ANALÝZA ČASOVÝCH RADOV DUSIČNANOV POMOCOU NELINEÁRNYCH MODELOV

Peter Valent[1], Ján Szolgay[1], Magda Komorniková[2]

Abstrakt

V príspevku sme analyzovali časové rady priemerných mesačných koncentrácií dusičnanov v riekach Hron a Váh. Dané časové rady sme najskôr rozložili na trendovú, sezónnu, cyklickú a reziduálnu zložku, ktorú sme potom použili na modelovanie lineárnymi modelmi triedy ARMA. V druhej časti sme pôvodné časové rady modelovali nelineárnymi viacrežimovými modelmi SETAR (z angl. self-exciting threshold autoregressive) a MSW (z angl. Markov-switching). Porovnaním oboch tried modelov sme zistili, že modelmi MSW dokážeme výrazne vylepšiť tradične používané modely ARMA, keď v niektorých prípadoch bolo zlepšenie až o 70 %.

Kľúčové slová: dusičnany, ARMA modely, MSW modely, SETAR modely

1. Úvod

Dusičnany, ako jedny z najdôležitejších látok spôsobujúcich znečistenie a eutrofizáciu povrchových a podzemných vôd sa v posledných desaťročiach stali predmetom záujmu v mnohých rozvinutých štátoch. Zvýšený obsah dusičnanov vo vodách má za následok vážne problémy s kvalitou vody ako napríklad premnoženie fytoplanktónu, úbytok rozpusteného kyslíka alebo zníženú transparentnosť čo obyčajne vedie k strate mnohých rastlinných a živočíšnych druhov, a teda aj k degradácii biologickej diverzity. Jednoduchým nástrojom na zhodnotenie úspešnosti a vplyvu týchto opatrení na dynamiku dusičnanov v povrchových a podzemných vodách je analýza časových radov nameraných koncentrácií, ktorá nám taktiež umožňuje lepšie pochopiť základný mechanizmus, pomocou ktorého sú pozorované veličiny tvorené.

Tradičným postupom pri analýze časových radov dusičnanov je dekompozícia časového radu na systematické zložky tvorené trendovou, sezónnou, cyklickou a reziduálnou zložkou, kde reziduálna zložka je ďalej modelovaná lineárnymi modelmi triedy ARMA (z angl. autoregressive moving average). Najväčším nedostatkom týchto modelov je ich neschopnosť zohľadniť rozdielne podmienky vplývajúce na tvoriaci mechanizmus pozorovanej veličiny, ako napríklad v tomto prípade veľkosť prietoku, množstvo zrážok alebo časť roku. Tento problém bol už čiastočne prekonaný viacrežimovými nelineárnymi modelmi používanými hlavne v ekonometrii. Cieľom tejto práce bolo porovnať lineárne a nelineárne modely a ich schopnosť opísať modelované časové rady.

2. Analýza časových radov

Keďže lineárnymi modelmi typu ARMA je možné modelovať iba stacionárne časové rady (s konštantným priemerom a rozptylom), je potrebné zaistiť ich stacionaritu. Na jej zabezpečenie sme použili rozklad časového radu na jednotlivé systematické zložky pozostávajúce z trendovej, sezónnej, cyklickej a zostatkovej reziduálnej zložky, ktorú sme ďalej použili ako základ pre lineárne modelovanie. Jednotlivé systematické zložky sú k sebe pripočítané a tvoria tzv. aditívny model, ktorý možno zapísať ako:

$$Y_t = T_t + S_t + C_t + R_t$$

kde Y predstavuje koncentráciu dusičnanov v čase t, T je trendová, S sezónna, C cyklická a R reziduálna zložka.

2.1 Trend

Trend je charakterizovaný ako dlhodobá zmena priemernej hodnoty časového radu [8]. Pri určovaní trendu treba byť obzvlášť opatrný najmä pri kratších časových radoch, kedy zdanlivý nárast (pokles) hodnôt môže byť iba prejavom periodicity pozorovanej pri väčšom množstve údajov. Na odstránenie trendovej zložky sme použili regresiu nasledovných funkcií: lineárnej, kvadratickej, exponenciálnej, logistickej a Gompertzovej, kde posledné dve funkcie majú tvar písmena S a sú dané vzťahmi:

$$T_{log} = \frac{a}{1 + be^{-ct}}$$

$$T_{Gomp} = a + be^{-ce^{-dt}}$$

kde *a*, *b*, *c*, *d* sú parametre funkcií. Funkciu reprezentujúcu trendovú zložku sme potom vybrali na základe minimálnej hodnoty sumy štvorcov reziduí (SSR).

