



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Acta oeconomica et informatica 2  
Nitra, Slovaca Universitas Agriculturae Nitriae, 2004, s. 49–52

## NOVÝ METODOLOGICKÝ PRÍSTUP K PROGNÓZOVANIU DEMOGRAFICKÝCH ČASOVÝCH RADOV A NEW METHODOLOGICAL APPROACH TO THE DEMOGRAPHIC TIME SERIES PROJECTION

Beáta STEHLÍKOVÁ

Slovenská poľnohospodárska univerzita v Nitre

In the past, conventional statistical techniques have been extensively used for forecasting. The most of the statistical methods have limited capabilities when modelling time series. The aim of the paper is to present a new nontraditional approach for time series projection. We use fuzzy c-mean cluster analysis to identify period of the time series. The neural network for projection is constructed with respect to this period. Suggested method is applied to some demographic indicators of the Slovak republic.

Key words: forecasting, fuzzy cluster analysis, neural network, demographic indicator

Neodmysliteľnou súčasťou demografie sú prognózy. Podľa Hrubého (1996) je budúcnosť každej krajiny v mnohom závislá od vývoja stavu obyvateľov ako základu jestvovania každého štátu. Demografické prognózy sú nezastupiteľné pri plánovaní rozvoja obcí, miest a oblastí, pri rozmiestňovaní investícií, v rozvoji dopravy, pri budovaní energetických sietí a podobne. Vývoj stavu obyvateľstva je dôležitý aj pre rozvoj poľnohospodárstva, dôležitého pre zásobovanie budúceho obyvateľstva potravinami. Podľa Vaňu (1998) po zmene politických pomerov na začiatku 90. rokov dochádza k zásadným zmenám v demografickom správaní obyvateľstva.

### Materiál a metódy

Fuzzy c zhlukovú analýzu popisujú mnohí autori, napríklad Klír (1995). Optimálny počet zhlukov sa môže voliť na základe viacerých kritérií. V príspevku sa volí na základe maximálnej hodnoty koeficienta separácie.

Zavedme niekoľko pojmov spájajúcich výsledky fuzzy c zhlukovej analýzy s analýzou časového radu. Rok  $t$  patrí z hľadiska ukazovateľa  $x_t$  do vývojovej etapy  $K$ , ak jeho hodnota funkcie príslušnosti  $x_t$  do  $K$ -teho zhluku je aspoň 0,80. Inak je rok nevyhraný. Rok  $t$  nazveme rokom zvratu (rokom zlomu) vo vývoji sledovaného ukazovateľa z etapy  $K$  do etapy  $Q$ , ak  $Kx_{t-1} > Qx_{t-1}$  a  $Kx_t < Qx_t$ , pričom  $Kx_{t-1}$  je maximálna hodnota funkcie príslušnosti pre  $x_{t-1}$  a  $Qx_t$  je maximálna hodnota funkcie príslušnosti pre  $x_t$ . Uvedené termíny môžeme zovšeobecniť prirodzeným spôs-

bom pre  $p$  ukazovateľov. Potom namiesto ukazovateľa  $x$  uvažujeme  $p$  rozmerný vektor  $x$ . Za periódu skúmaného javu zvolíme hodnotu  $z$  z intervalu od počtu vyhraných prvkov najväčšieho zhluku po počet prvkov majúcich maximálnu hodnotu funkcie príslušnosti v tomto zhluku. Prakticky sa ukazuje ako optimum počet prvkov od jedného bodu zvratu po nasledujúci. Fuzzy c zhluková analýza bola počítaná pomocou softvéru Syntax. Neuronové siete sa v súčasnosti úspešne používajú pre analýzu a prognózovanie časových radov. V našom prípade bola skonštruovaná pomocou softvéru Statistica-Neural Networks.

