & r(k-1) \\ r(1) & 1 & r(1) & \ldots & r(k-2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r(k-2) & \ldots & r(1) & 1 & r(1) \\ r(k-1) & r(k-2) & \ldots & r(1) & 1 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_{k-1} \\ a_k \end{pmatrix} = \begin{pmatrix} r(1) \\ r(2) \\ \ldots \\ r(k-1) \\ r(k) \end{pmatrix} \quad (5)$$

where:

r (1) – correlation coefficient between the data of the station i and those of the surrounding station 1

r (k – 1) – correlation coefficient between the data of the station i and those of the surrounding station (k – 1), etc.

In such a way, we obtain i^{th} row of the matrix W. If the Eq. (5) is solved for each raingauge station, we obtain the whole matrix W. A particular autoregressive parameter equals zero in the case that there is no correlation among the stations involved.

Table 1. The basic characteristics of the Dyje catchment

Catchment area (km^2)	1756
Highest elevation (m a.s.l.)	Javořice – 836
Lowest elevation (m a.s.l.)	Podhradí, water gauge – 349
Average annual outflow (m^3/s)	8.5
100 years return time period peak discharge (Q_{100}) (m^3/s)	390

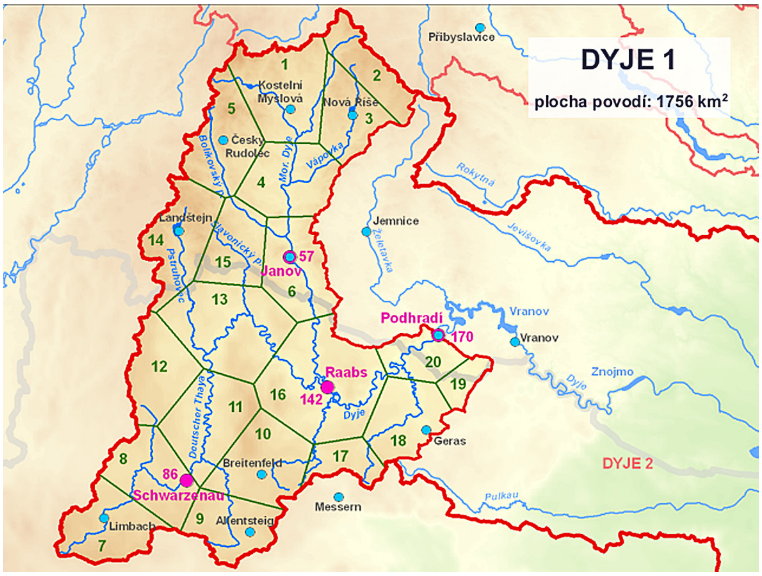

Figure 2. The map of the upper part of the Dyje catchment (with the closing profile Podhradí). For the purpose of hydrological modelling, the catchment is divided into 20 Thiessen polygons, within which the input quantity (precipitation, temperature or snow level) is considered as a constant value

