



**Posudek doktorandské disertační práce**

*Lukáše Popa*

## Analýza nárazů větru na území České republiky

Předložená práce je věnována především analýze extrémních nárazů větru na území České republiky. Její součástí je ale i zpracování vysokofrekvenčního měření rychlosti větru na stanici Kopisty, provedení homogenizace řad nejvyšších denních nárazů větru na 29 českých stanicích a vytvoření mapy extrémního větru pro ČR včetně odhadu její přesnosti.

Práce se skládá ze čtyř kapitol, závěrečného shrnutí a čtyř příloh. První kapitola stanoví cíle, popisuje strukturu práce, druhá kapitola se zaměřuje na měření větru v ČR, třetí se věnuje především analýze extrémních hodnot a odhadu doby návratu a čtvrtá kapitola konstruuje mapu extrémního větru v ČR. Jako matematický statistik (ve své odborné práci spolupracující i s klimatologem) se budu ve svém posudku zaměřovat především na stránku analýzy zkoumaných dat a na vhodnost použitých matematicko-statistických nástrojů a metod.

Práce využívá nástrojů statistické teorie extrémních hodnot, jejíž základy vznikaly v první polovině 20. století, ale jako svébytná disciplína existuje cca 30-40 let a v posledních letech dochází k jejímu velkému rozmachu. Získané teoretické výsledky postupně obohatily a obohacují řadu oborů jako je například klimatologie. Tyto obory zpětně přinášají řadu problémů, které teoretická statistika řeší. Z tohoto pohledu se proto domnívám, že práce řeší skutečně aktuální téma. Navíc se disertace opírá a navazuje na řadu výsledků, které byly publikovány teprve nedávno v kvalitní odborné literatuře.

Z pohledu matematické statistiky lze za nejzajímavější a nejcennější část disertace pokládat třetí kapitolu. Kromě rešeršních částí zahrnující základní poznatky o teorii extrémních hodnot, metodách a aproximaci extrémních hodnot a době návratu tato kapitola obsahuje rozsáhlou diskuzi použití zobecněného rozdělení extrémních hodnot (GEV rozdělení) a různých přístupů k pojmu doba návratu včetně řady vlastních odvození a tvrzení. To považuji za velmi cenný příspěvek autora, protože teoretické matematicko-statistické výsledky jsou odvozeny vždy za jistých předpokladů a podmínek a bez jejich řádného prověření může vést použití těchto výsledků k chybným a často i nesmyslným závěrům. To je právě případ použití GEV rozdělení v oblasti analýzy extrémních hodnot, které je limitním rozdělením standardizovaných maxim, ale na rozdíl od analogicky používané techniky asymptotického rozdělení průměru (normální rozdělení), je tato konvergence ke GEV rozdělení velmi pomalá a je proto nutné v reálných situacích při práci s daty s použitím tohoto rozdělení postupovat velmi obezřetně. Autor práce si je toho velmi dobře vědom a řadu možných úskalí v souvislosti s analýzou extrémních nárazů větru zkoumá a dospívá tak k zajímavým výsledkům.

Tato třetí kapitola je doplněna rozsáhlou numerickou studií autora pro aproximaci dat Gumbelovým a GEV rozdělením pomocí metody maximální věrohodnosti a metody L-momentů. Získaná zjištění jsou pak vhodným způsobem využita při odhadu chyby extrémní větrné mapy ve čtvrté kapitole. Provedení simulace nebylo rutinní záležitostí. Bylo například potřeba použít numerických metod pro hledání minima v metodě maximální věrohodnosti, což však přináší obvykle problém, že v některých případech dat algoritmus nekonverguje, což je nutné si uvědomit a řešit. Autor si je těchto obtíží vědom a vhodně je řeší. Simulační studii hodnotím jako velmi zdařilou a domnívám se, že přináší cenné výsledky.

Jak jsem uvedl výše, práce obsahuje i několik vlastních tvrzení a odvození autora. V těchto odvozeních a důkazech tvrzení jsem nenalezl žádná závažná pochybení a podle mého mínění jsou provedeny korektně. Stejně tak statistický aparát při analyzování dat nárazů větrů je volen vhodně a používán korektně.

