UNCERTAINTY MODELLING IN RAINFALL-RUNOFF SIMULATIONS BASED ON PARALLEL MONTE CARLO METHOD

M. Golasowski, *M. Litschmannová*, *Š. Kuchař*, *M. Podhorányi*, *J. Martinovič*

Abstract: This article describes statistical evaluation of the computational model for precipitation forecast and proposes a method for uncertainty modelling of rainfall-runoff models in the Floreon$^+$ system based on this evaluation. The Monte-Carlo simulation method is used for estimating possible river discharge and provides several confidence intervals that can support the decisions in operational disaster management. Experiments with other parameters of the model and their influence on final river discharge are also discussed.

1. Introduction

Rainfall-runoff (R-R) model is a dynamic mathematical model that transforms rainfall to the flow at the catchment outlet. The main purpose of the model is to describe rainfall-runoff relations of a catchment area. Common outputs of the model are surface runoff hydrographs, which depict relations between discharge Q and time t. In many countries, R-R models are usually used for predicting surface runoff within river catchments. One of their inputs is an information about weather conditions in the near future. These data are provided by numerical weather forecast models, such as ALADIN (Aire Limite, Adaptation Dynamique, Development International) [5]. However, weather forecast models are affected by errors that can severely affect precision of the modelling results and the magnitude of the error depends on a lot of factors such as wind, topography, temperature or humidity. One of the most sensitive components of the R-R models is the forecast rainfall intensity. As the rainfall intensity is a key element of rainfall-runoff modelling, we

---

*Martin Golasowski – Corresponding Author, Jan Martinovič, VŠB – Technical University of Ostrava, FEECS, Department of Computer Science, and IT4Innovations National Supercomputing Center, E-mail: `martin.golasowski@vsb.cz`, `jan.martinovic@vsb.cz`

†Martina Litschmannová, VŠB – Technical University of Ostrava, FEECS, Department of Applied Mathematics, E-mail: `martina.litschmannova@vsb.cz`

‡Štěpán Kuchař, Michal Podhorányi, VŠB – Technical University of Ostrava, IT4Innovations National Supercomputing Center, E-mail: `stepan.kuchar@vsb.cz`, `michal.podhoranyi@vsb.cz`

have to take uncertainties of the rainfall forecast model into account. Historical comparison (observed state against forecast) and assessment of rainfall intensity is one of the main goals of this article. We focused on developing a method that could provide additional information about rainfall intensity uncertainty to the R-R models and how to enhance the results of these models based on this information.

Such enhancement is very important for decision support in disaster management as it provides probabilistic view on possible scenarios that can happen in the near future. If the rainfall forecast model underestimates the precipitation rate, then the rainfall-runoff results will be very optimistic and the disaster management system will not alert the authorities about the upcoming flood until it is too late. As we are also developing a flood prediction system called Floreon$^+$ [11, 20] for local and regional authorities in the Moravian-Silesian region of the Czech Republic, the results will be used in the flood warning process. Such deployment demands high precision and detailed information, but also high availability and fast response mainly during emergency situations when the flooding is imminent or ongoing.

Overview of similar articles and studies related to the R-R uncertainty modelling is located in Section 2. Following sections 3 and 4 provide brief introduction to the R-R process and describe possible sources of uncertainty in its modelling methods. These sections also describe the *Math1D* model, our in-house R-R model used for experiments in this article. Sections 5 and 6 are focused on statistical analysis of the precipitation forecast errors. Application of the Monte Carlo method for modelling the forecast error is proposed in Section 7. Section 8 describes parallel implementation of the proposed modelling method. Results of the experiments and demonstration of the proposed modelling method are located in Section 9. The last Section 10 concludes this article and discusses presented results and future research.

2.   Related work

Precision of the precipitation forecast generated by a more general ECMWF model was analysed by Kobold and Sušelj [8]. Precision of the forecast for selected catchments during major precipitation events occurring in Northern Slovenia was analysed using the forecast as input for the R-R model. This article describes errors in precipitation forecasts as a major source of possible uncertainty and shows that the ECMWF model tends to underestimate the real situation by 60% in average.

Analysis of possible uncertainties in the output of regional climate models (RCM) was performed by Holtanova et al. [6]. The analysis showed that the largest source of uncertainty in RCM is the driving global climate model and that performance of the models differs through different seasons.

Crhová et al. [4] provided another study focused on the performance of the ALADIN model in the Czech Republic. Study compares the ALADIN model with thirteen other RCMs. Results of this analysis point out that ALADIN performs best in the terms of the monthly precipitation forecast totals for Czech Republic. ALADIN however tends to overestimate precipitation totals according to the study. This contradicts our results presented later in Section 5, which showed that the model tends to slightly underestimate the precipitation totals. This could be a result of a comparably smaller but more recent data set used in our analysis, which

is focused only on a small region of the Czech Republic, while the other study [4] used historical data between years 1961–1990 for a much larger area.

2.1 Rainfall-runoff uncertainty estimation methods

Montanari [13] recognizes four types of uncertainty modelling methods. The first type estimates uncertainty by utilizing approximate analytical methods. However those methods can be difficult to implement in practice due to the fact that statistical properties of the models and of the system itself tend to be hard to determine [14, 18].

Another type of methods focuses on the statistical analysis of errors computed by comparing simulation results with observed data. Precision of this approach could be affected by precision of the observed data themselves or by possible unreliable behaviour on scenarios that were not present in the initial data set. The second cause is crucial for simulation result reliability of short-term predictions in disaster management [14].

The third type of methods uses random sampling of the input space to determine statistics of the model output. If statistical properties of the input space are known, then a random sets of the input parameters can be generated and used as input for simulations. Outputs of these uncertainty simulations can then in turn be used for further analysis.

This article focuses on utilization of this type of methods while maintaining time-efficiency for their application in disaster management environment by using tools and methods available in high performance computing (HPC).

The last type of methods involves random set theory or fuzzy set theory and tries to see the R-R modelling problem as a possibilistic rather than probabilistic one [13].

2.2 Methods based on the analysis of model inputs

One of the widely used methods for analysing model inputs is the Generalized Likelihood Uncertainty Estimation (GLUE) method proposed by Beven and Binley in 1992 [12].

