Hybridní lineární stochastický předpovědní model pro řízení zásobní funkce nádrže

Tomas Kozel[1]

[1] University of Technology, Faculty of Civil Engineering, Institute of Landscape Water Management, Veveří 331/95 Brno, Czech Republic

kozel.t@fce.vutbr.cz

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 better by stochastic then deterministic forecasting. Discharge in measurement profile could be categorized as random process. Content of article is construction and application of forecasting model for managed large open water reservoir with supply function. Hybrid model is based on combination of autoregressive linear and zone models. The model forecast values of average monthly flow from combination of historic values of average monthly flow and random numbers. Part of data was sorted to one moving zone. The zone is created around last measurement average monthly flow. Matrix of correlation was assembled only from data belonging to zone. Regressive coefficients are computed by Yule-Worker equations (Yule, 1927 and Worker, 1931). The model was compiled for forecast of 1 to 12 month with using backward month flows from 2 to 11 months for model construction. Data was got ridded 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. The data were with monthly step and forecast is recurring. 90 years long real flow series was used for compile of the model. First 75 years were used for calibration of model (matrix input-output relationship), 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, was used application to management of artificially made reservoir. Course of water reservoir management using Genetic algorithm (GE) + real flow series was compared with Fuzzy model (Fuzzy) + forecast made by hybrid model. During evaluation process was founding the best size of zone. Results show that the highest number of input did not give the best results and ideal size of zone is in interval from 10 to 15, when course of management was almost same for all numbers from interval. Resulted course of management was compared with course, which was obtained from using GE + real flow series. Comparing results showed that fuzzy model with forecasted values has been able to manage main malfunction and artificially disorders made by model were founded essential, after values of water volume during management were evaluated. Forecasting model in combination with fuzzy model provide very good results in management of water reservoir with storage function and can be recommended for this purpose.

1. Úvod

Hlavním cílem výzkumu bylo sestavení stochastického předpovědního modelu (generátoru krátkých umělých řad) pro řízení zásobní funkce nádrže. Výhodou stochastické předpovědi je vějíř možných budoucích průtoků vody v měrném profilu (přítoků vody do nádrže), které jsou schopny lépe popsat budoucí vývoj náhodných procesů, mezi které může být zařazen průtok v měrném profilu.

Předpovědní model je vytvořen pro předpovídání přítoků vody do izolované vodní nádrže se zásobní funkcí, které je řízena s měsíčním krokem. Základním požadavkem na předpovědní model je schopnost předpovídat dlouhá málovodá období, která jsou kritická pro řízení zásobní funkce nádrže. Povodňové průtoky nepůsobí problémy při řízení zásobní funkce, a proto model může při jejich předpovídání dosáhnout větších chyb.

2. Data

Pro kalibraci a validaci modelu byl zvolen měrný profil Bílovice nad Svitavou. Reálná řada je 90 let dlouhá. Měrný profil je velmi málo ovlivněn řízením nádrže a je velmi blízký neovlivněnému toku. Prvních 75 let reálné řady je použito pro kalibraci předpovědního modelu a posledních 15 let je použito pro validaci. S daty se pracovalo na úrovni měsíců. Každý měsíc má jiné rozdělení pravděpodobnosti, a proto bylo přistoupeno k práci s daty na úrovni hladiny normálního normovaného rozdělení (hladina *Z*). Všechna data byla pomocí dvoustupňové transformace převedena z hladiny *Q* (hodnoty reálné řady).

První krokem transformace bylo odstranění asymetrie, ke kterému byl použit Box-Coxův [1] vztah

$$Y_{i,j} = \frac{x_{i,j}^{\ r_j} - 1}{r_j}, \qquad (1)$$

kde $Y_{i,j}$ je transformovaný člen reálné řady na rozdělení bez asymetrie, $x_{i,j}$ je člen reálné řady, který je transformován rovnicí, r_j je transformační koeficient pro příslušný měsíc, *i* je číslo měsíce a *j* je číslo pořadí členu v i-tém měsíci (roky). Na hladině *Y* je předpokládáno normálního rozdělení, a proto druhý stupeň transformace je proveden pomocí tradičních transformačních vztahu mezi normálním rozdělení (hladina *Y*) a normálním normovaným rozdělením (hladina *Z*).

