

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ
FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY
INSTITUTE OF MATHEMATICS

ING. JAN HOLEŠOVSKÝ

METODY ODHADU PARAMETRŮ ROZDĚLENÍ EXTRÉMNÍHO TYPU S APLIKACEMI EXTREME VALUE DISTRIBUTION PARAMETER ESTIMATION AND ITS APPLICATION

zkrácená verze PhD Thesis

Studijní obor: Aplikovaná matematika
Vedoucí práce: doc. RNDr. Jaroslav Michálek, CSc.
Oponenti: prof. RNDr. Jaromír Antoch, CSc.
prof. RNDr. Jan Pícek, CSc.

KLÍČOVÁ SLOVA

Rozdělení extrémních hodnot, index extrémní hodnoty, zobecněné Paretoovo rozdělení, extrémální index, neparametrické odhady.

KEY WORDS

Extreme value distribution, extreme value index, generalized Pareto distribution, extremal index, nonparametric estimation.

MÍSTO ULOŽENÍ DISERTAČNÍ PRÁCE

Areálová knihovna Fakulty strojního inženýrství VUT v Brně

OBSAH

Úvod	5
Cíle dizertační práce	5
1 Základy teorie extrémních hodnot	6
2 Odhady parametrů rozdělení extrémních hodnot	8
3 Metody výběru prahových hodnot	11
4 Analýza extrémních srážkových úhrnů	14
5 Teorie extrémních hodnot pro stacionární řady a její použití	18
Závěr	22
Literatura	22
Autorovy publikace	26
Autorův životopis	28
Abstrakt	30

ÚVOD

Teorie extrémních hodnot (zkráceně EV teorie z angl. Extreme Value) je oblast matematické statistiky zaměřená na rozvoj metod určených pro popis a modelování řídce se vyskytujících, extrémních jevů. Potřeba užití adekvátních technik predikce málo častých událostí se objevuje v mnoha oblastech lidské činnosti. Zejména se jedná o nutnost odhadu vlastností chvostů daných rozdělení pravděpodobnosti, neboť tyto mají zásadní vliv na selhání či poruchy sledovaného procesu. Obvykle je pro takový proces vyžadována předpověď dosažení nějaké vysoké, předem stanovené úrovně, přičemž tato se často nachází mimo oblast pozorovaných dat. Ačkoliv jsou tyto predikce vždy postaveny na důvěře ve správnost konkrétního modelu, EV teorie dnes představuje pravděpodobně nejadekvátnější prostředek pro popis vlastností extrémů náhodných veličin.

Historicky se jako první úloha EV teorie uvádí Bernoulliho problém (1709) nalezení pravého koncového bodu rovnoměrného spojitého rozdělení. Významný rozvoj EV teorie byl však zaznamenán až v polovině 20. století (viz práce [13, 17]). V současnosti lze největší zásluhy vývoje EV teorie přisoudit praktickým potřebám reálných úloh zejména v oblasti hydrologie a ekonomie. Poslední dobou se přitom v literatuře stále více objevují aplikace z nejrůznějších vědních oblastí i mimo tyto tradiční obory. Z moderních řešených problematik je možné zmínit například modelování výšky přílivu, rychlosti větru, analýzu rozsáhlých požárů, znečištění životního prostředí, redukci šumu obrazových dat nebo posouzení kapacity telekomunikačních sítí. Citace konkrétních publikací je možné nalézt v autorově dizertační práci. Přes celou řadu aktuálních článků a monografií však patří EV teorie ke stále se rozvíjejícím oborům matematické statistiky s mnoha dosud nedorozřešenými tématy. Vybranými oblastmi, které zůstávají nadále otevřeny dalšímu výzkumu, se zabývá tato práce.

CÍLE DIZERTAČNÍ PRÁCE

- Shrnout metody odhadu parametrů EV rozdělení pro jednorozměrná pozorování s důrazem na neparametrické techniky.
- Popsat přístupy odhadu parametrů EV rozdělení pro korelovaná pozorování a porovnat je s běžně užívanými aproximativními technikami pro nezávislá pozorování.
- Srovnat předešlé metody pomocí simulovaných dat, využít je při modelování reálných situací a provést diskuzi vhodnosti představených technik.
- Navržené postupy počítačově implementovat.

1. ZÁKLADY TEORIE EXTRÉMNÍCH HODNOT

ROZDĚLENÍ EXTRÉMNÍCH HODNOT

Nechť je dán náhodný výběr $X_1, \dots, X_n$ z rozdělení s distribuční funkcí $F(x)$. EV teorie je zaměřena na asymptotické chování extrémních hodnot, tj. hodnot, které jsou pozorovány s malou pravděpodobností. Označí-li se $M_n = \max\{X_1, \dots, X_n\}$ výběrové maximum, pak je nejprve cílem stanovení limitního rozdělení veličiny M_n pro $n \rightarrow \infty$. V tomto ohledu je EV teorie obdobou asymptotické teorie výběrových průměrů, která je popsána pomocí centrální limitní věty.

Bud' $F_{M_n}(x) := P(M_n \leq x)$ distribuční funkce výběrového maxima M_n . S využitím předpokladu nezávislosti veličin $X_1, \dots, X_n$ lze odvodit následující tvar funkce $F_{M_n}(x)$

$$F_{M_n}(x) = P(\max\{X_1, \dots, X_n\} \leq x) = P\left(\bigcap_{i=1}^n [X_i \leq x]\right) = F^n(x). \quad (1)$$

Označme $x^* := \sup\{x : F(x) < 1\}$ pravý koncový bod rozdělení s distribuční funkcí $F(x)$, tedy x^* je největší možná hodnota, které může náhodná veličina X nabývat.¹ Zaměřme se na limitní rozdělení výběrového maxima. Ze vztahu (1) je patrné, že pro $n \rightarrow \infty$

$$F_{M_n}(x) = F^n(x) \rightarrow G_{\text{deg}}(x),$$

kde $G_{\text{deg}}(x)$ je degenerovaná funkce v x^* , tj. $G_{\text{deg}}(x) = 0$ pro $x < x^*$ a $G_{\text{deg}}(x) = 1$ pro $x \geq x^*$. Pro $x^* < \infty$ je $G_{\text{deg}}(x)$ distribuční funkcí. V případě, že pravý koncový bod není konečný, je limitní funkce $G_{\text{deg}}(x)$ identicky rovna nule. Poslední jmenovaná situace přitom nastává v celé řadě běžně používaných rozdělení (mimo jiné v případě normálního rozdělení), což jistě nepodává o chování výběrových maxim dobrou představu. Z tohoto důvodu je stejně jako v případě centrální limitní věty potřeba výběrová maxima řádně normalizovat.

Nechť existují posloupnosti reálných konstant $a_n > 0, b_n$ ($n \in \mathbb{N}$) tak, že veličina $(M_n - b_n)/a_n$ má nedegenerované limitní rozdělení, tj.

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad (2)$$

pro každý bod spojitosti funkce $G(x)$, kde $G(x)$ je nějaká nedegenerovaná distribuční funkce. Rozdělení s distribuční funkcí $G(x)$ z podmínky (2) se nazývá *rozdělením extrémních hodnot* (zkráceně EV rozdělení). Podmínka (2) se nazývá *podmínkou atraktivity* a třída všech distribučních funkcí $F(x)$ (resp. příslušných rozdělení) vyhovující podmínce (2) se pak nazývá *obor atraktivity distribuční funkce $G(x)$* .

Podobně jako hraje při dokazování centrální limitní věty roli charakteristická funkce, má v EV teorii obdobný význam tzv. kvantilová funkce chvostu. Její zavedení se provede pomocí tzv. zleva spojitě inverzní funkce, kterou popisuje následující definice.

DEFINICE 1.1. Nechť f je neklesající reálná funkce. Pak funkci $f^{\leftarrow}$ danou vztahem

$$f^{\leftarrow}(x) = \inf\{y \in \mathbb{R} : f(y) \geq x\}. \quad (3)$$

nazýváme *zleva spojitou inverzní funkcí k funkci $f(x)$* .

¹Přirozeně v případě $x^* = \infty$ nemůže náhodná veličina X nabývat hodnoty x^* . V tomto případě máme na mysli, že realizace veličiny X může být libovolné reálné číslo.

Lze ukázat, že $(f^{\leftarrow})^{\leftarrow}(x) \leq f(x)$ (viz [41]), přičemž rovnost zde nastává v případě, kdy funkce f je rostoucí, neboť pak je $f^{\leftarrow}$ identická s f^{-1} . Zřejmě, je-li F distribuční funkcí nějaké náhodné veličiny, pak $F^{\leftarrow}$ je běžně užívanou kvantilovou funkcí. Dále bude pomocí pojmu zleva spojitě inverze zavedena kvantilová funkce chvostu.

DEFINICE 1.2. Nechť F je distribuční funkce náhodné veličiny X . Pak funkci U danou vztahem

$$U(t) = \left(\frac{1}{1 - F(t)} \right)^{\leftarrow} \quad (4)$$

definovanou pro $t > 1$, nazýváme *kvantilovou funkcí chvostu náhodné veličiny X* .

Články Fishera a Tippetta (1928, [13]) a Gnedenka (1943, [17]) byly jedny z prvních, které se zabývaly rozdělením výběrových maxim. Historicky byly postupně odvozeny tři třídy možných nedegenerovaných limitních rozdělení splňující rovnost (2). Tyto třídy jsou známé jako Gumbelova, Fréchetova a Weibullova třída. Následující tvrzení, také označované jako *Základní věta teorie extrémních hodnot*, popisuje třídu EV rozdělení obecněji pomocí parametrizace odvozené nezávisle von Misesem (1954, [47]) a Jenkinsonem (1955, [31]).

VĚTA 1.3 (FISHER A TIPPETT (1928), GNEDENKO (1943)). Nechť je splněna podmínka (2) pro nějaké reálné konstanty $a_n > 0, b_n$ ($n \in \mathbb{N}$). Pak distribuční funkce EV rozdělení je ve tvaru $G_\xi(\sigma x + \mu)$ pro $\mu \in \mathbb{R}, \sigma > 0$ a $\xi \in \mathbb{R}$, kde

$$G_\xi(x) = \begin{cases} \exp\left[-(1 + \xi x)^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-e^{-x}\right], & \xi = 0, \end{cases} \quad (5)$$

definovaná na $\{x : 1 + \xi x > 0\}$.

Vzhledem k von Misesově parametrizaci shrnující všechny třídy EV rozdělení, je rozdělení s distribuční funkcí (5) označováno jako *zobecněné EV rozdělení* (zkráceně GEV rozdělení z angl. Generalized Extreme Value). Vyjma parametru polohy μ a parametru měřítka $\sigma > 0$ tvoří tedy GEV rozdělení jednoparametrickou třídu rozdělení s parametrem tvaru ξ .

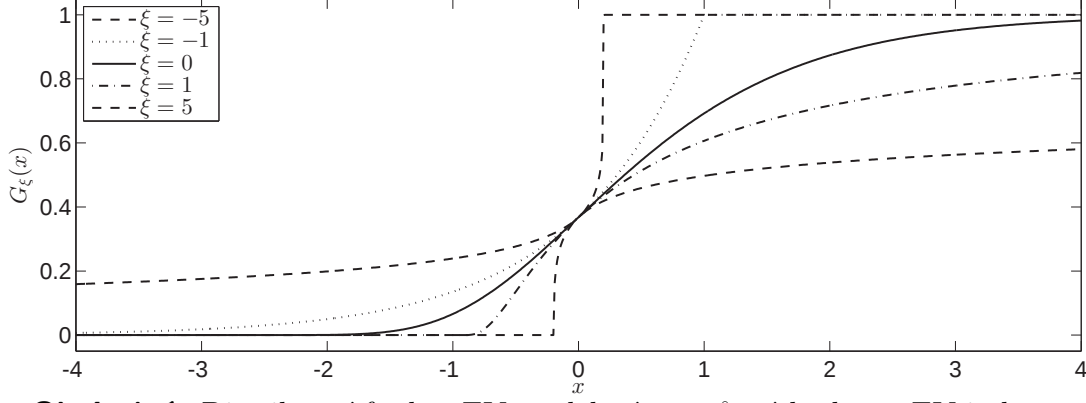
Parametr ξ má vzhledem k (5) zásadní vliv na chvosty EV rozdělení, jak je naznačeno níže na obrázku 1, a je tak často označován jako tzv. *index extrémní hodnoty* (zkráceně EV index). Dále budeme označením $F \in \mathcal{D}(G_\xi)$ rozumět, že distribuční funkce F patří do oboru atraktivity EV rozdělení s EV indexem ξ . Snadno se ukáže, že obor atraktivity je dán jednoznačně.