This method compares performance of different models and different sets of its parameters by comparing their individual results against observed data. Comparison of simulated hydrographs is done using common likelihood measures such as the *Nash-Sutcliffe coefficient* [15]. Measure is determined for each time-step value and then normalized to have the cumulative sum of 1.0. Using cumulative probabilities as weights, the cumulative distribution function can be constructed and confidence intervals can be easily derived. This method provides tools for comparing different modelling solutions and their parameters between each other [12]. A different method is the Bayesian Forecasting System (BFS) which was applied by Krzysztofowicz on flood forecasting systems in 1999 [10]. This approach recognizes two sources of uncertainty. The first source is the hydrological uncertainty, which accounts for parameter lumping, model structure, data errors and other related sources. The second source is the precipitation uncertainty, that is defined by precipitation forecast errors. These sources are then combined together

by the integrator processor which integrates both sources in real-time and produces probability distribution of predicted R-R model outputs.

Our proposed method models the input forecast data by non-parametric methods using kernel density estimation. This approach is useful because no prior knowledge about the forecast error probability distribution is needed and only general assumptions about its properties have to made. However, a time-expensive generation of input data sets and model execution can be seen as a trade-off, as the resulting confidence intervals are obtained via Monte Carlo simulation with a considerable amount of iterations. Parallel implementation of this method is described later in Section 8 as a possible solution for speeding up the process and maintaining fast response times of these simulations.

3. Rainfall-runoff process and its simulation

In this section, we introduce some important terms used when discussing the rainfall-runoff process. An accurate estimate of runoff from rainfall is one of the most important parts of the flood prediction process. Runoff is usually defined as water that flows over the soil surface toward the catchment outlet rather than infiltrating into the soil [1, 19]. Runoff is divided into three components:

- *baseflow* – runoff from the ground water,

- *interflow* – water that runs below the soil surface but does not reach the level of the ground water,

- overland flow – water flows across the soil surface.

The whole rainfall-runoff process is affected by many factors, e.g. basin properties such as basin size, basin shape, stream meanders, slope, roughness or soil properties including soil texture and composition. For all parts of the rainfall-runoff process, different modelling methods are usually used. These methods and their parameters are able to describe elements of the process with varying levels of precision.

The purpose of R-R models is to describe the rainfall-runoff process of the given catchment. The model inputs usually consist of physical parameters (e.g. topography, loss coefficients) and weather conditions (e.g. rainfall intensity) of the examined area. Standard outputs of the model are surface runoff hydrographs that are available for every computational unit (sub-basin in the case of semi-distributed models). In general, we can divide R-R models into three main groups:

- *Lumped models* describe the catchment as a single value with a single rainfall input.

- *Semi–distributed models* are based on a geomorphological approach. These models are focused on efficient description of the drainage system. The main computational units are sub-basins.

- *Distributed models* divide the whole catchment into fine-grained cells (mostly regular raster cells) and each cell has its own parameters and inputs. The computation is then performed for each cell and also depends on the neighbours of the processed cell.

Additional categories of models are described in [2, 19].

The *Math1D* model is our experimental in-house semi-distributed rainfall-runoff model that will be used for validating the proposed method. It uses the *SCS-CN method* [1] for transforming rainfall to runoff with its two main parameters – *initial abstraction Ia* that defines the amount of water in the soil at the start of the simulation, and *curve number CN* approximated from the hydrological soil group, land use and hydrological conditions of the modelled catchments. The contribution from river segments to a sub-basin outlet is computed using the kinematic wave approximation parametrized by the *Manning's roughness coefficient N*, which approximates physical properties of the river channel [19].

4. Sources of uncertainty

Depending on whether the model is used for prediction or plain simulation, different types of uncertainty can be involved. As stated by Montanari in [13], uncertainty of rainfall-runoff models is a very broad term and there are various types of uncertainty that have to be addressed separately. It could be a result of inaccuracy of the input data or inaccuracy of the parameters or a design flaw in the model itself.

Possible design flaws of the model itself originate from the approximation of real world processes. Such approximations are often results of a trade-off between the required level of accuracy and the amount of resources and time needed for effective execution of the given modelling method. Aside from the uncertainty produced by methods and parameters of the model, the input data can also bring inaccuracy to the results. The observed and forecast precipitation rates along with other weather conditions (e.g. air temperature, wind speed) are used as input data to the model. In this article, we only analysed precipitation data as the most important input to the model. We used observed rainfall as a reference for the analysis of the forecast model error.

5. Statistical evaluation of the ALADIN model

ALADIN is a numerical model used for short term forecasting of climate situation over Europe. It is developed by French consortium ALADIN in cooperation with local meteorological institutions. Its modification for the Czech Republic is developed in cooperation with the Czech Hydrometeorological Institute (CHMI). The model is typically used for short term forecast with good results typically within 72-hours period [5].

Its short-term precipitation forecasts are used in the Floreon$^+$ system as an input to the rainfall-runoff models. The focus of this evaluation is to determine statistical characteristics of possible deviations between the forecast and observed precipitation rates.

Analysis was performed on a database of historical observed and forecast precipitations stretching over a 5 year period covering years 2007–2012. The data were predicted and measured at 36 meteorological gauges located on catchments of the four main rivers located in the Moravian-Silesian region of the Czech Republic. Observed data were provided by the Czech Odra catchment management office in the form of hourly precipitation rates. Forecast data were provided by CHMI in sets generated every 6 hours, where each set contained precipitation forecast for the next 48-hour period. We performed the statistical analysis using the R statistical software [16]. Values of the forecast error S_{delta} were defined by

$$S_{\text{delta}} = S_{\text{o}} - S_{\text{f}} \tag{1}$$

as a difference between the observed precipitation rates S_{o} and forecast precipitation rates S_{f} for each hour and each available meteorological gauge.

5.1 Results of the evaluation

Values of selected data set characteristics are presented in Tab. I. Positive median of the data set shows that the data contained more positive errors than negative ones, which points to a possible underestimation of the forecasts (this can also be seen on the scatter plot in Fig. 1). Large standard deviation, which is three times larger than the average, points to a visible occurrence of positive outliers in the data set. Minimal value of the error (-9.1 mm/h) occurred when the model predicted 9.1 mm/h of rainfall but no real precipitations were observed. On the other hand, the highest error (38.2 mm/h) occurred when no rainfall was predicted but a heavy rain was observed. High skewness and especially very high kurtosis can be seen on the histogram in Fig. 2 and shows that errors of the forecast cannot be possibly described by a Gaussian distribution. This is the reason why robust statistical methods are often used for the analysis.