3. Model

Předpovědní model je hybridem mezi lineárním regresním modelem a zonálním modelem [2,3]. Běžný zonální model nemusí mít dostatek dat v jednotlivých zónách, a proto je výhodnější sestavit zónu v okolí zvolené hodnoty. Do zóny je uvažován stejný počet dat z okolí hodnoty, která jsou menší (větší) než zvolená hodnota. Uniformně rozdělený počet dat v jednotlivých zónách v zonálním modelu, může způsobit nepřesné nalezení závislosti, protože v okolí vybrané hodnoty může být jednostranně nakloněno v případě, že vybraná hodnota je okrajovou hodnotou v zóně. Uvedený problém je opět vyřešen s pomocí plovoucí zóny, protože zvolená hodnota je uprostřed zóny (medián zóny). Výjimka je tvořena okrajovými zónami, kde není možné zajistit, aby zvolená hodnota byla mediánem zóny. V uvedeném případě lze model prohlásit za zonální.

Samotné vytvoření modelu je provedeno ve dvou krocích. Před započetím konstrukce zóny je potřeba provést transformaci dat z hladiny *Q* na hladinu *Z* pomocí výše uvedeného postupu. Veškerá data jsou seřazena podle zvoleného měsíce (poslední reálný průměrný měsíční průtok vody v měrném profilu) od nejmenšího po největší. Poté je sestavena zóna v okolí zvolené hodnoty. Druhým krokem je nalezení regresních koeficientů lineárního modelu, který je aplikován na data ze sestavené zóny.

V druhém kroku je aplikován lineární autoregresní model, který předpovídá hodnoty průměrných měsíčních průtoků na základě lineární kombinace hodnot předchozích průměrných měsíčních průtoků (hladina *Z*), autoregresních koeficientů a náhodných čísel. Pro stanovení autoregresních koeficientů byly použity Yule-Walkerovy rovnice [4,5]. Po jejich vyřešení dostaneme koeficienty, které spolu s průtoky transformovanými na normované normální rozdělení tvoří dvojice, s jejichž pomocí podle rovnice (2) dostaneme předpovídaný průtok. Při samotném výpočtu model použije předepsaný počet měsíců zpětně (2 - 11) a vypočte rekurentní předpověď na požadovaný počet měsíců dopředu (1 – 12). Při opakování je měněna pouze hodnota $rnd_{i,j}$ v rovnici (2). Uvedená veličina je náhodně generována, proto dostaneme pokaždé jinou hodnotu předpovědi (myšlenka metody Monte-Carlo). Předpovědi jsou následně zpětně transformovány na rozdělení, které odpovídá měsíci, pro který byla předpověď vypočtena. Na obrázku 1 je uvedeno schéma modelu. Pokud je požadovaná předpověď delší než 1 měsíc, předpověď se posune o jeden krok (měsíc) vpřed a celý výpočet opakuje. Z výše uvedeného textu vyplývá, že pokud je požadována delší předpověď než 1 měsíc, dostávají se do modelu mimo data reálné řady i předpovědi poskytnuté modelem rovnice (2). Model nemění velikost matice vstupující do Yule-Walkerovy rovnice.

