DOI: 10.31577/ahs-2021-0022.02.0031

ACTA HYDROLOGICA SLOVACA

Volume 22, No. 2, 2021, 276 – 283

Statistical post-processing of short-term hydrological ensemble forecasts using the application of the dressing method

Tomáš VLASÁK*, Jakub KREJČÍ

Probabilistic hydrological forecasts used in forecasting offices are often based only on different variants of precipitation forecast, which are the dominant source of forecast uncertainty during flood periods. The proposed method called dressing extends the uncertainty of meteorological forecast input by estimating the uncertainty of hydrological modeling using statistical analysis of deviations derived from simulated and observed flows. Adjustment of probabilistic flow forecasts is applied by post-processing without interfering with the hydrological model itself. The method is focused primarily on runoff phases, where heavy precipitation is not expected and the dispersion of the original ensemble is insufficient. A comparison of the success of short-term operative ensemble predictions of river discharge in the upper Vltava basin before and after adjusting by the dressing method showed a clear improvement in statistics.

KEY WORDS: hydrological forecast, ensemble forecast, dressing, post-processing

Introduction

Simplification of reality in prediction models, inaccurate input data and other sources of uncertainty lead to predictions that always, more or less, differ from observation. Lack of accuracy of forecasts is the most important limitation in their use, and one solution is to quantify their uncertainty. Therefore, flow forecasts often include two basic products: 1.) **deterministic forecast**, single flow calculated from one selected set of causes (precipitation, saturation, etc.) and 2.) **ensemble** (**probabilistic) forecast**, model calculation is repeated for different scenarios of inputs and settings of the hydrological model. Ensemble forecast allows forecasters to estimate the risks (probabilities) of exceeding specific threshold. It also makes it possible to extend the time advance of forecasts and use them more effectively not only in flood protection but also at low flow rates.

In the case of hydrological river flow prediction, the ensemble forecasts are very often based solely on different variants of precipitation and temperature fore-casting. The uncertainty of hydrological modelling (observed inputs, initial conditions, model parameters, etc.) is omitted. This simplification is acceptable in flood forecasting of upper basins when the effect of the uncertainty of the precipitation forecast is so dominant that the expression of uncertainty by the ensemble forecast calculated in this way is acceptable. However, hydrological forecasts are gradually being used for purposes other than flood protection. Probability predictions are also important for dam manipulation planning, hydropower management or for river water use in times of drought, even in times of insignificant fluctuations or decrease inflows. In addition, a functional ensemble system in times of average flows is important for gaining confidence in probabilistic predictions as a whole. Probabilistic predictions should therefore contain quantified information on the uncertainty of the whole prediction system, not just precipitation forecast.

The presented method includes the uncertainty of hydrological modelling into the calculation of the ensemble hydrological forecast. It is primarily intended for the improvement of probabilistic forecasts based exclusively on precipitation variants. The method was inspired by the dressing method published by Pagano et al. (2012). It is based on the analysis of historical deviations of simulated and observed flows and the subsequent construction of error models. The method was tested in order to increase the success of operational ensemble predictions which serve as an irreplaceable source of information for river navigation in the Elbe and for the management of water reservoirs with regard to optimizing electricity production and minimizing the impact of drought. It is applied as a post-processing procedure, which means adjusting the hydrological forecast after its output from the hydrological model. The advantage of post-processing is easy implementation into operation without disrupting other established procedures.

Hydrological forecast uncertainty

Understanding the reasons why hydrological forecasts deviate from observations is key step in developing the success of both deterministic and probabilistic forecasts. Krzysztofowicz (1999) decomposes the total uncertainty into input uncertainty and hydrological uncertainty.

The uncertainty of the inputs is solved by pre-processing methods, which precede the calculation itself in the prediction model. The observed elements (precipitation, temperatures, flows) are usually not subject to such a significant error. Their uncertainty is usually neglected, or they are reduced or quantified using some of the pre-processing methods, for example (Schaake et al., 2004). The dominant source of uncertainty in the period of heavy rainfall is the quantitative precipitation forecast from numerical weather forecasting models. Different precipitation variants are therefore fundamental and often also the only quantified uncertainty for the probabilistic hydrological forecast. Probabilistic hydrological forecasts based only on different precipitation variants suffer mainly from an insufficient variance of variants during the precipitation-poor period. In these cases, the more significant is hydrological uncertainty. The distinction between meteorological and hydrological uncertainty and independent work with them was used, among others, in the work of Demargne et al. (2013) and Verkade et al. (2017).