Characteristics	Value
Data set size	70804
Average (mm/h)	0.5
Lower quartile (mm/h)	−0.1
Median (mm/h)	0.1
Upper quartile (mm/h)	0.6
Minimum (mm/h)	−9.2
Maximum (mm/h)	38.2
Skewness	6.2
Kurtosis	71.3
Std. deviation (mm/h)	1.6

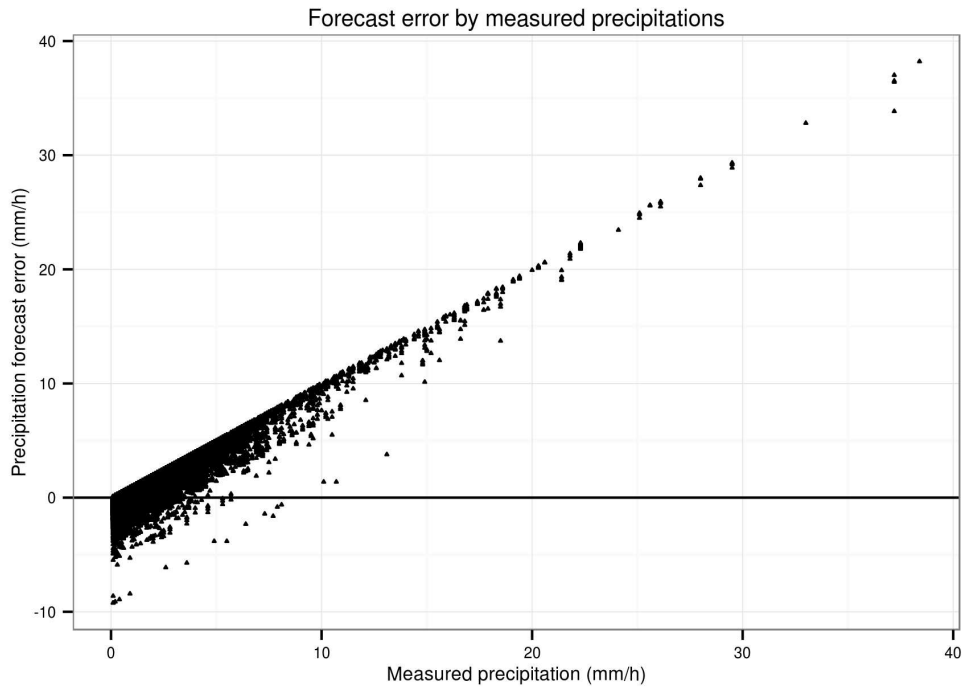**Tab. I** *Data set characteristics of the forecast error.*

Fig. 1 *Forecast error by observed precipitation rate.*

5.2 Median confidence interval

One of such robust methods of statistical inference is the construction of median confidence interval of the data set [17]. Using this method, assumptions can be made about statistical properties of all forecast errors generated by the numerical model. The 95% median confidence interval MCI is determined by its lower and upper bounds, which are estimated by the following formula [17]:

$$\text{MCI} = \left[x_{0.5} - 1.57 \frac{(\hat{x}_{0.75} - \hat{x}_{0.25})}{\sqrt{n}}, x_{0.5} + 1.57 \frac{(\hat{x}_{0.75} - \hat{x}_{0.25})}{\sqrt{n}} \right], \tag{2}$$

where $\hat{x}_{0.25}, \hat{x}_{0.5}, \hat{x}_{0.75}$ are respective p-quantiles and n is size of the input data set.

When we substitute the respective p-quantiles with values from Tab. I to Formula (2), we can estimate the bounds of the interval for the used data set as

$$\text{MCI} = \left[0.1 - 1.57 \frac{(0.6 - (-0.1))}{\sqrt{70804}}, 0.1 + 1.57 \frac{(0.6 - (-0.1))}{\sqrt{70804}} \right] \tag{3}$$

After solving the equation with substituted values, the final value is

$$\text{MCI} = \left[0.095, 0.104 \right] \tag{4}$$

This means that the model generates precise forecasts with a small underestimation of the real precipitation rates and with a very small variance of the error median.
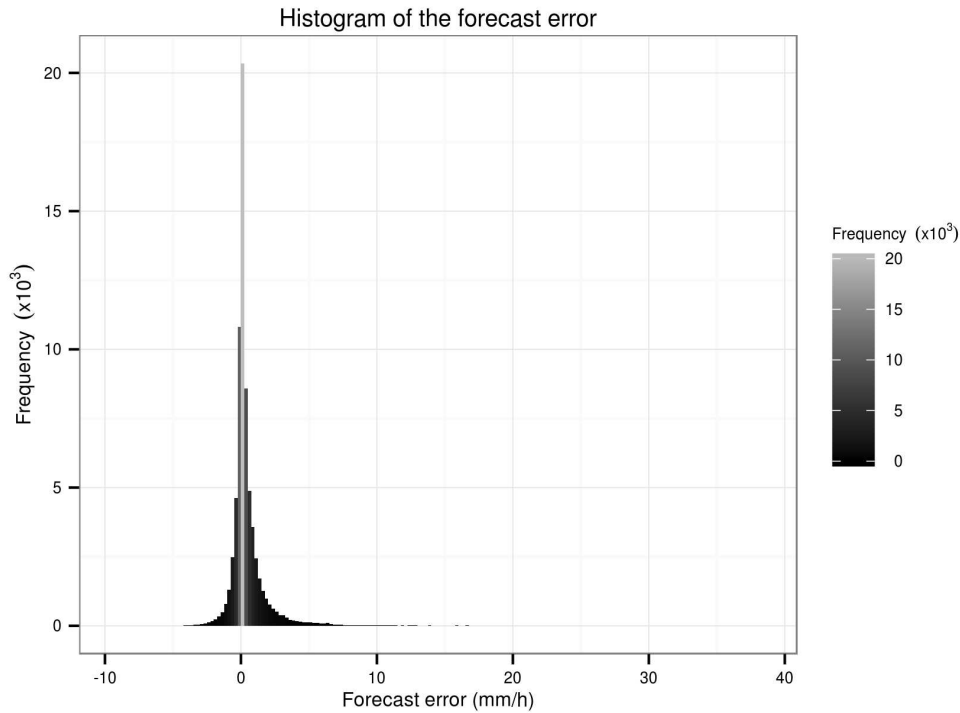
Fig. 2 *Forecast error histogram.*

5.3 Analysis of variance

As the forecast error could depend on a station (meteorological gauge) for which the prediction was generated or on a time offset of the forecast (i.e. difference between time for which the prediction was made and a time when the prediction was generated), we performed the analysis of variance of the forecast error. Both of these dependencies are undesirable as existence of both of them can lead to inconsistent forecasts and further difficulties in the uncertainty modelling process. The dependencies divide the data set into two dependency groups whose members will be subject to the variance analysis. The first group consists of 36 stations for which the forecast is computed. The second group consist of 8 individual time intervals (each 6 hours up to 48 hours in the future) in which the forecast is computed.As the data does not come from the Gaussian distribution, standard ANOVA methods cannot be used. We have therefore used a non-parametric Kruskal-Wallis test [3]. We were testing a *null hypothesis* assuming that no significant variation of the error median exists between at least three members of the tested dependency group.