2.2 Sezónna zložka

Sezónna zložka opisuje určité periodické zmeny v časovom rade, pričom perióda je rovná nejakej štandardnej dĺžke časového radu [3]. Sezónna zložka je tiež charakterizovaná ako korelačná závislosť medzi *i*-tym a (*i*+*k*)-tym prvkom časového radu, kde *k* je posunutie. Periódu sezónnej zložky sme stanovili pomocou výberovej autokorelačnej funkcie z korelogramu (pozri [8] a [1]). Na odstránenie sezónnej zložky z detrendovaného časového radu sme použili harmonickú funkciu s 12mesačnou periódou zapísanu v tvare

$$S_t = a + b\cos\left(\frac{2\pi}{12}t\right) + c\sin\left(\frac{2\pi}{12}t\right)$$

kde *a*, *b*, *c* sú parametre funkcie.

2.3 Cyklická zložka

Niektoré časové rady môžu okrem už odhadnutej periódy sezónnej zložky obsahovať aj inú, skrytú periodicitu nekorešpondujúcu s kalendárnymi jednotkami. Periódy cyklickej zložky sme odhadli metódami spektrálnej analýzy (pozri [8], [9], [1] a [7]), pričom tie štatisticky významné sme vybrali na základe Fisherovho testu periodicity [4]. Cyklickú zložku sme potom odstránili z časového radu po odstránení trendovej a sezónnej zložky pomocou nasledovnej harmonickej funkcie:

$$C_t = \alpha_1 + \alpha_2\cos\left(\frac{2\pi}{P_1}t\right) + \alpha_3\sin\left(\frac{2\pi}{P_1}t\right) + \cdots \alpha_{i+1}\cos\left(\frac{2\pi}{P_i}t\right) + \alpha_{i+2}\sin\left(\frac{2\pi}{P_i}t\right)$$

kde $\alpha_{\{1,\,2,\,\ldots,\,i+2\}}$ sú parametre funkcie a $P_{\{1,\,2,\,\ldots,\,i\}}$ sú štatisticky významné periódy.

Po odstránení všetkých systematických zložiek z pôvodného časového radu sme takto vzniknutý časový rad reziduí testovali na stacionaritu Kwiatkowski-Philips-Schmidt-Shinovým testom (KPSS) opísaným v [10]. Časový rad reziduí sme ďalej modelovali lineárnymi modelmi.

2.4 Lineárne modely triedy ARMA

Lineárne modely triedy ARMA sú v tejto práci zastúpené modelmi AR, MA a ich kombináciou modelmi ARMA opísanými v práci Boxa a Jenkinsa [1], ktorí vypracovali metodológiu na odhad ich parametrov. Jednotlivé modely sme odhadli Hannan-Rissanenovou procedúrou, kde najlepší model je vybraný na základe najmenšej hodnoty AIC (Akaikovo informačné kritérium) a BIC (Bayesovo informačné kritérium). Jednotlivé modely AR, MA a ARMA možno napísať v tvaroch:

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + z_t$$

$$x_t = z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}$$

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}$$

kde p a q sú rády modelov, $\varphi_{\{1, 2, ..., p\}}$ a $\theta_{\{1, 2, ..., q\}}$ sú ich parametre a z_t je rad nekorelovaných rovnako rozdelených náhodných premenných s nulovou strednou hodnotou a rozptylom σ^2.

2.5 Nelineárne modely SETAR a MSW

Modely SETAR a MSW patria do skupiny tzv. viacrežimových modelov, kde každý model obsahuje niekoľko režimov opísaných vlastným autoregresným modelom. Modely SETAR sú špeciálnym prípadom modelov TAR (z angl. threshold autoregressive), kde režim podľa ktorého sa hodnota v čase t počíta určuje predchádzajúca hodnota pozorovaného časového radu y_{t-d} (d>0 je celé číslo predstavujúce oneskorenie) a prahová hodnota c. V prípade dvojrežimového modelu prvý režim nastane, ak $y_{t-d} \leq c$ a druhý ak $y_{t-d} > c$. Trojrežimový SETAR model môžeme zapísať ako

