

MODELOVÁNÍ KVALITY OVZDUŠÍ POMOCÍ KOHONENOVÝCH SAMOORGANIZUJÍCÍCH SE MAP

Vladimír Olej, Petr Hájek, Jiří Křupka, Ilona Obršálová

Univerzita Pardubice, Fakulta ekonomicko-správní, Ústav systémového inženýrství a informatiky, Ústav veřejné správy a práva

Abstract: *The paper presents a design of parameters for air quality modelling and the classification of districts into classes according to their pollution. Further, it presents a model design, data pre-processing, the designs of various structures of Kohonen's Self-organizing Feature Maps (unsupervised methods), the clustering by K-means algorithm and the classification.*

Key words: *Air quality, modelling, Kohonen's self-organizing feature maps, K-means algorithm, classification.*

1 Úvod

Pod pojmem znečišťování ovzduší je možné zahrnout celou škálu činností, při nichž dochází k vnášení látek nebo energie do atmosféry. Jinými slovy, znečišťování ovzduší znamená vypouštění hmotných látek v tuhém, kapalném nebo plynném skupenství z různých zdrojů do ovzduší, které buď přímo nebo po chemických změnách negativně ovlivňují kvalitu a složení ovzduší [4]. Ochranou ovzduší se rozumí soubor technických i administrativních opatření [4], která přímo nebo nepřímo směřují k omezení prudkého nárůstu znečišťování ovzduší. Mezi technická opatření se zahrnují opatření technologická, surovinová, optimalizační či omezující. Do administrativních je možné zahrnout opatření legislativní, správní, ekonomická, kontrolní a další. Spolu s rostoucím znečišťováním ovzduší roste i význam jeho ochrany. Vývoj kvality ovzduší (klasifikaci zkoumaných oblastí $o_i^t \in O$ do tříd $\omega_{i,j}^t \in \Omega$ podle hodnot jejich znečištění) je možné uskutečňovat různými metodami. Jedná se například o fuzzy inferenční systémy [3], metody učení bez učitele [1,2] a neuro-fuzzy systémy [3], které jsou vhodné pro modelování vývoje kvality ovzduší. Neuronové sítě [1,2] jsou vhodné pro schopnosti se učit, zevšeobecňovat a dále modelovat nelineární vztahy. Vývoj kvality ovzduší je možné považovat za klasifikační problém, který lze řešit různými strukturami neuronových sítí. Výstupem neuronové sítě je v případě klasifikace přiřazení i -té zkoumané oblasti $o_i^t \in O$, $O = \{o_1^t, o_2^t, \dots, o_i^t, \dots, o_n^t\}$ v čase t do j -té třídy $\omega_{i,j}^t \in \Omega$, $\Omega = \{\omega_{1,j}^t, \omega_{2,j}^t, \dots, \omega_{i,j}^t, \dots, \omega_{n,j}^t\}$.

V článku je uveden návrh parametrů vývoje kvality ovzduší, kdy jsou vybrány pouze ty parametry, mezi nimiž existují nízké korelační vztahy. Vstupní data jsou pak reprezentovány maticí \mathbf{P} , kde vektory \mathbf{p}_i charakterizují zkoumané oblasti $o_i^t \in O$. Dále je uveden popis Kohonenových samoorganizujících se map (KSOM), které jsou vhodné pro klasifikaci v tom případě, když třídy $\omega_{i,j}^t \in \Omega$ nejsou předem známé. Původnost článku spočívá v návrhu modelu vývoje kvality ovzduší. Modelování je realizováno pomocí metod učení bez učitele (kombinací KSOM a algoritmu K-průměrů). Závěrečná část článku obsahuje analýzu výsledků a prezentaci klasifikace zkoumaných oblastí $o_i^t \in O$ do tříd $\omega_{i,j}^t \in \Omega$.