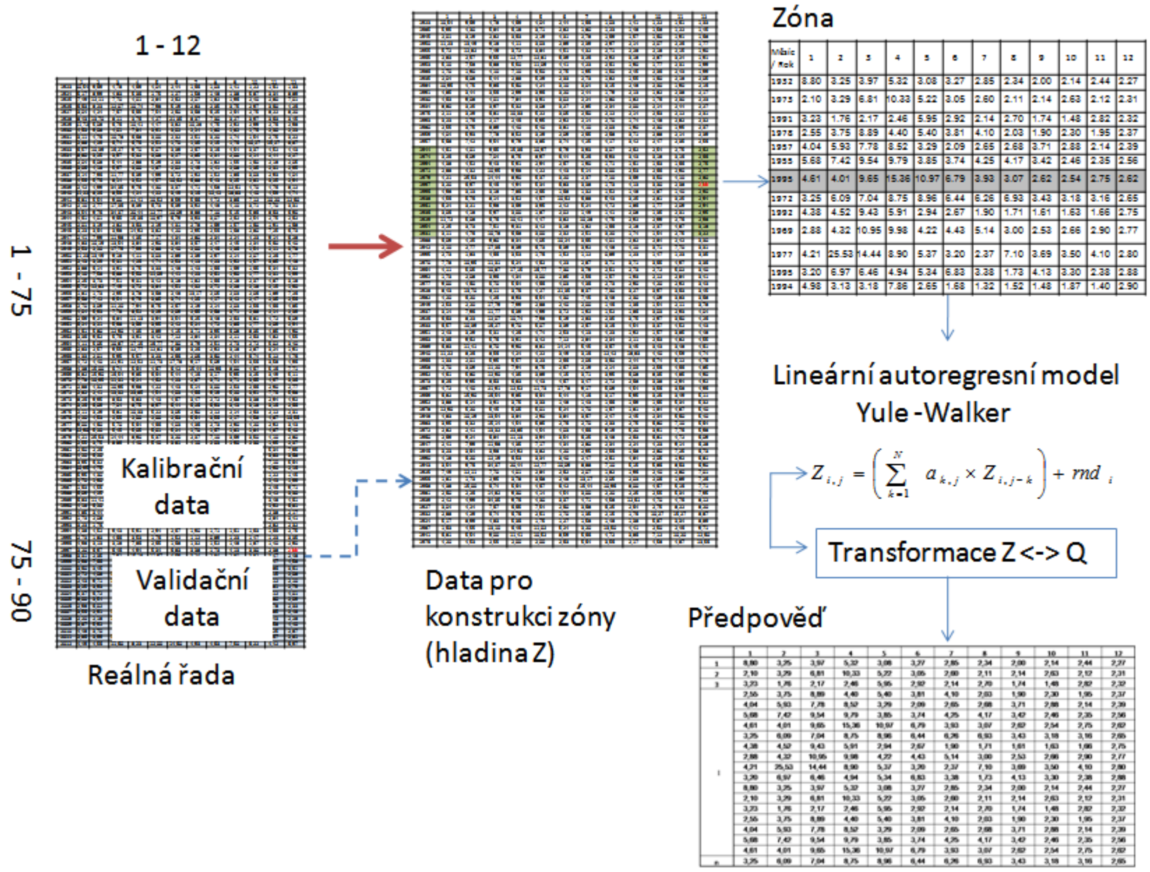
Základní rovnice lineárního autoregresního modelu má tvar:

$$Z_{i,j} = a_1 \cdot Z_{i,j-1} + a_2 \cdot Z_{i,j-2} + a_3 \cdot Z_{i,j-3} + ... + a_k \cdot Z_{i,j-k} + rnd_{i,j}\,, \qquad (2)$$

v uvedeném vztahu značí a_1 až a_k regresní koeficienty, $Z_{i,j-1}$ až $Z_{i,j-k}$ jsou předchozí hodnoty průměrných měsíčních průtoků transformované na hladinu *Z*, $rnd_{i,j}$ je náhodné číslo generované z normovaného normálního rozdělení, *j* je pořadí měsíce (pokud předpovědi přímo navazují na reálnou řadu a $Z_{i,j-1}$ až $Z_{i,j-k}$ jsou posledními členy reálné průtokové řady transformovaní na hladinu *Z*) a *i* je číslo vydávané předpovědi Yule-Walkerovy rovnice