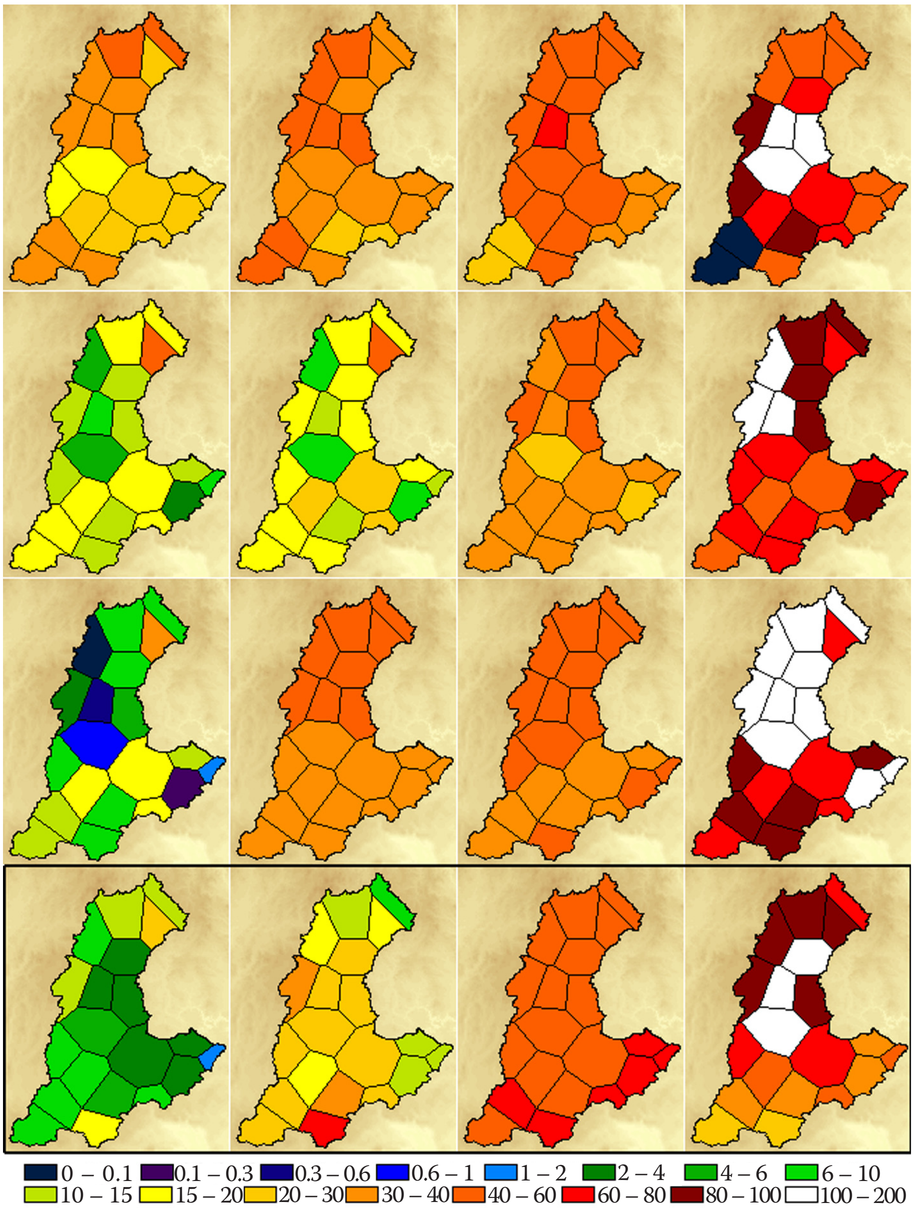

Figure 3. The example of generated and measured 48-hour rainfall sums (mm) in the Dyje catchment; the first three rows depict the randomly selected generated rainfall events derived from the predicted precipitation on 29th June 2006 06 UTC; the last row depicts the measured 48-hour rainfall sums (from left to right 28.–30. 7. 2006, 26.–28. 6. 2006, 6.–8. 8. 2006, 29.–1. 7. 2006)

