Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra Geoinformatiky

Jan GELETIČ

# CHARAKTERISTIKA PŘÍRODNÍHO PROSTŘEDÍ MODELOVÉ LOKALITY HALENKOVICE NA ZÁKLADĚ ANALÝZY ČASOVÝCH ŘAD

Bakalářská práce

Vedoucí práce: Mgr. Pavel Tuček

Olomouc 2008

**Prohlášení**

Prohlašuji, že jsem vytvořil tuto bakalářskou práci samostatně pod vedením Mgr. Pavla Tučka a že jsem v seznamu použité literatury uvedl všechny zdroje použité při zpracování práce.

V Olomouci dne 28. května 2008

**Poděkování**

Rád bych na tomto místě poděkoval svému vedoucímu bakalářské práce Mgr. Pavlovi Tučkovi za spolupráci a čas, který mi věnoval. Dále patří velký dík ČHMÚ a zvláště panu Petrovi Malinovi za poskytnutí kvalitních dat pro srovnání. Také bych chtěl poděkovat svým rodičům a sestře za trpělivost, toleranci a lásku.

# Obsah

# Úvod

Mnoho mých spolužáků se snaží matematice vyhnout, já se ji však již od střední školy rád věnuji i ve svém volném čase. Proto jsem měl o tématu a hlavní náplni své práce jasno již v prvním ročníku, po absolvování předmětu geostatistika. Chtěl jsem přijít na kloub datům, o kterých každý ví, že to tak je, ale většinou netuší proč. Proto mi nabídka od RNDr. Bíla Ph.D. na zpracování časových řad z modelové lokality Halenkovice padla jako ulitá a s chutí jsem se pod vedením Mgr. Tučka pustil do práce.

Co se modelové lokality v Halenkovicích týče, bylo zde již zpracováno velké množství prací, zejména se však jednalo o práce technického charakteru zabíhající z převážné části do oboru speciální nebo inženýrské geodézie, geomorfologie a geologie. Ty měly za úkol zjistit aktivitu sesuvu, přesnost měření nebo modelovat vlastní sesuv. Ovšem zpracování naměřených dat ze statistického hlediska zde doposud zpracováno nebylo.

Na závěr díky spolupráci s panem Petrem Malinou z Košíků bylo možné vzájemně porovnat naměřená data. Data byla navíc srovnána i s daty z Atlasu podnebí Česka, ale v tomto případě se jedná o dlouhodobé průměry za několik let, proto jsou výsledky do jisté míry zkresleny.

# Kapitola 1

# Cíl práce

Cílem bakalářské práce je:

- Veškerá naměřená data spojit do datasetů
- Opravit data od výpadků a chyb měření
- Statisticky popsat data
- Shladit data do hodinových, denních a měsíčních průměrů
- Graficky vizualizovat data
- Zpracovat data z hlediska analýzy časových řad
- Porovnat němřená data s dalšími dostupnými referenčními daty (Atlas podnebí Česka, amatérská meteostanice v Košíkách)
- Vytvoření webových stránek o bakalářské práci

# Kapitola 2

# Vymezení a charakteristika území

Sesuv v Halenkovicích se nachází v západní části katastrálního území Halenkovice, okres Zlín, kraj Zlínský. Jedná se o rozsáhlejší sesuvné území s aktivním sesuvem. Termínem sesuvné území je označována část svahu postiženého sesouváním. Jedná se o vlastní aktivní sesuv a širší území vzniklého sesuvu, jehož náchylnost k sesouvání se projevila staršími pohyby. Tento sesuv byl aktivován po povodních v roce 1997, od roku 2000 je podrobně monitorován. Dnes patří k nejlépe prostudovaným sesuvům v České republice.


Obr 1.: Přehledná mapka okolí

## 2.1 Sestava stanic v Halenkovicích

Zde uvádím jen 3 nejdůležitější stanice, ze kterých pochází naměřená data.

**ALA1**

Složení: datalogger, 5 x teplota půdy, 2 x vlhkost půdy

Teplota: 5, 10, 20, 35 a 50 mm

Vlhkost: 25 a 50 mm

Napájení: 12 V akumulátor

Četnost měření: každých 15 minut

Výpadky měření: nezaznamenány, jen občasné nulové hodnoty u teploměru v hloubce 35cm

**ALA - průtok**

složení: datalogger, 5 x teplota půdy, 2 x vlhkost půdy, průtok z horizontálního vrtu

Teplota: 5, 10, 20, 35 a 50 mm

Vlhkost: 25 a 50 mm

Průtok: napojení na horizontální vrt

Napájení: 12 V akumulátor

Četnost měření: každých 15 minut

Výpadky měření: od 9:30 19.8. do 11:15 31.8., občasné nulové hodnoty u teploměru v hloubce 35cm, průtokoměr měřil jen necelého půl roku, proto nebyl brán při zpracování v úvahu

Obr 2.: Stanice ALA s průtokoměrem

**Halenkovice 1 (také jako Fiedler)**

M4016 - 16 kanálová stanice se solárním panelem

Složení: Srážkoměr 500 $cm^2$, 4 x teplota: 2m, přízemní, 10 a 60 cm pod povrchem, 8x tenzometry ve dvou hnízdech nad sebou v délkách: 15, 30, 45 a 60 cm

Napájení: 1x olověný akumulátor podporovaný solárním panelem

Četnost měření: srážkoměr měří pouze při dešti čas překlopení člunku; nastavení ostatních zařízení je na 15 minut

Výpadky měření: od 1.1. do 1.3. (způsobeno zakrytím srážkoměru po dobu zimy), od 20:00 8.3. do 10:00 27.3.

Obr 3.: Stanice Fiedler se srážkoměrem

# Kapitola 3

# Popisná statistika

Pojmem **popisná statistika** se rozumí základní statistické charakteristiky, kterými jsou například aritmetický průměr, medián, modus, rozptyl, směrodatná odchylka, kovariance, korelace atd. O zákadních číselných charakteristikách je pojednáno v následující podkapitole. Matematicky složitější statistiky (křivost, špičatost, atd.) jsou popsány v [3]. Popis jednotlivých charakteristik byl zpracován podle [2] a [3]. Já jsem tyto charakteristiky zpracovával pomocí programu **project R**, ale lze použít i funkci *Popisná statistika*, kterou disponuje řešení od firmy Microsoft, konkrétně tedy program **Excel** z řady produktů Microsoft Office. Produkt společnosti OpenOffice.org ve verzi 2.3.1., **Calc**, tuto možnost nemá, obsahuje pouze jednotlivé charakteristiky.

## 3.1 Aritmetický průměr

**Aritmetický průměr** je statistická veličina, která v jistém smyslu vyjadřuje typickou hodnotu popisující soubor mnoha hodnot. Aritmetický průměr se obvykle značí vodorovným pruhem nad názvem proměnné, popř. řeckým písmenem $\mu$. Definice aritmetického průměru je $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$, tzn. součet všech hodnot vydělený jejich počtem. V každodenní praxi se obvykle obecným slovem průměr myslí právě aritmetický průměr.

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k. \tag{3.1}$$

Aritmetický průměr jako veličinu v této práci nemíním používat, protože jeho hodnoty výrazně zkreslují sledovaný jev. Proto je vhodnější použít např. vážený aritmetický průměr nebo klouzavé průměry. Pro úpravu časových řad jsem použil právě metodu klouzavých průměrů.