Prezentovaný nový metodologický postup je aplikovaný na vybrané demografické indikátory Slovenskej republiky zverejnené Štatistickým úradom Slovenskej republiky na webovej stránke za roky 1950–2000. Indikátory sú vybrané tak, že prognózy dvoch z nich sú priamo konfrontovateľné s výsledkami prognóz Infostat. Na tomto mieste treba upozorniť na možnosť aplikácie popísanej metodologického postupu aj na malé územné celky. Rovnako ju možno použiť aj pre ľubovoľné ekonomické ukazovatele vykazujúcu určitú periodicitu s periódou nie nutne tak veľkou, ako ukázkové demografické indikátory. Ich voľba bola podmienená výskumným zameraním autorky.

### Výsledky a diskusia

Výsledky aplikácie popísanej metódy prognózovania časových radov pre niekoľko vybraných demografických ukazovateľov sú uvedené v tabuľkách 1. až 4.

Tabuľka 1 Hodnoty koeficienta separácie

Model (1)	Dva zhluky (2)	Tri zhluky (3)	Štyri zhluky (4)	Päť zhlukov (5)	Šesť zhlukov (6)
Zaťaženosť mladou populáciou (7)	44,54766	43,40971	42,12566	40,90714	38,23024
Zaťaženosť starou populáciou (8)	47,02359	44,54761	43,93450	42,23819	41,20997
Priemerný vek žien pri pôrode (9)	46,64786	45,03579	44,43221	43,57439	42,30231
Počet obyvateľov (10)	43,40988	43,57714	40,55891	40,98146	40,57154

Table 1 Values of the coefficient of separation

(1) model, (2) 2 clusters, (3) 3 clusters, (4) 4 clusters, (5) 5 clusters, (6) 6 clusters, (7) the youth dependency ratio, (8) the old dependency ratio, (9) mean age at childbirth, (10) number of inhabitants

Tabuľka 2 Hodnoty chyby najmenších štvorcov

Model (1)	Trénovacie údaje (2)	Verifikačné údaje (3)	Testovacie údaje (4)
Zaťaženosť mladou populáciou (5)	0,06893	0,08435	0,05516
Zaťaženosť starou populáciou (6)	0,03373	0,001704	0,202824
Priemerný vek žien pri pôrode (7)	0,02277	0,001814	0,02588
Počet obyvateľov (8)	0,04739	0,04373	0,08412

Table 2 Values of the RMS error  
(1) model, (2) training data set, (3) verification data set, (4) test data set, (5) the youth dependency ratio, (6) the old dependency ratio, (7) mean age at childbirth, (8) number of inhabitants

### Podrobnejšia analýza výsledkov

Hodnota koeficienta separácie fuzzy c zhlukovej analýzy pre zaťaženosť mladou populáciou bola najvyššia pre dva zhluky. Vo vývoji zaťaženosť mladou populáciou sme podľa výsledkov fuzzy c zhlukovej analýzy identifikovali dve etapy vývoja. Do prvého zhluku patria roky 1950–1974. Roky 1975–80 sú nevyhranené. Najnevyhranejší je rok 1976, ktorého funkcia príslušnosti do prvého zhluku je 0,5394 a do druhého 0,4606. Bodom zlomu je rok 1977, ktorého funkcia príslušnosti do prvého zhluku je 0,3833 a do druhého 0,6167. Roky 1981–2000 patria do druhého zhluku. Výsledky korešponujú s populačným vývojom na Slovensku. V období 70. rokov došlo na Slovensku k zvýšeniu úrovne pôrodnosti aj plodnosti. Hlavnou príčinou je nástup početných populačných ročníkov žien a v dôsledku pronatálnych ekonomických opatrení štátu. Pre predikciu vývoja indexu zaťaženosť mladou populáciou sme použili neurónovú sieť. Najvhodnejšia sa ukázala neurónová sieť RBF (neurónová sieť s radiálnou báзовou funkciou) s jednou skrytou vrstvou so štyrmi neurónmi. Má veľmi dobrú výkonnosť – regresný pomer je 0,044435 a korelácia je 0,999085. Jej podrobné charakteristiky sú v nasledujúcich tabuľkách. Pri špecifikovaní parametrov neurónovej siete sme využili výsledok fuzzy c-zhlukovej analýzy a určili sme periódu tridsať rokov. Neurónová sieť umožňuje predikciu hodnôt. V rokoch 1980–2000 bola najpresnejšie aproximovaná hodnota pre rok 1992 s chybou -0,004061 a najmenej presne pre rok 1984 s chybou 0,133467. Predikované a skutočné hodnoty za obdobie 1980–2000 sa líšia iba nepatrne.