PODMÍNKA DRUHÉHO ŘÁDU

Jak již bylo naznačeno, při dokazování hraje v EV teorii zásadní roli kvantilová funkce chvostu $U(t)$. Obvykle se proto využívá podmínka atraktivity (2) v jiném tvaru. Nechť existuje kladná funkce $a(t)$ taková, že pro $x > 0$ platí

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\xi(x) := \frac{x^\xi - 1}{\xi}. \quad (6)$$

Pak je tato podmínka ekvivalentní podmínce (2). Přitom pro $\xi = 0$ chápeme pravou stranu (6) jako $D_\xi(x) = \ln x$. Funkce $a(t)$ z podmínky (6) je často označována jako tzv. *pomocná funkce prvního řádu*.



Obrázek 1: Distribuční funkce EV rozdělení pro různé hodnoty EV indexu.

Často je nutné charakterizovat konvergenci ve vztahu (2), resp. (6). Proto je zaváděna tzv. podmínka druhého řádu. Splnění této podmínky je obvykle nutným požadavkem pro existenci asymptotických rozdělení semiparametrických odhadů parametrů EV rozdělení popsaných v dalším odstavci.

DEFINICE 1.4. Řekneme, že kvantilová funkce chvostu $U(t)$ (resp. příslušné rozdělení pravděpodobnosti) splňuje podmínku druhého řádu, pokud existuje nějaká kladná funkce $a(t)$ a kladná či záporná funkce $A(t)$ taková, že $\lim_{t \rightarrow \infty} A(t) = 0$ a pro $x > 0$ platí

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - D_\xi(x)}{A(t)} = H_{\xi, \rho}(x) := \frac{1}{\rho} \left(\frac{x^{\xi+\rho} - 1}{\xi + \rho} - \frac{x^\xi - 1}{\xi} \right). \quad (7)$$

Parametr $\rho \leq 0$ se nazývá *parametr druhého řádu* a funkce $A(t)$ se pak označuje jako tzv. *pomocná funkce druhého řádu*.

V případě, že $\xi = 0$ nebo $\rho = 0$, chápeme funkci $H_{\xi, \rho}(x)$ ve tvaru, který se obdrží jejím limitním přechodem pro $\xi \rightarrow 0$, resp. $\rho \rightarrow 0$. Zejména je třeba upozornit na skutečnost, že funkce $H_{\xi, \rho}(x)$ není násobkem funkce $D_\xi(x)$. Funkce $a(t)$ z podmínky (7) přitom může (ale také nemusí) být totožná s funkcí $a(t)$ z podmínky atraktivity (6).

Konkrétní hodnoty parametru druhého řádu ρ a tvary pomocných funkcí $a(t)$, $A(t)$ podmínky prvního a druhého řádu jsou v dizertační práci pro celou řadu rozdělení detailně odvozeny.

2. ODHADY PARAMETRŮ ROZDĚLENÍ EXTRÉMNÍCH HODNOT

V této části budou popsány vybrané metody odhadu parametrů EV rozdělení. Nejdříve budou představeny základní přístupy vyhodnocení dat, a sice tzv. model blokových maxim a prahový model.

Model blokových maxim je postaven na základním tvrzení EV teorie (věta 1.3). Nechť je dán náhodný výběr $X_1, \dots, X_k$ rozsahu $k = mn$ s neznámou distribuční funkcí $F \in \mathcal{D}(G_\xi)$. Daný náhodný výběr se nejprve rozdělí na m bloků délky n , z nichž se posléze vyberou jen bloková maxima $M_{n,i}, i = 1, \dots, m$, tedy

$$M_{n,i} = \max\{X_{(i-1)n+1}, \dots, X_{in}\}, \quad i = 1, \dots, m.$$

Pro dostatečně velkou délku jednotlivých bloků se platnost tvrzení 1.3 chápe asymptoticky, a předpokládá se tedy, že veličiny $M_{n,i}, \dots, M_{n,m}$ tvoří náhodný výběr z GEV rozdělení $G_\xi(x)$. Zde je pak nutné uvažovat GEV rozdělení včetně parametrů polohy $\mu \in \mathbb{R}$ a měřítka $\sigma > 0$ s distribuční funkcí ve tvaru

$$G(x) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & \text{pro } \xi \neq 0 \text{ a } 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0, \\ \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, & \text{pro } \xi = 0, \end{cases} \quad (8)$$

kde EV index $\xi \in \mathbb{R}$ je parametr tvaru. Odhady parametrů μ, σ, ξ je možné získat například užitím metody maximální věrohodnosti (MV metody).

Metodě blokových maxim je často vytýkáno (viz např. [36]), že zanedbává případná jiná velká pozorování, která ovšem nedosahují hodnot příslušných blokových maxim. Nadále se tak zaměříme výhradně na použití prahového modelu popsaného níže.

Prahový model, též POT model (z angl. Peaks-Over-Threshold), je založen na asymptotickém chování náhodné veličiny X překračující nějakou prahovou hodnotu u . Lze dokázat, že platí

$$\lim_{u \uparrow x^*} P(X - u > x \mid X > u) = \left(1 + \xi \frac{x}{\sigma_u}\right)^{-1/\xi} =: 1 - H(x), \quad (9)$$

kde $H(x)$ je distribuční funkce zobecněného Paretova rozdělení (zkráceně GP rozdělení z angl. Generalized Pareto) s parametrem tvaru $\xi \in \mathbb{R}$ a parametrem měřítka $\sigma_u > 0$ definovaná pro $\{x : x > 0, 1 + \xi x/\sigma_u > 0\}$. Tedy pro dostatečně velkou prahovou hodnotu u lze podmíněné rozdělení náhodné veličiny $Y = X - u$ za podmínky $X > u$ aproximovat GP rozdělením, přičemž parametr tvaru $\xi \in \mathbb{R}$ GP rozdělení odpovídá parametru tvaru $\xi \in \mathbb{R}$ příslušného limitního GEV rozdělení výběrových maxim a $\sigma_u = \sigma + \xi(u - \mu)$ [4].

Oproti metodě blokových maxim neopomíjí prahový model ostatní extrémní pozorování, která nedosahují velikosti blokových maxim, a zpravidla tak dovoluje zahrnout do analýzy soubor většího rozsahu. Na druhou stranu vhodná volba prahu je kritickou částí celého modelu. Nízká volba prahu u vede ke špatné aproximaci limitním vztahem (9) a parametry modelu jsou odhadnuty se značným vychýlením. Vysoká volba prahu u vede k malému počtu překročení a k nárůstu variability odhadů. Volba optimální prahové hodnoty, vyvažující vychýlení a variabilitu, však stále patří k nedořešeným otázkám uvedeného modelu. Techniky výběru vhodného prahu budou popsány v odstavci 3.

Jakmile je zvolena prahová hodnota, lze použít aproximaci nadprahových pozorování GP rozdělením a přistoupit ke stanovení odhadů jeho parametrů a dalších parametrických funkcí. Vzhledem k aplikační části této práce budou popsány odhady návratové úrovně z_r příslušné periodě návratu r . Návratová úroveň z_r je taková hodnota, která je v průměru překročena jednou za r pozorování, tj. jedná se o $(1 - r^{-1})$ kvantil modelového rozdělení. Při užití prahového modelu lze z_r vyjádřit ve tvaru

$$z_r = \begin{cases} u + \frac{\sigma_u}{\xi} \left[(\lambda_u r)^\xi - 1 \right], & \text{pro } \xi \neq 0, \\ u + \sigma_u \ln(\lambda_u r), & \text{pro } \xi = 0, \end{cases} \quad (10)$$

kde $\lambda_u := P(X > u)$. Často se vyžaduje určení tzv. N -leté návratové úrovně. Pak stačí jednoduše položit $r = n_y N$, kde n_y je průměrný počet pozorování za jeden rok.

Odhad $\hat{z}_r$ hodnoty z_r se dostane nahrazením jednotlivých parametrů v (10) jejich odhady. Níže budou popsány odhady získané MV metodou a vybrané semiparametrické

odhady. Odhad parametru λ_u se obdrží jako relativní četnost pozorování nad prahem u . Protože počet překročení prahu u je náhodná veličina s binomickým rozdělením $\text{Bi}(n, \lambda_u)$, je takto získaný odhad $\hat{\lambda}_u$ MV odhadem parametru λ_u . Na základě asymptotických vlastností je možné dále určit odhad rozptylu statistiky z_r , jak je blíže popsáno v dizertační práci.

ODHADY METODOU MAXIMÁLNÍ VĚROHODNOSTI

Odhady parametrů v blokovém i prahovém modelu bývají často určeny pomocí standardních postupů, zejména užitím MV metody (viz [3] nebo [8]). MV odhady parametrů GEV rozdělení (resp. GP rozdělení) lze však získat jen s jistými omezeními, protože třída GEV rozdělení není obecně regulární a jsou tak porušeny základní předpoklady MV teorie. Jak je ukázáno v [45], konvergence MV odhadů závisí na hodnotě EV indexu. Pokud $\xi > -1/2$, mají odhady získané MV metodou obvyklé asymptotické vlastnosti, tj. asymptotické rozdělení odhadů je normální. Pro $\xi = -1/2$ byly MV odhady odvozeny v [44] pro semiparametrický model. V případě, že $-1 < \xi < -1/2$, lze odhady získat běžným způsobem, ovšem tyto již nemají obvyklé asymptotické vlastnosti [39]; přesto však, jak je ukázáno v [49], pro $\xi > -1$ představují odhady získané MV metodou konzistentní odhady. Nakonec pro $\xi \leq -1$, kdy mají rozdělení velmi krátké pravé chvosty, je málo pravděpodobné, že budou MV odhady dosaženy s dostatečnou přesností.

Dále budou popsány jen odhady parametrů v prahovém modelu, MV odhady parametrů modelu blokových maxim jsou rozebrány v dizertační práci. Nechť je dán náhodný výběr $X_1, \dots, X_n$ z GP rozdělení s distribuční funkcí $H(x)$ ze vztahu (9) a předpokládejme, že byly získány realizace $X_1 = x_1, \dots, X_n = x_n$. Logaritmická věrohodnostní funkce GP rozdělení je ve tvaru

$$l(\sigma_u, \xi) = \begin{cases} -n \ln \sigma_u - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln\left(1 + \xi \frac{x_i}{\sigma_u}\right), & \text{pro } \xi \neq 0, \\ -n \ln \sigma_u - \frac{1}{\sigma_u} \sum_{i=1}^n x_i, & \text{pro } \xi = 0. \end{cases} \quad (11)$$

Maximalizací výrazu (11), případně řešením systému věrohodnostních rovnic, obdržíme MV odhady příslušných parametrů. Protože systém věrohodnostních rovnic nelze řešit analyticky, je nutné použít vhodné numerické metody, např. proceduru založenou na Nelderově-Meadově simplexovém algoritmu [33] implementovanou v řadě softwarů. MV odhad návratové úrovně z_r se obdrží nahrazením parametrů ξ, σ_u, λ_u ve vztahu (10) jejich MV odhady.

