

# Stochastické předpovědi přítoků do nádrže se zásobní funkcí

Tomáš Kozel

## Abstract

The main advantage of stochastic forecasting is fan of possible value, which deterministic method of forecasting could not give us. Future development of random process is described much better by stochastic than deterministic forecasting. We can categorize discharge in measurement profile as random process. Content of article is development of forecasting model for managed large open water reservoir with supply function. 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. Autoregressive coefficient was calculated from Burgess equations (Burgess 1972). The model was compiled for forecast of 1 to 12 month. Data was got rid of asymmetry with help of Box-Cox rule (Box, Cox, 1964), value  $r$  was found by optimization. In next step were data transform to standard normal distribution. Our data were with monthly step and forecasting was recurrent. We used 90 year long real flow series for compile of the model. First 75 years were used for calibration of model (autoregressive coefficient), last 15 years were used only for validation. Outputs of model were compared with real flow series. For comparison between real flow series (100% successfully of forecast) and forecasts, we used as values of forecast average, median, modus and miscellaneous quintiles. Results were statistically evaluated on monthly level. The main criterion of success was average relative error between real and forecasting flow. On the other hand flow in month, which were forecasted recurrently from forecasting flow give smaller error than flow forecasted from real flow. Flows forecasted by the model give very fine results in drought period. Higher errors were reached in months with higher average monthly flows. This higher flow is caused by floods. The floods are very complexly predictable. If we evaluate all months together, we will

decreased precision of outputs, but in months with higher average monthly flows is enough water. This is reason, why we could not give this time period same importance as drought periods. Due to good results in drought periods we can use the model for managed large open water reservoir with supply function.

**Anotace:**

Príspevek obsahuje popis a konstrukci stochastického predpovedního modelu pro řízení nádrže se zásobní funkcí. Jedná se o adaptivní lineární rekurentní model, který obsahuje Burgessovy rovnice. Závěrem bylo provedeno vyhodnocení kvality předpovědi pro jednotlivé měsíce.

**Klíčová slova:** stochastický, předpověď, průměrný měsíční průtok

**Anotation:**

Content of article is development of forecasting model for managed large open water reservoir with supply function. Model is adaptive recurrently linear, which content Burgess equations. Outputs of forecasting model were evaluated for each month.

**Keywords:** stochastic, forecasting, average monthly flow

## 1. ÚVOD

Cílem výzkumu bylo vytvoření stochastického předpovědního modelu pro nádrž se zásobní funkcí. Výhodou stochastického předpovědního modelu je vytvoření vějíře možných hodnot, který lépe charakterizuje budoucí průběh náhodných procesů. Mezi náhodné procesy můžeme zařadit i průtoky vody v měrném profilu. Pro vytvoření stochastického modelu i následnou práci s ním bylo použito prostředí programu Matlab 2010. Hlavním požadavkem na model byla jeho schopnost předpovídat zejména málovodé a průměrné měsíce, které jsou kritické z hlediska řízení nádrže se zásobní funkcí. Model podléhá principu adaptivity, a proto se při každém časovém posunu zapomíná předchozí výpočet a model startuje znovu.

## 2 PŘÍPRAVA A PRÁCE S DATY

Pro sestavení i následnou validaci byla použita 75 let dlouhá řada průměrných měsíčních průtoků. Řada vznikla měřením v měrném profilu Bílovice, který se nachází na řece Svitavě. Nad profilem se nenachází žádné velké nádrže, které by ovlivňovaly přirozený průběh řady. byla rozdělena na dvě části. Prvních 60 let bylo použito při konstrukci modelu. Posledních 15 let bylo použito pro validaci modelu. Data byla dále roztržiděna, podle měsíce v němž se vyskytla. Pro použití v modelu musela být data transformována na normované normální rozdělení, protože se však u dat vyskytovala výrazná asymetrie, musela být data nejdříve transformována tak aby byla asymetrie odstraněna. Pro tuto transformaci bylo použito Box-Cox pravidlo (Box, Cox, 1964), protože průtoky mohou nabývat jen kladných hodnot byla vybrána rovnice 1. Problémem se při použití pravidla ukázal parametr  $r$ , a proto muselo být přistoupeno k jeho optimalizaci, která byla provedena za pomoci mřížkové metody. Z optimalizace byl získán parametr  $r$  a výsledná asymetrie nabývala hodnot v rozmezí -0.01 až 0.01.

$$\text{Eq1. } Y_i = \frac{x^r - 1}{r}$$

Kde  $Y_i$  je transformovaný průtok na rozdělení bez asymetrie,  $x$  je průtok který je transformován a  $r$  je koeficient transformace pro vybraný měsíc.