Tabuľka 3 Charakteristiky regresie

Model (1)	Kritérium (2)	Trénovacie údaje (3)	Verifikačné údaje (4)	Testovacie údaje (5)
Zaťaženosť mladou populáciou (6)	chyba priemeru (10)	-1,03E-14	-0,03613	-0,02046
	chyba štandardnej odchýlky (11)	0,07229	0,08349	0,05916
	korelácia (12)	0,999906	0,999085	0,999949
Zaťaženosť starou populáciou (7)	chyba priemeru (10)	-0,002057	-0,001126	0,018350
	chyba štandardnej odchýlky (11)	0,035490	0,001808	0,221271
	korelácia (12)	0,999200	1,000000	0,980242
Priemerný vek žien pri pôrode (8)	chyba priemeru (10)	-0,00022	-0,00022	-0,02346
	chyba štandardnej odchýlky (11)	0,024	0,002079	0,01263
	korelácia (12)	0,996821	0,999989	0,999694
Počet obyvateľov (9)	chyba priemeru (10)	1 828,155	59,28067	169,4546
	chyba štandardnej odchýlky (11)	4 111,195	13 744,03	1 2416,83
	korelácia (12)	0,999929	0,998084	0,99919

Table 3 Characterization of the regression  
(1) model, (2) criterion, (3) training data set, (4) verification data set, (5) test data set, (6) the youth dependency ratio, (7) the old dependency ratio, (8) mean age at childbirth, (9) number of inhabitants, (10) error mean, (11) error standard deviation, (12) correlation

Tabuľka 4 Prognózy vybraných demografických indikátorov

Rok (1)	Model A (2)	Model B (3)	Model C (4)	Model D (5)
2001	45,031	18,878	26,391	5 381
2002	44,020	18,838	26,545	5 379
2003	43,417	18,778	26,667	5 378
2004	43,172	18,692	26,687	5 377
2005	43,206	18,576	26,765	5 376
2006	43,439	18,430	26,812	5 375
2007	43,786	18,281	26,810	5 373
2008	44,189	18,140	26,830	5 372
2009	44,617	18,022	26,928	5 371
2010	45,049	17,937	27,045	5 370

A – zaťaženosť mladou populáciou, B – zaťaženosť starou populáciou, C – priemerný vek žien pri pôrode, D – počet obyvateľov

A – the youth dependency ratio, B – the old dependency ratio, C – mean age at childbirth, D – number of inhabitants

Table 4 Predictions of the some demographic indicators

(1) year, (2) model A, (3) model B, (4) model C, (5) model D

Z predikcie vidíme, že index zaťaženosť mladou populáciou by mal postupne klesať do roku 2007. V nasledujúcich rokoch 2009 a 2010 by sa mal mierne zvýšiť. Zistiť možné dôvody a tak prognózu podoprieť očakávanými zmenami faktorov vplyvajúcich na tento index by si však žiadalo hlbšiu analýzu.

Hodnota koeficienta separácie fuzzy c zhlukovej analýzy v prípade analýzy vývoja zaťaženosť starou populáciou bola najvyššia pre dva zhluky. Vo vývoji zaťaženosť starou populáciou sme podľa výsledkov fuzzy c zhlukovej analýzy identifikovali dve etapy vývoja. Do prvého zhluku patria roky 1950–1964. Roky 1965–1967 sú nevyhranené. Najnevyhranejší je rok 1966, ktorého funkcia príslušnosti do prvého zhluku je 0,535 a do druhého 0,465. Bodom zlomu je rok 1967, ktorého funkcia príslušnosti do prvého zhluku je 0,3488 a do druhého 0,6512. Roky 1968–2000 patria do druhého zhluku. Pre predikciu vývoja indexu zaťaženosť mladou populáciou sme použili neurónovú sieť. Najvhodnejšia sa ukázala neurónová sieť MLP s jednou skrytou vrstvou s tromi neurónmi. Má veľmi dobrú výkonnosť – regresný pomer je 0,025572 a korelácia je 1,000000. Pri špecifikovaní parametrov neurónovej siete sme využili výsledok fuzzy