SEMIPARAMETRICKÉ PŘÍSTUPY K ODHADŮM PARAMETRŮ ROZDĚLENÍ EXTRÉMNÍCH HODNOT

V poslední době se pozornost zaměřila především na neparametrické přístupy odhadu parametrů EV rozdělení. Řada autorů se zabývala odhady parametrů metodou pravděpodobnostně vážených momentů [6, 7, 11] nebo metodou L-momentů [1, 29]. Zde bude pozornost věnována semiparametrickým přístupům odhadu EV indexu, a sice momentovému odhadu v prahovém modelu. Mějme náhodný výběr $X_1, \dots, X_n$ a označme $X_{(1)}, \dots, X_{(n)}$ uspořádaný náhodný výběr, kde $X_{(1)} \leq \dots \leq X_{(n)}$ a $X_{(i)}$ označuje i -tou pořadovou statistiku. Dosud bylo popsáno asymptotické chování výběrového maxima $M_n = X_{(n)}$ pro

$n \rightarrow \infty$. Obecně je možné popsat limitní vlastnosti pořadové statistiky $X_{(n-k)}$ pro libovolné pevné k [10]. Lze však také uvažovat pořadové statistiky při volbě $k = k(n) \rightarrow \infty$ pro $n \rightarrow \infty$. Přidáme-li navíc podmínku $k(n)/n \rightarrow 0$, pak při takové volbě k lze řádnou normalizací obdržet asymptoticky normálně rozdělené statistiky, které mohou sloužit k odhadům parametrů rozdělení v prahovém modelu.

Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení s distribuční funkcí $F \in \mathcal{D}(G_\xi)$, $\xi \in \mathbb{R}$ a $x^* > 0$. Momentový odhad $\hat{\xi}_M(k)$ EV indexu ξ je ve tvaru

$$\hat{\xi}_M(k) := M_n^{(1)}(k) + 1 - \frac{1}{2} \left(1 - \frac{\left(M_n^{(1)}(k) \right)^2}{M_n^{(2)}(k)} \right)^{-1}, \quad (12)$$

kde $M_n^{(j)}(k) := \frac{1}{k} \sum_{i=0}^{k-1} (\ln X_{(n-i)} - \ln X_{(n-k)})^j$.

Za předpokladů splnění podmínky druhého řádu lze u semiparametrických odhadů $\hat{\xi}(k)$ EV indexu ukázat, že tyto odhady mají asymptoticky normální rozdělení, tedy lze psát

$$\sqrt{k}(\hat{\xi}(k) - \xi) \xrightarrow{d} \sqrt{\text{var}_\xi} N + \lambda b_{\xi, \rho}, \quad (13)$$

kde $N \sim N(0, 1)$, var_ξ je složka variability závisící na ξ a $\lambda b_{\xi, \rho}$ je složka vychýlení závisící na ξ a parametru druhého řádu $\rho \leq 0$. Detailní popis je možné nalézt přímo v dizertační práci, popř. v [10]. Tedy limitním rozdělením $\hat{\xi}_M$ je normální rozdělení s možným nenulovým vychýlením, které závisí na původním rozdělení F skrze parametry ξ a ρ . Při aproximaci vztahu (13) pro nějaké konečné n je však důležité vhodně určit hodnotu k a tím i podíl n/k . Volba vhodného k přitom odpovídá volbě vhodného prahu u v (parametrickém) prahovém modelu s tím rozdílem, že nyní je jako práh uvažována hodnota pořadové statistiky $X_{(n-k)}$. Tedy s rostoucím k varianční složka var_ξ klesá a složka vychýlení $\lambda b_{\xi, \rho}$ roste.

Na základě aproximace pomocí vztahu (6) se získá odhad návratové úrovně z_r ve tvaru

$$z_r = U(r) \approx U\left(\frac{n}{k}\right) + \frac{a\left(\frac{n}{k}\right)}{\xi} \left[\left(\frac{kr}{n}\right)^\xi - 1 \right],$$

kde $\hat{U}(n/k) = X_{(n-k)}$ je odhad hodnoty $U(n/k)$ a pravděpodobnost překročení prahu $\lambda_u := P(X > U(n/k))$ byla odhadnuta jako relativní četnost pozorování nad prahem $X_{(n-k)}$, tedy $\hat{\lambda}_u = k/n$, což odpovídá parametrickému modelu. Odhad členu $a(n/k)$ a další detaily jsou popsány v dizertační práci. Pomocí odhadu varianční matice těchto odhadů je dále možné určit rozptyl statistiky z_r a případně její interval spolehlivosti.

3. METODY VÝBĚRU PRAHOVÝCH HODNOT

Vhodná volba prahové hodnoty je stále nedořešeným problémem prahového modelu. Některé starší i současné metody jsou shrnuty v publikaci [43], v praktických situacích přitom stále bývá volba prahu prováděna na základě explorativních technik a zkušeností v daném oboru. Nejčastěji bývají užívány dvě grafické metody spočívající ve volbě vhodného prahu u_0 tak, aby bylo dosaženo určité stability empirické střední hodnoty (Mean

Residual Life plot) či odhadů parametrů ξ, σ_u . Detailně jsou grafické metody představeny v dizertační práci. V současnosti však roste zájem o metody adaptivního určování prahu, které nabízí automatizaci volby prahu a o vhodnosti prahu se snaží rozhodnout na základě vyvinutých statistických kritérií. Tyto techniky jsou zpravidla výpočetně náročnější.

ADAPTIVNÍ VOLBA PRAHU V SEMIPARAMETRICKÉM MODELU

Významnou třídu adaptivních metod tvoří dvojité bootstrapová technika. Nechť je dán náhodný výběr z rozdělení, pro kterou je pomocná funkce druhého řádu (7) ve tvaru $A(t) = ct^\rho$ pro $c \neq 0, \rho < 0$. Nejčastěji bývá uvažována tzv. Hallova třída rozdělení [12, 21], tedy rozdělení s kvantilovou funkcí chvostu $U(t) = Ct^\xi(1 + Dt^\rho + o(t^\rho))$, kde $C > 0$.

Uvažuje-li se semiparametrický momentový odhad $\widehat{\xi}_M(k)$ EV indexu, je pak optimální počet $k_0 = k_0(n)$ horních pořadových statistik vybrán tak, aby byla minimalizována střední kvadratická chyba odhadu, tj. $k_0 \in \operatorname{argmin}_k E\left(\widehat{\xi}_M(k) - \xi\right)^2$. Za předpokladu Hallovy třídy rozdělení a vyjádření odhadu $\widehat{\xi}_M(k)$ ve tvaru (13) lze odvodit rovnost

$$k_0(n) = \left\lfloor \left(\frac{\operatorname{var}_\xi}{-2\rho c^2 b_{\xi,\rho}^2} \right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)} \right\rfloor. \quad (14)$$

Zvolený odhad hodnoty k_0 závisí, mimo jiné, na parametru druhého řádu ρ . Obtíže spojené s odhadem ρ jsou v literatuře popsány [19, 20], často je pak nutné splnění podmínky třetího řádu [14]. Dvojité bootstrapová technika přitom umožňuje se, vhodnou kombinací odhadů, zcela vyhnout odhadu parametru ρ .

Nejprve se neznámý EV index nahradí jiným vhodným odhadem $\widehat{\xi}_{\text{aux}}$ (více viz dizertační práci nebo [12, 21]). Následně je z empirické distribuční funkce $F_n(x)$ B -krát nezávisle bootstrapován náhodný výběr rozsahu $n_1 < n$, přičemž pro každý takový výběr jsou pro $k = 2, \dots, n_1$ určeny bootstrapové odhady $\widehat{\xi}^*(k, b), \widehat{\xi}_{\text{aux}}^*(k, b)$ a dále

$$q_{n_1,b}^*(k) = \left(\widehat{\xi}^*(k, b) - \widehat{\xi}_{\text{aux}}^*(k, b) \right)^2, \quad \text{kde } b = 1, \dots, B.$$

Bootstrapový odhad asymptotické střední kvadratické chyby (AMSE)

$$\widehat{\text{AMSE}}^*(n_1, k) = \frac{1}{B} \sum_{b=1}^B q_{n_1,b}^*(k)$$

je minimalizován vzhledem ke k . Nechť je minima dosaženo v bodě $k_0^*(n_1)$. Celý postup je pak opakován pro rozsah bootstrapového výběru $n_2 = \lceil (n_1)^2/n \rceil$ a je získána hodnota $k_0^*(n_2)$. Dvojité bootstrapový odhad optimálního k_0 se následně získá položením (důkaz pro $\widehat{\xi}_M(k)$ viz [12])

$$\widehat{k}_0^* = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)}. \quad (15)$$

ADAPTIVNÍ VOLBA PRAHU ZALOŽENÁ NA VĚROHODNOSTI

Jiný přístup, tzv. *Multiple-Threshold Generalized Pareto model* (zkráceně MT-GP model), byl rozvinut v práci [38]. Tento v principu vychází z grafické metody určení prahu pomocí stability odhadů EV indexu. Zavedením diskretizace pomocí m prahových hodnot $u_1 < \dots < u_m$ je EV index ξ modelován jako po částech konstantní funkce se změnami v bodech $u_i, i = 2, \dots, m$.

Nechť je dána náhodná veličina X a buď $Y = X - u_1$ překročení prahu u_1 . Označme $v_i = u_i - u_1$ pro $i = 1, \dots, m$ a $w_i = v_{i+1} - v_i$ pro $i = 1, \dots, m - 1$. Nejprve se provede odvození hustoty $f_Y(y)$ veličiny Y . Aby nedocházelo k nespojitostem v $f_Y(y)$, je vhodné uvažovat parametry měřítka ve tvaru $\sigma_{i+1} = \sigma_i + \xi_i w_i$, kde $\sigma_1 > 0$. Původní model s $2m$ parametry je tak možné parametrizovat jen pomocí $m + 1$ parametrů $\theta = (\sigma_1, \xi_1, \dots, \xi_m)$.

Buď $p_i := P(Y > v_i)$. Je-li dán náhodný výběr $Y_1, \dots, Y_n$ překročení prahu u_1 z hustoty $f_Y(y)$, pak lze logaritmickou věrohodnostní funkci zapsat ve tvaru (viz [38])

$$l(\sigma_1, \xi_1, \dots, \xi_m) = \sum_{i=1}^n \sum_{j=1}^m \mathbf{1}_{[v_j < y_i < v_{j+1}]} \left\{ \ln p_j - \ln \sigma_j - \left(1 + \frac{1}{\xi_j} \right) \ln \left[1 + \xi_j \frac{y_i - v_j}{\sigma_j} \right] \right\}. \quad (16)$$

Celý model včetně hustoty $f_Y(y)$ je v dizertační práci detailně odvozen.

Uvažujme nyní situaci, kdy chceme ověřit, zda je parametr tvaru ξ identický na všech intervalech $(v_i, v_{i+1}), i = 1, \dots, m$, tj. testujeme nulovou hypotézu $H : \xi_1 = \dots = \xi_m$. Zamítnutí této hypotézy by ukazovalo, že pro $u \geq u_1$ nebylo dosaženo uspokojivé stability EV indexu a bylo by tak nutné uvažovat jako vhodný práh nějakou hodnotu větší než u_1 . Označme dále $\hat{\theta}$ MV odhad parametru θ získaný z (16) a $\hat{\theta}_0$ MV odhad získaný za platnosti nulové hypotézy, tj. za platnosti $\xi_1 = \dots = \xi_m$. Nulová hypotéza se testuje na základě statistik

$$LM(\hat{\theta}_0) = \frac{1}{n} \mathbf{U}^T(\hat{\theta}_0) \mathbf{J}^{-1}(\hat{\theta}_0) \mathbf{U}(\hat{\theta}_0), \quad (17)$$

$$LR(\hat{\theta}_0) = 2 \left(l(\hat{\theta}) - l(\hat{\theta}_0) \right), \quad (18)$$

kde $\mathbf{U}(\hat{\theta}_0)$ je skórový vektor příslušný hustotě $f_Y(y)$ a $\mathbf{J}(\hat{\theta}_0)$ je očekávaná Fisherova informační matice. Za předpokladu, že $\xi_m > -1/2$ a tedy MV odhady mají obvyklé asymptotické vlastnosti [45], mají uvedené statistiky asymptoticky χ_{m-1}^2 rozdělení. (Jedná se o asymptotické testy s rušivými parametry popsané viz např. [3].) Odvození tvarů statistik $U(\hat{\theta}_0)$ a $\mathbf{J}(\hat{\theta}_0)$ lze nalézt v dodatku dizertační práce.

