

VÍCEROZMĚRNÉ METODY PŘEDPOVĚDÍ KLIMATICKÝCH ANOMÁLIÍ

PRACOVNÍŠTĚ: Katedra meteorologie a ochrany prostředí MFF UK
VEDOUCÍ DISERTAČNÍ PRÁCE: doc. RNDr. Jaroslava Kalvová, CSc.

Chování klimatického systému ve vyšších zeměpisných šířkách vykazuje řadu znaků typických pro deterministicky chaotické systémy. To zahrnuje i omezenou možnost úspěšné deterministické předpovědi stavu atmosféry, jejímž důsledkem je praktická nepředpověditelnost počasí na víc než několik málo dní dopředu. Zároveň se ale ukazuje, že mj. díky vlivu pomaleji proměnlivých složek klimatického systému existuje v některých případech možnost sestrojít výhled podstatně dále do budoucna. Právě touto problematikou se zabývá předkládaná doktorská disertační práce mgr. S. Kliegrové, a to konkrétně možnosti využití empirických statistických modelů ke konstrukci dlouhodobých předpovědí teplot v České republice. Kromě vlastní předpověditelnosti je pozornost věnována rovněž identifikaci vhodných typů přenosových funkcí, především porovnání výsledků získaných pomocí vícenásobné lineární regrese a pomocí různých typů neuronových sítí.

Práce je psána v českém jazyce a její celkem 77 stran je rozděleno do pěti kapitol, doplněných obsahem, úvodem, seznamem literatury a obrázkovou a tabelární přílohou.

OBSAH PRÁCE

Po úvodu a přehledu problematiky dlouhodobé předpovědi (kap. 1) následuje shrnutí používaných metod (kap. 2). Jsou stručně představeny základní postupy předzpracování dat (2.1), modely založené na vícenásobné lineární regresi (2.2.1) a neuronové sítě (2.2.2). V sekci 2.3 jsou pak uvedeny a diskutovány různé postupy kvantifikace úspěšnosti předpovědi pro následné analýzy je (vedle kontingenčních tabulek) využíváno hlavně LEPS-skóre, aplikované pro rozdělení hodnot do tří kategorií (podnormální, normální, nadnormální hodnoty). V kapitole 3 jsou nejprve popsány používané datové soubory prediktandu (průměrné měsíční teploty vzduchu pro oblast České republiky) a prediktorů (hodnoty NAO indexu a teploty povrchu oceánu v různých částech Atlantiku), následně je provedena analýza lineárních korelací teplot v ČR s (časově zpožděnými) hodnotami NAO indexu, předchozími hodnotami teploty v ČR a předchozími teplotami povrchu oceánu. Vlastní realizace předpovědních modelů je popsána v kapitole 4, individuálně pro jednotlivé měsíce roku. V závěru práce jsou pak výsledky diskutovány s ohledem na praktickou použitelnost získaných modelů, vhodnost jednotlivých postupů a možnosti dalšího směřování výzkumu v této oblasti.

FORMÁLNÍ ZPRACOVÁNÍ

K formální stránce práce nemám žádné závažnější připomínky. Práce je logicky členěná, rozsahově přiměřená, počet drobných chyb a překlepů není velký, jen v některých partiích text trpí buď nedostatkem či naopak přebytkem mezer mezi slovy či slovy a interpunkčními znaménky. V seznamu literatury chybí práce Kennedy a Gentle (1980), zmiňovaná na str. 14; sekce 3.3 odkazovaná na str.19 neexistuje. Osobně bych uvítal, kdyby mapy korelací typu obr. 3.2.2 obsahovaly i škálu hodnot.

VĚCNÁ STRÁNKA PRÁCE

Celkově se podle mého názoru jedná o zajímavý příspěvek ke studiu možností dlouhodobé statistické předpovědi teplotních charakteristik. Použitelnost získaných modelů pro praktickou předpověď může být sice zatím omezená (díky značně rozkolísané úspěšnosti pro různé měsíce), ale jde o dobrý základ pro případný navazující výzkum této problematiky. K věcnému obsahu práce nemám podstatnější námitky, snad jen několik postřehů a dotazů:

U některých typů neuronových sítí, včetně vícevrstvých perceptronů učených pomocí algoritmů založených na zpětném šíření chyby, závisí výsledný tvar přenosové funkce na počáteční inicializaci vah, tj. trénováním sítě z různých počátečních podmínek lze dospět k různým výsledkům. Rád bych se zeptal, zda byla nějakým způsobem testována citlivost výsledků na volbu inicializace vah sítě, resp. zda bylo toto nějak zohledněno při identifikaci nejvhodnějších sítí (popsané na str. 19) a při vlastní realizaci předpovědních modelů v kap. 4?