c-zhlukovej analýzy a určili sme periódu tridsaťtri rokov. Neurónová sieť umožňuje predikciu hodnôt. V rokoch 1983–2000 bola najpresnejšie aproximovaná hodnota pre rok 2000 s chybou 0,012, a najmenej presne pre rok 1984 s chybou 0,464. Predikované a skutočné hodnoty za obdobie 1984–2000 sa líšia iba nepatrne. Z predikcie vidíme, že index zafarbenosti mladou populáciou by mal postupne klesať do roku 2007. V nasledujúcich rokoch 2009 a 2010 by sa mal mierne zvýšiť. Uvedené skutočnosti sú v súlade s prognózami Štatistického úradu SR.

Hodnota koeficienta separácie fuzzy c zhlukovej analýzy bola v prípade analýzy priemerného veku žien pri pôrode najvyššia pre dva zhluky. Vo vývoji priemerného veku žien pri pôrode sme podľa výsledkov fuzzy c zhlukovej analýzy identifikovali dve etapy vývoja. Do prvého zhluku patria roky 1950–1965. Roky 1966, 1967 sú nevyhranené. Bodom zlomu je rok 1967, ktorého funkcia príslušnosti do prvého zhluku je 0,3929 a do druhého 0,6071. Roky 1968–1998 patria do druhého zhluku. Roky 1999 a 2000 sú opäť nevyhranené. Pre predikciu vývoja indexu zafarbenosti mladou populáciou sme použili neurónovú sieť. Najvhodnejšia sa ukázala neurónová sieť MLP s jednou skrytou vrstvou s desiatimi neurónmi. Má veľmi dobrú výkonnosť. Jej podrobné charakteristiky sú

v nasledujúcich tabuľkách. Pri špecifikovaní parametrov neurónovej siete sme využili výsledok fuzzy c-zhlukovej analýzy a určili sme periódu tridsaťtri rokov. V rokoch 1983–2000 bola najpresnejšie aproximovaná hodnota pre rok 1985 s chybou -0,001 a najmenej presne pre rok 1991 s chybou 0,043. Predikované a skutočné hodnoty za obdobie 1983–2000 sa líšia iba nepatrne. V porovnaní s prognózou Infostatu sa prognóza pomocou neurónových sietí blíži k veľmi nízkemu scenáru v rokoch 2000–2005. Do roku 2013 je nižší ako priemerný vek uvažovaný v jednotlivých scenároch. V rokoch 2015–2020 je medzi priemerným vekom uvažovanom vo vysokom a strednom scenári. Podrobnejšie to vidieť na obrázku 7. Dá sa konštatovať, že model potvrdil všeobecne prijatú tézu o zvyšovaní priemerného veku žien pri pôrode. Uvažuje však v jednotlivých obdobiach iným tempom rastu.

Posledný, štvrtý príklad sa od ostatných líši počtom etáp vývoja. Hodnota koeficienta separácie fuzzy c zhlukovej analýzy bola v prípade prognózy počtu obyvateľov najvyššia pre tri zhluky. Vo vývoji počtu obyvateľov sme podľa výsledkov fuzzy c zhlukovej analýzy identifikovali tri etapy vývoja. Do prvého zhluku patria roky 1950–1959. Roky 1960–1963 sú nevyhranené. Bodom zlomu je rok 1962, ktorého funkcia príslušnosti do prvého

Figure 1 Vývoj a predikcia zafarbenosti mladou populáciou  
Obrázok 1 Development and prediction of the the youth dependency ratio