$$\begin{pmatrix} 1 & \rho(1) & ... & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & ... & \rho(k-2) \\ ... & ... & ... & ... & ... \\ \rho(1) & ... & \rho(1) & 1 & \rho(1) \\ \rho(k-1) & \rho(k-2) & ... & \rho(1) & 1 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ ... \\ a_{k-1} \\ a_k \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ ... \\ \rho(k-1) \\ \rho(k) \end{pmatrix}, \qquad (3)$$

kde a_1 až a_k jsou hledané regresní koeficienty a $\rho(1)$ až $\rho(k)$ jsou příslušné korelační koeficienty.



Obr. 1. Schéma hybridního modelu

4. Výsledky a diskuze

Model byl sestaven pro účel řízení zásobní funkce nádrže, a proto bylo přistoupeno k jeho vyhodnocení pomocí aplikace na řízení zásobní funkce nádrže. V měrném profilu Bílovice nad Svitavou se však žádná nádrž nenachází, a proto bylo přistoupeno k vytvoření fiktivní nádrže. Nádrž byla sestavena, tak aby vykazovala poruchy v podobě nedodávek vody. Nádrž má jeden přítok Q, jeden odtok vody O a objem vody v nádrži na počátku každého kroku řešení $V_{\tau-1}$ (počáteční podmínka) pro všechny kroky řešení. Přítok Q (okrajové podmínky) a odtok O jsou popsány řadou přítoků Q_τ (Q_1, Q_2,..., Q_N) a řízených odtoků O_τ (O_1, O_2,..., O_N), N kde je délka předpovědi a τ je časový krok předpovědi. Chování nádrže je popsáno rovnicí, protože je řešení prováděno s měsíčním krokem je vhodné použít diferenční tvar rovnice

$$Q^\tau - O^\tau = \frac{V^\tau - V^{\tau-1}}{\Delta t}. \qquad (4)$$

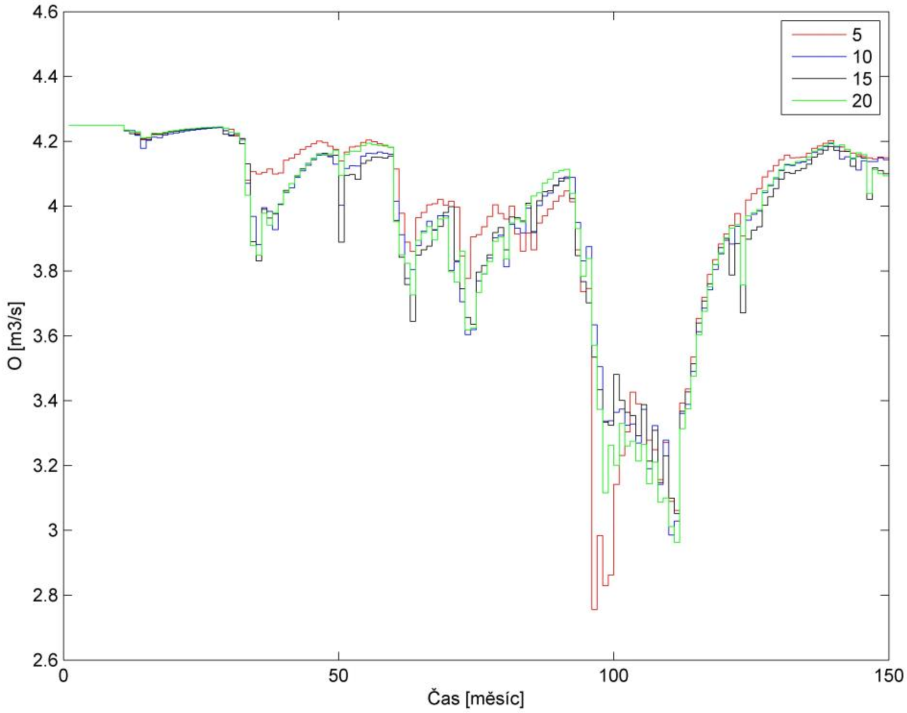
Řízení je prováděno na řídící hodnotu odtoku O_p, která byla zvolena jako konstantní pro všechny měsíce. Odtok vody leží v intervalu $(0,O_p>$, nižší hodnoty odtoku než O_p způsobí poruchu v řízení. Cílem řízení je vytváření dlouhých a mělkých poruch, které jsou výhodnější při řízení zásobní funkce nádrže než poruchy krátké a hluboké. Okrajové podmínky Q jsou při výpočtu nahrazeny krátkými předpověďmi. Na konci každého kroku je spočtena nová hodnota objemu, která v následujícím kroku použita jako počáteční podmínka $V_{\tau-1}$.

