Available online at www.sciencedirect.com

ScienceDirect

Procedia Earth and Planetary Science 15 (2015) 940 – 944

Procedia
Earth and Planetary Science

World Multidisciplinary Earth Sciences Symposium, WMESS 2015

Stochastic Forecast of Flow Reservoir Behaviour

Tomáš Kozel[a]*, Miloš Starý[a]

[a] Brno university of technology, faculty of Civil engineering, Institute of Landscape water management, Veveří 331/95, 602 00 Brno

Abstract

The main advantage of stochastic forecasting of flow reservoir behaviour is the fan of a possible value, which deterministic methods of forecasting could not give us. Future development of random process is described well by first stochastic then deterministic forecasting. We can categorize the discharge in measurement profile as a random process. The contents of this article is the development of a forecasting model for the management of large open water reservoirs with supply function. The model is based on a linear autoregressive model, which forecasts values of average monthly flow from a linear combination of previous values of average monthly flow, autoregressive coefficients and random numbers. The autoregressive coefficient was calculated from the Yule-Walker equations [2, 3]. The model was compiled for the forecasts in the range of 1 to 12 month with a backward correlation of 2 to 11 months. The data was freed of asymmetry with the help of the Box-Cox rule [1], the value *r* was found by optimization. In the next step, the data was transform to a standard normal distribution. Our data was with monthly step and forecasting was recurrent. We used a 90-year long real flow series ~~for~~ to compile the model. The first 75 years were used for the calibration of the model (autoregressive coefficient), the last 15 years were used only for validation. The model outputs were compared with the real flow series. For comparison between real flow series (100% successful of forecast) and forecasts, we used as values of forecast average, median, modus and miscellaneous quintiles. Results were statistically evaluated on a monthly level. The main criterion of success was the absolute error between real and forecasted flow. Results show that the longest backward correlation did not give the best results. On the other hand, the flow in months, which were forecasted recurrently, give a smaller error than flow forecasted from real flow. For each length of forecast, even for backward size of correlation, different values of quintiles were reached, for which forecasting values gave the smallest error, [4, 5]. Flows forecasted by the model give very fine results in drought periods. Higher errors were reached in months with higher average flows. This higher flow was caused by floods. The floods are predictable. Due to good results in drought, periods we can use the model managed large open water reservoirs with supply function.

Peer-review under responsibilty of the Organizing Commitee of WMESS 2015.

Keywords: stochastic; forecasting; average monthly flow.

* Corresponding author. Tel.: +420 54114 7774.
E-mail address: kozel.t@fce.vutbr.cz

Peer-review under responsibilty of the Organizing Commitee of WMESS 2015.
doi:10.1016/j.proeps.2015.08.150

1. Introduction

The main objective researched was construction of stochastic forecasting model for large open water reservoir with storage function. Advantage of stochastic forecasting model is fan of possibilities, which is able better describe future development of random process, than deterministic model. Flow in measure profile could be classified as random process. For construction of stochastic forecasting model was used interface of program Matlab 2010. The main task for model was ability of forecasting drought month and month with average flow, which are critically important for management of large open water reservoir with storage function. The model works with principle of adaptability, what for model for each time step movement is forgotten previous calculations and calculation of model is started again, but with data for new time step.

2. Data preparation

For construction of forecasting model was used 90 yearlong real flow series average monthly flows. The flow series was measured in measuring profile Bílovice, which is situated on river Svitava. Above the profile are not situated large open water reservoir, and therefore nature development of flow series is not affected by management of large open water reservoir. The first 75 years were used for construction of forecasting model. The last 15 years were used for validation of the model. Data were sorted according month, when they were measured. The data had to be transformed into the standard norm al distribution, but value of asymmetry was non-zero for each distribution of data set, and therefore asymmetry had to be eliminated. For elimination of asymmetry was used Box-Cox rule [1], because average monthly flow could be only positive number, for transformation were used Eq. 1. Problem were values of parameter *r,* which had to be optimized. Grid method was chosen as method of optimization. Optimization was calculated value of parameter *r* and resulting value of asymmetry after transformation data set was in range -0.01 to 0.01.

$$Yi = \frac{x^r - 1}{r} \qquad (1)$$

Where *Yi* is transformed average monthly flow without asymmetry, *x* is average monthly flow, which will be transformed by equation and *r* is coefficient of transformation for chosen month.
Data sets were transformed on standard normal distribution by appropriate transformation equations in next step.