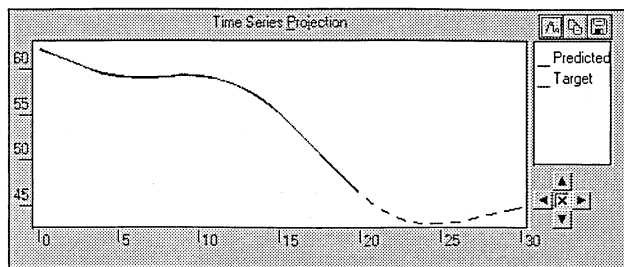


Figure 2 Neurónová sieť RBF s jednou skrytou vrstvou so štyrmi neurónmi  
Obrázok 2 RBF neuron network with one hidden layer with four neurons

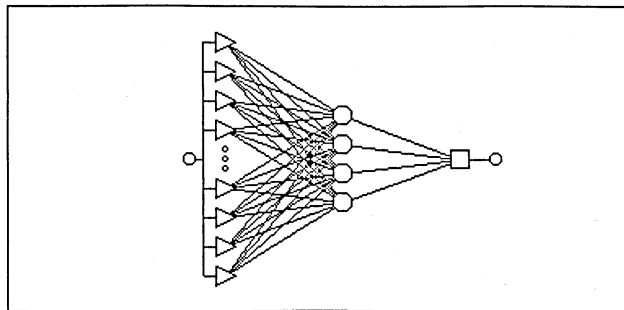


Figure 3 Vývoj a predikcia zafarbenosti starou populáciou  
Obrázok 3 Development and prediction of the the old dependency ratio

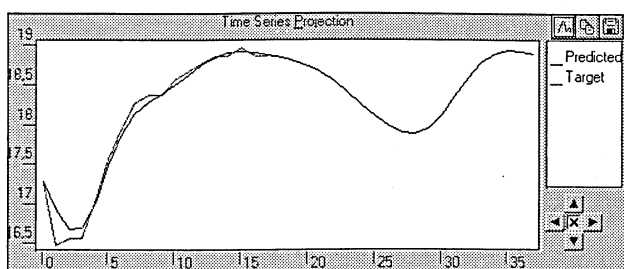


Figure 4 Neurónová sieť MLP s jednou skrytou vrstvou s tromi neurónmi  
Obrázok 4 MLP neuron network with one hidden layer with three neurons

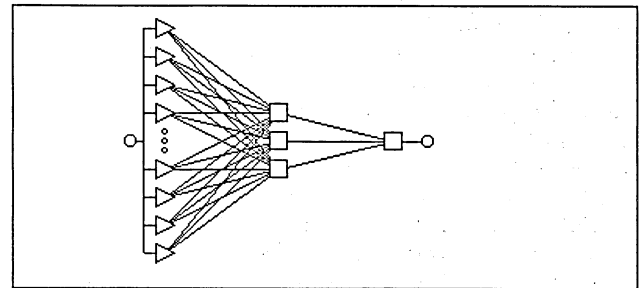


Figure 5 Vývoj a predikcia priemerného veku žien pri pôrode  
Obrázok 5 Development and prediction of the mean age at childbirth