2 Návrh parametrů pro modelování vývoje kvality ovzduší

Mezi parametry, které je možno použít pro modelování vývoje kvality ovzduší patří škodliviny v ovzduší. Látky znečišťující ovzduší (škodliviny) jsou označovány jako látky vnesené do vnějšího ovzduší nebo v něm druhotně vznikající, které buď přímo, nebo po fyzikální či chemické přeměně, popř. ve spolupůsobení s jinými látkami mají škodlivý vliv na

životné prostředí. Kromě znečišťujících látek mají na výsledné znečištění ovzduší významný vliv rovněž další složky podporující znečišťování nebo zvyšující jeho účinky. Jsou to např. ozon či sluneční záření, síla a směr větru, vlhkost, tlak a další. Vývoj kvality ovzduší ovlivňují nejen parametry týkající se škodlivin v ovzduší, ale také parametry meteorologické. Působení obou druhů parametrů může mít za následek zvýšení znečištění ovzduší a tedy i dopad na zdraví člověka. Návrh parametrů pro modelování vývoje kvality ovzduší, založený na předchozí korelační analýze a doporučení významných expertů v daném oboru je uveden v Tab. 1.

Tab. 1: Parametry pro modelování vývoje kvality ovzduší

Parametry	
Škodliviny	$x_1 = \text{SO}_2$, SO_2 je oxid siřičitý.
	$x_2 = \text{O}_3$, O_3 je ozon.
	$x_3 = \text{NO}$, NO_2 (NO_x) je oxid dusnatý, oxid dusičitý (oxidy dusíku).
	$x_4 = \text{CO}$, CO je oxid uhelnatý.
	$x_5 = \text{PM}_{10}$, PM_{10} je prašný aerosol (prach).
Meteorolo	$x_6 = \text{SV}$, SV je síla větru.
	$x_7 = \text{SmV}$, SmV je směr větru.
	$x_8 = T_3$, T_3 je teplota 3m nad zemským povrchem.
	$x_9 = \text{RV}$, RV je relativní vlhkost.
	$x_{10} = T$, T je tlak.
	$x_{11} = \text{SZ}$, SZ je sluneční záření.

Na základě uvedených skutečností je možné navrhnout následující datovou matici \mathbf{P}

$$\mathbf{P} = \begin{matrix} & \begin{matrix} x_1^t & \dots & x_k^t & \dots & x_m^t & \omega_{i,j}^t \end{matrix} \\ \begin{matrix} o_1^t \\ \dots \\ o_i^t \\ \dots \\ o_n^t \end{matrix} & \left| \begin{matrix} x_{1,1}^t & \dots & x_{1,k}^t & \dots & x_{1,m}^t & \omega_{1,j}^t \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i,1}^t & \dots & x_{i,k}^t & \dots & x_{i,m}^t & \omega_{i,j}^t \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n,1}^t & \dots & x_{n,k}^t & \dots & x_{n,m}^t & \omega_{n,j}^t \end{matrix} \right. \end{matrix},$$

kde $o_i^t \in \mathbf{O}$, $\mathbf{O} = \{o_1^t, o_2^t, \dots, o_i^t, \dots, o_n^t\}$ jsou objekty (zkoumané oblasti) v čase t , x_k^t je k -tý parametr v čase t , $x_{i,k}^t$ je hodnota parametru x_k^t pro i -tý objekt $o_i^t \in \mathbf{O}$, $\omega_{i,j}^t$ je j -tá třída přiřazená i -tému objektu $o_i^t \in \mathbf{O}$, $\mathbf{p}_i^t = (x_{i,1}^t, x_{i,2}^t, \dots, x_{i,k}^t, \dots, x_{i,m}^t)$ je i -tý vektor, $\mathbf{x}^t = (x_1^t, x_2^t, \dots, x_k^t, \dots, x_m^t)$ je vektor parametrů.