Práce je dobře čitelná, graficky dobře upravená a je psána v češtině, která odpovídá odbornému jazyku v časopisech s tématicky blízkými články. Protože není odborná terminologie některých novějších pojmů z teorie extrémních hodnot v češtině ustálená a je velmi málo frekventovaná, musel autor používat v řadě případů termíny vlastní. U některých těchto pojmů bych však doporučoval jiný termín, který je podle mě vhodnější a přeci jenom používanější, např. doména atrakce (domain of attraction). Domnívám se, že se ve statistice spíše užívá pojem sféra přitažlivosti. Za mnohem problematičtější však považuji použití termínu průměrný. V české statistické terminologii, pokud hovoříme o náhodné veličině, se spíše používá vyjádření střední (hodnota, doba, chyba). Např. v části 3.3.6 je pojednáno o průměrné době návratu. Vzhledem k tomu, že v důkazu se používá výpočet střední hodnoty této doby, by bylo patrně vhodnější použít pojem střední doba návratu. Podobně nešikovná formulace je na straně 69, kdy autor uvádí "L-momenta všech řádů existují, pokud je průměr rozdělení konečný". Pojem průměr rozdělení je podle mne nevhodný a mnohem smysluplnější by bylo použití formulace střední hodnota, tj. výše uvedené tvrzení pak lze psát ve tvaru "L-momenta všech řádů existují, pokud existuje první moment rozdělení".

V práci jsem našel také několik překlepů, nejednoznačností používaného značení, drobných nepřesností či přehlédnutí v logice vývoje výkladu, které mohou ztížit čtenářovu orientaci, např.:

- str. 49, (3.3.5): Nemělo by místo  $G^{-1}((G_n^{\frac{1}{n}}(x))^{1/n})$  být jen  $G^{-1}(G_n^{\frac{1}{n}}(x))$  ?
- str. 50, (3.3.10): Na tomto řádku je pro označení distribuční funkce normovaného normálního rozdělení použitý symbol  $\phi$  a o řádek níže symbol  $\varphi$ .
- str. 51: Autor uvádí: "... Nemožnost ověření podmínek konvergence blokových maxim ke GEV rozdělení vede k otázce ...". Domnívám se, že několik vhodných testů se pro tyto účely se ve statistické literatuře objevilo, např. viz článek Dietrich, D., de Haan, L., and Husler, J. (2002). Testing extreme value conditions. *Extremes*, 5:71.
- str. 51: Autor uvádí: "... Potom jeho bloková maxima budou konvergovat ke Gumbelovu rozdělení ... Ze vzorce (3.3.2) ovšem plyne, že rozdělení blokových maxim je jednoznačně určeno rozdělením původních hodnot. Určení parametrů limitního Gumbelova rozdělení a parametrů rozdělení původních hodnot tedy představuje ekvivalentní úlohu ...". Vzorec (3.3.2) říká, že distribuční funkce maxima  $n$  stejně rozdělených náhodných veličin je  $n$ -tá mocnina distribuční funkce těchto náhodných veličin. Abychom řešili ekvivalentní úlohu, měli bychom se také dívat na limitní rozdělení této  $n$ -té mocniny, což je funkce nabývající pouze dvou hodnot 0 a 1 se skokem v bodě  $x_{sup}$ . Domnívám se, že tedy nejde o ekvivalentní úlohu v matematickém slova smyslu. Problém je v tom,

že matematicky přesně bloková maxima nekonvergují ke Gumbelově rozdělení, ale konvergují k němu standardizovaná maxima.

- str. 54: Na některých místech textu autor řadu rychlostí **z**hladí na jiných však **sh**ladí.
- str. 103, (4.1.5): Ve vzorci je uveden člen  $\hat{y}_l$ , v dalším textu se však vyskytuje  $\hat{y}_J$ .
- ...

V rámci obhajoby práce bych se autora rád zeptal na problematiku rozhodování toho, zdali data mají Gumbelovo rozdělení nebo mají těžší (respektive lehčí) chvost než toto rozdělení (část 4.1.8). Proč pro řešení tohoto problému autor nepoužil žádný standardní postup jako je test poměrem věrohodnosti, ale navrhuje "jen" své vlastní tři postupy ?

Nedomnívám se však, že by výše uvedené kritické poznámky výrazněji ovlivnily celkově velmi dobrou úroveň disertační práce. Mohou být brány v úvahu zejména při další publikaci výsledků.

Závěrem konstatuji, že autor touto prací prokázal hlubokou znalost tématu a schopnost samostatné tvůrčí vědecké práce. Domnívám se, že předložená práce přináší nové hodnotné výsledky týkající se analýzy extrémních nárazů větru v ČR a že splňuje všechny požadavky kladené na práci disertační, a proto ji doporučuji přijmout k obhajobě.

Liberec, 30. července 2015

prof. RNDr. Jan Pícek, CSc.