Hydrological uncertainty is usually adjusted by post-processing methods, which stand between the output of the forecast from the model and its final publication for users. Statistical post-processing is simply a model that uses the relationship between the prediction and the observed element (Fig. 1). There are a number of statistical post-processing methods, from a simple percentile method through more complex statistical procedures such as the Kalman filter or the Bayesian method to the application of neural networks. An overview of post-processing methods in hydrology was published for example, by Li (2017).

The dressing method combines the already created hydrological ensemble forecast, which is based on the probabilistic prediction of precipitation, with the statistical distribution of deviations of hydrological modelling, and thus achieves a comprehensive description of the entire uncertainty of the hydrological forecast.

Materials and Methods

Hydrological forecasting system AquaLog (Krejčí and Zezulák, 2009) was used for the calculation of forecasts needed for method design and assessment. This system is the main tool for real-time hydrological forecasting in the Czech Republic in the Labe basin. AquaLog model consists of continuous SAC-SMA (Burnash, 1995) precipitation-runoff component and its operation is largely automated, excepts for assimilation of simulated flow to the last measured discharge. The upper Vltava river basin (tributary of the Elbe) was selected for testing the method. The catchment with area of 12105 km^2 is divided into 45 sub-basin delimited by water gauging stations with the observed discharge (Fig. 2).

Three statistical methods commonly used in the field of ensemble predictions verification were used for evaluating the success of the dressing method. They focus on the reliability, the skill and the conditional verification of ensemble prediction. The rank histogram (sometimes called Talagrand diagram) was used for assessing the spread of the prediction ensemble in relation to real observational variability. The Brier score is a suitable criterion for verifying a categorical prediction from the point of view of the accuracy of a probabilistic prediction when we examine whether a defined phenomenon did/didn't occur. It answers the question of how big the probability prediction error is (0 if it does not happen and 1 if it does happen). The benefit of the last used ROC (relative operation characteristics) criteria lies in its ability to distinguish between the occurrence and non-occurrence of a particular event for a given condition. All the mention methods are in detail described in WMO (2021). Basic interpretation of rank histogram and ROC plot used in Results chapter is shown on the Fig. 3.

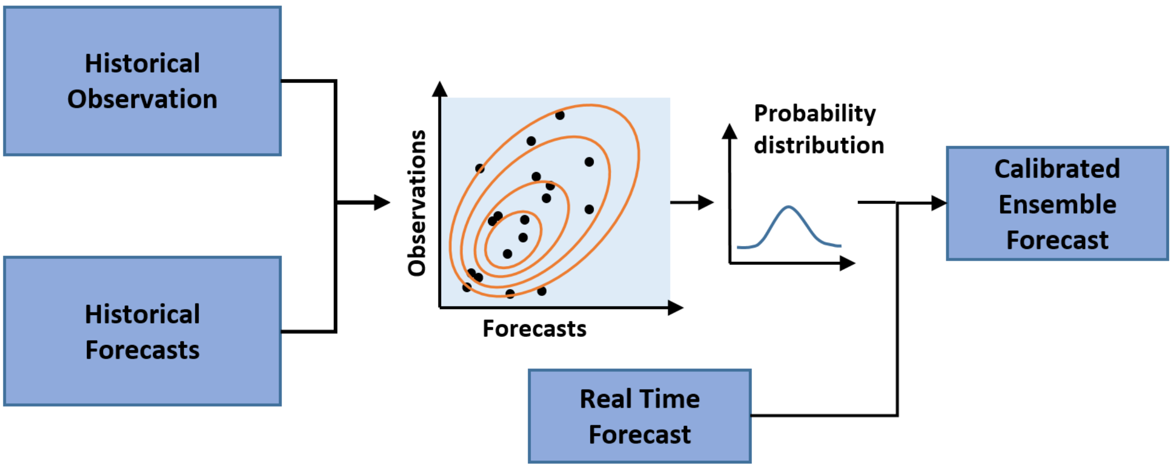

Fig. 1. Scheme of statistical post-processing of hydrological forecast (Li, 2017).