Kvalita ovzduší (Tab. 2) slouží k hodnocení stavu ovzduší na základě výsledků měření hmotnostních koncentrací látek v ovzduší. Hodnocení zohledňuje možný vliv na zdravotní stav obyvatelstva [4]. Nové limitní hodnoty z nařízení vlády České republiky č. 350/2002 Sb., (č. 429/2005 Sb.), kterým se stanoví imisní limity, podmínky a způsob sledování, posuzování, hodnocení a řízení kvality ovzduší, jsou uváděny spolu s příslušnými mezemi tolerance zvlášť

pro ochranu zdraví a zvláště pro ochranu vegetace a ekosystémů. Rozptylové podmínky závisí zejména na proudění vzduchu, a to v horizontálním i vertikálním směru [4] (Tab. 3).

Tab. 2: Kvalita ovzduší

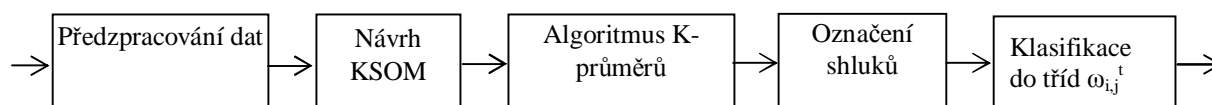
Kvalita ovzduší	SO ₂	NO ₂	CO	O ₃	PM ₁₀
	1h [$\mu\text{g}\cdot\text{m}^{-3}$]		8h [$\mu\text{g}\cdot\text{m}^{-3}$]	1h [$\mu\text{g}\cdot\text{m}^{-3}$]	
Velmi dobrá	0-25	0-25	0-1000	0-33	0-15
Dobrá	25-50	25-50	1000-2000	33-65	15-30
Uspokojivá	50-120	50-100	2000-4000	65-120	30-50
Vyhovující	120-250	100-200	4000-10000	120-180	50-70
Špatná	250-500	200-400	10000-30000	180-240	70-150
Velmi špatná	>500	>400	>30000	>240	>150

Tab. 3: Rozptylové podmínky

Rozptylové podmínky	Charakteristika
Dobré	Ve výšce do (1000-1500)m nad terénem se nevyskytuje zádržná vrstva, která by omezovala rozptyl škodlivin.
Mírně nepříznivé	Vyskytuje se zádržná vrstva, která v závislosti na rychlosti větru pod svou hranicí omezuje možnost rozptylu škodlivin, ale nesplňuje parametry nepříznivých ani dobrých rozptylových podmínek.
Nepříznivé	Stav, kdy rozptyl příměsí v atmosféře je téměř znemožněn a který v oblasti se zdroji znečištění dává předpoklad k déle trvajícím významnému překročení imisních limitů. Tento stav rozptylových podmínek nastává, když je mohutná zádržná vrstva ve výšce do 1000m nad terénem v kombinaci se slabým nebo žádným prouděním.

3 Návrh modelu vývoje kvality ovzduší

Navržený model realizuje modelování kvality ovzduší. Předzpracování dat umožňuje vhodnou enviromentální interpretaci výsledků. Pomocí metod učení bez učitele jsou objekty (zkoumané oblasti města Pardubice) přiřazeny do shluků. Shluky jsou označeny třídami $\omega_{i,j}^t \in \Omega$. Označení shluků je založeno na popisu tříd $\omega_{i,j}^t \in \Omega$ uvedených v Tab. 2 a Tab. 3. Předzpracování dat je realizováno metodou standardizace, čímž je dosaženo odstranění závislosti na jednotkách. Z metod učení bez učitele byla na základě analýzy zvolena kombinace KSOM a algoritmu K-průměrů. Navržený model klasifikace objektů $o_i^t \in O$ do tříd $\omega_{i,j}^t \in \Omega$ je uveden na Obr. 1.