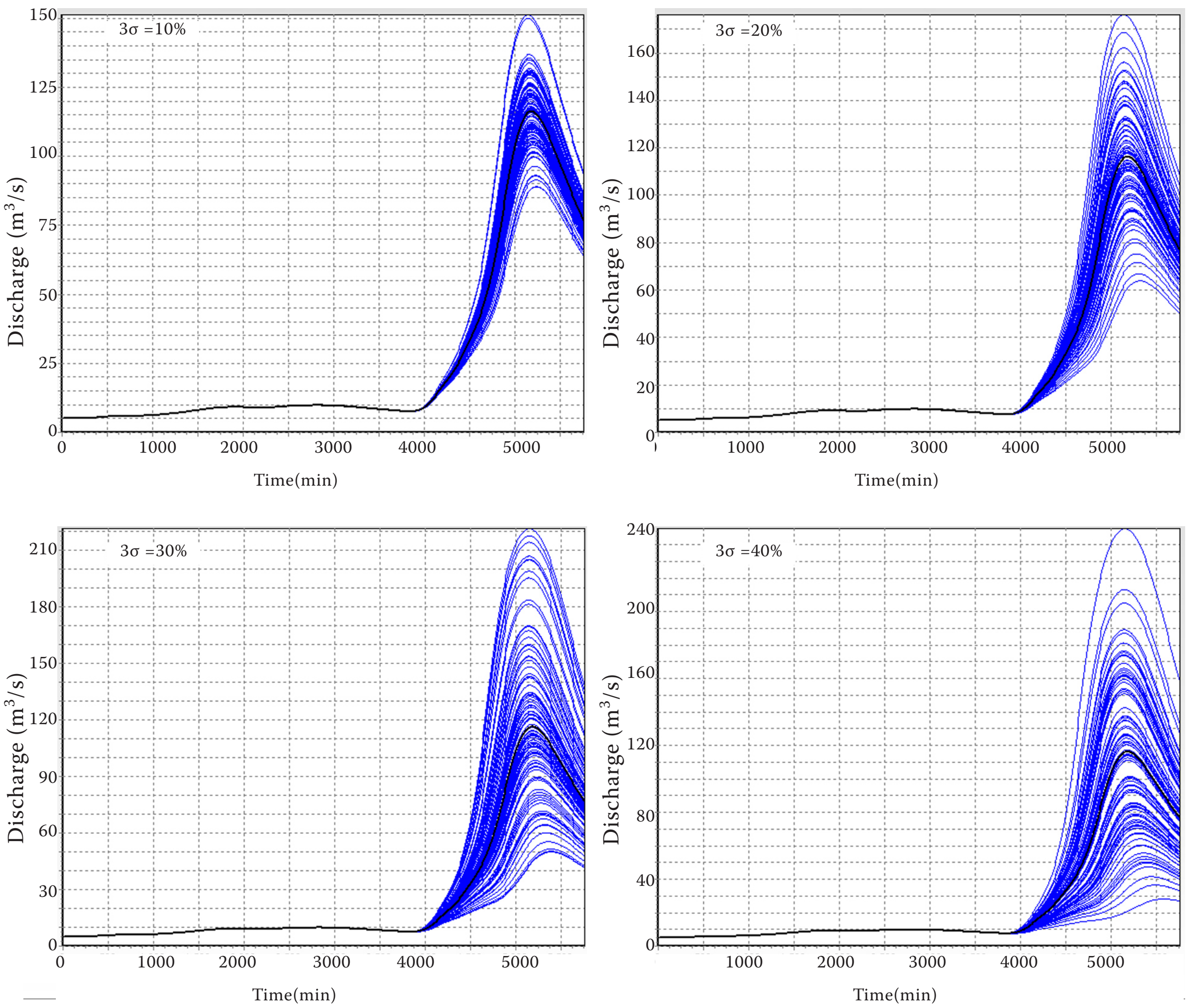

Figure 4. Stochastic discharge forecasts based on 100 generated precipitation scenarios. The expanded standard deviation (3σ) equals 10, 20, 30 and 40% of the mean (deterministic) predicted precipitation, respectively

The matrix *W* is different for different types of meteorological events. For example, the stratiform precipitation type is characterized by stronger correlation relationships then the convective one. Based on the historical data, various types of matrices *W* can be calculated. The choice of the proper matrix *W* according to the expected precipitation event can be very important in the discharge forecast process. Its relevance will be one of the subjects of future work in which the proposed generator will be tested in operation.

RESULTS AND DISCUSSION

The proposed generator was programmed by the authors in Delphi. It was tested for the Dyje catchment (Figure 2). The basic parameters of the catchment upstream of the Podhradí profile are given in Table 1. Deterministic discharge forecasts based on the HYDROG model have been issued daily for several river profiles since 2003.