## 3.2 Vážený průměr

Vážený průměr zobecňuje aritmetický průměr a poskytuje charakteristiku statistického souboru v případě, že hodnoty v tomto souboru mají různou důležitost, různou váhu. Používá se zejména při počítání celkového aritmetického průmeru souboru složeného z více podsouborů. Pro výpočet váženého průměru potřebujeme jednak hodnoty, jejichž průměr chceme spočítat, a zároveň jejich váhy.

Máme-li soubor $n$ hodnot $X = \{x_1, \ldots, x_n\}$ a k nim soubor $n$ vah $W = \{w_1, \ldots, w_n\}$, je vážený průměr dán vzorcem

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \tag{3.2}$$

nebo analogicky bez sumárního zápisu

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n}{w_1 + w_2 + w_3 + ... + w_n}. \tag{3.3}$$

Pokud jsou všechny váhy stejné, je vážený průměr totožný s aritmetickým průměrem. Vážený průměr má několik nezvyklých vlastností, které jsou například výjádřeny v Simpsonově paradoxu (více o této problematice je dostupné v [2] a [3]). Vážené verze jiných průměrů lze také spočítat.

## 3.3 Rozptyl

**Rozptyl** (též střední kvadratická odchylka, střední kvadratická fluktuace, variance nebo také disperze) se používá v teorii pravděpodobnosti a statistice. Je to druhý centrální moment náhodné veličiny. Jedná se o charakteristiku variability rozdělení pravděpodobnosti náhodné veličiny, která vyjadřuje variabilitu rozdělení souboru náhodných hodnot kolem její střední hodnoty, a je dána vzrcem:

$$\sigma^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2. \tag{3.4}$$

Rozptyl je definován jako střední hodnota kvadrátů odchylek od střední hodnoty. Odchylku od střední hodnoty, která má rozměr stejný jako náhodná veličina, zachycuje směrodatná odchylka.

## 3.4 Směrodatná odchylka

**Směrodatná odchylka** je v teorii pravděpodobnosti a statistice často používanou mírou statistické disperze. Jedná se o kvadratický průměr odchylek hodnot znaku od jejich aritmetického průměru.

$$\sigma = \sqrt{\frac{1}{n-1}\sum (x_i - \overline{x})^2} \tag{3.5}$$

Směrodatná odchylka se obvykle definuje jako odmocnina z rozptylu náhodné veličiny $X$, tzn.

$$\sigma = \sqrt{\sigma^2}. \tag{3.6}$$

Zhruba řečeno vypovídá o tom, jak moc se od sebe navzájem liší typické případy v souboru zkoumaných čísel. Je-li malá, jsou si prvky souboru většinou navzájem podobné, a naopak velká směrodatná odchylka signalizuje velké vzájemné odlišnosti. Pomocí pravidel $1\sigma$ a $2\sigma$ (viz níže) lze přibližně určit, jak daleko jsou čísla v souboru vzdálená od průměru, resp. hodnoty náhodné veličiny vzdálené od střední hodnoty. Směrodatná odchylka je nejužívanější míra variability.

**Pravidlo $1\sigma$ a $2\sigma$**

Jedná se o empirické pravidlo, jehož platnost závisí na konkrétním případu, proto je formulováno obecně. Lze je však velmi dobře použít pro základní orientaci v rozložení hodnot souboru nebo náhodné veličiny.

Jestli jde o soubor hodnot, pak se většina hodnot neodlišuje od průměru o více než jednu směrodatnou odchylku a skoro všechny hodnoty jsou v pásmu do dvou směrodatných odchylek od průměru.

Jestli jde o náhodnou veličinu, pak pravděpodobnost, že se hodnota náhodné veličiny bude od střední hodnoty lišit nejvýše o jednu směrodatnou odchylku, je přibližně 68%; pravděpodbnost, že se hodnota bude lišit nejvýše o dvě směrodatné odchylky je přibližně 99%.

$$\langle \overline{x} - \sigma; \overline{x} + \sigma \rangle \approx 68\% \tag{3.7}$$

$$\langle \overline{x} - 2\sigma; \overline{x} + 2\sigma \rangle \approx 99\% \tag{3.8}$$

## 3.5 Korelační koeficient

**Korelace** je ve statistice vzájemný vztah mezi znaky či veličinami. Korelační koeficient může nabývat hodnot v intervalu $\langle -1; 1 \rangle$. Hodnota korelačního koeficientu $-1$ značí zcela

nepřímou závislost, tedy čím více se zvětší hodnoty v první skupině znaků, tím více se zmenší hodnoty v druhé "skupině znaků", např. vztah mezi uplynulým a zbývajícím časem. Hodnota korelačního koeficientu +1 značí zcela přímou závislost, např. vztah mezi rychlostí bicyklu a frekvencí otáček kola bicyklu. Pokud je korelační koeficient roven 0, pak mezi znaky není žádná statisticky zjistitelná závislost, např. vztah mezi vlhkostí půdy a aktivitou sesuvu.

**Výpočet Pearsonova korelačního koeficientu**

Vypočteme aritmetické průměry souborů $X$ a $Y$, vynásobíme sumy čtverců odchylek od těchto průměrů obou souborů – tím jsme spočetli tzv. **kovarianci**,

$$cov(X,Y) = E\left[(X - EX)\,(Y - EY)\right], \tag{3.9}$$

což je však absolutní veličina. Pro výpočet relativní veličiny pak kovarianci dělíme odmocninou násobku rozptylu souboru $X$ a souboru $Y$:

$$\rho_{(X,Y)} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}. \tag{3.10}$$

Protože $\mu_X = E(X)$, $\mu_{X^2} = E(X^2) - E^2(X)$ a obdobně pro $Y$, můžeme psát:

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\,\sqrt{E(Y^2) - E^2(Y)}}. \tag{3.11}$$

# Kapitola 4

# Časové řady

## 4.1 Úvod

**Časová řada** je podle [1] a [5] chronologicky uspořádaná posloupnost hodnot určitého číselného ukazatele. Tento ukazatel musí být v čase vymezen věcně a prostorově shodně. Prakticky to znamená, že časová řada je náhodně naměřená řada čísel v určitých intervalech za určitou dobu; tuto řadu tvoří hodnoty nějaké (např. meteorologické) veličiny, které jsou uspořádány od nejstarších po nejmladší nebo naopak.

Ukázkou časové řady může být například měsíční chod srážek na stanici v Košíkách za rok 2007.