Dále byly data transformovány na normované normální rozdělení pomocí vhodných transformačních vztahů.

### 3 MODEL

Model je lineární autoregresní model, který předpovídá hodnoty průměrných měsíčních průtoků na základě lineární kombinace hodnot předchozího průměrného měsíčního průtoku, autoregresních koeficientů a náhodných čísel. Pro stanovení autoregresních koeficientů byly použity Burgessovy rovnice (Burgess, 1972). Po jejich vyřešení dostaneme koeficienty, které spolu s průtoky transformovanými na normované normální rozdělení tvoří dvojice z jejichž pomocí podle rovnice 2 dostaneme předpovídaný průtok. Při samotném výpočtu model použije hodnotu předchozího průměrného průtoku v předchozím měsíci a vypočte rekurentní předpověď na požadovaný počet měsíců dopředu (1 – 12). Celý proces je pro každou předpovídaný měsíc opakován 1000 krát. Hodnota 1000 byla zvolena, proto aby byl proveden dostatečný počet opakování pro vytvoření empirické čáry pravděpodobnosti a výpočet nevyžadoval enormní nároky na čas. Při opakování je měněna pouze hodnota  $rnd$  v rovnici 2, protože je uvedený parametr měněn, dostaneme pokaždé jinou hodnotu předpovědi. Předpovědi jsou následně zpětně transformovány na rozdělení, které odpovídá měsíci, pro který byla předpověď vypočtena. Poté se posune o jeden krok (měsíc) vpřed a celý výpočet opakuje, pokud je požadovaná předpověď delší než 1 měsíc. Z výše uvedeného textu vyplývá, že pokud požaduje delší předpověď než je 1 měsíc, tak se nám do modelu mimo historicky naměřených dat dostávají i předpovědi poskytnuté modelem, protože se nemění velikost matice vstupující do Burgessových rovnic.

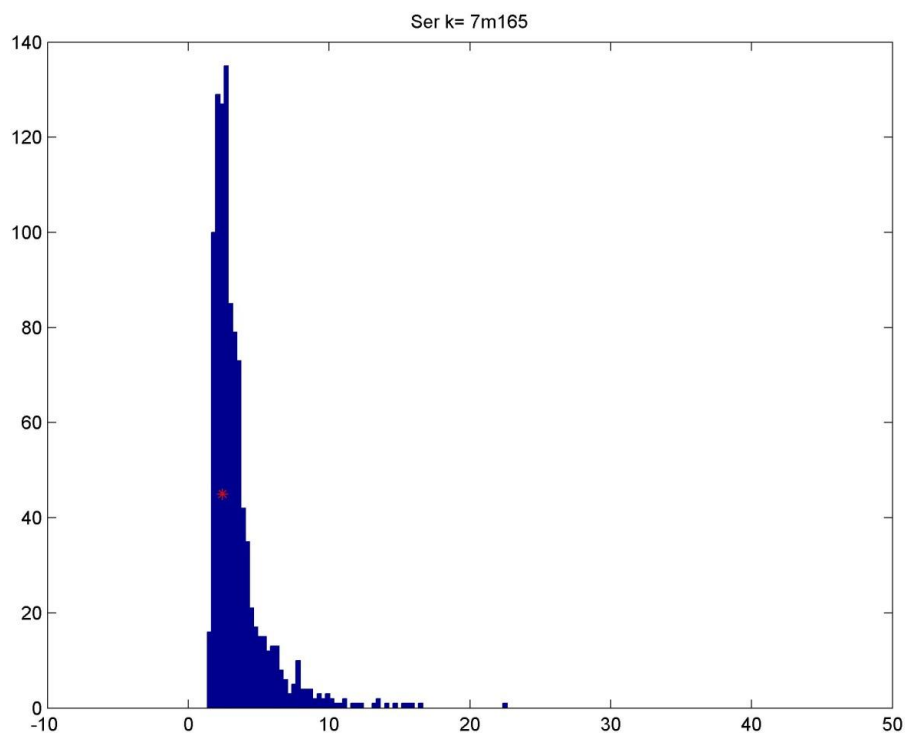
$$\text{Eq2} \quad Q = \rho_i \times Z_{i-1} + (1 - \rho_i)^{1/2} \times rnd$$

Kde  $Q$  je předpovídaná hodnota průtoku,  $\rho$  je koeficient regrese a  $Z$  je průtok měsíce uvažovaného zpětně, transformovaný na normované normální rozdělení a  $rnd$  je náhodné číslo z normovaného normálního rozdělení.