A test of the generator was based on the predicted rainfall event starting from 29. 6. 2009 06 UTC. The correlation matrix *W* was derived from the historical hourly precipitation sums (2002 to 2009) higher than 0.5 mm and lower than 5 mm. Quantitative precipitation estimates based on the combination of radar and raingauge measurement were used (Šálek *et al.* 2004). The examples of generated 48-hour precipitation sums patterns are depicted on Figure 3. The examples of several types of the measured 48-hour precipitation sums are also presented on Figure 3. Comparing

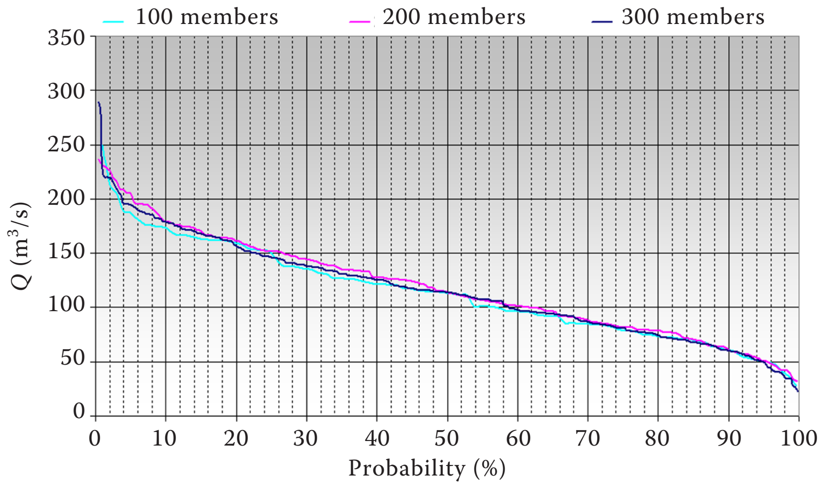

Figure 5. Comparison of the peak discharge exceeding curves derived from stochastic discharge forecasts based on 100, 200 and 300 generated rainfall scenarios. The expanded standard deviation (3σ) was equal to 40% of the mean (deterministic) predicted precipitation

both types of precipitation patterns, we see that generated precipitation sums resemble the real ones quite well.

In the next example, the predicted precipitation input data (100 scenarios) were generated with the expanded standard deviation 3σ considered as 10%, 20%, 30% and 40% of the mean (μ_x), e.g. of the deterministic precipitation forecast (which was about 50 mm per 24 h). The resulting stochastic discharge forecasts are depicted on Figure 4. Obviously,

