

**POROVNANIE NEURÓNOVÝCH A REGRESNÝCH MODELOV NA PREDPOVEĎ
KONCENTRÁCIÍ PRÍZEMNÉHO OZÓNU
COMPARISON BETWEEN NEURAL AND REGRESSION MODELS FOR FORECASTING
OF SURFACE OZONE CONCENTRATIONS**

Kremler Martin¹, Valach Fridrich²

¹ Univerzita Komenského v Bratislave, Geomagnetické observatórium GFÚ SAV v Hurbanove²

Abstract

Surface ozone is one of the most serious problems of the regional air pollution over Europe. At present the most of European countries are developing European system for forecasting of ozone episodes and real-time data exchange. In paper are presented first preliminary results of simple regression models for forecasting of daily maximum surface ozone concentrations for the station Prievidza. These are also compared with neural network model results. Data from the period 1992 – 2001 were used. Three types of models were realized (maximum ozone concentrations of the previous day, maximum daily air temperatures or relative air humidity at 2:00 p.m. are inputs). Statistic characteristics of the results from 600 models are calculated. Uncertainties of the results range from 5,468 to 10,34 ppb. Results from neural models have lower uncertainties.

Key words: surface ozone, forecast, regression model, neural model

Úvod

V troposfére sa nachádza asi 10% z celkového množstva ozónu. Troposferický (prízemný) ozón sa na rozdiel od stratosferického radí ku škodlivinám. Vo vyšších koncentráciách je škodlivý pre ľudské zdravie i vegetáciu. Za posledné desaťročia jeho koncentrácie značne vzrástli po celej severnej pologuli, čím sa stal jedným z najvýznamnejších problémov regionálneho znečistenia ovzdušia. Venuje sa mu preto značná pozornosť.

Pod vedením Európskej agentúry pre životné prostredie sa rozvíja európsky systém krátkodobej predpovede ozónových epizód (obdobia s vysokými koncentraciami ozónu) a výmeny dát v reálnom čase. Do iniciatívy sa zapojili mnohé štáty Európy, no Slovensko zatiaľ nie. Koncentrácie ozónu sú ovplyvňované prekurzormi (oxidy dusíka, oxid uhoľnatý, metán a iné) a meteorologickými podmienkami.

Predpovede koncentrácií prízemného ozónu sú založené na empirických, štatistických a kauzálnych modeloch alebo na ich kombinácii. Najjednoduchším empirickým modelom je model založený na predpoklade, že predpovedaná koncentrácia je rovná aktuálnej. Expertný názor, založený na nameraných údajoch a osobných skúsenostiach, môže značne zlepšiť takúto empirickú predpoveď.

V štatistických modeloch je predpoveď vytvorená z aktuálne nameraných hodnôt koncentrácií ozónu v kombinácii so štatistickou informáciou o najpravdepodobnejšom vývoji koncentrácií pri predpovedaných meteorologických podmienkach. Štatistická informácia sa získava z databázy meraní

z niekoľkých rokov. K týmto modelom patria jednoduché regresné modely, modely založené na klasifikačných stromoch, modely založené na Kalmanových filtroch a modely na báze umelých neurónových sietí. Posledne zmieňované majú schopnosť samoučenia sa v zmysle toho, že vedia optimalizovať predpovede na základe predchádzajúcich skúseností. Najčastejšími prediktormi v rôznych štatistických modeloch sú: teplota vzduchu, vlhkosť vzduchu a koncentrácia ozónu z predchádzajúceho dňa.

V kauzálnych chemických modeloch sa koncentrácie ozónu počítajú na základe emisií prekursorov a hodnôt predpovedaných meteorologických podmienok berúc do úvahy prebiehajúce atmosferické procesy (rozptyl, depozícia, ...). Chemické predpovedné modely sa v princípe skladajú z dvoch častí: meteorologického modelu a chemického modulu. Prvý modul počíta prenos a rozptyl ozónu a jeho prekursorov v atmosfére. Druhý modul zas počíta chemickú transformáciu a depozíciu sledovaných škodlivín, jeho základom je chemická schéma (chemický mechanizmus).

V práci (Valach a Kremler, 2003) boli po prvýkrát na Slovensku použité k predpovedi maximálnych koncentrácií prízemného ozónu neurónové modely. V tento príspevok sa venuje jednoduchým regresným predpovedným modelom a porovnaniu dosiahnutých výsledkov s výsledkami neurónových modelov.