Extreme values that are present in the examined data sample can have a significant influence on the results of performed statistical tests. As can be seen on the histogram in Fig. 2, most of the values are evenly distributed around the average forecast error (0.5 mm/h) where a lower quartile of the sample corresponds to value −0.1 mm/h and an upper quartile to value 0.6 mm/h. However, extreme values that can be seen on the scatter plot in Fig. 1 are randomly distributed

Intensity categories	mm/h
Very light	$0 - 0.1$
Light	$0.1 - 3$
Moderate	$3 - 8$
Heavy	$8 - 40$
Violent	> 40

Tab. II *Precipitation intensity intervals as used by CHMI [21].*

over a range which is several times larger than the interquartile range (the interval bounded by lower and upper quartile) of the sample. To lower the possible effects of the extremes on the test, we have executed these tests on a randomly selected sets of $n = 40$ observations obtained for each of the examined groups.

The tests proved that statistically significant variance of the forecast error does not exist between members of both aforementioned dependency groups, therefore there is no significant dependency of the error on either of the mentioned criteria. In case of the station dependency, null hypothesis was not rejected based on p-value = 0.560. The precipitation forecast error does not significantly vary between observed stations. In case of the time offset dependency, null hypothesis was not rejected based on p-value = 0.465. The forecast error does not significantly vary between the tested time offsets of the forecast.

5.4 Precipitation intensity intervals

According to the scatter plot in Fig. 1, the forecast error depends on the observed precipitation intensity. To account for this dependency, the range of the precipitation intensities in the data set was divided into several intervals. Determining proper size and number of the intervals is a subject of further statistical analysis, but for the experiments in this article, we used general categorization developed by the Czech Hydrometeorological Institute [21]. Individual categories of precipitation intensity are summarized in Tab. II. We have divided the data set into five intervals according to the categorization, but two of these intervals were not included to the analysis. The first category was omitted because it describes very small rainfall intensities that are hard to observe correctly due to the inherent error of the meteorological gauges. This would lead to imprecise analysis of the forecast error. The last category was omitted as well because the data set did not contain any values for this category. Histograms of data divided into the intervals are presented in Fig. 3.

6. Modelling of the precipitation forecast error

Sampling of the model input space cannot be done without certain knowledge of its statistical properties. In this case it is the probability distribution of the precipitation forecast error that can be sampled to create time series of the input precipitation data for each time step and each station. Sets of such generated time series can then be used as an input to the Monte Carlo algorithm.
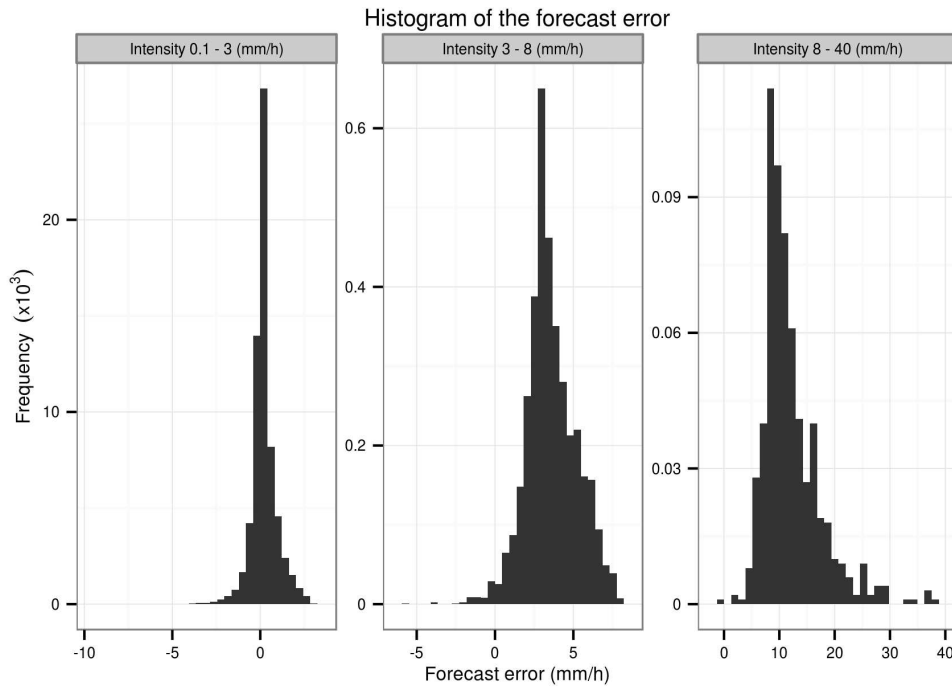
Histogram of the forecast error

Fig. 3 *Forecast error histogram grouped by limits.*

6.1 Non-parametric methods for probability density estimation

A non-parametric approach for estimating the probability density function of the data set is useful when the probability distribution of examined data does not correspond to any known distribution. The distribution of the data can be estimated empirically without restrictions coming from the shape and properties of known probability distributions (such as Gaussian $\mathcal{N}(\mu, \sigma^2)$) [22].

Constructing a histogram from the data set is a common way of estimating a shape of the probability distribution. Individual values from the data set are divided into bins of given width. Width of each bin corresponds to an interval of values within the data set. Height of a particular bin is relative to size of the data set and depends on the number of values that belong to the associated interval. This method is used to obtain a crude visual representation of the probability distribution. An example of a histogram is presented in Fig. 2. One of the disadvantages of the histogram is its finite granularity, which depends on the size of the data set and on the range of the values. The smaller the bin width is, the less values can be associated with a given bin. Therefore, some bins can be empty and discontinuity is introduced to the histogram. Estimation of probability density $\hat{f}(x)$ in a given point using histogram is mathematically expressed as

$$\hat{f}(x) = \frac{n_p}{h \cdot n} \qquad (5)$$

where n_p is the number of values in one bin, h is the width of the bin and n is the size of the data set.

6.2 Kernel density estimation

Another non-parametric method is the kernel density estimation [22]. Estimated probability density functions of the precipitation forecast error for each of the intensity intervals are shown in Fig. 4. Probability density $\hat{f}(x)$ in a given point is determined using the kernel density estimation as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} w(x - x_i, h), \qquad (6)$$

where w is the weighting function, h is the width of the smoothing window (*bandwidth*) and n is the size of the data set.

A weighting function (kernel) is placed over each value of the data set and the estimated probability density in a given point is then equal to a sum of each of the overlapping kernels. The kernel itself determines shape of the estimated probability density and could be represented by basically any function satisfying properties of a probability density function [22]. The bandwidth parameter h determines the amount of smoothing applied to the estimated curve. Too large values of the h parameter can result in over-smoothed estimates that do not reflect small variations in the real probability density function. On the other hand, small values of the bandwidth parameter can result in inaccurate density estimation affected by unnecessary noise and variations [22].