Obr. 2: Model vývoje kvality ovzduší

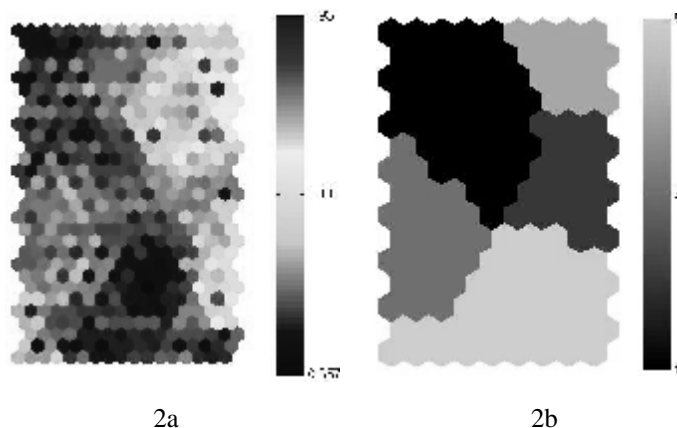
Kohonenovy samoorganizující se mapy [2] jsou založeny na kompetiční strategii učení. Vstupní vrstva slouží k distribuci vstupních vzorů \mathbf{p}_i^t , $i=1,2, \dots, n$. Neurony v kompetiční vrstvě jsou reprezentanty vstupních vzorů a jsou organizovány do topologické struktury. Ta určuje, které neurony spolu sousedí. Nejprve jsou vypočteny Euklidove vzdálenosti d_j mezi vzorem \mathbf{p}_i^t a váhami synapsí $\mathbf{w}_{i,j}$ všech neuronů v kompetiční vrstvě. Je vybrán ten vítězný neuron s indexem j^* , pro který je Euklidova vzdálenost d_j od vzoru \mathbf{p}_i^t minimální. Výstup tohoto neuronu je aktivní, zatímco výstupy ostatních neuronů jsou neaktivní. Cílem učení KSOM je aproximovat hustotu pravděpodobnosti vstupních vektorů $\mathbf{p}_i^t \in \mathbb{R}^n$ pomocí konečného počtu reprezentantů $\mathbf{w}_j \in \mathbb{R}^n$, kde $j=1,2, \dots, s$. Po nalezení reprezentantů \mathbf{w}_j je každému vzoru \mathbf{p}_i^t přiřazen reprezentant \mathbf{w}_{j^*} vítězného neuronu. V procesu učení je definována funkce okolí $h(j^*,j)$, která určuje rozsah spolupráce mezi neurony, tj. kolik reprezentantů \mathbf{w}_j v okolí vítězného neuronu bude adaptováno, a do jaké míry. Po nalezení vítězných neuronů je realizována adaptace vah synapsí $\mathbf{w}_{i,j}$. Principem sekvenčního trénovacího algoritmu [2] je ta skutečnost, že reprezentanti \mathbf{w}_{j^*} vítězného neuronu a jeho topologického okolí se posouvají směrem k aktuálnímu vstupnímu vektoru \mathbf{p}_i^t podle vztahu

$$\mathbf{w}_{i,j}(t+1) = \mathbf{w}_{i,j}(t) + \eta(t)h(j^*,j)[\mathbf{p}_i^t(t) - \mathbf{w}_{i,j}(t)], \quad (1)$$

kde $\eta(t) \in (0,1)$ je rychlost učení.