Údaje a metodika

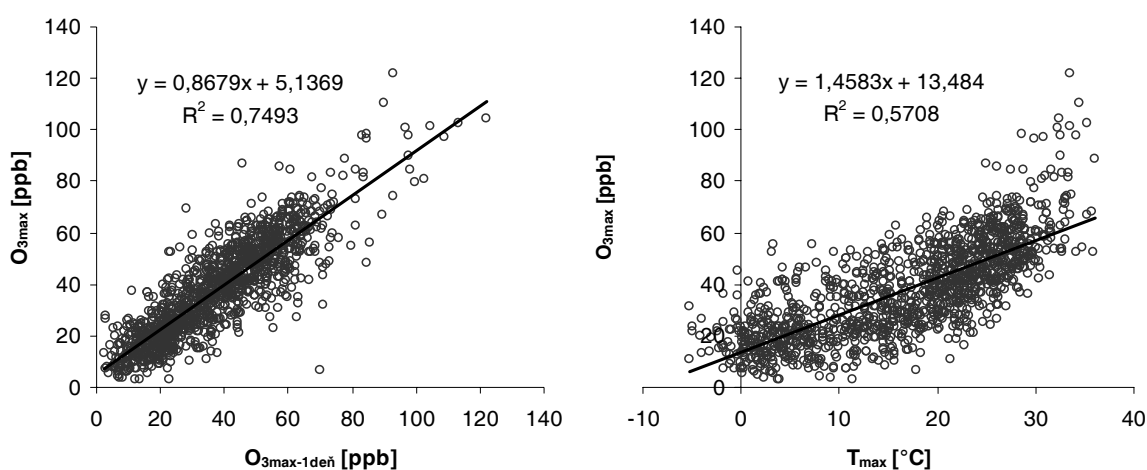
V práci sú použité údaje z Prievidze z obdobia 1992-2001: koncentrácie prízemného ozónu zo stanice merajúcej znečistenie ovzdušia a teplota a relatívna vlhkosť vzduchu z klimatologickej stanice. Z dostupných údajov bolo pripravených 1928 vzorov (1928 dní s kompletnými údajmi: maximálna denná koncentrácia prízemného ozónu, maximálna koncentrácia z predchádzajúceho dňa, maximálna denná teplota vzduchu a relatívna vlhkosť vzduchu o 14. hodine) pre dva navrhnuté neurónové modely. Tie boli náhodným výberom rozdelené na tri skupiny: prvá skupina (1300 vzorov) slúžila na tréning neurónovej siete, druhá (314 vzorov) na výber najlepšej siete a tretia (314 vzorov) na prezentovanie úspešnosti modelov.

Za účelom porovnania dosiahnutých výsledkov neurónových modelov s jednoduchými regresnými modelmi bolo zo základného súboru s 1928 údajmi náhodne vybraných 100 rôznych sád údajov po 1300 údajov a ďalších 100 sád po 1614 údajov (aj prvá aj druhá skupina pri neurónových modeloch vlastne slúži k výberu najvhodnejšieho modelu). Pre každú z týchto 100 údajových sád boli vytvorené tri jednoduché regresné modely. Prvý predpovedal maximálne koncentrácie prízemného ozónu na základe maximálnych koncentrácií z predchádzajúceho dňa, druhý z maximálnych teplôt vzduchu aktuálneho dňa a tretí pomocou relatívnych vlhkostí o 14. hodine daného dňa. Spolu sa použilo 600 regresných modelov. Každý bol otestovaný na skupine 314 údajov, ktoré ostali zo základného súboru po výbere sád s 1300 (resp. 1614) údajmi, pričom bola vyhodnotená pravdepodobná chyba modelu.

Výsledky a diskusia

Na obrázku 1 je závislosť maximálnych denných koncentrácií prízemného ozónu od maximálnych koncentrácií z predchádzajúceho dňa a od maximálnych denných teplôt vzduchu (1 prípad náhodného výberu 1300 údajov). Ako z neho vidno, tesnejšia je závislosť od koncentrácií minulého dňa ako od teploty, štvorce korelačných koeficientov sú 0,7493 a 0,5708. V prvom prípade sa na základe rozloženia údajových dvojíc v grafe dá tvrdiť, že ide o lineárnu závislosť, pri teplote sa to už tak jednoznačne tvrdiť nedá. Lineárna krivka nevystihuje najlepšie vzájomnú závislosť medzi maximálnymi koncentraciami a teplotami. V tomto konkrétnom prípade neboli uspokojivé výsledky dosiahnuté ani pri prekladaní údajových bodov inými krivkami. Tento fakt podporuje použitie neurónových sietí, pri ktorých nie je potrebné vedieť, aká závislosť je medzi skúmanými prvkami.