## 4.2 Význam analýzy časových řad

Analýza časových řad včetně předpovědi jejich chování jsou podle [5] velice významné prvky současné statistiky. Úspěšně se vyrovnává s popisem dynamických systémů, se kterými často přicházíme do styku. Data, která vytváří časovou řadu, vznikají jako chronologicky uspořádaná pozorování. Data ve formátu časových řad vznikají v mnoha vědách, například úhrnná hodnota srážek a denní chod teploty v meteorologii, vývoj ceny ropy v ekonomii, vývoj koncentrace emisí v ovzduší v ekologii, počet obyvatelstva v demografii nebo objem produkce obilnin v zemědělství. Cílem analýzy časové řady je většinou konstrukce odpovídajícího modelu. To umožní především **porozumnět mechanismu**, na jehož základě jsou generovány sledované údaje (např. porozumnět cyklickému chování v objemu zemědělské produkce). Můžeme si představit, že znalost modelu odpovídá znalosti algoritmu, podle něhož data generuje počítač, přičemž do tohoto algoritmu jsou

**Tabulka 4.1: Měsíční chod srážek na stanici v Košíkách za rok 2007**

| Měsíc | Srážky (v mm) |
|---|---|
| leden | 59,9 |
| únor | 25,0 |
| březen | 83,7 |
| duben | 7,2 |
| květen | 62,0 |
| červen | 155,7 |
| červenec | 61,8 |
| srpen | 132,4 |
| září | 162,8 |
| říjen | 47,9 |
| listopad | 51,9 |
| prosinec | 36,7 |

*Zdroj dat: http://www.hpa1.unas.cz, on-line 27.2.2008*

také zapojena náhodně generována čísla, která dodávají celému procesu náhodný charakter. Jsme sice schopni typ těchto generátorů a jejich zapojení do systému přesně specifikovat, ale nejsme na druhé straně v žádném případě schopni stanovit konkrétní numerické hodnoty produkované těmito generátory v jednotlivých časových okamžicích. Znalost modelu dále umožňuje předpověď budoucího vývoje systému (např. předpověď chodu teploty v následujích letech). Konstrukce modelu také umožní do jisté míry **řídit a optimalizovat** činnost příslušného systému vhodnou volbou vstupních parametrů a počátečních podmínek.

Je nutné ještě zdůraznit, že pod pojmem časová řada rozumíme **statistickou časovou řadu**, např. typu $y_t = \beta_0 + \beta_1 t + \epsilon_t$ ($t$ označuje čas, $\beta_0$ a $\beta_1$ jsou parametry tzv. lineárního trendu a $\epsilon_t$ je náhodná veličina). Na rozdíl od **deterministické časové řady**, např. typu $y_t = cos\,(2\pi f t)$ ($f$ je parametr raprezentující tzv. frekvenci), jejíž chování lze striktně popsat matematickým vzorcem, takže lze např. zkonstruovat její přesnou předpověď.

## 4.3 Typy časových řad

Přestože je podle [6] jakékoli rozdělení časových řad do jednotlivých kategorií poněkud umělé, umožní nám uvědomit si jejich určitá specifika, která musíme brát při jejich analýze

v úvahu. Všechny časové řady se dělí na dva základní typy: časové řady **stochastické** a **deterministické**. Deterministické časové řady v sobě neobsahují žádný prvek náhody; při znalosti toho, jak jsou generovány, je můžeme dokonale a bezchybně předpovídat. Příkladem deterministické časové řady je třeba posloupnost hodnot funkce $\sin(x)$. Naproti tomu stochastické časové řady v sobě prvek náhodnosti obsahují.

Další způsob třídění je následující: časové řady se svou povahou dělí na **ekvidistantní** a **neekvidistantní**. Práce s neekvidistantními časovými řadami je poněkud složitější, protože musíme provádět korekce kvůli proměnlivé délce kroku. Krokem se rozumí interval měření, tedy např. hodina, den, týden, měsíc, rok, atd. Dále je vhodné si uvědomit, zda se jedná o krátkodobou nebo o dlouhodobou časovou řadu (u obou typů nás mohou zajímat rozdílné faktory. Zatímco u krátkodobých (např. měsíčních) časových řad nás často zajímají spíše sezónní vlivy, u dlouhodobých spíše existence dlouhodobých trendů.

Poslední důležité dělení člení časové řady na **řady absolutních ukazatelů** a **řady odvozených charakteristik**. Časová řada absolutních ukazatelů je původní řada, tak jak vznikla pozorováním nebo měřením (např. hodinové ruční měření srážek). Časová řada odvozených charakteristik je naproti tomu nějakým způsobem transformovaná, např. na časovou řadu průměrných produkcí na jednoho zaměstnance. Obvyklý je také výpočet indexů, průměrování a podobně. Je nutné vzít v úvahu, že některé transformace mění charakter časové řady.

## 4.4 Specifické problémy analýzy časových řad

### 4.4.1 Volba časových bodů

Diskrétní časové řady jsou podle [5] tvořeny v určitých nespojitých časových bodech, mohou vznikat trojím způsobem:

- jsou **diskrétní přímo svou povahou** (např. úroda obilí v jednotlivých letech)
- **diskretizací spojité časové řady** (např. teplota na daném místě v danou dobu)
- **akumulací** hodnot za dané časové období (např. denní množství srážek; místo akumulace hodnot se také často provádí jejich **průměrování**)

Je zřejmé, že v některých případech nemáme možnost volby časových bodů pozorování. Pokud ale tuto možnost máme, musíme této volbě věnovat jistou péči a pokusit se často najít kompromis mezi protichůdnými požadavky. Například na jedné straně je

z hlediska numerické jednoduchosti výpočtů při analýze časových řad nežádoucí přespříliš "zhušťovat" počet pozorování. Na druhé straně však pozorování nesmí být natolik řídká, aby nám mohl uniknout některý charakteristický rys dané řady. Jestliže nás zajímají např. změny během průběhu roku, tj. tzv. sezónní fluktuace řady, musíme mít k dispozici alespoň několik pozorování během roku. Co se týče délky intervalu mezi sousedními pozorováními, je obvyklé pracovat s pozorováními v ekvidistantních (stejně vzdálených) časových bodech.

### 4.4.2 Problémy s kalendářem

Za tyto problémy je podle [5] zodpovědný člověk, který zavinil např. tyto problémy:

- různě dlouhé kalendářní měsíce,
- počet víkendů v měsíci (4 nebo 5),
- různý počet pracovních dní,
- pohyblivé svátky (např. Velikonoce),
- přestupný rok.

Tyto problémy je celkem jednoduché očitit, např. počet dní v měsíci, kdy uvažujeme "standardní" měsíc o délce 30 dnů. Za tímto účelem musíme hodnoty za leden vynásobit 30/31, za únor 30/28, atd.

### 4.4.3 Problémy s nesrovnalostí jednotlivých měření

Nesrovnalost některých měření může také souviset s tím, že některé přístroje např. neodeslaly do databáze všechny hodnoty, takže naměřené hodnoty za příslušné období (např. měsíc) se týkají např. 5 přístojů a za další měsíc už 7 přístrojů.

### 4.4.4 Problémy s délkou časových řad

Délkou časové řady budu v dalším textu vždy rozumnět příslušný počet $n$ těch měření, které danou řadu vytvářejí. Proto např. řada měsíčních měření za deset let bude mít délku 120. Je samozřejmé, že s rostoucí délkou časové řady se zvětšuje množství informace informace pro její analýzu. Je zde však nutné upozornit na to, že např. zdvojnásobení

počtu uskutečněných měření nemusí zdvojnásobit množství informace obsažené v těchto měřeních. Délka řady tedy není jednoznačnou mírou informace obsažené v řadě - k tomu je nutné navíc uvažovat ještě vnitřní strukturu řady.