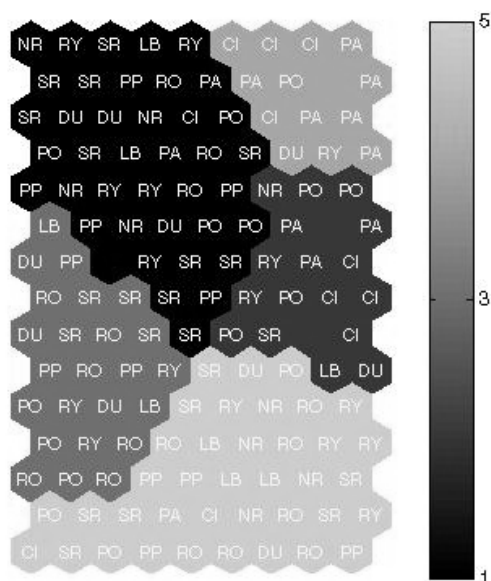
4 Analýza výsledků

Cílem modelování vývoje kvality ovzduší je klasifikace zkoumaných oblastí $o_i^t \in O$ v čase t do tříd $\omega_{i,j}^t \in \Omega$ podle jejich kvality ovzduší. Návrh struktury KSOM je založen na množství realizovaných experimentů. Použití KSOM vede k nalezení struktury v datech (Obr. 2a). Matice U prezentuje čtvercové Euklidove vzdálenosti d mezi reprezentanty \mathbf{w}_j . Algoritmus K-průměrů pak realizuje shlukování naučené KSOM takovým způsobem, jak je to prezentováno na Obr. 2b. Algoritmus K-průměrů patří mezi nehierarchické algoritmy shlukové analýzy, kde vzory $\mathbf{p}_1^t, \mathbf{p}_2^t, \dots, \mathbf{p}_i^t, \dots, \mathbf{p}_n^t$ ($n=720$) jsou přiřazeny do shluků $c_1^t, c_2^t, \dots, c_i^t, \dots, c_q^t$. Počet shluků $q=5$ je určen na základě indexů kvality shlukování [5].



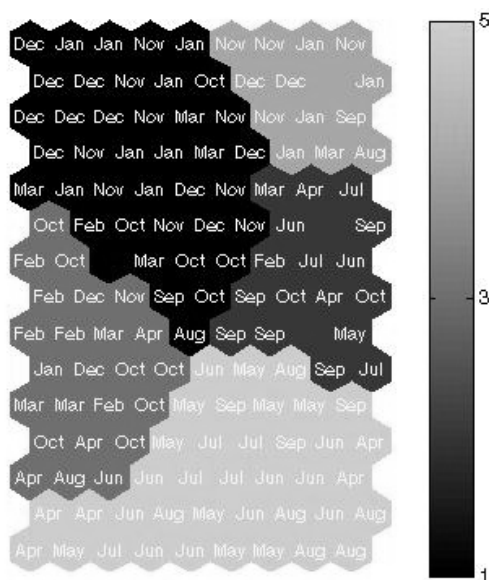
Obr. 3a: Matice U reprezentující čtvercové Euklidove vzdálenosti, Obr. 2b: Shlukování KSOM pomocí algoritmu K-průměrů

Proces shlukování je takto realizován ve dvou úrovních. Nejprve je n objektů redukováno do s reprezentantů $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_s$ pomocí KSOM. Potom je s reprezentantů shlukováno do q shluků. Největší vliv na vývoj kvality ovzduší města Pardubice má zkoumaná oblast (lokality) (Obr. 3) a měsíc (tj. roční období) (Obr. 4).



Obr. 3: Shlukování pomocí KSOM s využitím algoritmu K-průměrů (lokality)

Legenda: Zastávky autobusů (Cihelna (CI), Dubina (DU), Polabiny (PO), Rosice (RO), Rybitví (RY), Srnojedy (SR)), křižovatky (Palacha-Pichlova (PP), Náměstí Republiky (NR)), Lázně Bohdaneč (LB), chemická továrna Paramo (PA).