Hlavním kriteriem pro posouzení úspěšnosti řízení byla zvolena suma odchylek druhých mocnin od ideálního řízení. Jako ideální průběh řízení jsou uvažovány výsledky řízení poskytnuté genetickými algoritmy, které používají jako předpovědi výseky reálné řady (100% přesnost předpovědi) a používaly pro optimalizaci rovnici 5, π kde je hodnota optimalizovaného kritéria [6].

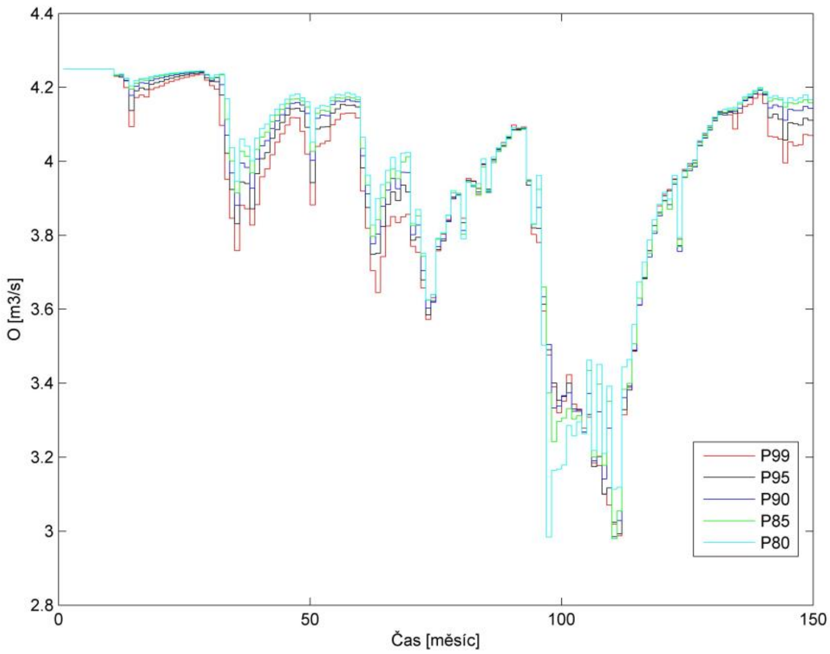
$$\pi = \sum_{\tau=1}^{N} \left(O_p - O^\tau\right)^2 \rightarrow MIN. \qquad (5)$$

Samotné stochastické řízení využívá adaptivní fuzzy model [7]. Pouze první člen řady řízených odtoků je použit pro vlastní řízení. Pro řízení byla použita délka předpovědi 4 měsíce a celkový počet předpovědí pro každý časový krok výpočtu byl 500. Fuzzy model byl zvolen z důvodu velké časové náročnosti potřebné genetickými algoritmy pro stochastické výpočty.