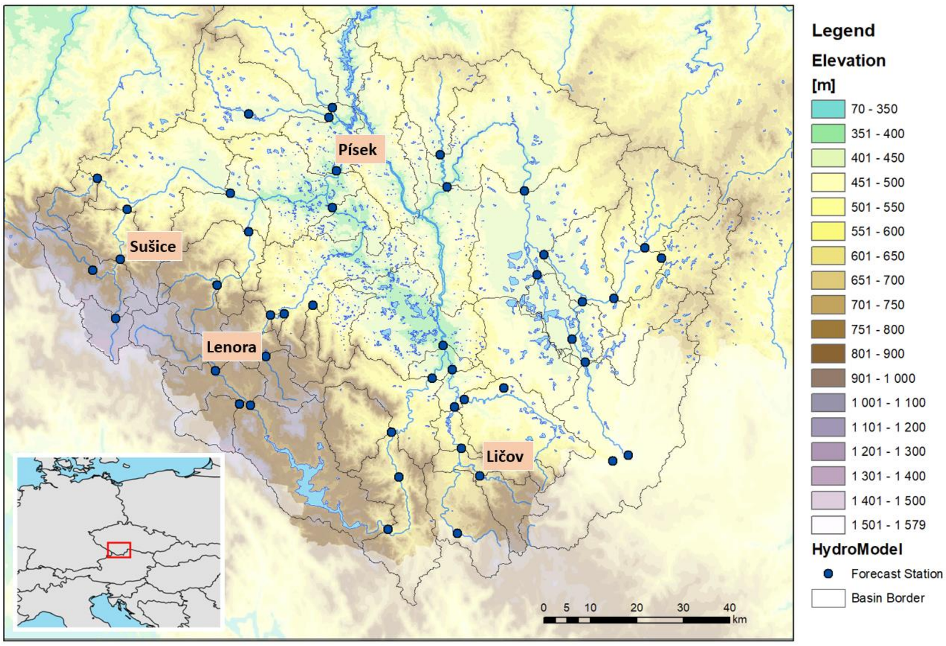

Fig. 2. The upper Vltava river basin with forecasted water gauging stations. The stations with label are mentioned in the chapter Method calibration and results.