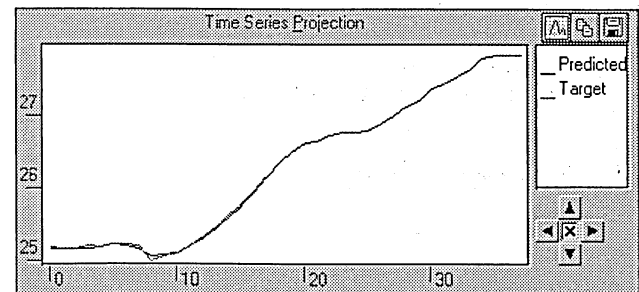
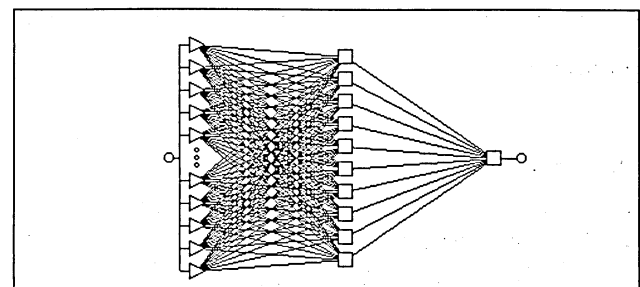
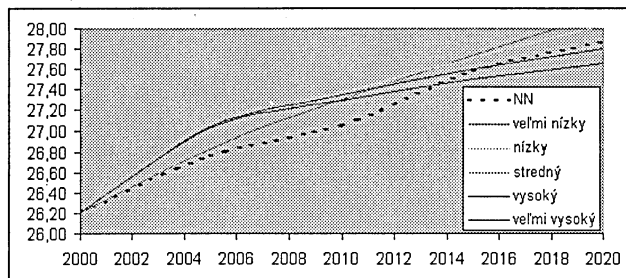


Figure 6 Neurónová sieť MLP s jednou skrytou vrstvou s desiatimi neurónmi  
Obrázok 6 MLP neuron network with one hidden layer with ten neurons



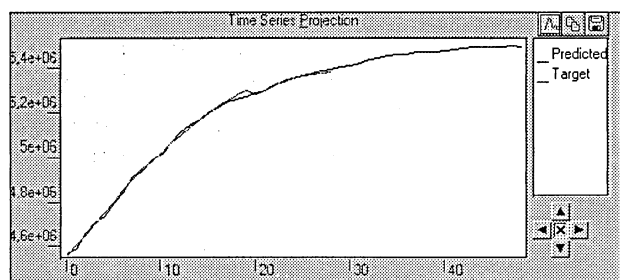
**Figure 7** Porovnanie predpovede priemerného veku žien pri pôrode so scenárom Infostat

**Obrázok 7** Comparison of the prediction of the the mean age at childbirth with Infostat's scenario



**Figure 8** Vývoj a predikcia počtu obyvateľov

**Obrázok 8** Development and prediction of the number of inhabitants



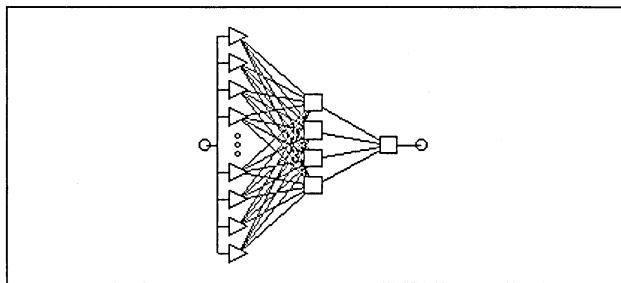
zhľuku je 0,2450, do druhého 0,0598 a tretieho 0,6952. Roky 1964–1975 patria do tretieho zhľuku. Roky 1976–1981 sú opäť nevyhranené. Bodom zlomu je 1979. Jeho funkcia príslušnosti do prvého zhľuku je 0,0420, druhého 0,5733 a tretieho 0,3847. Pre predikciu vývoja indexu zaťaženia mladou populáciou sme použili neurónovú sieť. Najvhodnejšia sa ukázala neurónová sieť MLP s jednou skrytou vrstvou so štyrmi neurónmi. Má veľmi dobrú výkonnosť. Pri špecifikovaní parametrov neurónovej siete sme využili výsledok fuzzy c-zhlukovej analýzy a určili sme periódu dvadsaťdva rokov. V porovnaní s prognózou Infostat sa prognóza pomocou neurónových sietí v prognózovaných rokoch nachádza medzi vysokým a stredným variantom.