# Kapitola 5

# Základní přístupy k analýze časových řad

Volba metody pro analýzu časové řady závisí na mnoha faktorech, nejdůležitějšími jsou tyto:

**účel analýzy**, kterým je většinou rozpoznání mechanismu generování hodnot časové řady a předpovídání jejího budoucího vývoje,

**typ časové řady**, protože některé metody jsou vhodné, jak uvidíme dále, jen pro časové řady vymezeného typu (např. nebudeme konstruovat Boxův-Jenkinsnův model pro ekonomickou řadu ročních pozorování o délce dvaceti let, která vykazuje zjevný lineární růst),

**zkušenost statistika**, který celou analýzu časové řady provádí; s tím také souvisí použitý software a hardware.

## 5.1 Dekompozice časové řady

Cílem dekompozice časových řad je podle [1] a [5] rozložit časovou řadu na "základní složky": trend, cyklickou, sezónní a náhodnou složku. Předpokládá se, že je snazší identifikovat jednotlivé složky odděleně než celou časovou řadu naráz. Je také snažší takovou řadu extrapolovat - stačí extrapolovat její jednotlivé složky a potom je vhodným způsobem opět "složit" do jedné časové řady. Dekompozice také umožňuje sledovat vývoj časové řady očištěné od některé složky - například sezonní (jedná se o tzv. sezonní očištění).

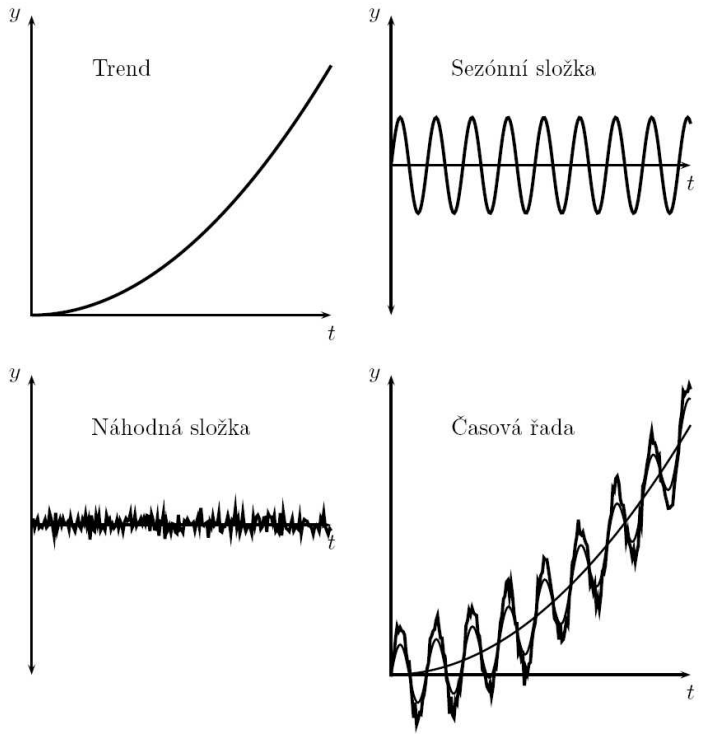
Při dekompocizi časových řad se implicitně předpokládá model časové řady, který nezávisí na žádných vysvětlujících proměnných, pouze na čase (alternativně je možno říct,

že čas a jeho transformace jsou jedinými vysvětlujícími proměnnými). Jedná se tedy o nejjednodušší modely časové řady, které se používají především pro krátkodobé předpovědi, jako nástřelné modely pro expertní odhad (aby expert mohl říct: "Takto se to vyvíjet nebude, protože tady začne působit takový a takový vliv.") a pro základní srovnání úspěšnosti vyspělejších metod.

Za hlavní složky časové řady jsou podle [1] považovány trend, sezonní, cyklická a náhodná složka.

- **Trend.** Trend ($T_t$) odpovídá hlavním tendencím dlouhodobého vývoje statistického ukazatele, který časová řada popisuje.
- **Sezonní složka.** Sezonní složka ($S_t$) odpovídá periodicky se opakujícím odchylkám od trendu, ke kterým dochází pravidelně v rámci každého roku (tedy s roční periodou). Mezi hlavní vlivy, které utvářejí sezonní složku, patří střídání ročních období, pravidelně se opakující svátky, různé délky jednotlivých měsíců apod.
- **Cyklická složka.** Cyklická složka ($C_t$) je nejspornější částí časové řady. Odpovídá dlouhodobým, často nepravidelným cyklům s proměnlivou periodou (délkou) i amplitudou ("výškou"). Modelování cyklické složky je proto poměrně obtížné. V krátkém období je možné její vliv zanedbat. Protože se dekompoziční metody používají především na krátkodobé a střednědobé předpovědi, bývá cyklická složka někdy zanedbána, tj. zahrnuta do trendu.
- **Náhodná složka.** Náhodná složka ($E_t$) je také nazývána reziduální, zbytková, iregulární, nesystematická. Jde o náhodné pohyby bez systematického charakteru. Zahrnuje také chyby měření a chyby ze zaokrouhlování při výpočtech. Při dekompozici časových řad se předpokládá, že se jedná o bílý šum, často dokonce nekorelovaný normálně rozdělený **bílý šum**. Bílý šum se často označuje jako i-tý člen časové řady s charakterem nezávislých realizací normálně rozložené náhodné veličiny se střední hodnotou $\mu = 0$ a konstantním rozptylem. Taková řada je svým způsobem "nejnáhodnejší" ze všech "rozumných" časových řad, protože o jejím příštím členu v podstatě nevíme na základě předchozího průbehu víc, než že půjde o "nějaké číslo kolem nuly". Název bílý šum vznikl z toho, že tato časový řada obsahuje rovnoměrný podíl frekvenčních složek všech vlnových délek, podobně jako bílé světlo obsahuje složky všech barev spektra.

Není nutné, aby byla v každé časové řadě zastoupena každá z těchto složek. Jednotlivé složky časové řady a jejich vizualizace je na Obr 1.



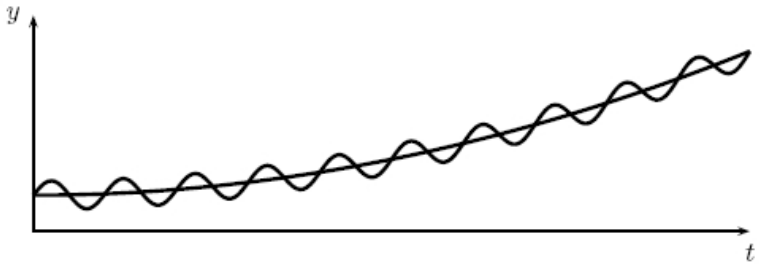

Obr 4.: Složky časové řady, zdroj: [6], str. 54

## 5.2 Typ dekompozice

Při dekompozici se podle [5] vychází ze tří různých modelů časové řady: aditivního, multiplikativního a smíšeného. Tyto modely specifikují, jakým způsobem jsou jednotlivé složky časové řady "skloubeny" dohromady.