## 4 VYHODNOCENÍ

Při vyhodnocování byly z validačního období odstraněny měsíce s vysokými hodnotami průměrných měsíčních průtoků, které byly prokazatelně ovlivněny povodněmi, které jsou statisticky nepředvídatelné. Za vysokou hodnotu se považoval dvojnásobek mediánu průměrného měsíčního průtoku  $Q_m$  v příslušném měsíci. Základním kritériem pro posouzení úspěšnosti byla průměrná absolutní chyba mezi reálným průměrným průtokem a předpovězeným průměrným průtokem pro každý měsíc v pořadí 1 až 12. Jaké kvantily z pravděpodobnostního rozdělení pravděpodobností předpovězených  $Q_m$  v daném měsíci použít? Bylo rozhodnuto použít průměr, medián, modus a kvantily od 0.05 s krokem 0.05 do 0.95. Model dosahoval v každém měsíci různých velikostí chyby. Velkým paradoxem však bylo, že pokud byl  $Q_m$  předpovídán rekurentně z již předpovězených  $Q_m$  vypočtených modelem dával lepší výsledky než hodnoty získané výpočtem z historicky naměřených dat. Tento výsledek se opakoval pro všechny srovnávací hodnoty (modus, medián, průměr, kvantily) v měsících březen, duben, říjen, listopad, prosinec. Nejlepších výsledků bylo dosaženo, pokud byl  $Q_m$  předpovídán 4 až 7 v pořadí. Při podrobnějším vyhodnocování bylo dospěno k závěru, že modus dosahuje velikosti sumy absolutní chyby pro vybraný měsíc ve validačním období v rozmezí  $0.6 - 1.1 \text{ m}^3/\text{s}$  (8 – 31%). Hodnoty výše zmíněné chyby se v málovodých obdobích pohybovaly v rozmezí  $0.6 - 0.8 \text{ m}^3/\text{s}$  (8 - 20%), což vzhledem k průměrným měsíčním průtokům  $3.5 - 6 \text{ m}^3/\text{s}$  pro uvedené období je úspěch.

Druhým krokem bylo srovnání zbylých ukazatelů. Medián dával lepší výsledky než průměr v průměrných i málovodých měsících. Naopak průměr dával lepší výsledky než medián pro velmi vodné měsíce. Nejlepší výsledky poskytnuté kvantily nabývaly průměrné absolutní chyby pro vybraný měsíc ve validačním období v rozmezí 0.4 – 0.85 m<sup>3</sup>/s (5 – 30 %), Hodnoty se v málovodých obdobích pohybovaly v rozmezí 0.4 – 0.65 m<sup>3</sup>/s (5 – 18 %). Nejlepších výsledků dosáhly kvantily v rozpětí 0.45 – 0.65. Pro málovodé měsíce to byly kvantily 0.55 až 0.65. Pro průměrné měsíce pak kvantily 0.5 až 0.6. Z výsledků vyplynulo, že každý měsíc má svůj kvantil, a pokud použijeme za výslednou předpověď 4. až 6. předpovídaný měsíc v pořadí, dostaneme lepší výsledky, než pokud je brán za předpověď 1. nebo druhý měsíc v pořadí (pouze již u výše zmíněných měsíců). Při procházení výsledků bylo zjištěno, že předpovězené  $Q_m$  jsou oproti reálným hodnotám podhodnocovány.