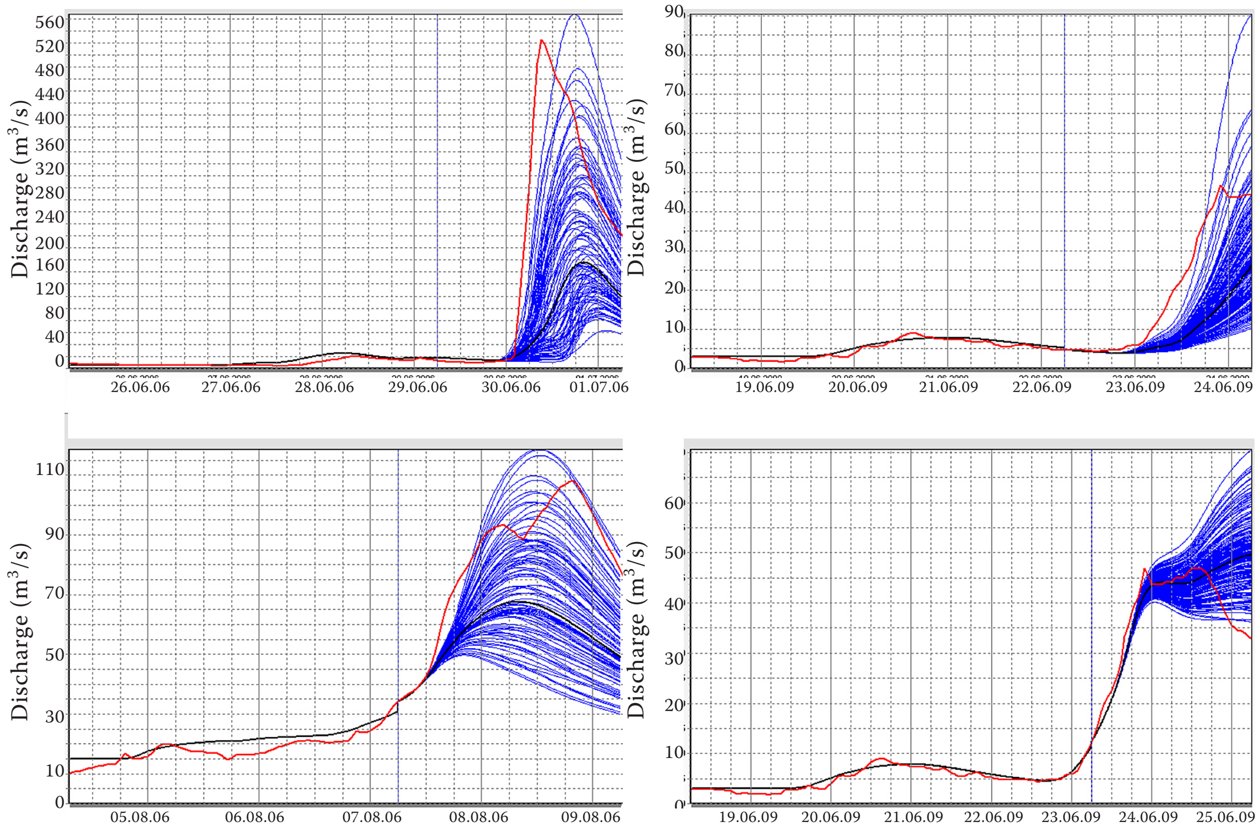

Figure 6. Examples of simulations of the operative stochastic flow forecast (100 runs) in the Podhradí river profile from 29. 6. 2006 06 UTC (left up, μ_X = 45 mm, $3\sigma_X$ = 45 mm), 7. 8. 2006 06 UTC (left bottom, μ_X = 10 mm, $3\sigma_X$ = 7 mm), 22. 6. 2009 06 UTC (right up, μ_X = 38 mm, $3\sigma_X$ = 16 mm) and 23. 6. 2009 06 UTC (right bottom, μ_X = 26 mm, $3\sigma_X$ = 3 mm); the precipitation inputs were generated with the given expanded standard deviations 3σ estimated by meteorologists for the actual synoptic situations; the black line expresses the deterministic discharge forecast, the red line means the actual development of the discharge (measurement)

the range of the predicted peak discharges grows together with the increasing expanded standard deviation – as it was expected.

The sensitivity of the proposed generator to the number of generated scenarios is depicted on Figure 5, where the peak discharge exceeding curves derived from 100, 200 and 300 members (discharge forecasts based on generated scenarios) are depicted. Only the extremes (the highest discharges) can be underestimated when the number of the rainfall scenarios is small (such as, for example, 100 generated members).

Figure 6 shows the examples of stochastic discharge forecasts in Podhradí based on real rainfall predictions given by meteorologists from CHMI (events from 29. 6. 2006 06 UTC, 7. 8. 2006 06 UTC, 22. 6. 2009 06 UTC and 23. 6. 2009 06 UTC). The black line depicts the deterministic discharge forecast (which usually comes from the ALADIN precipitation forecast, but can be corrected by meteorologists), the blue lines mean the discharge variants based on the generated rainfall scenarios (for which the expanded standard deviation was estimated by meteorologists from the synoptic situation and the precipitation predictions from all available NWP models). The red line expresses the actual development of the discharge. It is obvious that the deterministic discharge forecast cannot describe the uncertainties of the predicted rainfall.

CONCLUSION

The algorithm for the calculation of a short-time operative stochastic discharge forecast by the Monte Carlo method with the use of a hydrological prediction model was presented. The proposed generator of the input data for the hydrological model HYDROG was tested with acceptable results. The presented work is original and offers new possibilities in the operative discharge forecasting. However, this is only a first step. The real-time operation will bring new experience and also new requirements.