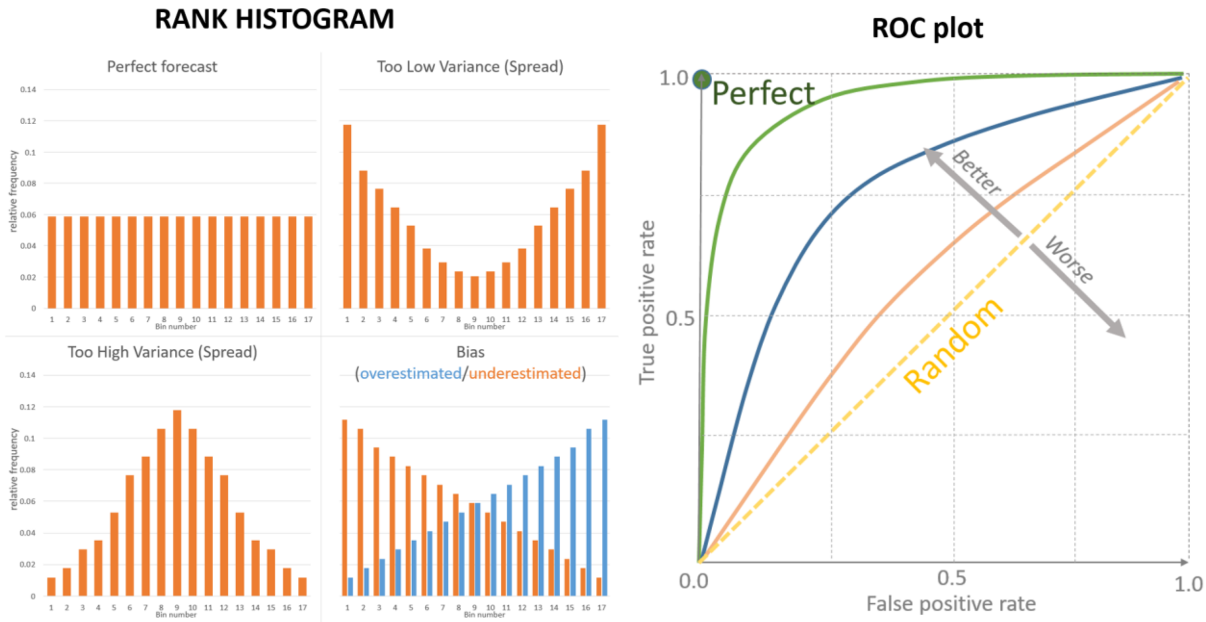

Fig. 3. Basic interpretation of rank histogram and ROC plot.

Dressing method

Dressing is designed to modify the probabilistic hydrological forecast calculated from the meteorological forecast ensemble. The uncertainty of hydrological modelling is expressed by an error model, which is derived from the statistical distribution of deviations between historical flow forecasts and the observed flow for different lead times.

Historical flow forecasts were calculated by replacing the predicted precipitation with observed precipitation to clear the influence of precipitation forecast uncertainty on forecast error. The method of calculating historical forecasts as well as a number of forecasts is essential for the successful application of the dressing method. Historical forecasts should well represent the uncertainty of hydrological modelling in real-time operations and should cover as many runoff variants as possible. Because the AquaLog hydrological forecasting system is built on continuous models, we assume that deviations of automatically calculated historical forecasts from the observed flow well represent the uncertainty of hydrological modelling. Uncertainty is expressed as a whole without distinguishing between individual sources of uncertainty (input data, initial conditions, model structure, operational control, etc.).

Error model

Although historical forecasts are not affected by uncertain precipitation prediction, the magnitude of the errors of historical forecasts significantly increases with the lead-time. This is due to two facts. (1) The forecast is in the last phase of the calculation assimilated to the last observed flow, which eliminates the error in short lead-time. (2) The forecast for downstream water gauging stations is in short advance based on a more reliable channel routing model with input of observed discharge from the upper station. After exceeding the travel time of water among two water gauging stations the observed discharge is replaced with simulated discharge, which contains errors from the less reliable rainfall-runoff model. It is obvious that specific error models for different lead-time as well as for different water gauging stations are required.
Error models were constructed according to the frequency of flow multiplicative deviations *Qdif* :

$$Qdif_p = \frac{Qobs_p}{Qsim_p} \qquad (1)$$

where
$Qsim_p$ – is the forecasted flow in prediction lead time $_p$,
$Qobs_p$ – is the observed flow in prediction lead time $_p$.

With a short lead time, most of the deviations *Qdif* derived from the historical forecasts are close to number one. With the increasing lead, the standard deviation, as well as the variance of deviations, increase (see Fig. 4). For some water gauging profiles, there is an uneven distribution of overestimated and underestimated forecasts in the error models. It indicates systematic bias, which is related to the calibration of the hydrological model. The error model created in this way adjusts the ensemble prediction in two ways. (1) It expands the variance of the hydrological ensemble calculated according to precipitation variants. (2) It corrects the systematic error of the hydrological model (bias).
Pagano (2012) uses one error model for each water gauging profile. The advantage of this approach is a small fluctuation of the error models because they are calculated from a large number of historical forecasts. One set of error models for each forecasting point also facilitates the application of the method to daily operation. In fact, it is clear that the uncertainty of the hydrological model differs for different runoff phases. The increase of forecast errors with a lead-time for the period without precipitation with steady river discharge and for the period when heavy precipitation is expected varies significantly.
The dynamic construction of the error model proved to be a suitable solution to this problem. For each hydrological forecast, a number of the most similar historical forecasts are selected. The specific error model is built from this selection. This means that the error model differs not only for each water gauging profile and the lead-time but also according to the type of runoff phase. The dressing method is combined with the method of the historical analogue (Li, 2017). Nash-Sutcliffe coefficient was chosen as a criterion for the selection of historical forecast analogues. Its calculation is based on equation (2):

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \qquad (2)$$

where
S_i – is the discharge of the current forecast at the time $_i$,
O_i– is the discharge of the historical forecast at the time $_i$,
$\bar{O}$ – is the average discharge of the historical forecast.