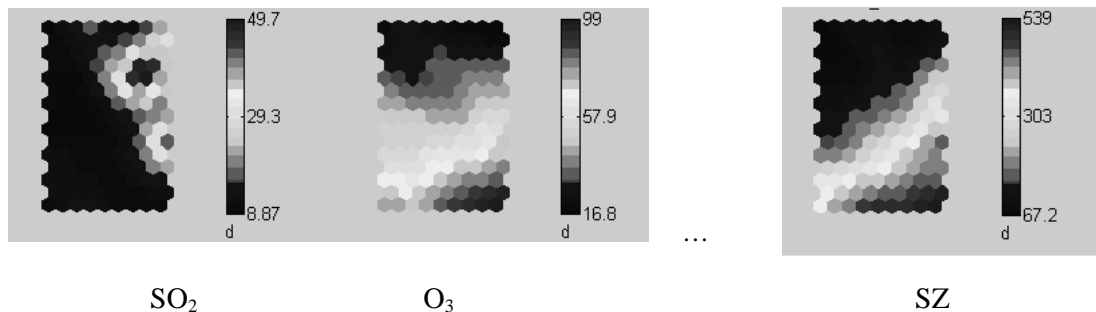


Obr. 4: Shlukování pomocí KSOFM s využitím algoritmu K-průměrů (měsíce)

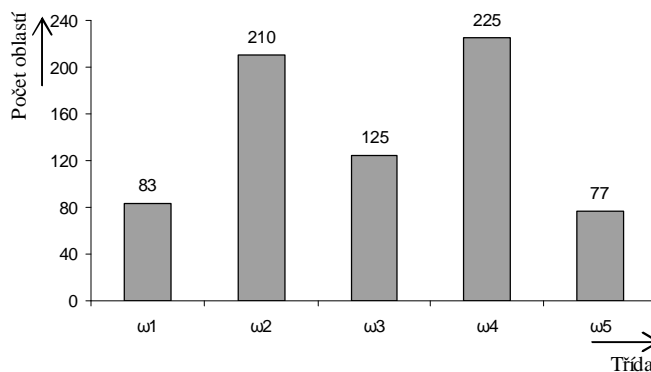
Legenda: Měsíce, leden (Jan), únor (Feb), březen (Mar), duben (Apr), květen (May), červen (Jun), červenec (Jul), srpen (Aug), září (Sep), říjen (Oct), listopad (Nov), prosinec (Dec).

Každému shluku je možné přiřadit souhrnný název na základě lokalit, které v něm převládají. Může to být např. zelená zóna (u shluku, kde převažují lokality jako jsou Lázně Bohdaneč nebo Srnojedy), dopravní křižovatky (Palacha-Pichlova, Náměstí Republiky) a další. Vliv roku na rozdělení shluků je minimální (hodnoty parametrů jsou v jednotlivých letech podobné, tzn. nedošlo k většímu výkyvu). Vliv měsíce, ve kterém jsou parametry měřeny je u některých shluků významný a u některých nevýznamný. Při interpretaci shluků se vychází z hodnot všech parametrů (Obr. 5). Výsledkem interpretace shluků je jejich zařazení do tříd $\omega_{i,j}^t \in \Omega$. Třídy jsou určeny pomocí kvality ovzduší (Tabulka 2) na základě rozmezí

hodnot znečišťujících látek. Všech pět shluků je označeno třídami $\omega_1^t, \omega_2^t, \dots, \omega_5^t$ a jejich četnosti výskytu tak, že třída ω_1^t reprezentuje nejméně znečištěné ovzduší a třída ω_5^t reprezentuje nejvíce znečištěné ovzduší. Četnosti výskytu (klasifikace zkoumaných oblastí $o_i^t \in O$ do tříd $\omega_{i,j}^t \in \Omega$ podle hodnot jejich znečištění) jsou znázorněny na Obr. 6. Charakteristika shluků pomocí parametrů je popsána v Tab. 4.