7. Uncertainty modelling using the Monte Carlo method

The Monte Carlo method is a well-known member of a group of algorithms focused on stochastic modelling [9]. We use this method for modelling of probabilistic properties of the R-R model input space. Output of the method is then used for constructing the confidence intervals (CI) in which the outputs of the model can occur.

The Monte Carlo approach is quite useful in this case since the input space of the R-R model is multi-dimensional. Each uncertain parameter (e.g. initial abstraction mentioned in Section 3) has its own probability distribution and adds another dimension to the sampled input space. These parameters are set for each sub-basin and channel of the modelled catchment. In the case of our experimental model of the Ostravice river, the basin is divided into 49 channels and their contributing sub-basins each with its own parameters. Besides channel and sub-basin parameters, the input of the model is largely influenced by the forecast precipitations.

As described in Section 3, the precipitation is represented as a time series of precipitation rates for each meteorological gauge, where each gauge belongs to a
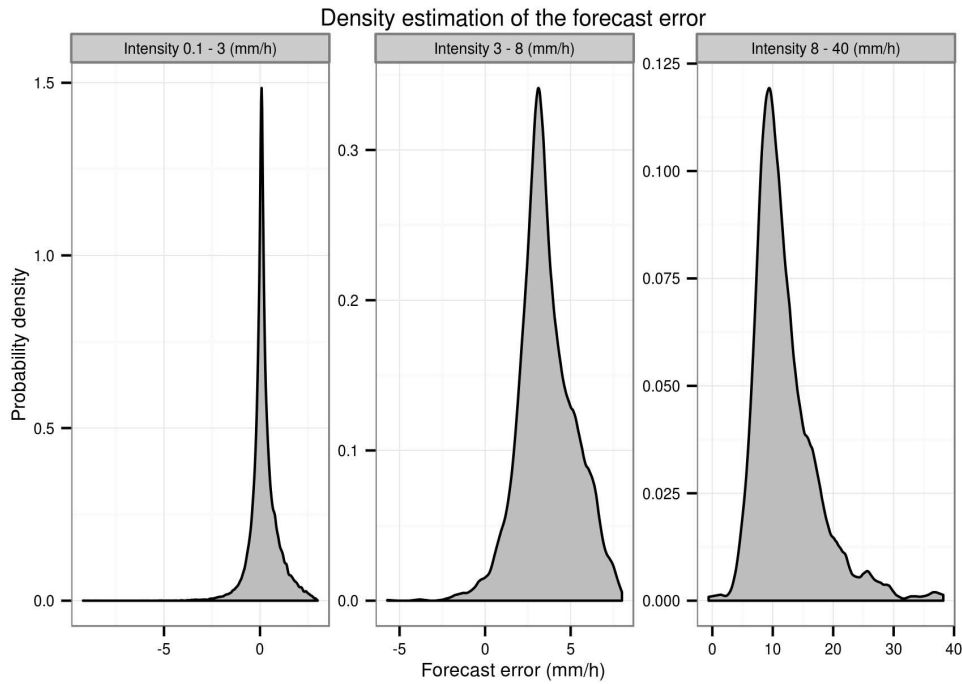
Density estimation of the forecast error

Fig. 4 *Estimated forecast error density for three precipitation intervals.*

particular sub-basin. Our approach presented in this article samples the values for each time step and each gauge independently from the probability distribution modelled by the kernel density estimation described in Section 6.

7.1 Sampling of the model input space

The sampling algorithm uses sampled curves obtained by the kernel density estimation. Sampled points of the probability density function curves are converted to the cumulative distributive function curves and stored in a binary tree. This tree stores the forecast error value in mm/h and its corresponding probability. A canonical probability $p \in (0, 1)$ is generated using a pseudo-random number generator and two nearest probability points are selected using the binary tree search. The resulting forecast error is then interpolated from the two coordinates on the cumulative probability curve represented by the selected probability points and their respective forecast errors. This interpolated forecast error is then used for generating precipitation forecasts affected by certain error.

As described at the beginning of this section, the input space of the R-R does not consist only of the input precipitation but also of a number of other parameters, some of which will be incorporated in the modelling process due to their possible uncertainty and large influence on the model output.

In case of the *Math1D* model we have incorporated uncertainty of the following parameters – the *CN number* (SCS-CN method), *initial abstraction* for the sub-

basins and the *Manning's roughness* coefficient (N) for river channels. Values of these parameters are usually estimated by an expert or obtained by a calibration process.

Input space of these parameters in our method can be described by known probability distributions and their parameters. Experiments with values sampled from different probability distributions and results obtained by their sampling are summarized in Section 9.

7.2 Statistical analysis of the results

Rainfall-runoff simulation is performed for each generated set of input parameters and precipitations and the resulting hydrographs are gathered for quantile selection. Results of the Monte Carlo simulations form a three-dimensional discrete function of modelled river discharges. The first dimension specifies a time step, the second dimension specifies a catchment channel and the third one specifies individual Monte Carlo iterations.

Several uncertainty hydrographs with different probabilities are created for each river channel. Single uncertainty hydrograph with specified probability p is formed by selecting p-percent quantiles from the Monte Carlo results for each time step of the simulation.

8. Parallelization strategy

Due to a significant computational demand of the proposed method, we have created a scalable parallel implementation of the modelling algorithm to ensure that the results can be provided in a short time frame by running it on a high performance computing cluster. This parallel implementation was created using the Message Passing Interface (MPI) and the OpenMP standard. The requested number of Monte Carlo samples is divided into chunks, which are then distributed among all available MPI processes. The *Math1D* model itself uses OpenMP threads for parallelization of convolution in a part used for estimating the base flow.

In the initialization phase of the algorithm, MPI processes are created with the initial model configuration. One master process generates all parameter sets for each of the Monte Carlo samples and distributes them in chunks to the respective slave processes. OpenMP threads are used for generating the value sets for each parameter. The total number of Monte Carlo samples is divided into chunks and each thread generates a chunk of the input space samples. When all sets are generated and distributed, each process performs its own number of Monte Carlo iterations. After each process finishes processing its own chunk, the resulting hydrographs are gathered to the master process. The master process then performs the selection of requested quantiles.