The unique error model for each forecast is more correct because it doesn't mix different runoff phases with different errors into one error model. The other advantage

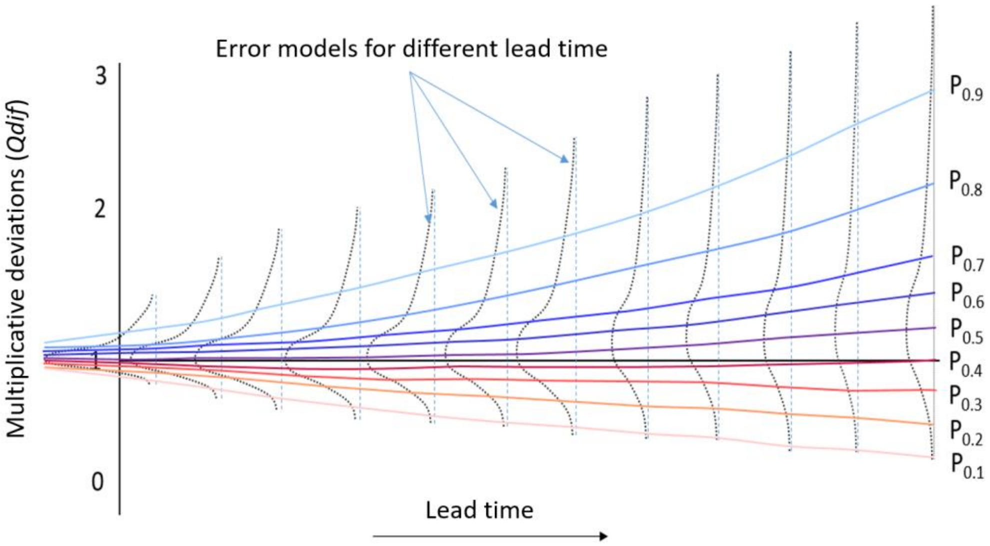

Fig. 4. Error model (distribution of frequency of multiplicative deviations Qdif) for different lead time. Coloured lines indicate percentiles of the distribution of deviations.

of dynamic construction of the error model is the elimination of forecasts which are not suitable for the dressing method, ie. forecasts that don't have a sufficient number of similar historical predictions (Fig.5). The time development of these forecasts is so dynamic that it is difficult to find enough similar historical forecasts. Error models built from lower numbers or less similar discharges can give too large variance and unreal estimation of forecast uncertainty. Testing the method proved that the vast majority of removed forecasts was from the high flow period when the dominant source of uncertainty is the quantitative precipitation forecast, and therefore the original ensemble, based on variants of precipitation forecasting, usually provides a sufficient estimate of forecast uncertainty.

The successful application of the method is related to the setting of the minimum value of the NS coefficient, the number of historical floods required for the error model and the length of the processed time series. The minimum value of the NS separates historical forecasts suitable and not suitable for the error model. The number of chosen historical forecasts determines the reliability of the error model and the length of the time series of the forecast has a similar effect. In the case of forecasts with a very short lead-time, the length of time series should be extended to include observed data because autocorrelation between the observed and the predicted flow is usually very strong.

Setting the method, which means a high degree of similarity between the adjusted forecast and historical forecasts, gives a better chance for a more accurate estimate of the uncertainty of hydrological modelling, on the other hand, it reduces the number of forecasts that can be processed by the method. Finding the optimal compromise between the number of processed predictions and the success of the method was the subject of method calibration.

From error model to ensemble forecast

The error model was expressed by 9 levels of probability of exceeding from the deviations *Qdif* ordered by size. The levels correspond to percentiles $P_{0.1}$; $P_{0.2}$; $P_{0.3}$ to $P_{0.9}$ (see Fig. 3). Each hydrological forecast (each member of the forecast ensemble) was divided into 9 forecasts multiplying the flow by nine *Qdif* values for each lead time of the forecast. This created a new ensemble nine-time larger than the original ensemble. For example in the case of hydrological forecast ensemble based on 17-member precipitation variants from ALADIN-LAEF system extended 153-member was created.