V případě zamítnutí nulové hypotézy $H : \xi_1 = \dots, \xi_m$ se přistupuje k postupnému testování stability EV indexu pro prahové hodnoty $(u_i, \dots, u_m)$, dokud použité kritérium nevede pro nějaké $i = 2, \dots, m - 1$ k jejímu přijetí. Při nulové hypotéze $H : \xi_i = \dots = \xi_m$ je tak asymptotické rozdělení statistik (17) a (18) χ_{m-i}^2 .

4. ANALÝZA EXTRÉMNÍCH SRÁŽKOVÝCH ÚHRNŮ

Poznatky z předešlých odstavců byly aplikovány k odhadu extrémních srážkových úhrnů v jihomoravském regionu. Tato analýza byla motivována nutností aktualizace hydrologických podkladů pro potřeby plánování a údržby městského odvodňovacího systému. Důraz je kladen na odhady tzv. *IDF křivek* (zkráceně z angl. *Intensity-Duration-Frequency*), které představují základní hydrologický nástroj pro vyhodnocení frekvence výskytu intenzivních dešťových srážek. Dodnes u nás používané odhady IDF křivek (viz [46]) se jeví jako zastaralé a nově vyvinuté statistické metody by tak mohly přispět k jejich zpřesnění.

SIMULAČNÍ STUDIE PRO POROVNÁNÍ ADAPTIVNÍCH METOD URČENÍ PRAHU

Pro porovnání adaptivních metod určení prahu byla provedena simulační studie [25]. Opakovaně byly generovány výběry s rozsahy $n = 500, 1000, 2000, 3000, 5000$ a 8000 z rozdělení, která svou povahou nejlépe odpovídají analyzovaným srážkovým měřením. Konkrétně bylo uvažováno Fréchetovo, GP a loglogistické rozdělení, patřící do Hallovy třídy rozdělení, a dále gama, Weibullovo a lognormální rozdělení.

Dvojitě bootstrapová technika byla uvažována v kombinaci s momentovým odhadem (12). Počet bootstrapových odhadů byl vždy nastaven na $B = 250$. Ze studie [21] vyplývá, že rozsah bootstrapových výběrů n_1 je vhodné volit co největší. Zde bylo uvažováno $n_1 = \lceil n^{0,995} \rceil$, kde n je délka dané řady. Pro stabilizaci odhadu $\hat{k}_0$ optimálního počtu horních pořadových statistik byla celá procedura vždy $1000 \times$ opakována, přičemž konečný odhad $\hat{k}_0$ byl stanoven jako medián takto získaných hodnot.

Metodou MT-GP byl vhodný práh u hledán užitím kritéria (17). Optimální hodnota u_0 byla uvažována mezi 0,5 % a 90 % empirickým kvantilem. Vzhledem k otevřeným závěrům v práci [38] byla zvolena diskretizace pomocí postupně $m = 10, 20$ a 40 prahových hodnot mezi uvedenými kvantily. Hodnota $u_0(m)$ byla pro každé m zvolena jako nejmenší práh, pro který nebyla testovaná hypotéza zamítnuta na hladině významnosti 0,05.

Závěry simulační studie [25] ukazují, že použití bootstrapové techniky vždy vede k menšímu rozsahu výběru uvažovaného POT modelem. Jistá nestabilita MT-GP byla pozorována pro loglogistické, Fréchetovo a lognormální rozdělení. Na druhou stranu tato technika vhodně identifikuje GP rozdělení a práh zde určuje velmi nízko. Oba přístupy byly porovnány vzhledem k odhadům EV indexu. Zde vyplývá, že v rámci Hallovy třídy rozdělení jsou obě metody srovnatelné. Mimo Hallovu třídu se pak ukazují odhady EV indexu méně vychýlené při užití bootstrapové metody, ačkoliv tato zde nemá jasné teoretické opodstatnění.

ODHADY EXTRÉMNÍCH DEŠŤOVÝCH UDÁLOSTÍ

Jako vstupní data pro analýzu slouží záznamy reálných dešťových srážek (s rozlišením 1 minuty) ze 6 stanic situovaných na území Jihomoravského kraje. Sledovaná období spolu s délkami úplných záznamů jsou shrnuta v tabulce 1.

Teorie extrémních hodnot je při posuzování hydrologických jevů často aplikována. Z měřených časových řad bývá však nejprve nutné vybrat taková pozorování, která můžeme pokládat za nezávislá. Jak uvádí např. práce [30], v zásadě se rozlišují dva

Tabulka 1: Přehled srážkoměrných stanic, jejich sledovaného období a délky řad s úplnými záznamy.

Stanice	Sledované období	Délka řady [roky]
Brno-Tuřany	1959–2000	41
Brno-Žabovřesky	1987–2003	16
Brno-Jundrov	1992–2003	11
Znojmo-Kuchařovice	1956–2003	27
Jevišovice	1961–2000	37
Vyškov	1961–1992	31

základní způsoby výběru prvků pro statistické zpracování. První je postaven na metodě blokových maxim, kdy jsou do statistického souboru zahrnuty jen největší prvky v daném bloku pozorování. V hydrologii bývají většinou uvažovány bloky o délce jednoho roku a získaná řada je pak označována jako *Annual Maxima Series*. Nevýhodou je ovšem fakt, že jsou zanedbávány případné další vysoké hodnoty, které ovšem nedosahují příslušných blokových maxim.

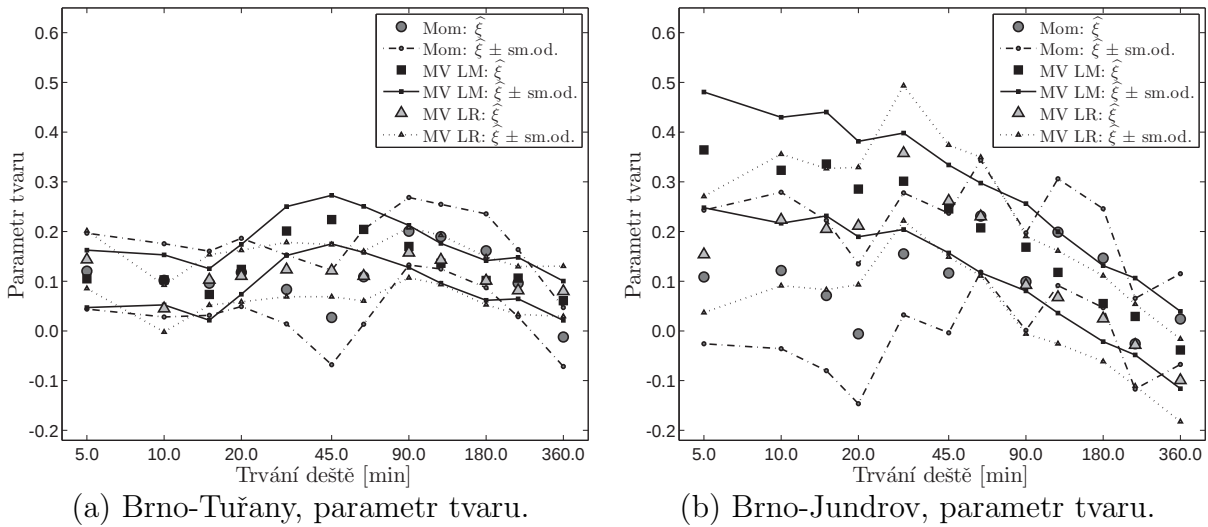
Jiný způsob výběru prvků je založen na prahovém POT modelu, kdy jsou do statistického souboru zahrnuty hodnoty překračující nějakou vysokou prahovou hodnotu. V současnosti pak POT model představuje stále více používanou techniku [5, 32, 48]. Vzhledem k délce dostupných dat bude tento přístup upřednostněn. Při analýze srážkových extrémů pomocí POT modelu se postupuje tak, že se z daných časových řad separují tzv. *dešťové události*. Tyto jsou mezi sebou uvažovány jako přibližně nezávislé. Aplikace různých metodik separace dešťových událostí závisí na konkrétní aplikační oblasti. Jak naznačuje práce [37], právě přístup separace uvedený v [35] se ukazuje jako nejvhodnější pro potřeby návrhu městského odvodnění v brněnském regionu. Následně se z každé události pro předem stanovené trvání deště vyberou největší pozorované srážkové úhrny. Podle metodiky [35] tak byly vyčleněny srážkové úhrny za dobu trvání deště 5, 10, 15, 20, 30, 45, 60, 90, 120, 180, 240 a 360 minut.

Ve spojitosti s analýzou extrémních hodnot byla provedena řada počítačových implementací (viz např. [16]), které jsou také součástí dizertační práce. Jedná se např. také o zde diskutované techniky vzorkování.

Při analýze byla pozornost zaměřena zejména na rozsah datových souborů. Srovnání technik volby prahu a odhadu IDF křivek je tak možné porovnat také vzhledem k délce dostupných časových řad. Vhodná volba prahu u_0 , resp. počtu nadprahových pozorování k_0 , byla stanovena grafickými i adaptivními metodami v kombinaci s MV a semiparametrickými odhady. Správnost zvolených prahů byla vždy ověřena vizuálně pomocí histogramů a Q-Q plotů a také pomocí testů dobré shody (Pearsonova χ^2 testu, Kolmogorovova-Smirnovova testu a Andersonova-Darlingova testu).

Ukázalo se, že GP rozdělení velmi dobře odpovídá empirickému rozdělení nadprahových hodnot ve všech studovaných případech, s výjimkou užití MT-GP metody založené na kritériu (18). Vzhledem k diskuzi automatizace postupu však nebyly tyto prahy nijak korigovány. Bylo zjištěno, že pro delší dostupné řady dvojité bootstrapová technika vybírá vždy nejmenší počet pozorování. Pro rostoucí délku deště jsou pak výběry podobné jako užitím grafických metod. Nejnížší prahová hodnota byla obecně stanovena MT-GP technikou. Zejména při užití statistiky (18) pak byly získány (patrně nevhodně) rozsáhlé nadprahové soubory. Více je k nalezení v publikaci [24].

Následně byly pro určené prahy stanoveny odhady parametrů GP rozdělení – semi-parametrické pro bootstrapovou volbu k_0 a MV odhady jinak. Odhady EV indexu jsou zobrazeny na obrázku 2. Jak bylo očekáváno, pro krátké řady, včetně té ze stanice Brno-Jundrov, jsou odhady zatíženy větší variabilitou. Při porovnání s momentovými odhady vykazují MV odhady obecně (ať už při užití MT-GP nebo grafických metod) menší variabilitu pro delší časové řady. Pro krátké řady pak dochází u MV odhadů k nárůstu variability a momentové odhady se zde ukazují jako vhodnější. Na druhou stranu je patrné vzájemné vychýlení všech odhadů, a to i v rámci MV metody v kombinaci s různými technikami určení prahu. Teoretické aspekty by měly vést k závěrům, že při vyšší hodnotě prahu dochází k menšímu vychýlení odhadů, avšak vzhledem ke značné variabilitě u krátkých časových řad není možné takové závěry jasně stanovit.



