

Posudek na doktorskou disertační práci Mgr. S. Kliegrové „Vícerozměrné metody předpovědi klimatických anomálií“

Práce Mgr. S. Kliegrové se zabývá problémem, který je aktuální jak z čistě vědeckého, tak praktického hlediska: dlouhodobou předpovědí měsíčních průměrných teplot na území ČR. Mgr. Kliegrová využívá pro předpověď statistických modelů, a to jak lineárních (násobnou lineární regresi), tak nelineárních (neuronové sítě se spojeným a kategoriálním výstupem). Jako potenciálních prediktorů používá průměrných hodnot teploty v ČR, indexu Severoatlantické oscilace (NAO) a teploty povrchu oceánu (SST) v severním Atlantiku v předchozích měsících. Z nich pro vstup do regresních modelů vybírá ty, jež s teplotami v ČR v daném měsíci vykazují statisticky významné korelace. Přitom SST jsou předzpracovány pomocí shlukové analýzy tak, že vlastní prediktory tvoří SST v oblastech homogenních z hlediska časového vývoje jejich korelací s teplotou v ČR. Předpovědní modely jsou konstruovány pro každý měsíc v roce. Výsledky jsou ověřovány křížovou validací a kvantifikovány pomocí LEPS skóre. Nelineární metody dávají lepší výsledky než lineární, nicméně předpovědi jsou prakticky použitelné jen pro některé měsíce. Z odborného pohledu je práce zajímavá mj. proto, že v oblasti Evropy je vliv El Niña / Jižní Oscilace velmi slabý, a proto dlouhodobá předpověditelnost vývoje počasí (klimatických anomálií) je značně problematická. Každou práci na dlouhodobých předpovědích v Evropě lze proto pokládat za svým způsobem odvážnou.

Předložená práce je kompaktní, logicky a přehledně členěná. Skládá se z pěti kapitol. V první kapitole autorka zavádí základní pojmy a prezentuje několik zdrojů dlouhodobých předpovědí, které jsou pro naši republiku dostupné a v různé míře se v předpovědní praxi využívají. Druhá kapitola stručně shrnuje použité statistické metody (analýzu hlavních komponent, shlukovou analýzu, násobnou regresi, neuronové sítě) a skóre použítá pro vyhodnocení výsledků. Třetí kapitola je věnována popisu vstupních dat a předzpracování prediktorů – tento nenápadný název zahrnuje jak výpočty korelací (se zpožděním až 12 měsíců) řad měsíčních teplot navzájem a s NAO indexem, tak shlukovou analýzu korelačních koeficientů mezi teplotou v ČR a SST v Severním Atlantiku, podepřenou analýzou korelací mezi SST a NAO. Čtvrtá kapitola popisuje získané statistické modely pro jednotlivé měsíce, charakterizuje jejich úspěšnost a diskutuje jejich použitelnost v praxi. Závěrečná kapitola shrnuje dosažené výsledky.

Práci hodnotím jako kvalitní, prokazující, že Mgr. Kliegrová problematice dobře rozumí, že prostudovala potřebnou literaturu a získala v ní přehled a že je schopna samostatně tvořivé práce. Dále uvedené své připomínky proto nepovažuji za nijak zásadní. Rovněž v dalším textu dávám námět na obecnější „vědeckou rozpravu“ o některých tématech, jichž se práce dotýká.

str. 14: odkaz na práci Kennedy a Gentle (1980) – v seznamu literatury chybí
str. 18, 3.odst.: zmíněna „dvě úskalí“, vyjmenováno a diskutováno je však jen jedno
str. 21 nahore: v příkladu má být 20% pravděpodobnost nadnormální (nikoliv podnormální) teploty
kap. 2.3: uvítal bych obecnější popis, jak se ze „skóre“ dostane „skill skóre“

popis LEPS skóre: zasloužil by si být úplnější; zejm. je nejasné, proč se uvažuje zrovna druhá derivace vztahu (2.3.7).

data – prediktandy: Prosim dodat vysvětlení, proč byly vyloučeny horské stanice. tab. 3.2.1., 3.2.2.: Jaká je použitá hladina statistické významnosti – to je důležitá informace, nikde jsem ji ale nenašel. Je počet významných korelací větší, než by odpovídalo náhodě? (Přepokládám, že hladina významnosti je 5%, a odpověď na mou otázku je tedy „ano“.)

str. 45 a dále: Hodnota LEPS = 0.4 je používána jako dolní hranice praktické použitelnosti předpovědi. Z čeho to vyplývá? (Stačil by odkaz na literaturu nebo stručná diskuse.)

prediktory v statist. modelech: domnívám se, že proměnné MEAN a FAC popisují v zásadě totéž; není jejich použití nadbytečné?

str. 47: prosím bližší vysvětlení tvrzení, že pro únor nebyl nalezen žádný uspokojivý model

Autorka používá některé anglikanismy, jež jsou snadno nahraditelné českými ekvivalenty: jako příklad uvádím „dataset“, „jet-stream“, „gridové“ body, °N.

Náměty k obecnější diskusi:

1. Kap. 3.2.3 obsahuje výsledky, které jsou podle mne velmi zajímavé i mimo kontext dlouhodobé předpovědi (vazby NAO → SST → teplota v ČR, a zejména vazba SST a NAO); jejich interpretace by si rozhodně zasloužila větší prostor než jen jeden odstavec na str. 42.

2. Jako potenciální prediktory jsou používány ty veličiny, které mají statisticky významné korelace s prediktandem (teplotou v ČR). Nemohou však být jako prediktor užitečné i takové veličiny, které významně korelovány nejsou? To platí zejm. pro nelineární metody. (Je vidět, že vyšší korelace nezajišťuje větší „užitečnost“ v předpovědním modelu – viz např. MEAN vs. FAC v prosinci.)

3. Není úplně jasně popsáno, jak přesně byla realizována křížová validace.

Předpokládám, že prediktory a architektura nelin. modelů byly vybírány na celém souboru; křížová validace potom byla použita na jejich vyhodnocení. Podle mne je křížová validace použitelná i tak, že prediktory jsou vybírány pro každý cross-validovaný soubor zvlášť (přičemž nevím, je-li to prakticky realizovatelné v nelin. modelu).

4. Čím může být způsobena velká meziměsíční proměnlivost úspěšnosti statistických předpovědí?

Na závěr znovu konstatuji, že předpožená disertace rozhodně prokazuje předpoklady Mgr. S.Kliegrové k samostatné vědecké práci.

Motyčky-Bukovec 22.7.2006