- **Aditivní model** předpokládá, že výsledná časová řada je součtem jednotlivých složek, tedy že

$$Y_t = T_t + C_t + S_t + E_t. \tag{5.1}$$



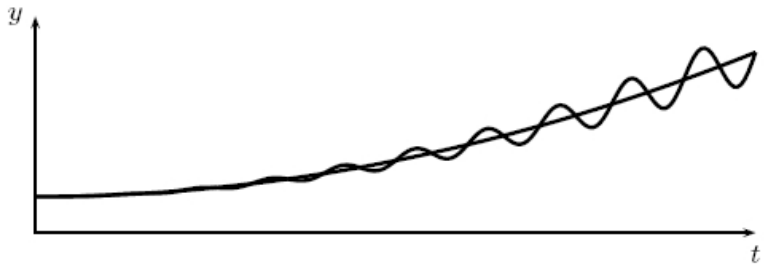

Obr 5.: Aditivní model časové řady, zdroj: [6], str. 55

V tomto modelu je každá ze složek uváděna v absolutní hodnotě a má stenou jednotku jako celá sledovaný ukazatel $Y_t$.

- **Multiplikativní model** na druhou stranu předpokládá, že výsledná časová řada je spíše součinem jednotlivých složek:

$$Y_t = T_t C_t S_t E_t. \tag{5.2}$$

U multiplikativních časových řad přisuzujeme absolutní hodnotu (a tedy i jednotku) pouze trendu; ostatní složky považujeme za bezrozměrné koeficienty.




Obr 6.: Multiplikativní model časové řady, zdroj: [6], str. 55

- **Smíšený model** je vlastně jen kombinací obou předchozích přístupů. Některé složky mohou být v součtu, jiné v součinu. Typickým příkladem může být třeba takovýto model časové řady:

$$Y_t = T_t C_t S_t + E_t. \tag{5.3}$$

Smíšené časové řady se identifikují nejobtížněji.

To, jaký model stanovíme, vychází z našich znalostí o statistickém ukazateli, který časová řada reprezentuje. Užitečný je také pohled na graf časové řady. Pokud vidíme, že se velikost amplitudy sezonní složky mění s velikostí trendu, můžeme usuzovat na multiplikativní časovou řadu.

## 5.3 Metody dekompozice

Při dekompozici se podle [5] snažíme nejdříve identifikovat trend a potom teprve sezonní složku. Někdy se však postupuje opačně: řada se nejdříve zbaví sezónních vlivů a pak se v takto očištěné řadě hledá trend nebo její závislost na jiných, vysvětlujících proměnných. K identifikaci trendu se používají především čtyři následující metody:

- proložení matematickou křivkou

- vyrovnání metodou klouzavých průměrů
- exponenciální vyrovnání

Prokládání časových řad zvolenou matematickou křivkou je souhrnně nazýváno neadaptivními metodami, klouzavé průměry a expoenenciální vyrovnání metodami adaptivními.

**Neadaptivní metody** jsou takové metody, které časovou řadu vysvětlí jako celek pomocí několika v čase konstatních parametrů. Takový model se jen velmi pomalu (nebo vůbec) přizpůsobuje změnám v charakteru časové řady. Je zřejmé, že tyto metody nelze používat v žádném případě na indentifikaci modelu veličiny, u níž není zaručena podmínka, že se vnější vlivy nemění. Na druhou stranu umožňují tyto metody (aspoň z technického hlediska) snadnou předpověď i pro delší období. Mezi neadaptivní modely patří i regresní a ekonometrické modely.

**Adaptivní metody** se naopak přizpůsobují změnám v charakteru analyzované veličiny poměrně rychle. Je to způsobeno jejich charakterem. Většinou se jimi zpracovávají postupně malé kusy řady nebo se používá metody "zapomínání" starých hodnot. Flexibilita těchto metod umožňuje rychle se adaptovat na změnu, poskytovat kvalitní krátkodobé předpovědi, ale většinou vylučuje možnost kvalitních dlouhodobých předpovědí. Mezi další adaptivní metody patří např. i Box-Jenkinsovské metody.

## 5.4 Vyrovnání trendu matematickou křivkou

První třídu metod, které se používají při dekompozici časové řady jsou podle [5] a [6] tzv. **neadaptivní metody**. Tyto metody vycházejí z předpokladu, že se trend po celou námi sledovanou dobu nemění a že je možné ho popsat některým typem matematické křivky. Celá úloha identifikace trendu se potom redukuje na výběr správného typu matematické křivky a odhadu jejích parametrů. Vycházíme přitom z jednoduchého modelu časové řady

$$Y_t = T_t + E_t. \tag{5.4}$$

kde $T_t$ je hodnota trendu, který závisí jen na čase, a $E_t$ hodnota reziduální složky (její další dekompozice, tj. hledání sezónnosti a testování náhodné složky je prováděno až v dalším kroku).

Předpověd budoucích hodnot trendu je dána prostou extrapolací dosazením příslušné hodnoty $t > n$ do vzorce matematické křivky, popisující trend.

Mezi základní typy křivek, které se při odhadu trendu používají, patří:

- polynomy (proložení konstantou, přímkou, kvadratickou funkcí, atd.)
- exponenciální a modifikovaná exponenciální funkce
- logistická křivka
- Gompertzova křivka

### 5.4.1 Konstantní trend

Nejjednodušší případ trendu je konstantní trend, kdy sledovaná veličina v zásadě ani neroste, ani neklesá, ale osciluje okolo své průměrné hodnoty. Tento trend tedy můžeme popsat základním vztahem

$$T_t = b_0 \quad t = 1, 2, \ldots, n. \tag{5.5}$$

Normální rovnice pro výpočet parametru $\hat{b}_0$ má následující tvar:

$$\frac{\partial \sum_{t=1}^{n} (y_t - \hat{b}_0)^2}{\partial \hat{b}_0} = 0. \tag{5.6}$$

Po výpočtu normálních rovnic získáme odhad $\hat{b}_0$ o parametru $b_0$ jako

$$\hat{b}_0 = \overline{y} = \frac{1}{n} \sum_{t=1}^{n} y_t. \tag{5.7}$$

Alternativně je možné použít lineárního regresního modelu. Matice plánu $\mathbf{X}$ bude v tomto případě sloupcový vektor o délce počtu pozorování $n$, tvořený jedničkami.

Předpovězená hodnota trendu $\hat{y}_T$ pro čas $T$ námi sledovaného ukazatele v čase $T$ má tvar

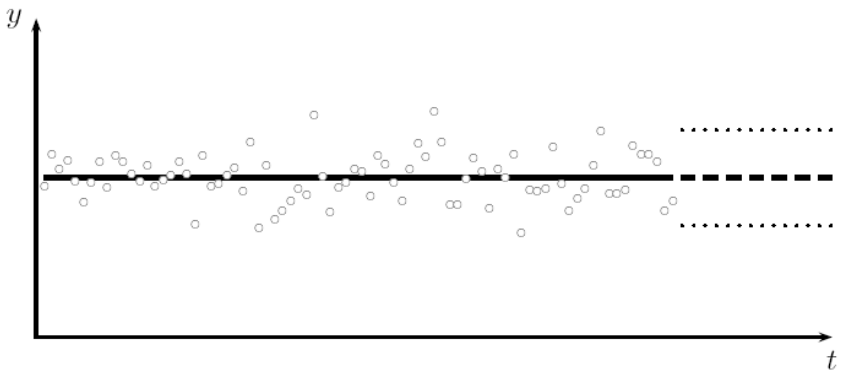
$$\hat{y}_T = \hat{b}_0. \tag{5.8}$$

K této bodové předpovědi se podle [6] konstruuje procentní předpovědní interval

$$\left( \hat{y}_T - t_{n-1}(\alpha)\, s \sqrt{1 + \frac{1}{n}}, \quad \hat{y}_T + t_{n-1}(\alpha)\, s \sqrt{1 + \frac{1}{n}} \right), \tag{5.9}$$