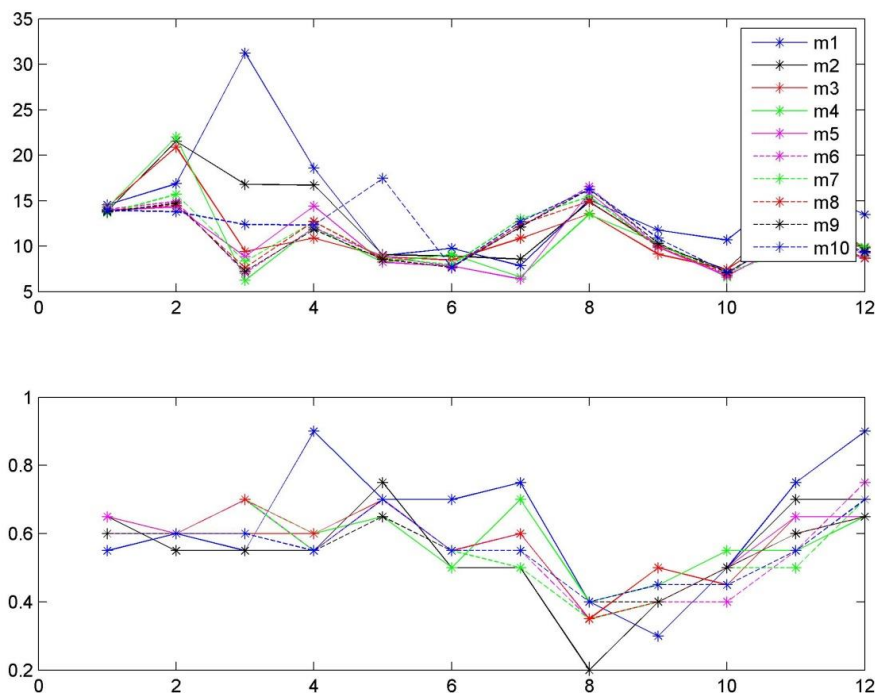
**Fig. 1.** Na obrázku je histogram, který má na ose x hodnotu průtoku  $\text{m}^3/\text{s}$  a na ose y celkový počet výskytů průtoků v předpovědním modelu. Červenou hvězdičkou je znázorněna hodnota reálného průtoku, který nastal. Obrázek reprezentuje výsledek modelu pro měsíc, který je předpovídán jako sedmý v pořadí a je uvažováno šest měsíců zpětně.

**Tab. 1.** Průměrná relativní chyba  $E_r$

Měsíc	1	2	3	4	5	6	7	8	9	10	11	12
Kvantil [%]	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

V tabulce je uvedena relativní průměrná absolutní chyba pro jednotlivé měsíce. Na druhém řádku je velikost chyby pro nejlepší kvantil a na třetím pro modus. Hodnoty v tabulce jsou zaokrouhleny na celá čísla.





**Fig. 2.** Na prvním subgrafu jsou na ose x měsíce (leden = 1, prosinec = 12) a na ose y je suma absolutní chyby mezi nejlepším kvantilem pro daný měsíc a naměřenou hodnotou v  $\text{m}^3/\text{s}$ . Každá čára reprezentuje, jako kolikátý měsíc v pořadí byl předpovídán (m1 = měsíc byl předpovídán jako první). Na druhém subgrafu je osa x shodná s prvním subgrafem a na ose y jsou hodnoty kvantilů. Každá čára reprezentuje jako kolikátý měsíc v pořadí byl předpovídán.

## 5 ZÁVĚR

Model dokázal velmi dobře předpovídat průměrné měsíční průtoky, zejména v málovodých a průměrněvodých měsících, které jsou stěžejní při řízení nádrže se zásobní funkcí. Ve vodných obdobích nabývala chyba vyšších hodnot, avšak ve vodných obdobích nejsou problémy s nedostatkem vody, a proto vyšší chybu lze tolerovat. Nádrž je schopna do určité míry výše zmíněné chyby kompenzovat, a proto lze model použít při řízení nádrže se zásobní funkcí. Model je schopný předpovídat  $Q_m$  s dobrými výsledky, a protože podhodnocuje  $Q_m$  mělo by nastat při jeho použití agresivnější řízení nádrže se zásobní funkcí, které povede k mělkým a

delším poruchám. Model se bude testovat při řízení nádrže, kde by měli být výše zmíněné předpoklady potvrzeny.

## **6 ACKNOWLEDGEMENTS**

Článek vznikl za přispění FAST-J-15-2798 “ŘÍZENÍ ZÁSOBNÍ FUNKCE NÁDRŽE PŘI UVAŽOVÁNÍ NEJISTOT HYDROLOGICKÝCH VSTUPECH S POUŽITÍM METOD UMĚLÉ INTELIGENCE”

## **7 REFERENCE**

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 26, 211-252.

Burgess, S.J. (1972) Some Problems with Log Normal Markov Runoff Models, *Journal of the Hydraulics*

Division, ASCE, Vol. 98, No. HY9, 1487

Walker, Gilbert (1931) "On Periodicity in Series of Related Terms", *Proceedings of the Royal Society of London, Ser. A*, Vol. 131, 518–532.

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. Environmental Security – Vol.71, Kluwer Academic Publishers, 2000, s.229-236, ISBN 0-7923-6452-X.*

McCouen, R.H. and Snyder, W.H. (1986) “Hydrologic Modeling, Statistical Methods and Applications”, *Pretince-Hall, New Jersey.*