The stochastic discharge forecast gives much more credible information about the discharge development in the catchment than the deterministic one. The probable range of the peak discharge values is important information for the crisis management. The peak discharge exceedance curve expresses the hazard associated with the decision (e.g. about a particular way of reservoir operation based on the selected inflow scenario).

The first results also show that this tool can be used in operation. The main advantage is the sufficient quickness of the calculation - the resulted forecast must be available in real time (the calculation time of 100 simulation runs is several minutes).

The presented generator makes it possible to generate not only rainfall scenarios, but also temperature and snow scenarios, which are necessary for winter flood forecasting (these results will be presented in another paper).

The proposed algorithm will now be tested in real time in the Brno regional office of the Czech Hydrometeorological Institute.

Acknowledgements. The authors thank the Czech Hydrometeorological Institute for data support.

References

ALADIN International Team (1997): The ALADIN project: Mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research. WMO Bulletin, **46**: 317–324.

Diomede T., Marsigli C., Nerozzi F., Paccagnella T., Montani A. (2006): Quantifying the discharge forecast uncertainty by different approaches to probabilistic quantitative precipitation forecast. Advances in Geosciences, **7**: 189–191.

Eshel G. (2009): The Yule Walker Equations for the AR Coefficients. Wharton University of Pennsylvania. Available at http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf (accessed April 8, 2010)

Pierce C., Bowler N., Seed A., Jones A., Jones D., Moore R. (2005): Use of a stochastic precipitation nowcast scheme for fluvial flood forecasting and warning. Atmospheric Science Letters, **6**: 78–83.

Ramos M.H., Bartholmes J., Thielen-del Pozo J. (2007): Development of decision support products based on ensemble weather forecasts in the European Flood Alert System. Atmospheric Science Letters, **8**: 113–119.

Rabuffetti D., Ravazzani G., Corbari C., Mancini M. (2008): Verification of operational Quantitative Discharge Forecast (QDF) for a regional warning system – the AMPHORE case studies in the upper Po River. Natural Hazards and Earth System Science, **8**: 161–173.

Smith L., Freeze R.A. (1979): Stochastic analysis of steady state groundwater flow in a bounded domain. 2. Two-dimensional simulations. Water Resources Research, **15**: 1543–1559.

Starý M. (1983): Silt cores reliability of rockfill dams. [Ph.D. Thesis.]. Brno University of Technology, Brno. (in Czech)

Starý M. (1985): Autocorrelation theory application for the generation of random parameters in discrete points of a plane area. In: Sborník VUT. Brno University of Technology, Brno, 800–805.

Starý M., Tureček B. (2000): Operative control and prediction of floods in the River Odra basin. In.: Flood Issues in Contemporary Water Management. NATO Science Series, 2. Enviromental Security, Vol. 71. Kluwer Academic Publishers, Dordrecht, 229–236.

Šálek M., Novák P., Seo D.-J. (2004): Operational application of combined radar and raingauges precipitation estimation at the CHMI. In: ERAD 2004 Proceedings. ERAD Publication Series, Vol. 2, Copernicus GmbH, Visby, 16–20.

Šálek M., Březková L., Novák P. (2006): The use of radar in hydrological modelling in the Czech Republic – case studies of flash floods. Natural Hazards and Earth System Sciences, **6**: 229–236.

COST 731: Concerted Research Action 731 – Propagation of uncertainty in advanced meteo-hydrological forecast systems. Available at http://cost731.bafg.de

Czech Hydrometeorological Institute. Available at www.chmi.cz

ECMWF: European Centre for Medium-Range Weather Forecasts. Available at http://www.ecmwf.int/

Received for publication May 3, 2009
Accepted after corrections April 8, 2010

Corresponding author:

Ing. Lucie Březková, Vysoké učení technické v Brně, Fakulta stavební, Ústav vodního hospodářství krajiny, Žižkova 17, 602 00 Brno, Česká republika
tel.: + 420 541 147 770, e-mail: lucie.brezkova@chmi.cz