### Záver

Navrhovaný nový metodologický postup je kombináciou dvoch metód fuzzy c-zhlukovej analýzy a neurónových sietí. Lahko sa dá presvedčiť, že neurónové siete skonštruované za účelom projekcie časového radu bez informácie periódy časového radu dávajú menej presné výsledky ako siete skonštruované navrhovaným postupom. Navrhovaný postup nevyžaduje od časového radu splnenie predpokladov, za ktorých boli klasické štatistické metódy odvodené. Nemá nároky ani na určenie tvaru predikčnej funkcie. Naviac fuzzy zhluková analýza aplikovaná na časový rad určuje body zvratu vo vývoji analyzovaného javu, čo môže byť v prípade demografických, ale hlavne ekonomických ukazovateľov veľmi cenná informácia. Výsledky predikcií, ktoré je možné porovnať s oficiálnymi prognózami Štatistického úradu Slovenskej republiky, nasvedčujú o vhodnosti použitia navrhovaného postupu. Prognózy SÚ SR sú založené na komplikovaných štatistických metódach za predpokladu detailného poznania prognózovaných javov. V našom prípade tieto znalosti nemusíme mať. Sú vytážené zo vstupných údajov neurónovou sieťou a fuzzy zhlukovou analýzou. Vzhľadom na skutočnosť, že výsledky boli získané použitím príslušného softvéru, môže najstredkaldaná metodológia širšie použitie.

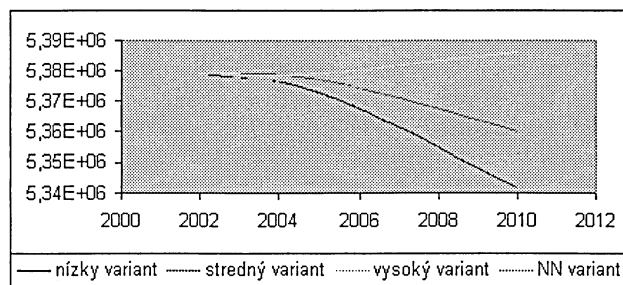
**Figure 9** Neurónová sieť MLP s jednou skrytou vrstvou so štyrmi neurónmi

**Obrázok 9** MLP neuron network with one hidden layer with four neurons



**Figure 10** Porovnanie predpovede počtu obyvateľov so scenárom Infostat

**Obrázok 10** Comparison of the prediction of the population with Infostat's scenario



### Súhrn

V minulosti, a najčastejšie aj v súčasnosti, sa pre predpovede používajú klasické štatistické metódy. Mnohé z nich však majú za predpokladu korektného použitia obmedzené využitie. Cieľom príspevku je prezentovať nový netradičný prístup pre projekciu časových radov. Pomocou fuzzy c-zhlukovej analýzy sa určí perióda hodnoteného časového radu. Pre projekciu sa použije neurónová sieť skonštruovaná vzhľadom na túto periódu. Navrhovaný postup je aplikovaný na vybrané demografické ukazovatele Slovenskej republiky.

**Kľúčové slová:** predpoveď, fuzzy zhluková analýza, neurónová sieť, demografický indikátor

### Literatúra

- BISHOP, C. 1995. Neural networks for Pattern Recognition. Oxford : University Press, 1995 ISBN 0-19-853849-9  
 HRUBÝ, J. 1996. Základy demografie pre manažéra na vidieku. Nitra : SPU, 1996. ISBN 80-7137-311-7  
 KLIR, R. - YUAN, B. O. 1995. Fuzzy sets and logic. London : Prentice Hall, 1995. ISBN 0-13-101171-5  
 NOVÁK, V. 1990. Fuzzy množiny a jejich aplikace. Praha : SNTL, 1990  
 PODANI, J. 1993. SYN-TAX-pc. Budapest : Scientia Publishing, 1993  
 VAŇO, B. a kol. 2002. Prognóza vývoja obyvateľstva SR do roku 2050. Bratislava : Infostat, 2002  
 VAŇO, B. a kol. 2001. Obyvateľstvo Slovenska 1945–2000. Bratislava : Infostat, 2001

Kontaktná adresa:  
 doc. RNDr. Beáta Stehlíková, CSc., FEM SPU Nitra, Tr. A. Hlinku 2, 949 76 Nitra, tel.: 037/65 08 153, e-mail: Beata.Stehlikova@uniag.sk