Volba prediktorů se v práci provádí na základě statistické významnosti lineární korelace mezi potenciálními prediktory a prediktandem. Vzhledem k vysokému počtu testovaných dvojic se ale může stát, že i v případě zcela nezávislých řad mohou být korelační koeficienty u určitého počtu případů formálně vyhodnoceny jako statisticky významné – u použitého testování na 95% hladině tak v průměru v jednom z dvaceti případů může dojít k odmítnutí hypotézy o nezávislosti, i pokud žádná reálná vazba neexistuje. Např. v tab. 3.2.1 je jako statisticky významně korelované vyhodnoceno 9 z celkem 132 časově posunutých dvojic řad, přičemž na základě čisté náhody by to mohlo být cca 7 párů. Pokud by tomu tak opravdu bylo, nalezené vztahy by byly (alespoň částečně) důsledkem pouhé nahodilosti a nepoužitelné pro konstrukci modelů pro skutečnou předpověď mimo analyzovaný časový interval. S tím souvisí i další možný problém – i pokud vyjadřují nalezené korelace reálné deterministické vazby pro sledovanou periodu jako celek, není to zárukou jejich stacionarity. Bylo by proto možná dobré se pokusit o ověření předpovědí na zcela nezávislém vzorku, a to včetně volby prediktorů prováděné pouze na základě dat z učebního souboru. Je mi ale jasné, že vzhledem k malému množství dostupných dat by se jednalo o značně problematický úkol – snad by bylo možné pracovat nikoliv separátně s jednotlivými měsíci, ale sdružit měsíční hodnoty do větších celků (např. ročních období).

Na str. 26 se konstatuje, že v červnu, srpnu a září sice existují statisticky významné korelace mezi hodnotami NAO a teplotami, ale nemají pro případný prognostický potenciál valný význam – co je tím míněno?

V tab. 3.2.2 došlo pravděpodobně k přehlédnutí statisticky významné korelace mezi T7 a T8.

Oblasti identifikované shlukovou analýzou pro volbu SST prediktorů jsou často relativně složitě tvaru, občas i nespojitě, navíc se silně liší mezi různými měsíci. Byly poloha a tvar nejvýznamnějších nalezených oblastí vysoké korelace (tj. těch následně použitých jako zdrojových oblastí prediktorů) citlivé na volbu nastavení shlukové analýzy, konkrétně na počet hledaných shluků?

Vzhledem k poměrně malému počtu učebních vzorků (36 v případě realizace regresních modelů a neuronových sítí v kapitole 4) a poměrně vysokému počtu volných parametrů některých sítí (např. v 4.2.9 byla v případě sítě se spojitými hodnotami na výstupu použita architektura se třemi vstupy a čtyřmi neurony ve skryté vrstvě, což představuje minimálně 16 vah/volných parametrů sítě) bude existovat značný sklon k přetrévání (tento problém je ostatně autorkou uváděn v sekci 2.2.2). Na str. 18, v souvislosti s volbou optimální architektury sítě, je zmiňováno dělení souboru na učební, verifikační a testovací část v poměru 1/2, 1/4, 1/4, ale není mi úplně jasné, jak bylo toto ošetřeno v případě vlastní aplikace neuronových sítí v kap. 4 bylo by to možné upřesnit?

Volba potenciálních prediktorů se provádí na základě statistické významnosti Pearsonova korelačního koeficientu a do vztahů v kap. 4 tak vstupují už jen prediktory zvolené na základě lineárních korelací. Přitom

rozdíl mezi výsledky lineární regrese a neuronových sítí (patrný i v případech, kdy jsou použity totožné prediktory) naznačuje možnou přítomnost detekovatelné nelineární komponenty. U nelineárních metod tak nemusí podle mého názoru volba prediktorů coby proměnných s vysokou lineární korelací s prediktandem být optimální, protože využitelná deterministická relace prediktor(y)-prediktand může být složitější. V budoucnu by proto možná stálo za pokus aplikovat i nějakou veličinu beroucí v potaz i nelineární statistické vazby.

ZÁVĚR:

Vzhledem ke složitosti a nelinearitě zkoumaných vazeb a omezenému množství dostupných dat se jedná o náročné a netriviální téma. Výsledky předkládané práce jsou zajímavým příspěvkem k pochopení struktury vazeb v klimatickém systému a jejich eventuelnímu praktickému využití při dlouhodobé předpovědi teplot. **Výše uvedené poznámky jsou jen dílčího rázu, předkládaná doktorská disertační práce mgr. S. Kliegrové podle mého názoru splňuje obsahové i formální požadavky kladené na tento typ práce a prokazuje předpoklady autorky k samostatné tvořivé činnosti. Proto ji jednoznačně doporučuji k obhajobě.**

V Praze 27.7.2006

Jiří Mikšovský *



* Katedra meteorologie a ochrany prostředí, Matematicko-fyzikální fakulta, Universita Karlova; jiri.miksovsky@mff.cuni.cz