Obr. 5: Hodnoty parametrů $x_1^t, x_2^t, \dots, x_{11}^t$ pro reprezentanty KSOM



Obr. 6: Klasifikace oblastí do tříd $\omega_{i,j}^t$

Tab. 4: Označení shluků třídami $\omega_{i,j}^t$ podle kvality ovzduší

Shluk	Parametry škodlivin a rozptylových podmínek ovzduší	$\omega_{i,j}^t$ $j=1,2, \dots, 5$
1 ■	Dobrá kvalita, dobré rozptylové podmínky, zdravé ovzduší.	$\omega_{i,2}^t$
2 ■	Uspokojivá kvalita, mírně nepříznivé rozptylové podmínky, zdravotně přijatelné ovzduší.	$\omega_{i,3}^t$
3 ■	Vynikající kvalita, mírně nepříznivé rozptylové podmínky, zdraví příznivé ovzduší.	$\omega_{i,1}^t$
4 ■	Vyhovující kvalita, mírně nepříznivé rozptylové podmínky, ovzduší ohrožující citlivé osoby.	$\omega_{i,4}^t$
5 ■	Špatná kvalita, nepříznivé rozptylové podmínky, ovzduší ohrožující celou populaci.	$\omega_{i,5}^t$

5 Závěr

Vzhledem k neznámé příslušnosti zkoumaných oblastí $o_i^t \in O$ do tříd $\omega_{i,j}^t \in \Omega$ podle hodnot jejich znečištění byla pro modelování kvality a ochrany ovzduší Pardubicka použita metoda učení bez učitele KSOM v kombinaci s algoritmem K-průměrů. Tato metoda umožňuje nalezení dobře oddělených shluků a jejich vizualizaci. Z měření dat pomocí mobilního

monitorovacího systému HORIBA není možné třídy $\omega_{i,j} \in \Omega$, do kterých oblasti patří, zjistit. V předzpracování dat byly zjištěny korelační závislosti mezi parametry NO a NO₂. Pro další práci byl proto použit parametr NO. Z analýzy výsledků vyplývá, že znečištění vybraných lokalit Pardubicka je možno zařadit do j=5 tříd. Každá třída je ohodnocena kvalitou ovzduší a rozptylovými podmínkami, přičemž kvalita ovzduší je rozdělena do pěti intervalů na vynikající, dobrou, uspokojivou, vyhovující a špatnou a rozptylové podmínky do tří intervalů, na dobré, mírně nepříznivé a nepříznivé. Navržený model byl realizován v programovém prostředí Matlab 7.1 pod operačním systémem MS Windows XP.

Poděkování

Tato práce byla podporovaná vědecko-výzkumným projektem Ministerstva životního prostředí České republiky, grant číslo: SP/4i2/60/07 s názvem Indikátory pro hodnocení a modelování interakcí mezi životním prostředím, ekonomikou a sociálními souvislostmi.

Použitá literatura:

- [1] HAYKIN, S. Neural Networks: A Comprehensive Foundation. 2nd edition. New Jersey: Prentice-Hall, Inc., 1999.
- [2] KOHONEN, T. Self-organizing Maps. 3rd edition. New York: Springer-Verlag Berlin Heidelberg, 2001.
- [3] OLEJ, V. Modelovanie ekonomických procesov na báze výpočtovej inteligencie. [Vedecká monografia], Hradec Králové: M&V, 2003.
- [4] Státní politika životního prostředí České republiky 2004-2010. Praha: Ministerstvo životního prostředí, 2004.
- [5] STEIN, B., MEYER ZU EISSEN, S., WISSBROCK, F. On Cluster Validity and the Information Need of Users. Proc. of the Int. Conf. on Artificial Intelligence and Applications (AIA 03), Benalmádena, Spain, (2003), pp.216-221.

Kontaktní adresa:

prof. Ing. Vladimír Olej, CSc., Ing. Petr Hájek, Ph.D., doc. Ing. Jiří Křupka, Ph.D.
Ústav systémového inženýrství a informatiky
doc. Ing. Ilona Obršálová, CSc.
Ústav veřejné správy a práva
Fakulta ekonomicko-správní
Univerzita Pardubice
Studentská 84, 532 10 Pardubice
email: vladimir.olej@upce.cz, petr.hajek@upce.cz, jiri.krupka@upce.cz,
ilona.obrsalova@upce.cz
tel.: +420 466 036 004