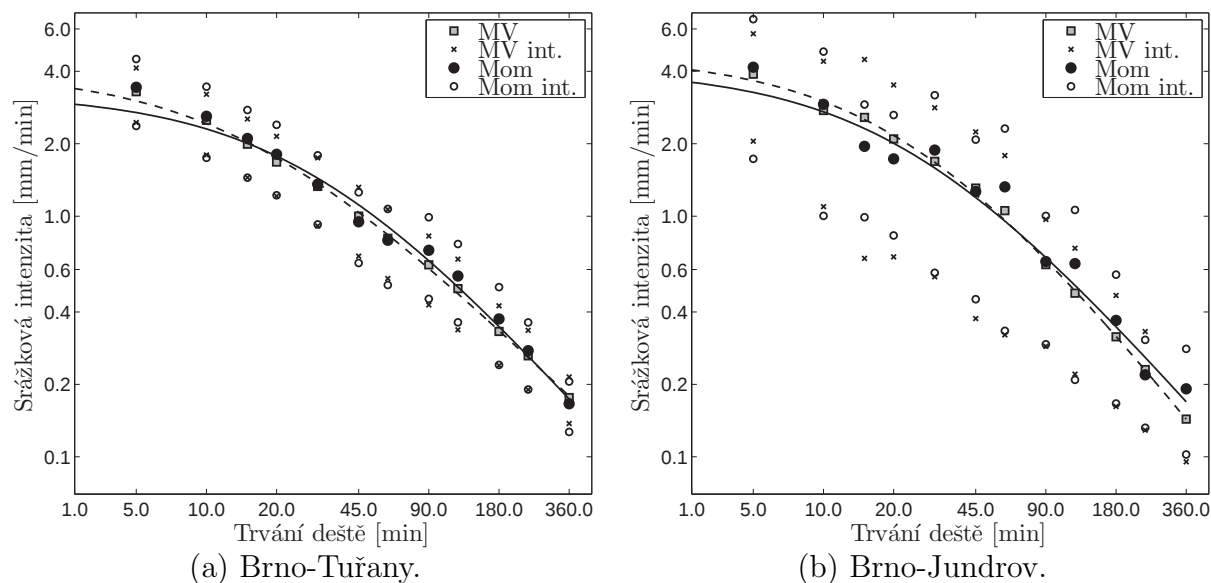
Obrázek 2: Parametr tvaru získaný semiparametrickými momentovými odhady (*Mom*) a MV metodou při volbě prahu pomocí přístupu MT-GP při použití testovacích statistik LM a LR pro stanice Brno-Tuřany a Brno-Jundrov. Směrodatné odchylky získány na základě asymptotické normality. Horizontální osa je zobrazena v logaritmickém měřítku.

Pomocí odhadů parametrů byly dále určeny odhady IDF křivek jakožto návratové úrovně srážkových intenzit. Intenzita je zde určena jako podíl srážkových úhrnů za danou dobu trvání deště. V hydrologické praxi bývá zvykem [30] získané návratové úrovně prokládat regresní křivkou

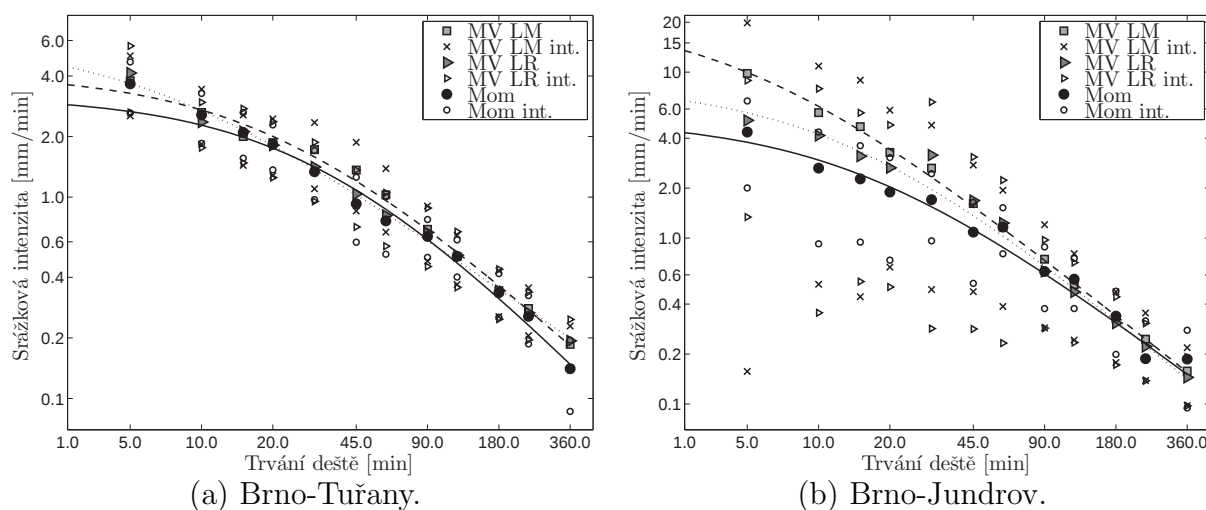
$$I(t) = \frac{\beta_1}{(1 + \beta_2 t)^{\beta_3}},$$

kde $I(t)$ je odhadnutá intenzita, t je doba trvání deště a $\beta_1, \beta_2, \beta_3$ jsou regresní koeficienty. Na obrázcích 3 a 4 je zobrazeno srovnání odhadů IDF křivek pro stanice Brno-Tuřany a Brno-Jundrov získaných jako 100leté návratové úrovně, přičemž tyto byly pomocí nelineární regrese proloženy výše uvedenou křivkou. Vyznačené 95% intervaly spolehlivosti byly určeny na základě asymptotické normality odhadů. Pro delší časové řady lze všechny odhady z praktického hlediska považovat za totožné. Rozdíly mezi metodami jsou patrné pro krátké řady, zejména při odhadech dlouhodobých událostí (viz ukázaná 100letá úroveň). Srovnáním momentového a MV odhadu při grafické volbě prahu se ukazuje (obr.3), že křivky získané MV metodou vykazují menší variabilitu pro kratší doby trvání deště. Pro delší doby trvání je pak variabilita menší pro momentové odhady. Na obrázku 4 je vidět, že volba prahu pomocí MT-GP metody má za následek velké vychýlení křivek oproti momentovému odhadu pro krátké doby trvání deště. Získané MV odhady jsou zde

prakticky nepoužitelné kvůli extrémní šířce intervalů spolehlivosti. Na druhou stranu se pro dlouhé doby trvání deště (240 a 360 minut) ukazují MV odhady jako přesnější. Širší pojednání ke srovnání všech zde diskutovaných přístupů lze nalézt v autorových publikacích [24] a [26], kde byly odhady IDF křivek pro brněnský region detailně studovány.



Obrázek 3: MV odhady při grafické volbě prahu (čárkovaně) a momentové odhady (plně) IDF křivek pro 100leté návratové úrovně. 95% intervaly spolehlivosti nebyly regresní křivkou proloženy. Osy v logaritmickém měřítku.



Obrázek 4: MV odhady při volbě prahu pomocí MT-GP metody (LM čárkovaně, LR tečkovaně) a momentové odhady (plně) IDF křivek pro 100-leté návratové úrovně. 95% intervaly spolehlivosti nebyly regresní křivkou proloženy. Osy v logaritmickém měřítku.

5. TEORIE EXTRÉMNÍCH HODNOT PRO STACIONÁRNÍ ŘADY A JEJÍ POUŽITÍ

Dosud představené metody statistického vyhodnocení extrémních hodnot byly založeny na předpokladu nezávislosti jednotlivých náhodných veličin. Vzájemná nezávislost však představuje v mnoha praktických situacích zcela nerealistický požadavek a bývá tak nutné použít dodatečné techniky vzorkování. Pro zobecnění EV teorie pro případ korelovaných náhodných veličin je třeba se omezit na (striktně) stacionární řady. Navíc je nutné přidat požadavek na omezení závislosti tak, aby závislost mezi veličinami X_i, X_j byla pro rostoucí vzdálenost $|i - j|$ nějakým způsobem limitována. V literatuře je pak nejběžněji uvažována Leadbetterova $D(u_n)$ podmínka [34] typu silného mixing. Vzhledem k povaze typických aplikačních problémů ale není tento požadavek významně omezující, protože závislost mezi časově či prostorově vzdálenými pozorováními je často zanedbatelná.

Nechť $X_1, \dots, X_n$ je stacionární řada splňující uvedenou $D(u_n)$ podmínku a $X_1^*, \dots, X_n^*$ náhodný výběr se stejným jednorozměrným marginálním rozdělením jako daná stacionární řada. Označme $\text{GEV}(\mu, \sigma, \xi)$ GEV rozdělení s parametry tvaru μ , měřítka σ a polohy ξ . Podobně jako v tvrzení 1.3 je možné ukázat, že limitním rozdělením (řádně normalizované) veličiny $M_n = \max\{X_1, \dots, X_n\}$ je opět $\text{GEV}(\mu_\theta, \sigma_\theta, \xi_\theta)$ rozdělení s distribuční funkcí $G_\theta(x)$. Je-li limitní rozdělení výběrového maxima $M_n^* = \max\{X_1^*, \dots, X_n^*\}$ nějaké $\text{GEV}(\mu, \sigma, \xi)$ rozdělení s distribuční funkcí $G(x)$, pak mezi oběma distribučními funkcemi platí vztah

$$G_\theta(x) = [G(x)]^\theta, \quad (19)$$

kde $\theta \in \langle 0, 1 \rangle$ je tzv. *extremální index*. Navíc platí $\xi_\theta = \xi$.

Parametr θ vyjadřuje míru krátkodobé závislosti na extrémních úrovních a pro posloupnost nezávislých veličin je zřejmě $\theta = 1$. Vliv krátkodobé závislosti lze v realizacích stacionárních řad pozorovat ve formě shlukování vysokých hodnot, tj. extrémní hodnoty bývají obvykle následovány dalšími. Extremální index je možné interpretovat několik způsoby, nejčastěji bývá uvažován výklad podle [34], kdy θ^{-1} vyjadřuje limitní střední hodnotu velikosti těchto shluků.

V praktických situacích, kdy jsou k dispozici pozorování nějaké stacionární řady, je možné postupovat dvěma způsoby: (i) extrémální index vhodně odhadnout nebo (ii) použít nějaké filtrační schéma, které by vliv θ eliminovalo. Druhý přístup, označován jako *declusterování*, bývá v aplikacích obvykle užíván. Vybrané techniky declusterování byly představeny v odstavci 4, kde byly použity po vzorkování přibližně nezávislých dešťových srážek. Použití těchto technik má ovšem za následek (často významnou) redukci původního datového souboru. Volba potřebných pomocných parametrů je navíc zatížena značnou mírou subjektivity.

Jako lepší se tak jeví odhadnout θ pomocí nějakého vhodného odhadu $\hat{\theta}$. Protože limitním rozdělením výběrových maxim stacionární řady je GEV rozdělení, lze k analýze použít standardní metody bez ohledu na závislost. Odhady parametrických funkcí se poté korigují podle získaného odhadu θ . Výhodou toho postupu je možnost využití všech dostupných pozorování, což může výrazně snížit variabilitu odhadu. Představme krátce odhady vybraných parametrických funkcí při použití prahového modelu, konkrétně odhady návratových úrovní. Z limitních vlastností výběrového maxima stacionární řady vyplývá, že můžeme aproximovat $P(M_n \leq x) \approx G^\theta(x) \approx F^{n\theta}(x)$. Hledaný $(1 - r^{-1})$ kvantil GP

rozdělení nezávislých pozorování tak základě pozorování celé stacionární řady odpovídá $(1 - r^{-1})^{\theta^{-1}}$ kvantilu. Namísto (10) se tedy dostane návratová úroveň z_r ve tvaru

$$z_r = \begin{cases} u + \frac{\sigma_u}{\xi} \left[\left\{ \lambda_u^{-1} \left[1 - (1 - r^{-1})^{\theta^{-1}} \right] \right\}^{-\xi} - 1 \right], & \text{pro } \xi \neq 0, \\ u + \sigma_u \ln \left\{ \lambda_u \left[1 - (1 - r^{-1})^{\theta^{-1}} \right]^{-1} \right\}, & \text{pro } \xi = 0. \end{cases} \quad (20)$$

Odhad $\widehat{z}_r$ se obdrží nahrazením uvedených parametrů $\xi, \sigma_u, \lambda_u, \theta$ jejich odhady.