Scalability of this approach is expected to be quite good since the individual Monte Carlo samples are independent of each other and the biggest part of the communication between nodes is done in the barrier part, where the results are gathered to a master process. The biggest portion of the execution time is used for running R-R simulations while sampling of the input space and quantile selection takes only a fraction of the whole execution time.
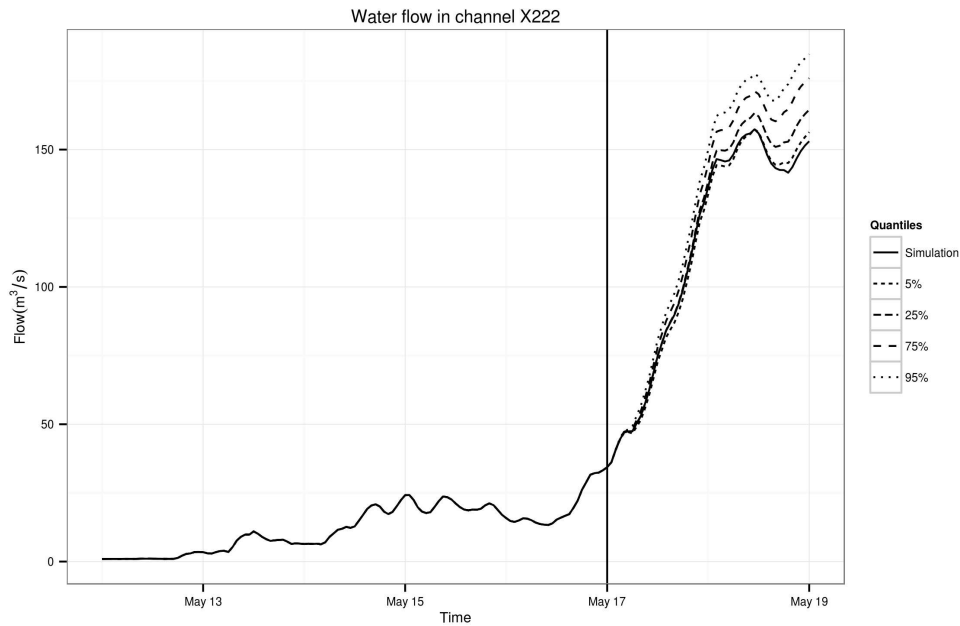
Water flow in channel X222

Fig. 5 *Simulation of the precipitation forecast uncertainty using 30000 Monte Carlo iterations.*

9. Experiments

Rainfall-runoff simulations were performed for a catchment area of the Ostravice river located in the Moravian-Silesian region of the Czech Republic. Simulations were executed with hourly precipitation data from a selected 7-day time frame. In the first 5 days, observed data were used only with parameter uncertainty applied. In the last 2 days, rainfall ALADIN forecast is used and its uncertainty is modelled by the methods described in previous sections.

The time period of the data is between 12–19 May 2010, during which a major rainfall event occurred in the modelled catchment and its influence on the water flow is clearly visible. Different line types in the output hydrographs show quantiles selected from the simulation output. All hydrographs show simulated water discharge on the catchment outlet located in the city of Ostrava.

9.1 Uncertainty of the ALADIN forecasts

The following simulations were performed using estimated probability distribution of the precipitation forecast. The boundary between observed precipitations (5 days) and forecast precipitations (2 days) in output hydrographs is marked by a black vertical line.

The selected quantiles in the simulation on Fig. 5 tend to have higher values than the original simulation. This is caused by the underestimation of the forecast model that was identified in Section 5. Generated samples for the Monte Carlo simulations followed this identified distribution and provided more precise estima-
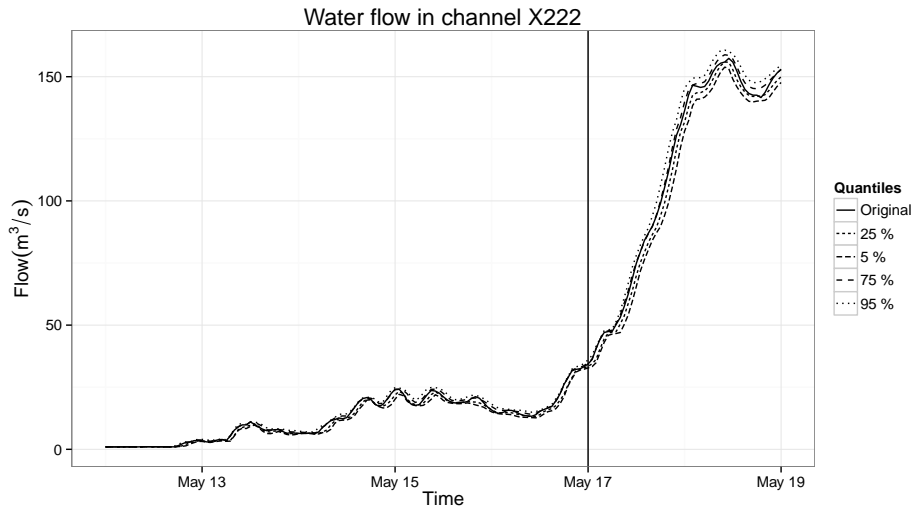
Fig. 6 *Simulation of the Manning's coefficient uncertainty using 30000 Monte Carlo iterations with uniform distribution, 20% deviation.*

tion of the resulting river discharge. This discovery could be also applied directly to the standard simulations without uncertainties by directly increasing the value of forecast precipitations by the median of the forecast error.

9.2 Schematization parameters

Next experiment was focused on the simulation of selected model parameters. However, used probability distributions are based only on expert assumptions and not on real data. The main purpose of this simulation is to look at possible effects of predictable changes of the parameter values on the simulation output. Results of this experiment are shown on Fig. 6–9.

9.3 Scalability of the parallel implementation

Parallel implementation of the simulation method was executed on different number of nodes with different number of iterations. Duration of individual experiments is presented in Tab. III. Experiments were executed on the Anselm supercomputer operated by IT4Innovations [7] consisting of 209 compute nodes, each running one MPI process with 16 OpenMP threads available. Saturation occurs when the simulation is executed on approx. 100 nodes at which point the overhead of MPI processes takes longer than the actual computation of small chunks of Monte Carlo samples.