However, the number of members of the hydrological forecast ensemble should not change after post-processing for two reasons. (1) Some forecasts are not suitable for the dressing method due to too few similar historical forecasts. (2) Post-processing, in general, should not affect further processing of forecasts (publications, archiving). For these reasons, the next step is to reduce the number of ensemble members to the original count. From the several tested procedures, a simple percentile selection method was finally chosen. The members of the extended ensemble were sorted by the size based on the selected criteria (average flow, or maximum flow, or a combination of multiple indicators) and every 9th member was chosen. The disadvantage of this approach is that new ensemble members don't have to be derived from the same member in the original ensemble. Therefore some variants, typically with secondary waves, may not appear in the new ensemble. However, the variance of the predictions according to the selected criterion (average flow, maximum flow, etc.) is expressed correctly.

Method calibration and results

Calibration and testing of the post-processing method dressing with a dynamically generated error model consisted of (1) finding optimal parameters for building the error models (2) comparison of the assessment of original and modified hydrological ensemble forecast.

The set of historical hydrological forecasts covered of 2780 episodes from the period 2012 to 2020. They were calculated for 40 forecasted points in the Upper Vltava river basin as a time series of discharge values with 1 hour time step and 66 hours lead time. The time series of predicted discharge started always at 7:00 AM, which is the time zero of real-time forecast. This may be important in building an error model because some forecast errors can be affected by the daily development of weather, especially air temperature. The minimum number of historical forecasts required for the building of the error model was set at 20 cases. Forecasts were compared without including any section of observed flows that precedes the predictions. Calibration was focused on finding the optimal size of the NS coefficient. For the calibration and the testing of the performance of the method, 270 ensemble hydrological forecasts calculated in real-time operation in the years 2020 to 2021 were used. These ensembles were based on 17 variants of precipitation from the ALADIN-LAEF forecast system with a time step of 1 hour and 66 hours lead-time.

With a high degree of similarity (NS>0.7) between the adjusted forecast and its historical analogues, the best statistics of improvement were obtained. Unfortunately, the rate of forecasts that were adjusted by post-processing fell to units of per cent. For the criterion of low degree of similarity (NS>0), between 95 and 99% of all forecasts have already been adjusted by dressing method. However, in this case, the variance of the error models was too large and they produced worse results, especially in the too-large spread of the adjusted ensemble of hydrological forecasts. The size of NS between 0.2 and 0.3 turned out to be the optimal value, which allowed the adjustment of approximately half of the predictions.

The success of river flow forecast can be viewed in different ways and there isn't one perfect evaluation criterion. Therefore three statistical methods were selected for verification of dressing. The positive effect of the adjustment of forecasts was reflected above all in the widening of the spread of ensemble members.

Insufficient spread of original ensemble forecast caused that there were too frequent cases where the observed flow was behind the edge of the ensemble members. It is manifested as too big bars in the rank histogram (Fig. 6). After applying dressing with appropriate parameters the frequency of position of observed discharge between members ensemble forecast was more equal.

Methods based on the evaluation of the probability of exceeding a certain discharge threshold showed significant improvement in the area of average and below-average flow. Furthermore, there was a high rate of adjusted forecasts in this interval of discharge. Towards higher flows, the rate of adjusted forecasts decreases and the effect of post-processing disappears (Fig. 7). The percentage of adjusted forecasts, as well as improvement rate, varies among water gauging profiles.