$$x_t = \begin{cases} \varphi_0^{(1)} + \varphi_1^{(1)} x_{t-1} + \cdots + \varphi_p^{(1)} x_{t-p} + z_t^{(1)} & ak \;\; x_{t-d} \leq c_1 \\ \varphi_0^{(2)} + \varphi_1^{(2)} x_{t-1} + \cdots + \varphi_p^{(2)} x_{t-p} + z_t^{(2)} & ak \;\; c_1 < x_{t-d} \leq c_2 \\ \varphi_0^{(3)} + \varphi_1^{(3)} x_{t-1} + \cdots + \varphi_p^{(3)} x_{t-p} + z_t^{(3)} & ak \;\; c_2 < x_{t-d} \end{cases}$$

kde $\varphi_{\{1, 2, ..., p\}}$ sú parametre AR modelov v jednotlivých režimoch (pre každý režim rozdielne), $c_{\{1, 2, ..., k-1\}}$ sú hraničné hodnoty, d je spomínané oneskorenie a z_t je proces bieleho šumu pre ktorý platí $z_t \sim N(0,\sigma^2)$. Viac informácií o modeloch SETAR v prácach [12], [2], [6] a [11].

Štruktúra MSW modelov je totožná s modelmi SETAR. Jediným rozdielom je spôsob prepínania medzi režimami, keď v prípade MSW modelov je režim určený na základe nepozorovanej veličiny s, nadobúdajúcej iba kladné celočíselné hodnoty {1, 2, ..., k}, kde k je počet režimov. Veličina s je pritom opísaná pomocou Markovovho reťazca prvého rádu, čo znamená, že pravdepodobnosť s akou stav s_t nadobudne hodnotu j je daná len hodnotou v predchádzajúcom stave:

$$P\left(s_t = j \mid s_{t-1} = i\right) = p_{ij}$$

kde i a j sú celé čísla z intervalu <1, k> a p_{ij} je pravdepodobnosť s akou režim i bude nasledovať režim j. Trojrežimový MSW model môžeme zapísať ako

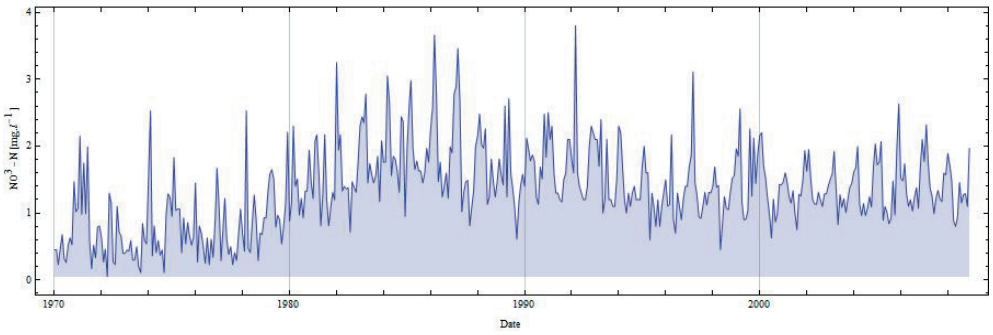
$$x_t = \begin{cases} \varphi_0^{(1)} + \varphi_1^{(1)} x_{t-1} + \cdots + \varphi_p^{(1)} x_{t-p} + z_t^{(1)} & ak \;\; s_t = 1 \\ \varphi_0^{(2)} + \varphi_1^{(2)} x_{t-1} + \cdots + \varphi_p^{(2)} x_{t-p} + z_t^{(2)} & ak \;\; s_t = 2 \\ \varphi_0^{(3)} + \varphi_1^{(3)} x_{t-1} + \cdots + \varphi_p^{(3)} x_{t-p} + z_t^{(3)} & ak \;\; s_t = 3 \end{cases}$$

kde $\varphi_{\{1, 2, ..., p\}}$ sú parametre AR modelov v jednotlivých režimoch a z_t je proces bieleho šumu pre ktorý platí $z_t \sim N(0,\sigma^2)$. Podrobnejšie o MSW modeloch v prácach [13], [5] a [6].

V práci uvažujeme len s dvoj a trojrežimovými modelmi SETAR a MSW.

3. Výsledky

Prvým analyzovaným časovým radom bol časový rad priemerných mesačných koncentrácií dusičnanového dusíka (NO_3-N) z rieky Hron pri Šalkovej z rokov 1970 až 2008 (Obr. 1). Odstránením systematických zložiek z pôvodného časového radu sme dostali stacionárny časový rad reziduí, keď ako najlepšia krivka opisujúca trendovú zložku bola Gompertzova krivka a spektrálnou analýzou sme určili dve štatisticky významné periódy: 468 a 11,7 mesiacov. Ďalším krokom bolo modelovanie lineárnymi a nelineárnymi modelmi. Keďže nelineárne modely je možno použiť aj na modelovanie nestacionárnych časových radov, modelmi SETAR a MSW sme modelovali pôvodný časový rad. Výsledky modelovania sú uvedené v Tab. 1, ktorá uvádza aj relatívne zlepšenie jednotlivých modelov k najjednoduchšiemu AR(1) modelu.