Během procesu vyhodnocování byla hledána vhodná velikost zóny a počet měsíců, které byly použity pro sestavení matice korelace. Výsledky na obrázku 2 ukázaly, kde na vertikální ose je řízený odtok vody *O* a na horizontální ose je uveden čas v měsících, že je velmi vhodné použít pro tvorbu předpovědi velikost zóny 10 – 15. Výsledky poskytnuté velikostí zóny podle uvedeného intervalu si byly velmi podobné, a proto byly prohlášeny za kvalitní. Pokud byla použita velikost zóny mimo uvedený interval, došlo ke zhoršení průběhu řízení. Nejlepší výsledky byly dosaženy při použití 4 měsíců zpětně pro tvorbu korelační matice. V poslední kroku vyhodnocení byla hledána vhodná pravděpodobnost překročení řízeného průměrného měsíčního odtoku vody z nádrže označovaného v obrázku 3 například jako P90. Výsledky na obrázku 3 ukázaly, popis os je shodný s obrázkem 2, že nejlepších výsledků bylo dosaženo pro pravděpodobnost P90.

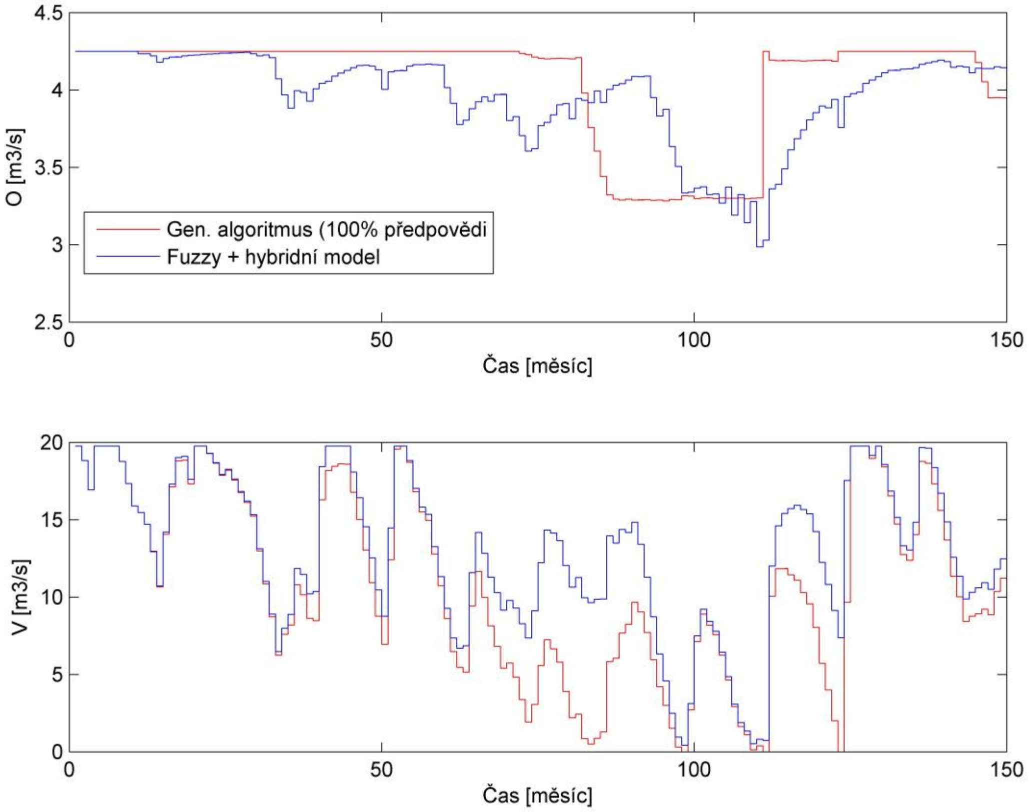


Obr. 2. Srovnání výsledků pro různé velikosti zón



Obr. 3. Srovnání výsledků pro různé pravděpodobnosti

Výsledky na obrázku vykreslují srovnání výsledků stochastického řízení fuzzy modely pro předpovědi poskytnuté hybridním modelem (4 měsíce zpětně, velikost zóny 12) a výsledků poskytnutých genetickými algoritmy používající 100% přesnost předpovědi. Na horním grafu obrázku 3, popis os je shodný s obrázky 2 a 3, jsou srovnány jednotlivé průběhy řízení. Z obrázku 3 je patrné, že stochastické řízení vytváří výrazně více poruch, oproti ideálnímu průběhu řízení. Na spodním grafu obrázku 3 je uveden průběh plnění a prázdnění fiktivní nádrže, na vertikální ose je uveden poměrný objem vody v nádrži a na horizontální čas v měsících, ze kterého jsou patrné příčiny jednotlivých poruch vytvořených navíc oproti ideálnímu průběhu řízení. Fuzzy model spolu s hybridním předpovědním modelem zavadí agresivnější způsob řízení, který se snaží udržet vyšší objemy vody.



Obr. 4. Srovnání výsledku stochastického řízení a ideálního řízení

5. Závěr

Výsledky ukázaly, že hybridní předpovědní model spolu s fuzzy modelem jsou schopny provést efektivní řízení zásobní funkce nádrže. Při použití velikosti plovoucí zóny z intervalu <10,15> a čtyř měsíců zpětně při konstrukci předpovědního modelu. Pokud zóna byla větší, byly závislosti zašuměny ve větším množství dat. U nižšího počtu dat v zóně model nedokázal nalézt vhodné závislosti z důvodu nedostatku dat. Řízení vykazovalo velmi dobré výsledky pro pravděpodobnost překročení řízeného odtoku 0.9.