V dizertační práci jsou diskutovány dvě významné metody odhadu extrémálního indexu podle Gomes $\widehat{\theta}_G$ [18] a Ancony-Navarreteho a Tawna $\widehat{\theta}_{AT}$ [2]. Oba přístupy jsou založeny na vztahu (19) a odhad θ určují porovnáním parametrů limitních $\text{GEV}(\mu, \sigma, \xi)$ a $\text{GEV}(\mu_\theta, \sigma_\theta, \xi_\theta)$ rozdělení. V dizertační práci lze dále nalézt diskuzi odhadů rozptýlů statistik $\widehat{\theta}_G, \widehat{\theta}_{AT}$, přičemž tyto byly i počítačově implementovány do prostředí Matlab.

Pro účely porovnání vlastností těchto odhadů byla provedena simulační studie [27], kde byly opakovaně generovány realizace různých stacionárních procesů a na základě těchto realizací odhadnut extrémální index. Ukazuje se, že odhad $\widehat{\theta}_G$ je méně vychýlený pro hodnoty θ v naprosté většině případů. Na druhou stranu pro $\theta \rightarrow 0$ či $\theta \rightarrow 1$ vychýlení $\widehat{\theta}_G$ roste a $\widehat{\theta}_{AT}$ je zde více robustním odhadem. Ačkoliv je variabilita $\widehat{\theta}_{AT}$ nepatrně menší, odhad podle Gomes je méně citlivý na volbu pomocných parametrů a bude proto dále upřednostněn.

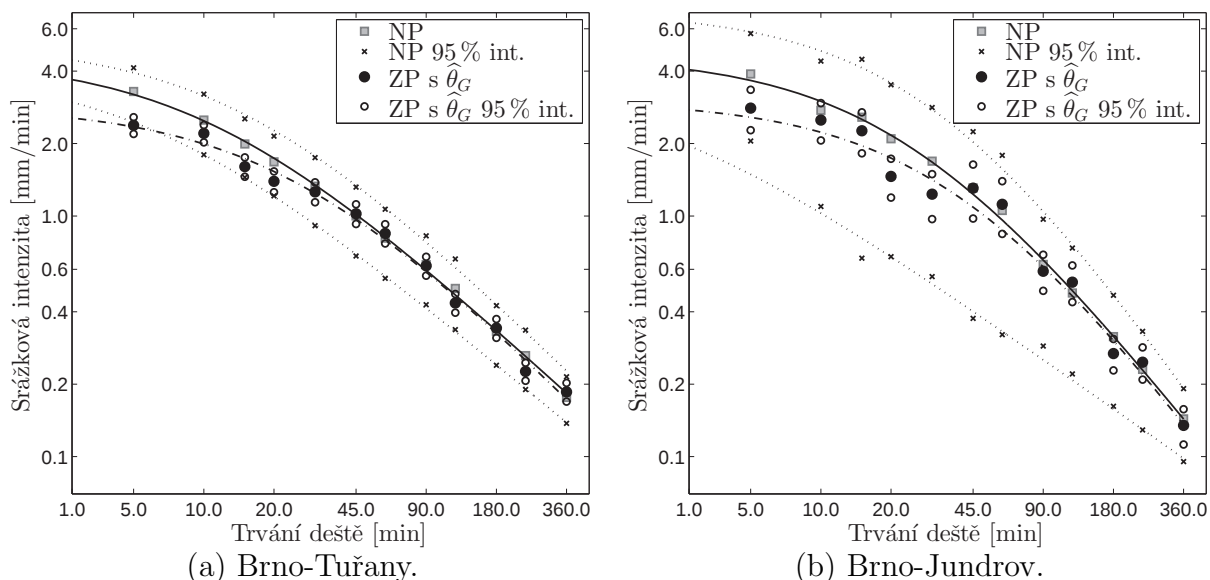
ANALÝZA SRÁŽKOVÝCH ÚHRNŮ POMOCÍ EV TEORIE STACIONÁRNÍCH ŘAD

Pomocí EV teorie stacionárních řad byla opětovně analyzována srážková data z odstavce 4. Pro dané doby trvání deště 5, 10, ..., 360 minut byly časové řady srážkových úhrnů určeny agregací pozorovaných minutových úhrnů pro po sobě jdoucí disjunktní periody. Volba pomocných parametrů při odhadu extrémálního indexu byla zvolena na základě simulační studie [27]. Protože nyní vychází analýza z celého datového souboru (bez nutnosti vzorkování), je potřeba znovu určit vhodné prahové hodnoty. K tomuto účelu byly použity základní grafické přístupy diskutované dříve.

Nově získané odhady byly porovnány s odhady původně získanými analýzou nezávislých měření, které byly obdrženy užitím vzorkovací techniky [35] a volbou prahové hodnoty grafickými metodami. Z teorie vyplývá, že $\xi_\theta = \xi$. Získané odhady parametrů ξ, ξ_θ se však zásadně odlišují, a to zejména pro krátké doby trvání deště. Toto platí i pro nejdelší dostupné řady, např. ze stanice Brno-Tuřany. Detailním rozbořem bylo zjištěno, že tyto odchylky není možné přisoudit rozdílným prahovým hodnotám obou modelů ani nevhodností odhadů θ . Nabízí se tak diskuze, zda je technika vzorkování [35] opravdu vhodná a zda jsou dešťové události separovány dostatečně dlouhými časovými intervaly.

Podobná situace je pozorovatelná pro získané odhady IDF křivek (viz obrázek 5). Zobrazené 100leté návratové úrovně vykazují velké vzájemné vychýlení pro krátké a střední doby trvání (do 45 minut) bez ohledu na rozsah datového souboru. Odhady IDF křivek určené na základě úplných časových řad dále vykazují významně menší variabilitu, a to opět i tam, kde byly k dispozici dostatečně dlouhé záznamy srážek. Běžně aplikované metody separace dešťových událostí pro vzorkování nezávislých měření se ukazují jako zásadně nevhodné. Jedno z možných doporučení při použití techniky vzorkování [35] by

tak mohla být úprava separačního schématu, aby byly dešťové události odděleny delším časovým obdobím.



Obrázek 5: MV odhady IDF křivek pro 100leté návratové úrovně získané analýzou nezávislých pozorování (NP; plná čára) a závislých pozorování s odhadem $\hat{\theta}_G$ (ZP; čerchovaná čára). IDF křivky byly proloženy regresní křivkou, 95% intervaly spolehlivosti byly proloženy jen pro případ NP. Osy jsou v logaritmickém měřítku.

VALIDACE ČASOVÝCH ŘAD

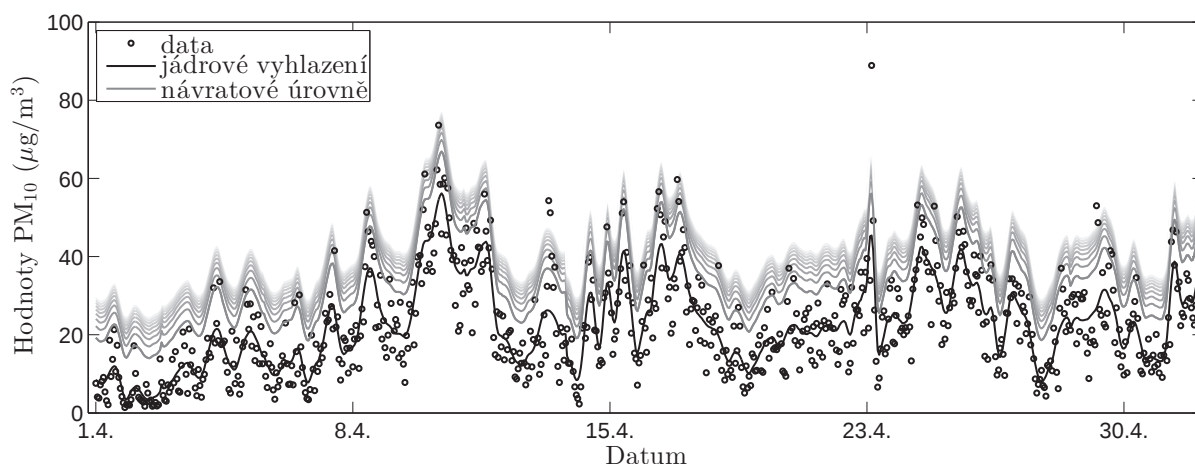
V environmentálních vědách, ale také v jiných technických oblastech, se často vyskytuje požadavek validace získaných měření časových řad. Typicky se jedná o nutnost identifikace odlehlých pozorování, jež nebyla způsobena vlastní sledovanou veličinou. Tato pozorování mohou být zapříčiněna např. chybami měřících přístrojů či jinými vnějšími vlivy. Běžně sice bývají získaná pozorování doprovázena také metadaty, která popisují vybrané podmínky měření, přesto je ale v praktických situacích často nutné provést dodatečnou validaci na základě empirických zkušeností a konkrétních znalostí v daném oboru. V tomto odstavci bude proto pozornost zaměřena na identifikaci odlehlých pozorování v časových řadách, která chápeme jako řídce se vyskytující extrémní jevy.

Konkrétně byly k analýze vybrány časové řady měření polétavého prachu PM_{10} , tedy pevných částic menších než $10 \mu m$. Měření PM_{10} byla automaticky prováděna na pěti lokacích v městě Brně umístěných v oblastech se zvýšenou zátěží na kvalitu ovzduší, přičemž časové rozlišení měřícího přístroje bylo nastaveno 1 hodinu. Konkrétně se jedná o stanice Arboretum, Lány, Svatoplukova, Výstaviště a Zvonařka. Dostupné byly časové řady za období od 11/2007 až 11/2015 pro poslední tři uvedené stanice, 11/2006 až 11/2015 pro stanici Arboretum a 9/2002 až 11/2015 pro stanici Lány.

Validaci dat za použití EV teorie není v literatuře dosud věnována velká pozornost. Základní shrnutí metod identifikace odlehlých hodnot publikovaných do roku 2004 lze nalézt v práci [28], užití EV teorie se pak věnují například publikace [40, 42]. Téměř vždy se však autoři omezují na případy Gaussovských procesů a předpokládají obor atraktivity v rámci Gumbelovy třídy EV rozdělení. Níže navržený postup je založen na odhadech parametrů GEV rozdělení stacionární řady. Při rozsáhlých a dlouhotrvajících měření

sledované náhodné veličiny většinou není splněn předpoklad stacionarity příslušné časové řady. Zde se omezíme na případy, kdy je stacionarita porušena jen z důvodu nestacionární střední hodnoty řady, tj. pro $X_1, \dots, X_n$ máme $m = m_t, 1 \leq t \leq n$, kde $m_t := EX_t$.

Pro odhad $\hat{m}_t$ střední hodnoty m_t byly použity techniky jádrového vyhlazení, konkrétně Gasserův-Müllerův odhad s Epanečnikovou jádrovou funkcí. Případná diskuze o vhodnosti tohoto postupu zde bude vynechána, detailní popis modelu je uveden v [23]. Zásadním problémem jádrových vyhlazovacích metod je volba vhodné šířky vyhlazovacího okna. Některé pokročilé postupy spočívají v adaptivní volbě lokální šířky okna v závislosti na průběhu konkrétní řady (viz např. [22]). Tento postup byl použit také v publikaci [23]. Odečtením odhadu $\hat{m}_t$ byla z původních měření získána přibližně stacionární řada, pro kterou je možné určit odhady návratových úrovní $\hat{z}_r$. Kombinací hodnot $\hat{m}_t + \hat{z}_r$ pro dané r a $1 \leq t \leq n$ se obdrží v jistém smyslu intervalový odhad pro extrémní hodnoty původní časové řady.