kde $t_{n-1}(\alpha)$ je kritická hodnota Studentova rozdělení na hladině významnosti $\alpha$ a $s$ je odhad směrodatné odchylky rozptylu náhodné složky $u_t$, získaný podle [6], tedy v tomto případě ze vzorce

$$s = \sqrt{\sum_{t=1}^{n} \frac{(y_t - \bar{y}_t)^2}{n-1}}. \tag{5.10}$$




Obr 7.: Konstantní trend, zdroj: [6], str. 62

### 5.4.2 Lineární trend

Často se můžeme domnívat, že se určitá veličina vyvíjí v čase zhruba lineárně - předpokládáme tedy lineární trend. Model tohoto trendu je možné popsat vztahem

$$T_t = b_0 + b_1 t \quad t = 1, 2, \ldots, n. \tag{5.11}$$

Normální rovnice pro výpočet parametrů $\hat{b}_0$ a $\hat{b}_1$ mají tvar:

$$\frac{\partial \sum_{t=1}^{n} (y_t - \hat{b}_0 - \hat{b}_1 t)^2}{\partial \hat{b}_0} = 0, \tag{5.12}$$

$$\frac{\partial \sum_{t=1}^{n} (y_t - \hat{b}_0 - \hat{b}_1 t)^2}{\partial \hat{b}_1} = 0. \tag{5.13}$$

Odhady $\hat{b}_0$ a $\hat{b}_1$ parametrů $b_0$ a $b_1$ můžeme získat výpočtem těchto rovnic, konkrétně tedy:

$$\hat{b}_1 = \frac{\sum_{t=1}^{n} t y_t - \overline{t} \sum_{t=1}^{n} y_t}{\sum_{t=1}^{n} t^2 - n\overline{t}^2}, \tag{5.14}$$

$$\hat{b}_0 = \overline{y} - \hat{b}_1 \overline{t}, \tag{5.15}$$

kde $\overline{y}$ je průměrná hodntota vysvětlované proměnné $y_t$ a $\overline{t}$ je průměrná hodnota časové proměnné.

Jinou možností, jak odhadnout parametry, je použít lineární regresní model. Matice plánu $\boldsymbol{X}$ bude mít v tomto případě dva sloupce, první bude sloupec složený z jedniček, druhý z hodnot časové proměnné $t$; délka obou sloupcových vektorů bude stejná jako

počet pozorování $n$ vysvětlované proměnné $y_t$. Vektor parametrů $\boldsymbol{b}$ obsahuje po řadě oba parametry, tedy

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}.$$

Vyrovnaná hodnota trendu i její předpověď $\hat{y}_T$ na čas $T > n$ je potom dána vztahem

$$\hat{b}_0 + \hat{b}_1 T. \tag{5.16}$$

Odpovídající předpovědní interval má tvar

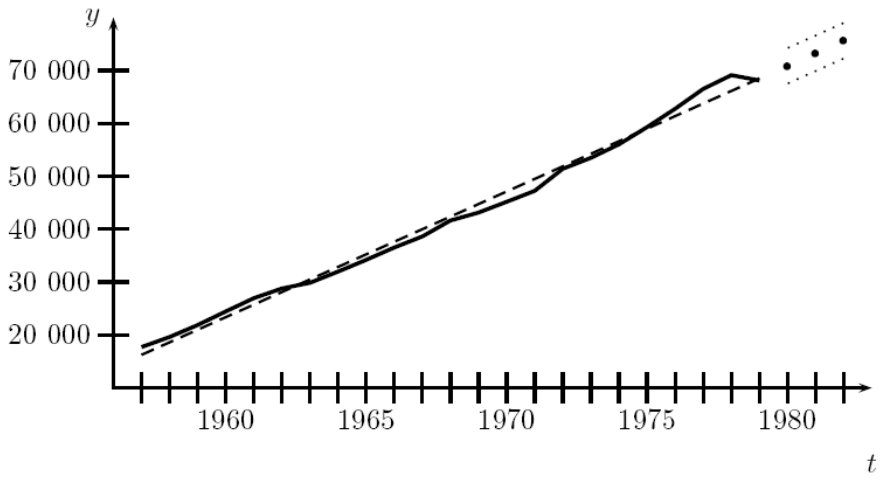
$$\left(\hat{y}_T - t_{n-2}\left(\alpha\right) s f_T, \quad \hat{y}_T + t_{n-2}\left(\alpha\right) s f_T\right), \tag{5.17}$$

kde $s$ je směrodatná odchylka náhodné složky vypočtená podle [6], tedy ze vzorce

$$s = \sqrt{\frac{\sum_{t=1}^{n} \left(y_t - \hat{y}_2\right)^2}{n-2}}, \tag{5.18}$$

a $f_T$ hodnota je

$$f_T = \sqrt{1 + \frac{1}{n} + \frac{\left(T - \overline{t}\right)^2}{\sum_{t=1}^{n} t^2 - n\overline{t}^2}}. \tag{5.19}$$




Obr 8.: Lineární trend, zdroj: [6], str. 64

### 5.4.3 Polynomiální trend

Polynomiální trend je zobecnění dvou předchozích uvedených trendů. Předpokládáme při něm, že trend je dán polynomem $k$-tého řádu, tedy modelem

$$T_t = b_0 + b_1 t + b_2 t^2 + \ldots + b_k t^k \quad t = 1, 2, \ldots, n. \tag{5.20}$$

Při obecném počtu parametrů (stupni) polynomu je výhodnější pracovat přímo s lineárním regresním modelem, ne s odhady na základě normálních rovnic. Matice plánu $\mathbf{X}$ má tvar

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 4 & \ldots & 2^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & n & n^2 & \ldots & n^k \end{pmatrix}, \tag{5.21}$$

kde $\boldsymbol{t}$ je sloupcový vektor času.

Odhad $\hat{b}$ a parametrů $b$ je potom možné najít jako

$$\boldsymbol{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}. \tag{5.22}$$

Vyrovnaná nebo předpovězená hodnota $\hat{y}_T$ v čase $T$ má tvar

$$\hat{y}_T = \left(1, T, T^2, \ldots, T^k\right)\hat{\boldsymbol{b}}, \tag{5.23}$$

její interval spolehlivosti pak tvar

$$\left(\hat{y}_T - t_{n-(k+1)}\left(\alpha\right) s f_T, \quad \hat{y}_T + t_{n-(k+1)}\left(\alpha\right) s f_T\right), \tag{5.24}$$

kde $s$ je směrodatná odchylka reziduí

$$s = \sqrt{\sum \frac{\left(y_t - \hat{y}_t\right)^2}{n - (k+1)}}, \tag{5.25}$$

a $f_T$ je statistika

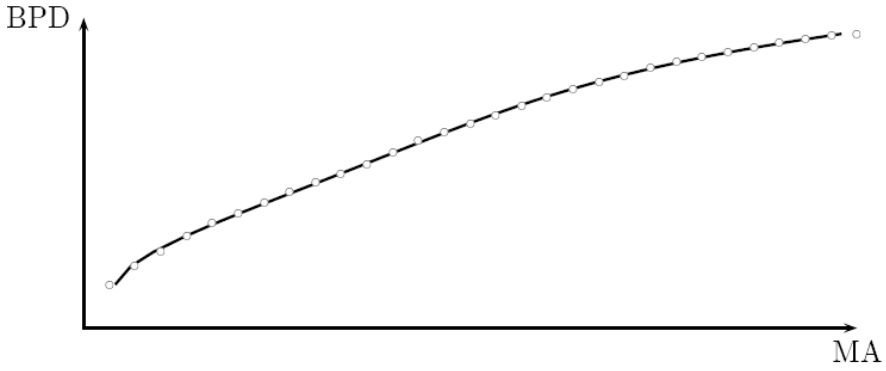
$$f_T = \sqrt{1 + (1, T, T^2, \ldots, T^k)(\mathbf{X}'\mathbf{X})^{-1}(1, T, T^2, \ldots, T^k)'}. \tag{5.26}$$