2.1 Model

Model is based on linear autoregressive model, which forecasting values of average monthly flow from linear combination previous values of average monthly flow, autoregressive coefficients and random numbers. Each autoregressive coefficient was calculated from Yule-Walker equations [2, 3]. Forecasting flow is calculated from Eq. 2, when matching pairs are substituted to this equation. The matching pairs are consisted of autoregressive coefficient and average monthly flow (transformed on standard normal distribution). Model is used required number of average monthly flow (2-11) backward and required length of forecast (1-12) is given by calculation. Whole process is repeated for each forecasting monthly flow by 1000 times. This value of repeating was chosen, because model needs enough numbers of repeating to construction of empirical curve of probability and number of repeating could not be too high due to consumption of time for calculation. During repeating is changing only value *rnd* in Eq. 2, because this parameter is changed, model is given each time different values of forecast. Forecast is transformed back to distribution, which is matching with distribution of month for whose forecast was calculated. Model is moved by one step (month) forward after all repeating and whole calculation is repeated with data set for next step, if number of forecasting flow is higher than 1. From text above is came out, if forest is longer than 1 month, than model is calculated with measured data and forecast which was given by model, because size of matrix, which is inputted to Yule -Walker Eq. 2 equations, is not changed.

$$Q = \left(\sum_{i=1}^{d} a_i \times Y_i \right) + rnd \qquad (2)$$

In euation *Q* is forecast of average monthly flow, *a* is regression coefficient, *Y* is backward average monthly flow transformed into standard normal distribution , which is matched with coefficient *a*, *d* is maximal number of backward average monthly flow, which were used by model and *rnd* is random number from standard normal distribution.

3. Results and discussions

Average monthly flow Q_m with high flow rate, which were influenced by flood, was removed, because floods are statistically unpredictable. Q_m with values of flow rate higher than twice of median for corresponding month was classified as month influenced by flood. Average absolute error *Er* between real Q_m and forecasted Q_m for each month in line 1 to 12 was chosen as the main criteria for successfulness of forecasting model. What does quintile from probabilistic distribution of probability forecasted Q_m should be use? Average, median, mode and quintiles from 0.05 with step 0.05 to 0.95 were chosen. Model was given for each month different values of error *Er.* Values of error *Er* was depended on number of Q_m backward, which forecasting model was used. The best results were given, when model was used 4 to 7 Q_m backward. Results were showed up, that error *Er* was reached the smallest value for each month for different number Q_m backward. For drought month was this value equal 7.

If Q_m was forecasted by the model recurrently and were used forecasted Q_m, which were calculated by the model, results were better than results of the model, which forecasting Q_m only with measured values. This was a paradox. The paradox was repeated for all tested values (average, media, etc.) in this month (March, April, October, November and December). The best results were given by model, if Q_m were forecasted as 4 to 7in line. For modus was error *Er* in range 0.6 – 1. 1 m^3/s (8 – 31%). The error *Er* for drought months was in range 0.6 – 0.8 m^3/s (8 - 20%). Real values of Q_m are assumed values 3.5 - 6 m^3/s in drought months for Svitava river. Median was given better results than average for drought months and for months with average Q_m, but the results were not better than results of modus. Average was given better results than median for months with higher values of Q_m. Quintiles were given the best results. Range of error Er for the best quintiles was 0.45 – 0.85 m^3/s (5 - 30%) and for drought months was this range 0.4 – 0.65 m^3/s (5 - 18%). As the best quintiles for drought months the results were showed up quintiles range 0.55 – 0.65 and for month with average Q_m was range 0.5 – 0.6. Each month has different value of quintile for which the model is given the smallest error *Er.* The results of the model are undervalued against real values Q_m.

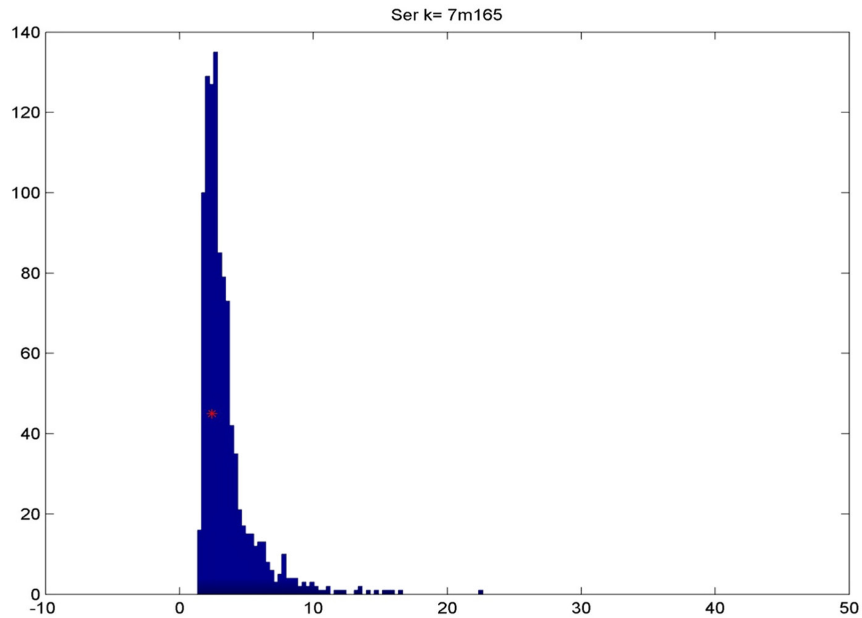

Fig.1. Histogram of forecasted Q_m, horizontal axis is representing values of Q_m m^3/s and vertical axis is representing frequency of Q_m, during repeating. Red mark is symbolized real value of Q_m.