Obrázek 6: Odhady návratových úrovní extrémních hodnot pro řadu PM_{10} při volbě prahové hodnoty jako 90 % výběrového kvantilu reziduí jádrového vyhlazení. Zobrazena je stanice Brno-Arboretum v průběhu měsíce dubna 2015. Uvažovány byly periody návratu příslušné $r = 24, 48, \dots, 240$ pozorováním.

Pro odhady extrémních hodnot byl zvolen prahový model v kombinaci s MV metodou pro odhady parametrů GP rozdělení. Pro snadnou možnost automatizace byla prahová hodnota zvolena jako 90 % výběrový kvantil. Shoda rozdělení nadprahových hodnot s GP rozdělením byla ověřena vizuálně i pomocí testů dobré shody, přičemž žádný z použitých testů nevedl k zamítnutí této shody na hladině významnosti 0,05. Pro odhad parametru θ byl uvažován odhad podle Gomes.

Na obrázku 6 jsou zobrazeny výsledky pro stanici Brno-Arboretum v průběhu měsíce dubna 2015. Získané návratové úrovně odpovídají periodě návratu $r = 24, 48, \dots, 240$ hodin. Stanovení periody návratu r je nutné založit na expertních znalostech. Celkově však byla pozorována velmi dobrá shoda mezi očekávanými a skutečnými počty překročení všech návratových úrovní, což svědčí o vhodnosti uvedeného postupu k označení extrémních hodnot pozorovaných se stanovenou frekvencí.

ZÁVĚR

Představená práce shrnuje soudobé poznatky teorie extrémních hodnot a podtrhuje tak jejich aktuálnost při řešení praktických inženýrských úloh. Převážně jsou diskutována dvě témata spojená s odhady frekvence výskytu extrémních jevů. V první řadě se jedná o problematiku spojenou s volbou vhodné prahové hodnoty v prahovém POT modelu. V současnosti roste zájem o adaptivní metody výběru vhodného prahu a vzhledem k tomuto tématu je práce zaměřena. Za účelem srovnání vybraných, v poslední době rozvinutých, technik byla provedena simulační studie. Skrze odhady EV indexu se ukazuje, že obě metody vedou ke srovnatelným výsledkům v rámci Hallový třídy rozdělení. Mimo Hallovu třídu se pak ukazuje, že dvojité bootstrapová metoda přispívá k menšímu vychýlení odhadů, ačkoliv zde nemá jasné teoretické opodstatnění. Srovnání metod výběru prahu a technik odhadu parametrů GP rozdělení bylo také provedeno vzhledem ke konkrétní reálné aplikaci odhadu extrémních dešťových srážek a odhadu IDF křivek. Cílem bylo navázat na předchozí studie zhotovené za účelem aktualizace IDF křivek, důležitého hydrologického nástroje. V tomto ohledu se ukazuje, že použití adaptivních technik určení prahu může vést ke snížení variability odhadů tam, kde jsou k dispozici jen krátké časové řady. Zejména lze tyto přístupy doporučit pro odhady dlouhodobých extrémních událostí.

V druhé části se práce zabývá zobecnění teorie extrémních hodnot pro stacionární řady. Zde bylo autorem provedeno srovnání dvou nejvýznamnějších odhadů rozvinutých v poslední době a následně byly těmito pokročilými technikami opětovně analyzována předchozí srážková data. Zásadní zjištění bylo odhaleno v nedostacích vzorkovacích technik, které bývají běžně užívány při vyhodnocení pozorování pomocí představené teorie nezávislých veličin. S ohledem k odhadům IDF křivek dochází aplikací těchto technik ke značnému vychýlení odhadů. Navíc použitím teorie pro stacionární řady lze významně redukovat variabilitu odhadů. V tomto směru jsou v práci učiněna některá praktická doporučení. V závěru práce je představena původní autorova aplikace teorie extrémních hodnot stacionárních řad pro účely validace časových řad. Aplikace je cílena na problematiku identifikace odlehlých pozorování při rozsáhlých měřeních, která je v tomto kontextu možno chápat jako extrémní jevy. Kombinací s neparametrickými metodami vyhlazování je možné určit odlehlá pozorování i pro některé nestacionární řady a představená metoda tak nabízí možnost automatizace a implementace do již zavedených postupů.

LITERATURA

- [1] ADAMOWSKI, K. Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *Journal of Hydrology*, 2000, vol. **229**, no. 3-4, pp. 219-231.
- [2] ANCONA-NAVARRETE, M.A., TAWN, J.A. A Comparison of Methods for Estimating the Extremal Index. *Extremes*, 2000, vol. **3**, no. 1, pp. 5-38.
- [3] ANDĚL, J. *Základy matematické statistiky*. 3. vyd. Praha: Matfyzpress, 2005. 358 s. ISBN 80-86732-40-1.
- [4] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., TEUGELS, J., DE WALL, D., FERRO, C. *Statistics of Extremes: Theory and Applications*. 1st edition. Chichester: Wiley, 2004. 514 p. ISBN: 978-0-471-97647-9.

- [5] BEN ZVI, A. Rainfall intensity-duration-frequency relationships derived from large partial duration series. *Journal of Hydrology*, 2009, vol. **367**, pp. 104-114.
- [6] CAEIRO, F., GOMES, M.I., VANDEWALLE, B. Semi-parametric probability-weighted-moments estimation revisited. *Methodology and Computing in Applied Probability*, 2014, vol. **16**, no. 1, pp. 1-29.
- [7] CAEIRO, F., GOMES, M.I. Semi-parametric tail inference through probability-weighted-moments. *Journal of Statistical Planning and Inference*, 2010, vol. **141**, pp. 937-950.
- [8] CASELLA, G., BERGER, R.L. *Statistical Inference*. 2nd edition. Pacific Grove: Thomson Learning, 2002. 660 p. ISBN-13 978-0-495-39187-6.
- [9] COLES, S. *An Introduction to Statistical Modeling of Extreme Values*. 3rd edition. London: Springer, 2001. ISBN 1-85233-459-2.
- [10] DE HAAN, L., FERREIRA, A. *Extreme Value Theory: An Introduction*. 1st edition. New York: Springer, 2006. ISBN-10 0-387-23946-4.
- [11] DIEBOLT, J., GUILLOU, A., RACHED, I. Approximation of the distribution of excesses through a generalized probability-weighted moments method. *Journal of Statistical Planning and Inference*, 2007, vol. **137**, no. 3, pp. 841-857.
- [12] DRAISMA, G., DE HAAN, L., PENG, L., PEREIRA, T.T. A Bootstrap-based Method to Achieve Optimality in Estimating the Extreme-Value Index. *Extremes*, 1999, vol. **2**, no. 4, pp. 367-404.
- [13] FISHER, R.A., TIPPETT, L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceeding of Cambridge Philosophical Society*, 1928, vol. **24**, pp. 180-190.
- [14] FRAGA ALVES, M.I., GOMES, M.I., DE HAAN, L. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 2003, vol. **60**, no. 1, pp. 193-213.
- [15] FUSEK, M. Rozdělení extrémních hodnot a jejich aplikace. Vysoké učení technické v Brně, Brno, 2013. Dizertační práce, 136 s.
- [16] FUSEK, M., HOLEŠOVSKÝ, J. *EVDest* [software]. Ver. 1.0, 2014 [citováno 9.1.2016]. Dostupné z <http://www.umat.feec.vutbr.cz/software/evdest.html>.
- [17] GNEDENKO, B.V. Sur la distribution limite du terme maximum d' une série aléatoire. *Annals of Mathematics*, 1943, vol. **44**, pp. 423-453.
- [18] GOMES, M.I. On the estimation of parameters of rare events in environmental time series. In *Statistics for the Environment 2: Water Related Issues*. Chichester: Wiley, 1993. pp. 225-241.
- [19] GOMES, M.I., DE HAAN, L., PENG, L. Semi-parametric estimation of the second order parameter - asymptotic and finite sample behavior. *Extremes*, 2002, vol. **5**, no. 4, pp. 387-414.
- [20] GOMES, M.I., MARTINS, M.J., NEVES, M. Alternatives to a Semi-Parametric Estimator of Parameters of rare Events—The Jackknife Methodology. *Extremes*, 2000, vol. **3**, no. 3, pp. 207-229.

- [21] GOMES, M.I., OLIVEIRA, O. The Bootstrap Methodology in Statistics of Extremes - Choice of the Optimal Sample Fraction. *Extremes*, 2001, vol. 4, no. 4, pp. 331-358.
- [22] HERRMANN, E. Local Bandwidth Choice in Kernel Regression Estimation. *Journal of Computational and Graphical Statistics*, 1997, vol. 6, no. 1, pp. 35-54.
- [23] HOLEŠOVSKÝ, J., ČAMPULOVÁ, M., MICHÁLEK, J. Time-series outlier identification based on extreme value theory for stationary processes. Přípraveno k odeslání, 2016.
- [24] HOLEŠOVSKÝ, J., FUSEK, M., BLACHUT, V., MICHÁLEK, J. Comparison of precipitation extremes estimation using parametric and nonparametric methods. *Hydrological Sciences Journal*, 2015. Přijato k publikaci. DOI 10.1080/02626667.2015.1111517.
- [25] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Modelling of precipitation extremes using parametric and nonparametric methods with automated threshold selection. *International Journal of Mathematics and Computers in Simulation*, 2015, vol. 9, pp. 94-102.
- [26] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Automated threshold selection for parametric and non-parametric estimates of intensity-duration-frequency curves. In *Proceedings of the 1st International Conference on Mathematical Methods & Computational Techniques in Science & Engineering, MMCTSE 2014*. Athens, Greece, 2014, pp. 114-118. ISBN 987-1-61804-256-9.
- [27] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Extreme value estimation for correlated observations. In *Proceedings of 20th International Conference on Soft Computing MENDEL 2014*. Brno, Czech Republic, 2014, pp. 359-364. ISBN 978-80-214-4984-8.
- [28] HODGE, V.J., AUSTIN, J. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 2004, vol. 22, pp. 85-126.
- [29] HOSKING, J.R.M., WALLIS, J.R. *Regional frequency analysis: An approach based on L-moments*. 1st edition. Cambridge: Cambridge University Press, 1997. 224 s. ISBN 05-214-3045-3.
- [30] CHOW, V.T., MAIDMENT, D.R., MAYES, L.W. *Applied Hydrology*. 1st edition. New York: McGraw-Hill, 1988. 572 s. ISBN 978-0070108103.
- [31] JENKINSON, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 1955, vol. 81, pp. 158-171.
- [32] KHALIQ, M.N., OUARDA, T.B.M.J., ONDO, J.-C., GACHON, P., BOBÉE, B. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observation: A review. *Journal of Hydrology*, 2006, vol. 329, pp. 534-552.
- [33] LAGARIAS, J.C., REEDS, J.A., WRIGHT, M.H., WRIGHT, P.E. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 1998, vol. 9, pp. 112-147.