Nutno připomenout, že vektor $(1, T, T^2, \ldots, T^k)$ je obecně jedním řádkem matice plánu (skutečné nebo předpovědní).

Vektor vyrovnaných hodnot $\hat{\boldsymbol{y}}$, resp. předpovědí $\hat{\boldsymbol{y}}_P$ lze pak samozřejmě získat ze vzahu

$$\hat{\boldsymbol{y}} = \mathbf{X}\hat{\boldsymbol{b}}, \quad \hat{\boldsymbol{y}}^P = \mathbf{X}^P\hat{\boldsymbol{b}}. \tag{5.27}$$

Při volbě stupně polynomu je třeba postupovat velice opatrně. Vyšší stupeň sice zajistí těsnější proložení empirických hodnot křivkou, vede však také k výrazné nestabilitě "trendu". Vyšší polynomy se většinou vůbec nehodí k extrapolacím.




Obr 9.: Polynomiální trend, zdroj: [6], str. 71

### 5.4.4 Exponenciální trend

Model exponenciálního trendu má tvar

$$T_t = b_0 b_1^t, \quad t = 1, 2, \ldots, n, b_1 > 0. \tag{5.28}$$

Je pro něj typické, že podíl dvou sousedních hodnot $T_t$ a $T_{t-1}$ je konstantní a je roven koeficientu $b_1$.

Při odhadu parametrů tohoto trendu začneme tím, že model převedeme logaritmováním na lineární regresní model do tvaru

$$\log T_t = \log b_0 + t \log b_1. \tag{5.29}$$

Takto upravený model je v podstatě modelem lineárního trendu. Můžeme ho transformovat do tvaru

$$T_t^* = b_0^* + b_1^* t, \tag{5.30}$$

kde $T_t^* = \log y_t$, $b_0^* = \log b_0$ a $b_1^* = \log b_1$, a odhadnout stejným postupem jako lineární a verifikovat trend. Odhad i verifikace se provádějí v logaritmovaném tvaru. Teprve výsledky, tj. odhady hodnot parametrů, vyrovnané hodnoty, předpovědi a konfidenční intervaly se opět odlogaritmovávají.

Při použití lineárního regresního modelu použijeme jako vysvětlovanou veličinu $\boldsymbol{y}$

a matici plánu $\mathbf{X}$ hodnoty

$$\boldsymbol{y} = \begin{pmatrix} \log y_1 \\ \log y_2 \\ \vdots \\ \log y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix}, \tag{5.31}$$

vektor parametrů $\hat{\boldsymbol{b}}$ bude mít tvar

$$\hat{\boldsymbol{b}} = \begin{pmatrix} \log \hat{b}_0 \\ \log \hat{b}_1 \end{pmatrix}. \tag{5.32}$$

Praktické zkušenosti ukazují, že mnohem lepší výsledky jsou však dosaženy při použití **vážené regrese**, která je podrobněji popsána v [6], str. 31-32. Logaritmování totiž snižuje relativní váhu "velkých" a zvyšuje relativní váhu "malých" pozorování oproti netransformovanému modelu. Parametry jsou tedy bez vážení odhadnuty tak, aby vystihly spíše malé (v našem případě staré) hodnoty na úkor větších a novějších. Vážení také eliminuje dopady na konfidenční interval. Jako vhodné váhy se ukazují $w_t$ dané vztahem

$$w_t = y_t^2.$$

Pokud chceme jednotlivým pozorováním přiřadit vlastní váhy $v_t$, které odpovídají např. jejich geografickému, můžeme je zařadit do předchozího vztahu:

$$w_t = y_t^2 v_t.$$

Pro odhad parametrů váženého modelu můžeme použít vzorce odvozené z normálních rovnic

$$\log \hat{b}_1 = \frac{\sum_{t=1}^{n} (t - \overline{t}) y_t^2 \log y_t}{\sum_{t=1}^{n} (t - \overline{t})^2 y_t^2}, \tag{5.33}$$

$$\log \hat{b}_0 = \overline{\log y} - \overline{t} \log \hat{b}_1, \tag{5.34}$$

kde

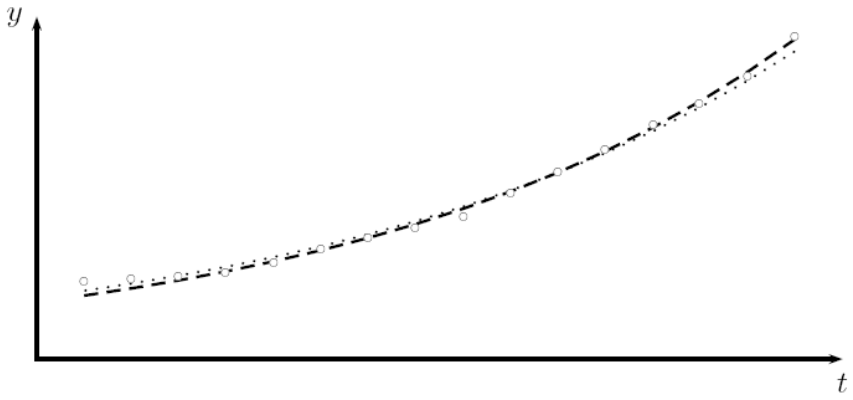
$$\overline{\log y} = \sum_{t=1}^{n} \frac{\log y_t}{n}, \quad \overline{t} = \sum_{t=1}^{n} \frac{t}{n}.$$

Odhady skutečných parametrů pak získáme odlogaritmováním hodnot $\log \hat{b}_0$ a $\log \hat{b}_1$.

Lineární regresní odhad váženého modelu se snadno provede s následující transformací matice plánu $\mathbf{X}^*$ a vektoru vysvětlované proměnné $\boldsymbol{y}^*$:

$$\boldsymbol{y}^* = \begin{pmatrix} y_1^2 \log y_1 \\ y_2^2 \log y_2 \\ \vdots \\ y_n^2 \log y_n \end{pmatrix}, \quad \mathbf{X}^* = \begin{pmatrix} y_1^2 & y_1^2 \\ y_2^2 & 2y_2^2 \\ \vdots & \vdots \\ y_n^2 & n y_n^2 \end{pmatrix}. \tag{5.35}$$

Předpovědi a konfidenční intervaly jsou (při výše uvedené transformaci) stejné jako u lineárního trendu.