Štatistické charakteristiky (priemer, minimum, maximum a vybrané percentily) 6 rozdielnych súborov po 100 regresných modeloch (v každom pre korelačný koeficient, jeho štvorec, lineárne regresné koeficienty a pravdepodobnú chybu) sú prehľadne uvedené v tabuľke 1. Najvyššie korelácie sú pri modeloch, do ktorých vstupuje maximálna koncentrácia ozónu z predchádzajúceho dňa (pohybujú sa v rozsahu 0,85 – 0,876), nižšie pri modeloch s maximálnou teplotou (0,741 – 0,766) a najnižšie pri modeloch s relatívnou vlhkosťou (-0,68 – -0,637). Skoro totožné výsledky vyšli pri modeloch s 1614 vstupnými údajmi. Opačná situácia je pri pravdepodobných chybách modelov, pri modeloch s ozónovými koncentraciami sa pohybujú od 5,468 do 6,968 ppb, pri modeloch s teplotou od 7,288 do 8,878 ppb a pri modeloch s relatívnou vlhkosťou od 8,064 do 10,34 ppb.



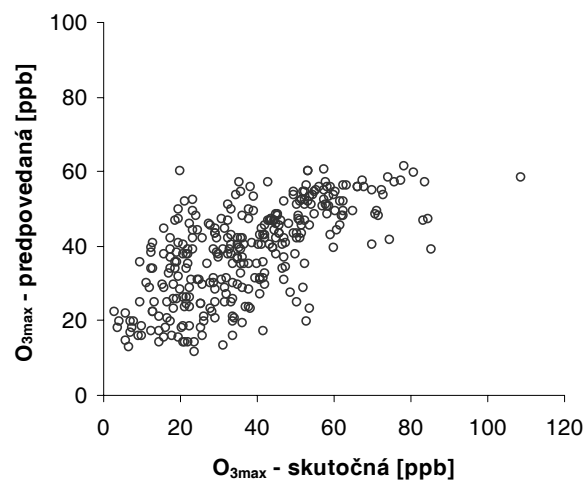
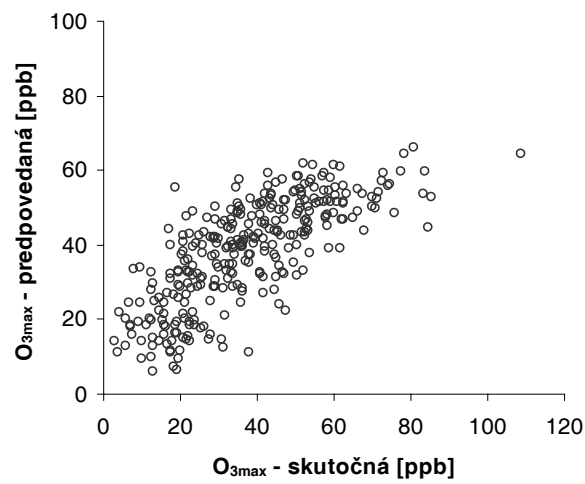
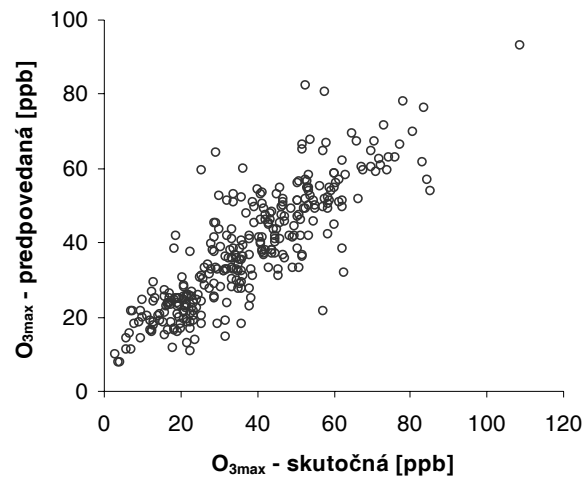
Obrázok 1 Závislosť maximálnych denných koncentrácií prízemného ozónu od maximálnych koncentrácií z predchádzajúceho dňa a od maximálnych teplôt vzduchu