- [34] LEADBETTER, M.R., LINDGREN, G., ROOTZÉN, H. *Extremes and Related Properties of Random Sequences and Series*. 1st edition. New York: Springer, 1983. 336 p. ISBN 978-1-4612-5451-5.
- [35] MADSEN, H., MIKKELSEN, P.S., ROSBJERG, D., HARREMÖES, P. Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series. *Water Resources Research*, 2002, vol. **38**, no. 11, pp. 21-1–21-11.
- [36] MADSEN, H., RASMUSSEN, P.F., ROSBJERG, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 1. At site modeling. *Water Resources Research*, 1997, vol. **33**, no. 4, pp. 747-757.
- [37] MICHÁLEK, J., FUSEK, M., HELLEBRAND, R. Srovnání statistických metod odhadů IDF křivek. In *Workshop Adolfa Patery 2012 Extrémní hydrologické jevy v povodích*. Brno, 2012, s. 163-170. ISBN 978-80-02-02423-1.
- [38] NORTHROP, P.J., COLEMAN, C.L. Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 2014, vol. **17**, pp. 289-303.
- [39] PENG, L., QI, Y. Maximum likelihood estimation of extreme value index for irregular cases. *Journal of Statistical Planning and Inference*, 2009, vol. **139**, pp. 3361-3376.
- [40] QUINTELA-DEL-RÍO, A., FRANCISCO-FERNÁNDEZ, M. Nonparametric functional data estimation applied to ozone data: Prediction and extreme value analysis. *Chemosphere*, 2011, vol. **82**, no. 6, pp. 800-808.
- [41] RESNICK, S.I. *Extreme Values, Regular Variation, and Point Processes*. 1st edition. New York: Springer, 1987. 320 p. ISBN 978-0-387-75952-4.
- [42] ROBERTS, S.J. Novelty detection using extreme value statistics. *IEEE Proceedings: Vision, Image and Signal Processing*, 1999, vol. **146**, no. 3, pp. 124-129.
- [43] SCARROTT, C., MACDONALD, A. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT*, 2012, vol. **10**, no. 1, pp. 33-60.
- [44] SMITH, R.L. Estimating tails of probability distribution. *Annals of Statistics*, 1987, vol. **15**, no. 3, pp. 1174-1207.
- [45] SMITH, R.L. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 1985, vol. **72**, no. 1, pp. 67-90.
- [46] TRUPL, J. *Intenzity krátkodobých dešťů v povodích Labe, Odry a Moravy*. Praha: Výzkumná ústav vodohospodářský, 1958. Práce a studie, 34 s.
- [47] VON MISES, R. La distribution de la plus grande de n valeurs. *American Mathematical Society, Selected Papers*, Providence, 1954.
- [48] WILLEMS, P. Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types. *Journal of Hydrology*, 2000, vol. **233**, no. 1-4, pp. 189-205.
- [49] ZHOU, C. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 2009, vol. **100**, pp. 794-815.

AUTOROVY PUBLIKACE

ČLÁNKY V ČASOPISECH

- [1] HOLEŠOVSKÝ, J., FUSEK, M., BLACHUT, V., MICHÁLEK, J. Comparison of precipitation extremes estimation using parametric and nonparametric methods. *Hydrological Sciences Journal*, 2015. Přijato k publikaci. DOI 10.1080/02626667.2015.1111517.
- [2] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Modelling of precipitation extremes using parametric and nonparametric methods with automated threshold selection. *International Journal of Mathematics and Computers in Simulation*, 2015, vol. **9**, pp. 94-102.
- [3] OLEJNÍČKOVÁ, Z., HOLEŠOVSKÝ, J., VÁVROVÁ, M., KRÁLOVÁ, Z., MICHÁLEK, J. Methylmercury in tissues of fish from the Svratka River, the Czech Republic. *Fresenius Environmental Bulletin*, 2014, vol. **23**, no. 12b, pp. 3319-3324.
- [4] ŠKARKOVÁ, P., ZLÁMALOVÁ GARGOŠOVÁ, H., HOLEŠOVSKÝ, J., VÁVROVÁ, M., MICHÁLEK, J., OLEJNÍČKOVÁ, Z. Application of statistical methods for ecotoxicological data evaluation. *Fresenius Environmental Bulletin*, 2015, vol. **24**, no. 5, pp. 1692-1698.

ČLÁNKY VE SBORNÍCÍCH KONFERENCÍ

- [5] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Extreme value estimation for correlated observations. In *Proceedings of 20th International Conference on Soft Computing MENDEL 2014*. Brno, Czech Republic, 2014, pp. 359-364. ISBN 978-80-214-4984-8.
- [6] HOLEŠOVSKÝ, J., POPELA, P., ROUPEC, J. Disruption in Congested Networks. In *Proceedings of 19th International Conference on Soft Computing MENDEL 2013*. Brno, Czech Republic, 2013, pp. 191-196. ISBN 978-80-214-4755-4.
- [7] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Automated threshold selection for parametric and non-parametric estimates of intensity-duration-frequency curves. In *Proceedings of the 1st International Conference on Mathematical Methods & Computational Techniques in Science & Engineering, MMCTSE 2014*. Athens, Greece, 2014, pp. 114-118. ISBN 987-1-61804-256-9.
- [8] HOLEŠOVSKÝ, J., POPELA, P. Stochastic Extensions of a Traffic Assignment Problem. In *XX International Conference PDMU-2012: Problems of Decision Making under Uncertainties, Proceeding - Applied Papers*. Brno, Czech Republic, 2012, pp. 61-70. ISBN 978-80-7231-897-1.

ABSTRAKTY VE SBORNÍCÍCH KONFERENCÍ

- [9] HOLEŠOVSKÝ, J. Comparison of Parametric and Nonparametric Methods for Estimation of Hydrological Extremes. In *Building Bridges: Probability, Statistics and Applications - Book of Abstracts*. Braunschweig, Germany, 2013, pp. 26-27.

- [10] HOLEŠOVSKÝ J., BLACHUT V., FUSEK M., MICHÁLEK J. Comparison of Parametric and Nonparametric Estimates of Extreme Value Distribution. In *Precipitation extremes in a changing climate*. Hejnice, Czech Republic, 2013, pp. 29.
- [11] HOLEŠOVSKÝ, J., MICHÁLEK, J. Estimation of Extreme Values: A Case Study for the Intensity-Duration-Frequency Curves Assessment. In *XXIV International Conference Problems of Decision Making under Uncertainties, PDMU-2014 - Abstracts, Series B*, Český Rudolec, Czech Republic, 2014, pp. 47.
- [12] HOLEŠOVSKÝ, J., MICHÁLEK, J. Extreme Value Estimation under the Presence of Short-Time Dependence. In *XX. Summer School of Biometrics*. Slavonice, Czech Republic, 2014, pp. 35.
- [13] KOSÁROVÁ, H., HOLEŠOVSKÝ, J., MATUŠKOVÁ, M., KOMENDOVÁ, R., VÁVROVÁ, M. Statistical Evaluation of the Data Determination of Platinum in Air in the City of Brno by ET-AAS after SPE Preconcentration. In *EMEC 2015 - Book of Abstracts*. Brno, Czech Republic, 2014, pp. 89.

SOFTWARE

- [14] FUSEK, M., HOLEŠOVSKÝ, J. *EVDest* [software]. Ver. 1.0, 2014 [citováno 9.1.2016]. Dostupné z <http://www.umat.feec.vutbr.cz/software/evdest.html>.

PŘIPRAVENO K PUBLIKACI VE VĚDECKÝCH ČASOPISECH

- [15] HOLEŠOVSKÝ, J., ČAMPULOVÁ, M., MICHÁLEK, J. Time-series outlier identification based on extreme value theory for stationary processes. Připraveno k odeslání, 2016.
- [16] HOLEŠOVSKÝ, J., FUSEK, M. Metody analýzy extrémních hodnot a jejich softwarová implementace. Odesláno k publikaci, 2015.

AUTORŮV ŽIVOTOPIS

ING. JAN HOLEŠOVSKÝ

E-mail holesovsky.j@fce.vutbr.cz
Adresa Vavřínecká 16, 62400 Brno, Česká republika
Datum narození 8. únor 1987, Brno, Česká republika

VZDĚLÁNÍ

Doktorské
studium
2012 - 2016

**Vysoké učení technické v Brně,
Fakulta strojního inženýrství,**
Obor: Aplikovaná matematika.
Dizertační práce: *Metody odhadu parametrů rozdělení
extrémního typu s aplikacemi.*
Vedoucí práce: doc. RNDr. Jaroslav Michálek, CSc.

Magisterské
studium
2010 - 2012

**Vysoké učení technické v Brně,
Fakulta strojního inženýrství,**
Obor: Matematické inženýrství.
Diplomová práce: *Traffic assignment optimization models.*
Vedoucí práce: RNDr. Pavel Popela, PhD.

2011

Molde University College, Norsko,
Studijní pobyt Erasmus.

Bakalářské
studium
2007 - 2010

**Vysoké učení technické v Brně,
Fakulta strojního inženýrství,**
Obor: Matematické inženýrství.
Bakalářská práce: *Obecné m-znakové kódy.*
Vedoucí práce: prof. RNDr. Ladislav Skula, DrSc.

PRACOVNÍ ZKUŠENOSTI

2015 - dosud

**Vysoké učení technické v Brně,
Fakulta stavební,**
Asistent, Ústav matematiky a deskriptivní geometrie.

2014

Centrum výzkumu globální změny AV ČR,
Příprava publikace František Zemek et al. (2014)
Airborne Remote Sensing. Brno: CVGZ AV ČR.

2014

**Vysoké učení technické v Brně,
Fakulta strojního inženýrství,**
Výuka předmětů zajišťovaných Ústavem matematiky.

DALŠÍ AKTIVITY

- | | |
|-------------|--|
| 2012 | Absolvování magisterského studia s vyznamenáním. |
| 2008 - 2010 | Člen studentské komory Akademického senátu FSI VUT v Brně. |

ÚČAST NA PROJEKTECH

- | | |
|------|---|
| 2015 | Využití moderních statistických metod pro hodnocení úrovně kontaminace životního prostředí. |
| 2014 | Posouzení kontaminace životního prostředí vybranými polutanty a jejich zhodnocení moderními statistickými metodami. |
| 2014 | Moderní metody aplikované matematiky pro využití v technických vědách. |
| 2013 | Aplikace řídkých řešení systému lineárních rovnic ve zpracování vícerozměrných dat. |
| 2013 | Aplikace moderních statistických metod pro hodnocení kontaminace životního prostředí. |

ABSTRAKT

Předložená práce je zaměřena na teorii extrémních hodnot a její užití v aplikačních úlohách. V první části je zavedeno rozdělení extrémních hodnot a popsány jeho vlastnosti. Na základě předložených tvrzení jsou diskutovány dva přístupy k analýze extrémních hodnot, a sice model blokových maxim a prahový model postavený na zobecněném Paretoově rozdělení. Ačkoliv je první jmenovaný v mnoha ohledech chápán jako robustnější, patří prahový model ke stále častěji užívaným přístupům. Samotná volba prahu, která má zásadní vliv na kvalitu odhadu, však pořád patří k nedořešeným problémům tohoto přístupu. Především na techniky určení vhodné prahové hodnoty je tato práce zaměřena. Z aplikačního hlediska jsou pak nejzajímavější adaptivní přístupy určení prahu, které danou volbu vhodně automatizují. Pro porovnání vybraných adaptivních technik byla provedena simulační studie a tyto byly dále použity pro analýzu srážkových úhrnů v jiho-moravském regionu. Dále se práce věnuje v poslední době rozvíjeným metodám odhadu extrémních hodnot stacionárních řad. V praxi je často nutné měřené časové řady vzorkovat na přibližně nezávislá pozorování. Použití teorie pro stacionární řady přitom tento problém redukce dat zcela eliminuje. Jak je ukázáno, běžně používané metody vzorkování se v tomto kontextu ukazují jako nevhodné a užití pokročilých technik pro stacionární řady vede k lepším odhadům extrémních hodnot.

ABSTRACT

The thesis is focused on extreme value theory and its applications. Initially, extreme value distribution is introduced and its properties are discussed. At this basis are described two models mostly used for an extreme value analysis, i.e. the block maxima model and the Pareto-distribution threshold model. The first one takes advantage in its robustness, however recently the threshold model is mostly preferred. Although the threshold choice strongly affects estimation quality of the model, an optimal threshold selection still belongs to unsolved issues of this approach. Therefore, the thesis is focused on techniques for proper threshold identification, mainly on adaptive methods suitable for use in practice. For this purpose a simulation study was performed and acquired knowledge was applied for analysis of precipitation records from South-Moravian region. Further on, the thesis also deals with extreme value estimation within a stationary series framework. Usually, an observed time series needs to be separated to obtain approximately independent observations. The use of the advanced theory for stationary series allows to avoid the entire separation procedure. In this context the commonly applied separation techniques turn out to be quite inappropriate in most cases and the estimates based on theory of stationary series are obtained with better precision.