Samotné výsledky řízení fuzzy modelu používajícího předpovědi vytvořené hybridním předpovědním modelem byly agresivnější než v případě ideálního průběhu řízení. Uvedená vlastnost by mohla být velmi přínosná, při zachování trendu poklesů průtoků, který by vedl k vytvoření vyšší napjatosti mezi přítoky a nadlepšeným odtokem vody z nádrže. Skutečnost, že řízení má tendenci udržovat vyšší hladinu vody v nádrži, by mohla být s výhodou využita při řízení hydroenergetické funkce nádrže.
Závěrem lze prohlásit, že hybridní model s plovoucí zónou je schopen v kombinaci s fuzzy modelem provést řízení zásobní funkce logicky, a proto jej lze doporučit pro další testování či použití.

Poděkování
Příspěvek vznikl za podpory výzkumu FAST-J-17-4044 Řízení zásobní funkce nádrží při uvažování nejistot hydrologických vstupech s použitím metod umělé inteligence s podporou stochastických předpovědních modelů.

Literatura

[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] MENŠÍK, P.; STARÝ, M.; MARTON, D., Using Predictive Models of Mean Monthly Flows for Operative Outflows Control from Large Open Reservoirs, příspěvek na konferenci Proceedings ITISE 2014, International work-conference on Time Series, ISBN 978-84-15814-97-9, Copicentro Granada S.L, Spain, Granada, 2014
[3] MARTON, D.; MENŠÍK, P.; STARÝ, M., Using Predictive Model for Strategic Control of Multi-reservoir System Storage Capacity, článek v Procedia Engineering, ISSN 1877-7058, Elsevier Science Publishers, Amsterdam, 2015
[4] 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.
[5] Walker, Gilbert (1931) "On Periodicity in Series of Related Terms", Proceedings of the Royal Society of London, Ser. A, Vol. 131, 518–532.
[6] HAUPT, R. L. and HAUPT, S. E. (2003) Appendix II: MATLAB Code, in Practical Genetic Algorithms, Second Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471671746.app2
[7] KOZEL, T.; STARÝ, M., STOCHASTIC MANAGEMENT OF THE OPEN LARGE WATER RESERVOIR WITH STORAGE FUNCTION WITH USING FUZZY LOGIC, příspěvek na konferenci SGEM Conference Proceedingsc, ISSN 1314-2704, ISBN 978-619-7105-61-2, STEF92 Technology Ltd., 51 Alexander Malinov Blvd., 1712, Sofia, Bulgaria, 2016