Individual runs of the model are independent of each other and most of the MPI communication occurs at the beginning of the simulation during scattering of the input data chunks to running processes and at the end during gathering of the final results. Variations of the speed-up factor can be attributed to different utilization of the internal cluster network and variations in the run-time environment during execution of the experiments.
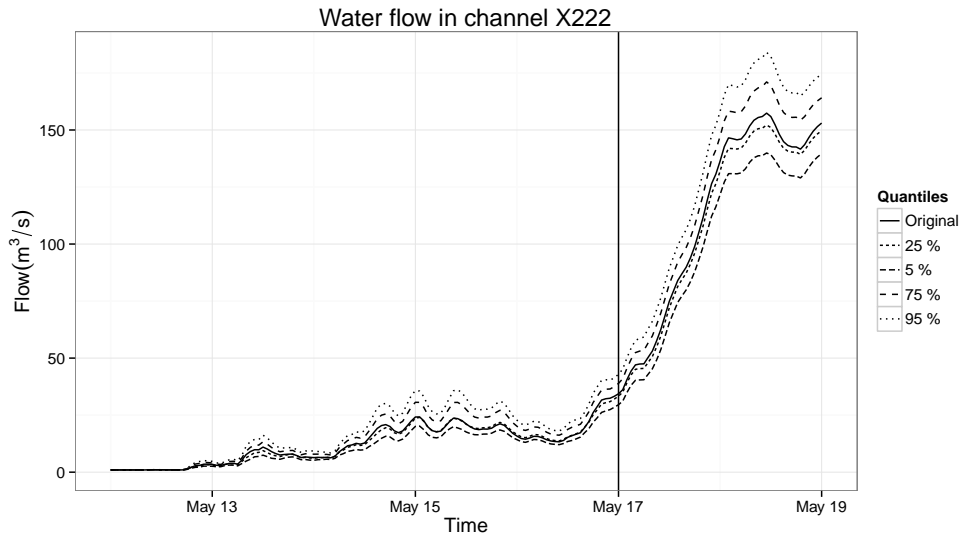
281

Water flow in channel X222

Fig. 7 *Simulation of the CN curve number uncertainty using 30000 Monte Carlo iterations with uniform distribution, 20% deviation.*
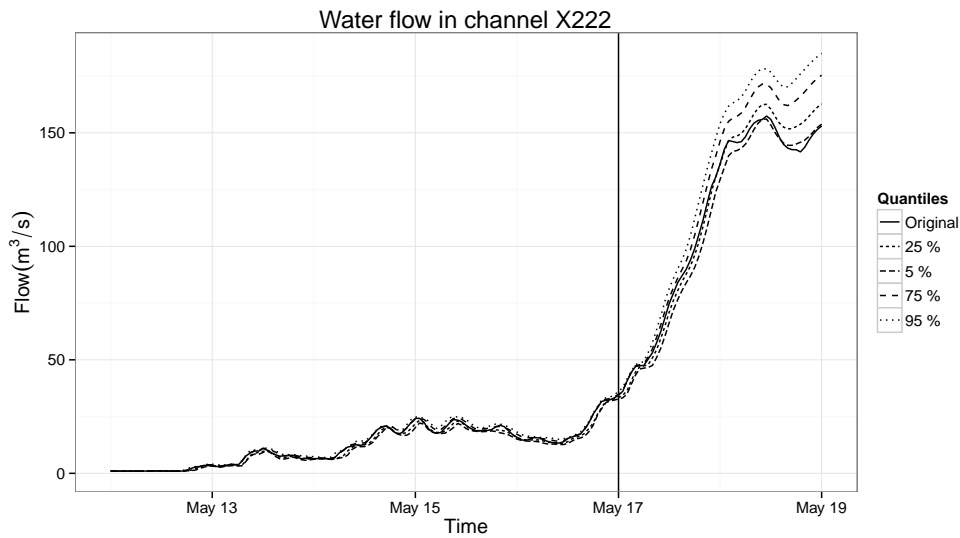
Water flow in channel X222

Fig. 8 *Simulation of the Manning's coefficient uncertainty together with precipitation forecast using 30000 Monte Carlo iterations with uniform distribution, 20% deviation.*

10.   Conclusion and future work

The main focus of this article was to present a method for modelling possible uncertainties in the rainfall-runoff model inputs. Precipitation forecast was identified as one of the main sources of possible errors in the R-R models, so we performed a
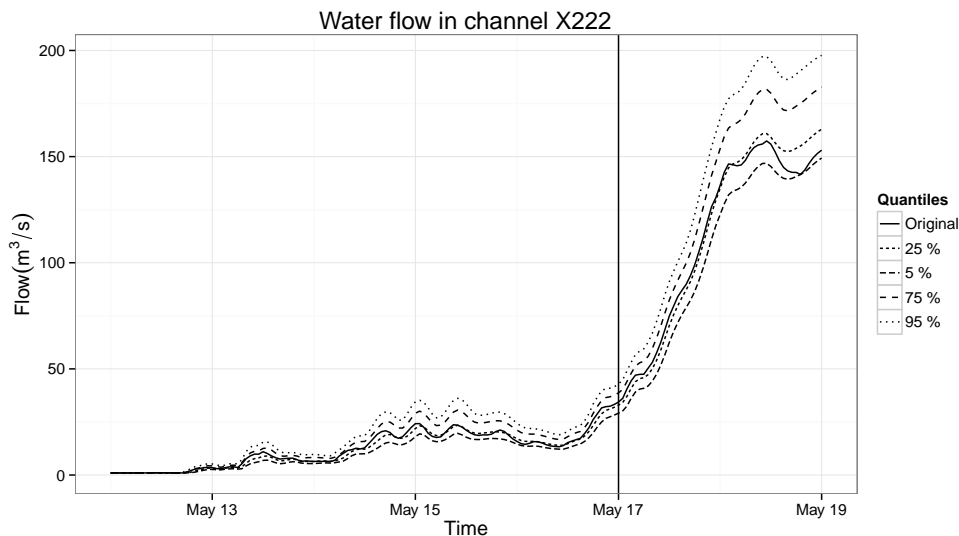
Fig. 9 *Simulation of the CN number, Manning's coefficient and precipitation forecast uncertainty using 75000 Monte Carlo iterations with uniform distribution of the parameters, 20% deviation.*

statistical analysis of the precipitation forecast produced by the numerical model ALADIN. Slight underestimation of the predicted rainfall was evident from the data set, where average forecast error was 0.5 mm/h. The 95% median confidence interval of the data set was estimated as $[0.095, 0.104]$.

This statistical analysis was then used to create a method for estimating the probability distributions of the forecast error by using kernel density estimation for several categories of rainfall intensity. These distributions describe the uncertainty of the forecast and can be used to model the input space for stochastic uncertainty modelling methods. Monte Carlo method was used to sample and run multiple rainfall-runoff simulations from the modelled input space. These results are then processed and analysed to provide additional information about possible flooding scenarios and their probability.

Even though the proposed method increases the precision and provides additional information to emergency committees, it is computationally intensive and can take a long time to execute. Fortunately the method can be effectively parallelized and deployed on an HPC cluster to significantly decrease the time needed for its computation. In this environment, the method can be used for decision support in disaster management even during critical situations.