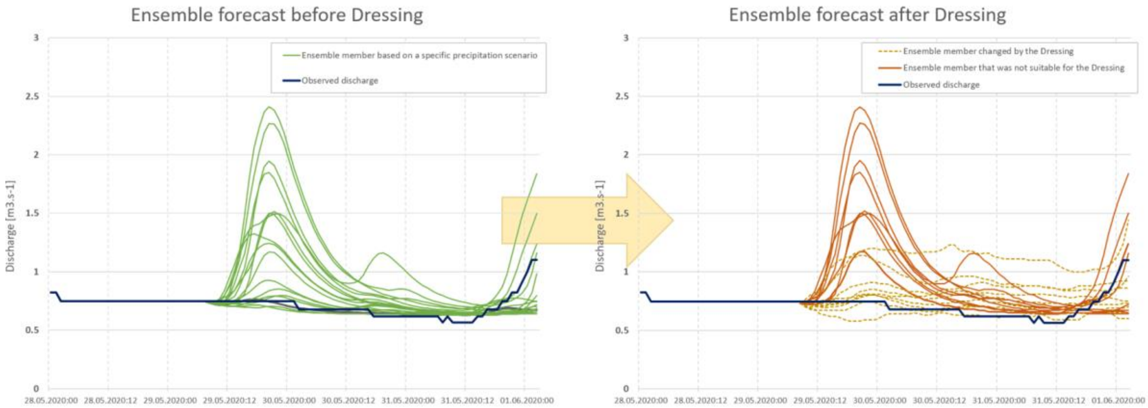

Fig. 5. Example of post-processing with the Dressing method where some of the original member weren't processed because of weak error model.

Fig. 6. Rank histograms of the frequency of the observed average discharge between 17 members of forecasted ensemble of average discharge. Comparison of real-time forecast and the forecast adjusted by post-processing.

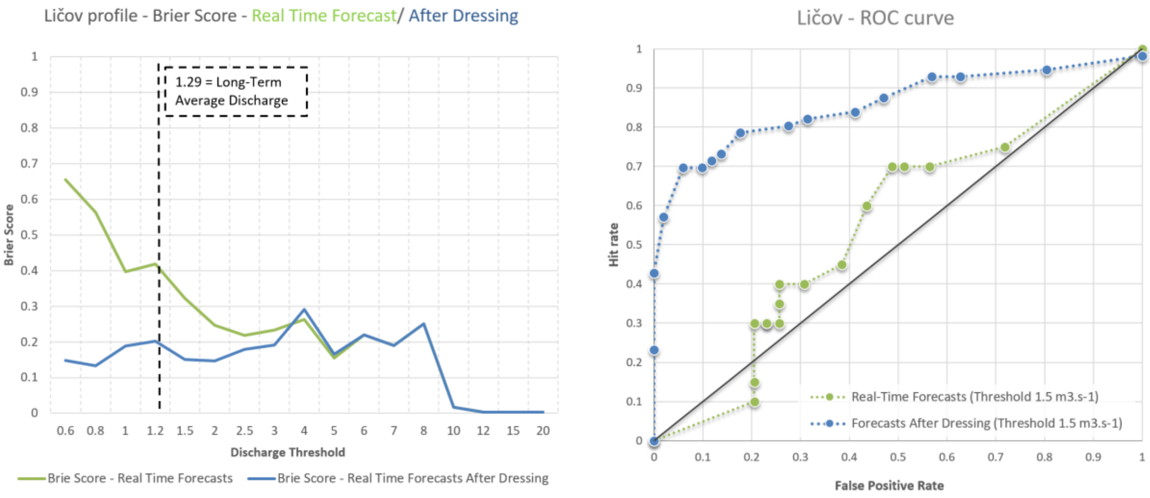

Fig. 7. Brie Score and ROC curve for ensemble hydrological forecasts for Ličov water gauging station (basin area 126 km^2). Comparison of evaluation of original real time forecasts and adjusted forecasts by dressing method.

It is related to different variability rainfall-runoff conditions and the influence of water constructions (weirs, dams).

Discussion

The estimation of hydrological modelling uncertainty using the dressing method depends largely on the extent of the archive of historical forecasts and on the fact whether historical forecasts correspond the uncertainty of hydrological modelling of real-time operations. The archive of 2780 historical forecasts for a period of 9 years provided a sufficient database for post-processing forecasts with little variability, mostly average or below-average flows. Better results can be expected by expanding this archive because it should cause more similarity between the current forecast and historical analogues.

Historical forecasts in which the precipitation forecast is replaced by observed precipitation cannot be calculated in real-time operation but must be prepared in automatic calculation afterwards. Forecasting systems where the operation is highly interactive, e.g the hydrologist significantly interferes with the setting of the initial conditions, or even the parameters of the hydrological model and adjusting the forecast are not suitable for the application of this method. This is because a significant part of the uncertainty of hydrological modelling is associated with hydrologist decision-making, which cannot be transferred to the automatic calculation of historical forecasts. However, the development of hydrological forecasting models, especially the increase of their spatial resolution, leads to more automatic real-time operations.