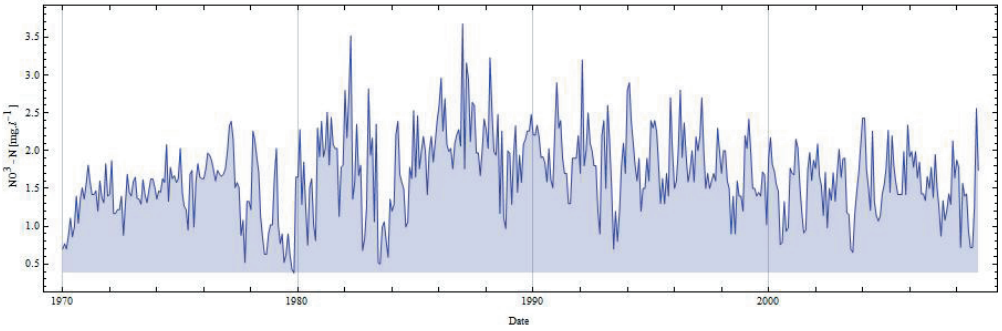


Obr. 1: Priemerná mesačné koncentrácie dusičnanov v rieke Hron pri Šalkovej z obdobia od 1970 do 2008.

Tab. 1: Čiastkové výsledky analýzy časového radu koncentrácie dusičnanov v rieke Hron. Relatívne zlepšenie k AR(1) je vypočítané podľa $(SSR_{AR(1)}-SSR)/SSR_{AR(1)}$. p predstavuje rád AR modelu v režimoch.

Komponent	SSR	Relatívne zlepšenie k AR(1) [%]
Systematické zložky		
Trend	125,3	
Sezónna	93,8	
Cyklická	80,3	
ARMA trieda		
AR (1)	79,1	0,0%
MA(1)	80,7	-2,0%
AR(2)	78,9	0,3%
ARMA(1,1)	114,6	-44,9%
AR(3)	78,3	1,0%
ARMA(1,2)	117,9	-49,1%
ARMA(2,1)	111,2	-40,6%
SETAR		
2 režimy	107,2	-35,5%
3 režimy	102,6	-29,7%
MSW		
2 režimy		
p=1	54,4	31,2%
p=2	50,7	35,9%
p=3	46,8	40,8%
p=4	48,8	38,3%
p=5	47,8	39,6%
3 režimy		
p=1	35,8	54,7%
p=2	28,8	63,6%
p=3	28,7	63,7%
p=4	29,2	63,1%
p=5	38,4	51,5%

Druhým analyzovaným časovým radom bol časový rad priemerných mesačných koncentrácií dusičnanového dusíka (NO_3-N) z rieky Váh pri Hlohovci z rokov 1970 až 2008 (Obr. 2). Odstránením systematických zložiek sme opäť dostali stacionárny časový rad s kvadratickou krivkou najlepšie vystihujúcou trendovú zložku a významnými periódami cyklickej zložky rovnými: 58,5; 234; 66,8 a 11,7 mesiacov. Výsledky modelovania lineárnymi a nelineárnymi modelmi sú uvedené v Tab. 2.



Obr. 2: Pôvodný časový rad priemerných mesačných koncentrácií dusičnanov na rieke Váh pri Hlohovci.

Tab. 2: Čiastkové výsledky analýz y časového radu koncentrácie dusičnanov v rieke Váh. Relatívne zlepšenie k AR(1) je vypočítané podľa $(SSR_{AR(1)}-SSR)/SSR_{AR(1)}$. p predstavuje rád AR modelu v režimoch.