Tabuľka 1 Štatistické charakteristiky 6 súborov po 100 regresných modeloch. r – korelácia, a, b – lineárne regresné koeficienty, ch – pravdepodobná chyba, p - percentil

		1300					1614				
		r	r ²	a	b	ch	r	r ²	a	b	ch
O₃max	priemer	0,863	0,745	0,861	5,233	6,322	0,863	0,745	0,863	5,126	6,287
	min	0,85	0,722	0,838	4,541	5,468	0,854	0,73	0,848	4,574	5,46
	5p	0,855	0,73	0,846	4,616	5,673	0,856	0,734	0,853	4,753	5,625
	10p	0,856	0,733	0,847	4,858	5,953	0,858	0,736	0,855	4,847	5,823
	25p	0,86	0,739	0,855	5,016	6,112	0,861	0,741	0,859	4,979	6,042
	50p	0,863	0,745	0,862	5,213	6,341	0,863	0,745	0,863	5,121	6,274
	75p	0,866	0,75	0,868	5,477	6,55	0,865	0,749	0,867	5,31	6,503
	90p	0,869	0,756	0,872	5,706	6,722	0,868	0,753	0,871	5,428	6,804
	95p	0,871	0,758	0,875	5,794	6,791	0,87	0,756	0,873	5,479	6,971
	max	0,876	0,767	0,878	5,887	6,968	0,872	0,76	0,881	5,671	7,169
Tmax	priemer	0,753	0,567	1,465	13,3	8,107	0,754	0,568	1,468	13,25	8,155
	min	0,741	0,548	1,393	12,58	7,288	0,746	0,556	1,42	12,56	7,192
	5p	0,744	0,554	1,427	12,74	7,576	0,748	0,559	1,446	12,88	7,629
	10p	0,746	0,557	1,44	12,77	7,618	0,748	0,56	1,449	12,96	7,721
	25p	0,749	0,561	1,448	12,97	7,888	0,751	0,564	1,458	13,07	7,942
	50p	0,753	0,566	1,468	13,31	8,118	0,753	0,567	1,47	13,25	8,129
	75p	0,757	0,573	1,483	13,62	8,292	0,756	0,572	1,478	13,4	8,464
	90p	0,761	0,579	1,492	13,76	8,605	0,759	0,577	1,487	13,56	8,603
	95p	0,764	0,583	1,495	13,92	8,71	0,76	0,578	1,491	13,68	8,786
	max	0,766	0,587	1,511	14,52	8,878	0,765	0,585	1,507	13,92	9,096
R14	priemer	-0,66	0,43	-0,64	75,29	9,284	-0,66	0,432	-0,64	75,35	9,313
	min	-0,68	0,406	-0,67	73,2	8,064	-0,67	0,412	-0,66	73,83	8,184
	5p	-0,67	0,41	-0,65	74,15	8,428	-0,67	0,42	-0,65	74,66	8,503
	10p	-0,66	0,416	-0,65	74,22	8,666	-0,66	0,423	-0,65	74,8	8,777
	25p	-0,66	0,421	-0,64	74,77	8,873	-0,66	0,427	-0,64	75,06	9,003
	50p	-0,66	0,431	-0,64	75,29	9,374	-0,66	0,431	-0,64	75,34	9,28
	75p	-0,65	0,439	-0,63	75,86	9,61	-0,65	0,437	-0,63	75,6	9,601
	90p	-0,65	0,442	-0,62	76,41	9,916	-0,65	0,441	-0,63	75,9	9,969
	95p	-0,64	0,446	-0,62	76,51	10,06	-0,65	0,443	-0,63	76,01	10,15
	max	-0,64	0,463	-0,61	76,93	10,34	-0,64	0,449	-0,62	76,82	10,68

Neurónové modely mali pravdepodobné chyby: model s 2 vstupmi 5,76 ppb a model s 3 vstupmi 5,52 ppb (Valach a Kremler, 2003). Regresné modely vychádzajúce z maximálnej dennej teploty vzduchu a z relatívnej vlhkosti majú pravdepodobné chyby 1,5 až 2-krát vyššie ako neurónové modely. Porovnateľné chyby majú len modely s maximálnymi koncentraciami ozónu z predošlého dňa na vstupe. Hodnoty chýb neurónových modelov sú nižšie ako 7. (resp. 3.) percentil rozloženia chýb regresných modelov tohto typu. Ťažko je porovnávať jednu hodnotu so štatistickými charakteristikami súboru pozostávajúceho zo 100 hodnôt, no v tomto prípade je dosť pravdepodobné, že neurónové modely poskytnú lepšie výsledky.