Tab.1. Relative error Er.

Month	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Best quiantile [%]	23*	21	5	11	20	18	14	30	15	16*	25	18
Modus [%]	25*	22	8	15	23	20	17	31	18	18*	28	22

In the Tab. 1 is relative error Er for d=7 (maximal number of backward Q_m), values with * are represented d=6. In second row are relative errors for the best quintile in each months. Values in table are rounded up to whole number.

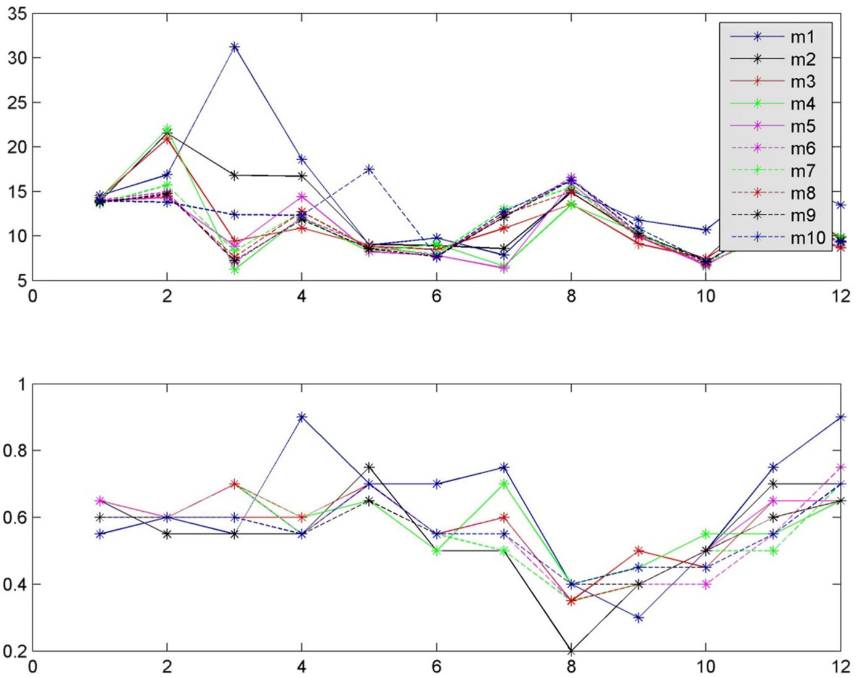

Fig.2. First subgraf, horizontal axis month (February = 1, December = 12), in vertical axis is sum of absolute error Er for the best quintile for each month in m^3/s. Each line is represented order in which model forecasted Q_m (m1= Q_m was forecasted as first in the order). In second subgraf is horizontal axis same as in the firs subgraf and vertical axis is represent values of quintiles', which were given the smallest error *Er*. Lines have same meaning as in the first subgraf.

4. Conclusions

The results of the forecasting model assumed a small error *Er* for drought month and for months with an average flow rate of Q_m. These months are critically important for the management of large open water reservoir with storage function. For month with higher flow rate of Q_m were values of the error *Er* higher, but large open water reservoir do not have problem with lack of water in reservoir for these months. For this reason, the management of the reservoir is able to tolerate a higher error *Er* for these months. A large open water reservoir is able to compensate the error *Er* with its volume of storage water, if values for the error *Er* are not too big.

The forecasting model is able predict values of Q_m with good results. Values of Q_m are undervalued compared to real values of Q_m, with due to this attribute, the forecasting model is good for the use in the management of the reservoir, because using of forecasting model should be lead to a more aggressive management of large open water reservoir with storage function. Aggressive management very often leads to longer but shallower failure. Forecasting model will be tested in the management of large open water reservoirs, where assumptions should be confirmed.

Acknowledgements

Research was supported by FAST-J-15-2798 "Management of storage tanks for the reasoning of uncertainty of hydrological inputs using artificial intelligence methods".

References

1. Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society*, Series B, *26*, 211-252.
2. Yule, G. Udny (1927) "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers", Philosophical Transactions of the Royal Society of London, Ser. A, Vol. 226, 267–298.
3. Walker, Gilbert (1931) "On Periodicity in Series of Related Terms", Proceedings of the Royal Society of London, Ser. A, Vol. 131, 518–532.
4. Starý, M. and Tureček, B. (2000) "Operative control and prediction of floods in the river Odra basin". Flood Issues in Contemporary Water Management, NATO Science Series, 2. Enviromental Security – Vol.71, Kluwer Academic Publishers, 2000, s.229-236, ISBN 0-7923-6452-X.
5. McCouen, R.H. and Snyder, W.H. (1986) "Hydrologic Modeling, Statistical Methods and Applications", Pretince-Hall, New Jersey.