Komponent	SSR	Relatívne zlepšenie k AR(1) [%]
Systematické zložky		
Trend	114,3	
Sezónna	91,8	
Cyklická	72,8	
ARMA trieda		
ARMA(1,1)	91,4	-36,8%
AR(2)	67,5	-1,0%
AR(1)	66,8	0,0%
ARMA(2,1)	81,6	-22,2%
ARMA(1,2)	88,9	-33,1%
MA(3)	71,0	-6,3%
SETAR		
2 režimy	85,6	-28,1%
3 režimy	83,5	-25,0%
MSW		
2 režimy		
p=1	41,9	37,3%
p=2	38,8	41,9%
p=3	38,1	43,0%
p=4	38,6	42,2%
p=5	96,1	-43,9%
3 režimy		
p=1	28,3	57,6%
p=2	20,2	69,8%
p=3	26,3	60,6%
p=4	26,7	60,0%
p=5	99,7	-49,3%

4. Záver

Odstránením systematických zložiek sa nám v obidvoch prípadoch podarilo zabezpečiť stacionaritu daného časového radu. V obidvoch časových radoch sa na základe minimálnej

hodnoty SSR javia ako najlepšie jednoduché AR modely nízkych rádov (pozri Tab. 1 a 2). Vizuálne posúdenie týchto modelov však ukazuje, že tieto modely nedokážu dobre opísať extrémne hodnoty, a teda vhodnejšími sa aj napriek vyšším hodnotám SSR javia modely ARMA. V prípade nelineárnych viacrežimových modelov sa nám iba modelmi MSW podarilo výrazne vylepšiť najjednoduchší AR(1), keď najmä trojrežimovými modelmi sme dosiahli až 70% zlepšenie. Neschopnosť modelov SETAR zlepšiť jednoduchý AR(1) model môže byť čiastočne vysvetlená ich aplikáciou na pôvodné časové rady, čím boli výrazne znevýhodnené v porovnaní s lineárnymi modelmi triedy ARMA.

Dosiahnuté výsledky ukázali, že nelineárne modely MSW dokázali významne vylepšiť modely triedy ARMA aj napriek tomu, že boli použité priamo na modelovanie originálnych časových radov. Práve táto ich vlastnosť, keď z analyzovaného časového radu nemusia byť odstránené systematické zložky mu dáva výhodu oproti lineárnym modelom triedy ARMA. Nevýhodou oboch nelineárnych modelov je ich zložitosť, ktorou kladú zvýšené požiadavky na užívateľove znalosti zo štatistiky. Množstvo parametrov, ktoré treba v oboch modeloch odhadnúť je taktiež náročné na výpočtovú techniku a je teda časovo náročné.

Literatúra

[1] BOX, G. E. P. – JENKINS, G. M., 1976. Time series analysis: forecasting and control. Oakland: Holden-Day.

[2] BROOKS, C., 2009. RATS Handbook to Accompany Introductory Econometrics for Finance. Cambridge: Cambridge University Press.

[3] CIPRA, T., 1986. Analýza časových řad s aplikacemi v ekonomii. Bratislava: SNTL – Nakladatelství technické literatury.

[4] FISHER, R. A., 1929. Tests of significance in harmonic analysis. In: *Proceedings of the Royal Society of London*. Series A, vol. 125, no. 796, pp 54-59.

[5] FRANSES, P. H. – VAN DIJK, D., 2003. Time Series Models in Empirical Finance. Cambridge: Cambridge University Press.

[6] HAMILTON, J. D., 1989. A New Approach to the Economic Analysis of Nonstationary Time-series and the Business Cycle. In: *Econometrica*, vol. 57, no. 2, pp 357-384.

[7] HAMILTON, J. D., 1994. Time Series Analysis. Princeton: Princeton University Press.

[8] CHAFIELD, C., 1989. The Analysis of Time Series: An Introduction (4th ed), Chapman and Hall, London.

[9] JENKINS, G. M., 1968. Spectral analysis and its applications. Oakland: Holden-Day.

[10] KWIATKOWSKI, D. – PHILLIPS, P. C. B. – SCHMIDT, P. – SHIN, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. In: *Journal of* Econometrics, vol. 54, pp 159-178.

[11] LENČUCHOVÁ, J., 2009. MSW modely časových radov. Diplomová práca. Bratislava: STU.

[12] TONG, H., 1990. Non-Linear Time Series: A Dynamical System Approach. Oxford: Oxford Univ. Press.

[13] TSAY, R. S., 2005. Analysis of Financial Time Series (2nd ed.). Hoboken (New Jersey): Wiley.

Kontakty na autora

1 Stavebná fakulta STU v Bratislave, katedra vodného hospodárstva krajiny, e-mail: pevalent@gmail.com, jan.szolgay@stuba.sk

2 Stavebná fakulta STU v Bratislave, katedra matematiky a deskriptívnej geometrie, e-mail: komornikova@math.sk