Na obrázku 2 sú grafy skutočných a regresným modelom predpovedaných maximálnych koncentrácií ozónu pre 3 modely s rôznymi vstupmi. Z druhého a tretieho grafy vidno, že tieto modely nevedia dobre predpovedať najmä vyššie koncentrácie prízemného ozónu (predpoveď je aj o 20 – 30 ppb nižšia). V praxi by boli preto nepoužiteľné, nakoľko najväčší dôraz sa kladie na to, aby boli čo možno najpresnejšie predpovedané práve vysoké koncentrácie ozónu.



Obrázok 2 Skutočné a modelom predpovedané maximálne koncentrácie prízemného ozónu, 1. graf – model, do ktorého vstupujú maximálne koncentrácie ozónu, 2. – vstupujú maximálne teploty, 3. – relatívna vlhkosť

Tieto prvé výsledky poukazujú na vhodnosť použitia neurónových modelov k predpovedi maximálnych koncentrácií prízemného ozónu. Bude však ešte potrebné urobiť porovnania s regresnými modelmi, do ktorých bude vstupovať viac parametrov (nie len jeden ako tomu bolo v tejto práci). Tiež by bolo vhodné vyskúšať viacero neurónových modelov, aj keď to bude dosť náročné na strojový čas, prípadne sa sústrediť len na letné obdobia s vysokými koncentraciami prízemného ozónu, kde by modely mali dávať ešte lepšie výsledky. V neposlednom rade by sa do testovania mali zahrnúť aj ďalšie slovenské stanice merajúce ozón.

PodĎakovanie

Práca bola vypracovaná v rámci projektov VEGA 1/8256/01 a 2/2009/22. Autor ďakuje grantovej agentúre VEGA za podporu výskumu.

Súhrn

Prízemný ozón je jedným z najväznejších problémov znečistenia ovzdušia nad Európou. V súčasnosti sú mnohé európske krajiny zapojené do programu vyvíjajúceho európsky systém na výmenu ozónových údajov v reálnom čase a predpoveď ozónových epizód. V príspevku sú prezentované prvé predbežné výsledky jednoduchých regresných modelov na predpoveď maximálnych denných koncentrácií prízemného ozónu pre stanicu Prievidza. Tie sú tiež porovnávané s výsledkami modelov založených na umelých neurónových sieťach. Údaje boli použité z obdobia 1992 – 2001. Skúmané boli tri typy modelov (vstupom do modelu boli maximálne koncentrácie ozónu z predchádzajúceho dňa, maximálne denné teploty vzduchu alebo relatívna vlhkosť o 14. hodine). Z výsledkov 600 použitých modelov boli vypočítané štatistické charakteristiky. Pravdepodobné chyby jednotlivých modelov sa pohybujú v rozsahu od 5,468 do 10,34 ppb. Výsledky z neurónových modelov majú nižšiu neurčitost'.

Kľúčové slová: prízemný ozón, predpoveď, regresný model, neurónový model

Literatúra

van Aalst, R.M. - Fiala, J. - Hunová, I. - Dostálová, Z., 1999: National ozone forecasting systems and international data exchange in Central Europe. PTL-ETC/AQ Report of TWG-DFO, Praha.

Pelikán, E. et al., 1998: Zpráva o průběhu a výsledcích řešení dílčího úkolu č. 2 grantového projektu MŽP ČR č. VaV/520/2/97. Predikce fotochemického smogu. Ústav informatiky AV ČR, Praha.

Valach, F. - Kremler, M., 2003: Koncentrácie prízemného ozónu modelované neurónovou sieťou. In: Bioklimatologické pracovné dni 2003, Račkova dolina.

Kontaktná adresa: RNDr. Martin Kremler, Katedra meteorológie a klimatológie FMFI UK, Mlynská dolina, 842 48 Bratislava, tel.: 02/602 95 216, e-mail: kremler@fmph.uniba.sk