Categorization of rainfall intensities was done to decrease the variance of the results for predefined intensity intervals, but the number of samples in each of these categories was significantly different. We would like to create a different categorization that will take the number of samples in each category into account. This categorization could lead to even better precision of the method as each category would be described by its specific probability distribution. Although only precipitation forecast error distribution was estimated, the proposed method can

Iterations	Nodes	Time (s)
	2	1981
	4	977
	8	685
	16	409
10000	32	167
	64	116
	100	114
	128	112
	4	1773
	8	1127
	16	564
20000	32	319
	64	203
	100	176
	128	171
	8	3266
	16	1743
	32	949
60000	64	584
	100	431
	128	404

Tab. III *Duration of experiments executed on different number of cluster nodes.*

be used to simulate uncertainty of other model parameters as well. But without exact knowledge of their respective probability distributions or data that can be used to identify these distributions, only expert estimations can be made.

Acknowledgement

References

[1] BEVEN K. *Rainfall-Runoff Modelling – The Primer*. 2nd ed. UK: Wiley-Blackwell, 2012, doi: 10.1002/9781119951001.

[2] BOGNER K., PAPPENBERGER F., CLOKE H.L. *Technical Note: The Normal Quantile Transformation and its application in a flood forecasting system.* Hydrology and Earth System Sciences Discussions. 2011, 8(5), pp. 9275–9297, doi: 10.5194/hessd-8-9275-2011.

[3] BRESLOW N. *A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship.* Biometrika, 1970, 57(3), pp. 579–594, doi: 10.1093/biomet/57.3.579.

[4] CRHOVA L., HOLTANOVA E., KALVOVA J., FARDA A. *Performance of ALADIN-Climate/CZ over the area of the Czech Republic in comparison with ENSEMBLES regional climate models.* Stud Geophys Geod. 2013, 58(1), pp. 148–169, doi: 10.1007/s11200-013-1107-0.

[5] FARDA A., DEUE M., SOMOT S., HORANYI A., SPIRIDONOV V., TOTH H. *Model ALADIN as regional climate model for Central and Eastern Europe.* Stud Geophys Geod. 2010, 54(20), pp. 313–332, doi: 10.1007/s11200-010-0017-7.

[6] HOLTANOVA E., KALVOVA J., MIKSOVSKY J., PISOFT P., MOTL M. *Analysis of uncertainties in regional climate model outputs over the Czech Republic.*Stud Geophys Geod. 2010, 54(3), pp. 513–528, doi: 10.1007/s11200-010-0030-x.

[7] IT4INNOVATIONS. *IT4Innovations Anselm Cluster Documentation* [online]. VSB – Technical University of Ostrava, 2014. [viewed 2014-09-01]. Available from: https://docs.it4i.cz/anselm-cluster-documentation/hardware-overview.

[8] KOBOLD M., SUSELJ K. *Precipitation forecasts and their uncertainty as input into hydrological models.* Hydrology and Earth System Sciences. 2005, 9(4), pp. 322–332, doi: 10.5194/hess-9-322-2005.

[9] KROESE D.P., TAIMRE T., BOTEV Z.I. Handbook of Monte Carlo Methods. USA and Canada: John Wiley & Sons, Inc., 2011, doi: 10.1002/9781118014967.

[10] KRZYSZTOFOWICZ R. *Bayesian theory of probabilistic forecasting via deterministic hydrologic model.* Water Resour. Res. 1999, 35(9), pp. 2739–2750, doi: 10.1029/1999wr900099.

[11] MARTINOVIC J., KUCHAR S., VONDRAK I., VONDRAK V., SIR B., UNUCKA J. Multiple Scenarios Computing In The Flood Prediction System FLOREON. In: A. BARGIELA, S.A. ALI, D. CROWLEY, E.J.H. KERCKHOFFS, eds. *Proceedings of the 24th European Conference on Modelling and Simulation, Simulation Meets Global Challenges (ECMS 2010)*, Kuala Lumpur. ECMS, 2010, pp. 182–188, doi: 10.7148/2010-0182-0188.

[12] MONTANARI A. *Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations.* Water Resour. Res. 2005, 41(8), doi: 10.1029/2004wr003826.

[13] MONTANARI A. *What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology.* Hydrol. Process. 2007, 21(6), pp. 841–845, doi: 10.1002/hyp.6623.

[14] MONTANARI A., BRATH A. *A stochastic approach for assessing the uncertainty of rainfall-runoff simulations.* Water Resour. Res. 2004, 40(1), doi: 10.1029/2003wr002540.

[15] NASH J., SUTCLIFFE J. *River flow forecasting through conceptual models part I – A discussion of principles.* Journal of Hydrology. 1970, 10(3), pp. 282–290, doi: 10.1016/0022-1694(70)90255-6.

[16] R CORE TEAM. *R: A Language and Environment for Statistical Computing* [online]. R Foundation for Statistical Computing, Vienna, Austria, 2014. [viewed 2014-09-01] Available from: http://www.R-project.org/.

[17] SCHWARZ C. J. *Sampling, regression, experimental design and analysis for environmental scientists, biologists, and resource managers.* Canada: Department of Statistics and Actuarial Science, Simon Fraser University 57, 2011.

[18] STEENBERGEN N.V., RONSYN J., WILLEMS P. *A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication.* Environmental Modelling & Software. 2012, 33, pp. 92–105, doi: 10.1016/j.envsoft.2012.01.013.

[19] VIEUX B.E. *Distributed Hydrologic Modeling Using GIS.* Springer Netherlands, 2004, doi: 10.1007/1-4020-2460-6.

[20] VONDRAK I., MARTINOVIC J., KOZUSZNIK J., STOLFA S., KOZUBEK T., KUBICEK P., VONDRAK V., UNUCKA J. A Description of a Highly Modular System for the Emergent Flood Prediction. In: *Proceedings of the 7th conference on Computer Information Systems and Industrial Management Applications (CISIM 2008)*, Ostrava, Czech Republic. IEEE, 2008, pp. 219–224. doi: 10.1109/cisim.2008.22.

[21] ZIDEK D., LIPINA P. *CHMI Manual for the observers of the meteorological stations* [online]. Methodological regulations [viewed 2014-09-01]. 2003, 13. In Czech. Available from: http://www.meteoopava.estranky.cz/file/7/navod-pro-pozorovatele.pdf.

[22] ZUCCHINI W. *Applied smoothing techniques. Part I: Kernel Density Estimation* [online]. Philadephia, PA: Temple University, 2003, pp. 15–19 [viewed 2014-09-01]. Available from: http://staff.ustc.edu.cn/ zwp/teach/Math-Stat/kernel.pdf.