Recalculation of historical forecasts even in very complex hydrological model is possible. In comparison with numerical meteorological models, which are extremely demanding on the computing capacity of computers it is relatively easy and quick to update the archive of historical forecasts in case of changes in the structure of the model or its parameters. These facts open up space for more frequent use of post-processing methods.

Conclusion

The post-processing method dressing with a dynamically compiled run-time error model is a functional tool for adjusting ensemble hydrological forecasts which are based only on ensemble precipitation forecasts. Methods increase the success of hydrological ensemble predictions by including uncertainty of hydrological modelling. This uncertainty is derived from deviations of historical forecasts with a similar pattern of simulated discharge and observation. Historical forecasts must represent solely possible errors of the hydrological forecasting system as a whole without the influence of precipitation forecast uncertainty.

Testing the effect of dressing on the short-term ensemble's hydrological forecasting method demonstrated a significant improvement in the success of the forecast adjustment. Above all, there was a positive spreading of the variance of the forecast ensemble and also a slight correction of the systematic bias of the flow from hydrological model resimulation. The change was particularly noticeable in the area of average and below-average flows, where hydrological modelling is the dominant source of uncertainty. For forecasts with higher flows and with rising river levels, there wasn't a sufficient number of similar situations in the database of historical forecasts and therefore no adjustment by the dressing method was possible. However, the most of rejected forecasts were runoff episodes where the dominant source of uncertainty is the precipitation forecast, which is covered in the ensemble's meteorological forecast input.

The method is suitable for the operational operation of hydrological services using automatic or semi-automatic forecasting systems. The application of the method into a hydrological forecasting system is simple and can be implemented without disrupting already established processes.

References

Burnash, R., J., C. (1995): The NWS river forecast system – Catchment modeling, in: Computer models of watershed hydrology, edited by: Singh, V. P., Water Resources Publicat., Highlands Ranch, Colorado, U.S.A., 311–366.

Demargne, J., Wu., L., Regonda, S., Brown, J., Lee, H., He, M., Seo, D. J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., Zhu, Y. (2013): The science of NOAA’s operational hydrologic ensemble forecast service. Bull. American Meteorological Society, 130611111953000. http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-12- 00081.1.

Krejčí, J., Zezulák, J. (2009): The use of hydrological system AquaLog for flood warning service in the Czech Republic. [online]. Regional Workshop on Hydrological Forecasting and Real Time Data Management At: Dubrovnik [cit. 30.6.2011] z https://www.researchgate.net/publication/301358002.

Krzysztofowicz, R (1999): Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resources Research 35 (9), 2739–2750.

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., Di, Z. (2017): A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. WIREs Water, e1246. https://doi.org/10.1002/wat2.1246.

Pagano, T., C., Shrestha, D., L., Wang, Q., J., Robertson, D., Hapuarachchi, P. (2012): Ensemble dressing for hydrological applications, In: Hydrological Processes, Special issue S173 Hydrological Ensemble Prediction Systems (HEPS).

Schaake, J., Perica, S., Mullusky, M., Demargne, J., Welles, E., Wu, L. (2004): Pre-processing of atmospheric forcing for ensemble streamflow prediction., In: 17th Conference on Probablity and Statistics in the Atmospheric Sciences, AMS.

Verkade, J. S., Brownd, J. D., Davids, F., Reggiani, P., Weertsaf, A. H. (2017): Estimating predictive hydrological uncertainty by dressing deterministic and ensemble forecasts; a comparison, with application to Meuse and Rhine. Journal of Hydrology, Volume 555, December 2017, Pages 257–277, https://doi.org/10.1016/j.jhydrol.2017.10.024

WMO (2021) Guidelines on Ensemble Prediction System Postprocessing, WMO no-1284, edition 2021, pages 46, ISBN 978-92-63-11254-5

RNDr. Tomáš Vlasák, PhD. (*corresponding author, email: tomas.vlasak@chmi.cz)
Ing. Jakub Krejčí, PhD., MSc.
Czech Hydrometeorological Institute
Na Šabatce 17
140 00 Praha 4
Czech Republic