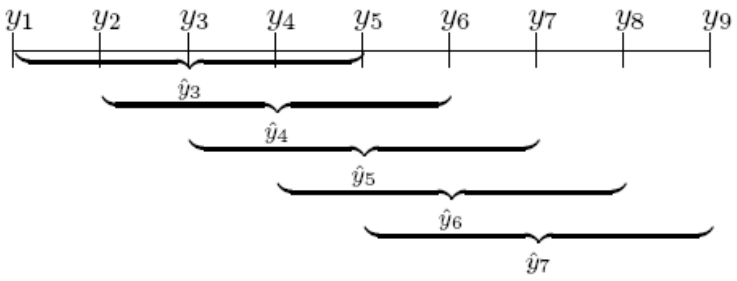

Obr 10.: Exponenciální trend, zdroj: [6], str. 77

## 5.5 Klouzavé průměry

Metoda klouzavých průměrů se podle [6] řadí mezi adaptivní přístupy k trendové složce. Adaptivní přístupy lze obecně charakterizovat tak, že jsou schopny pracovat s trendovými složkami, které mění v čase globálně svůj charakter, takže pro jejich popis nelze použít žádnou matematicky definovatelnou křivku s neměnnými parametry.

V některých případech se trend v námi pozorované časové řadě mění a není tedy možné vyrovnat jej na celé této časové řadě jednou matematickou křivkou. Jedno z možných řešení tohoto problému je rozdělit časovu řadu do jednotlivých úseků a potom vyrovnat každý z těchto úseků zvlášť vhodnou matematickou křivkou. Nicméně i to může být někdy velmi obtížné, ba dokonce neřešitelné, protože průběh trendu se může měnit spojitě. V tomto případě je vhodné použít nějakou adaptivní metodu.

Klouzavé průměrování je založeno na jednoduchém postupu. V časové řadě o $n$ prvcích nejdříve zprůměrujeme vhodně vybraným typem průměru (aritmetický, vážený aritmetický, atd.) prvních $2m + 1$ hodnot a takto získané hodnotě přiřadíme vhodný index (většinou se hodnota centruje do prostřed intervalu $2m+1$ průměrovaných hodnot (proto se zpravidla volí délka $2m+1$ liché tak, aby index padl na celé číslo); potom průměrujeme stejně dalších $2m + 1$ hodnot, posunutých o jeden člen. Vyrovnání prvních a posledních $m$ hodnot, stejně jako případné předpovědi trendu se provádějí zvláštními metodami.




Obr 11.: Ukázka principu klouzavých průměrů, zdroj: [6], str. 86

Co se týče vlastních průměrů, používá se většinou vážený aritmetický průměr, který se liší podle použitých vah.

### 5.5.1 Jednoduché klouzavé průměry

Nejjednodnodušší typ klouzavého průměru je prostý (nevážený) aritmetický průměr. V případě, že je centrován, tj. výsledek je umístěn doprostřed, platí při délce $2m+1$ vztah

$$\hat{y}_t = \frac{1}{2m+1}\left(y_{t-m} + y_{t-(m-1)} + \ldots + y_{t+m}\right). \tag{5.36}$$

Jiné často používané váhy jsou $1/(2(2m+1))$ pro obě krajní průměrované hodnoty a $2/(2(2m+1))$ pro ostatní průměrované hodnoty, tedy

$$\hat{y}_t = \frac{1}{2m+1}\left(y_{t-m} + 2y_{t-(m-1)} + 2y_{t-(m-2)} + \ldots + 2y_{t+(m-1)} + y_{t+m}\right). \tag{5.37}$$

Tento vztah vychází z centrování dvou sousedních prostých průměrů o sudé délce.

V některých specifických odvětvích geografické analýzy (např. v socio-ekonomických analýzách) se často používají necentrované průměry, ať už prosté

$$\hat{y}_t = \frac{1}{m}\left(y_t + y_{t-1} + \ldots + y_{t-(m-1)}\right) \tag{5.38}$$

nebo vážené, např. s váhami klesajícími do minulosti lineárně

$$\hat{y}_t = \frac{2}{m(m+1)}\left(y_t + y_{t-1} + \ldots + y_{t-(m-1)}\right) \tag{5.39}$$

nebo exponenciálně

$$\hat{y}_t = \frac{1}{\sum_{i=0}^{m-1}\alpha^i}\left(y_t + \alpha y_{t-1} + \alpha^2 y_{t-2} + \ldots + \alpha^{m-1} y_{t-(m-1)}\right), \quad 0 < \alpha < 1, \tag{5.40}$$

kde $\alpha$ je **koeficient zapomínání**. Výhodou těchto necentrovaných průměrů je fakt, že jejich délka $m$ nemusí být nutně lichá, a výpočetní jednoduchost; značnou nevýhodou je ale zkreslení (aspoň u prostého průměru).

U většiny těchto průměrů je velmi obtížné nebo i nemožné jak vyrovnávat krajní hodnoty, tak i extrapolovat trend do budoucna. Proto je zpravidla výhodnější používat výpočetně poněkud komplikovanější **polynomiální klouzavé průměry**.

### 5.5.2 Polynomiální klouzavé průměry

Metoda polynomiálních klouzavých průměrů vychází z předpokladu, že většinu funkcí je možné proložit polynomem vhodného řádu. Na rozdíl od polynomiálního trendu představeného v oddíle 3.6.1 se při použití této metody nesnažíme proložit polynomem celou časovou řadu naráz, ale postupně - klouzavě vždy $2m + 1$ členů tak, jak to odpovídá obrázku 8. Většinou se používají polynomy maximálně čtvrtého nebo pátého řádu.

Znamená to tedy, že vezmeme prvních $2m + 1$ pozorování a proložíme je zvoleným polynomem, tj. odhadneme parametry polynomu tak, aby součet čtverců reziduí byl minimální. Potom se posuneme o jedno pozorování a provedeme totéž. V každém kroku spočítáme vyrovnanou hodnotu uprostřed vyrovnávané oblasti - ta je pak vyrovnáním klouzavými průměry.

Výpočetně vycházíme z metody nejmenších čtverců nebo přímo lineárního regresního modelu. Řekněme, že se snažíme prokládat trend klouzavými průměry o délce $2m + 1$ a o stupni polynomu $k$. Potom hledáme takové odhady parametrů $b_0, b_1, \ldots, b_k$, které minimalizují výraz

$$\sum_{\tau=-m}^{m} \left( y_{t+\tau} - \left( \hat{b}_0 + \hat{b}_1 \tau + \hat{b}_2 \tau^2 + \ldots + \hat{b}_k \tau^k \right) \right)^2 \rightarrow \min . \tag{5.41}$$

Opět zde vycházíme z centrování okolo hodnoty $y_t$, prokládáme tedy hodnoty

$$y_{t-m}, y_{t-m+1}, \ldots, y_t, y_{t+1}, \ldots, y_{t+m}. \tag{5.42}$$

Dále tuto problematiku již nemá význam kvůli značné složitosti rozebírat, podrobněji je tato látka popsána v [5], str. 42-57 a [6], str. 87-92.

## 5.6 Analýza sezónní složky

Cíle analýzy sezónní složky mohou být podle [5] alternativně dva:

- **Získat dodatečné informace o vývoji časové řady.** Odhad sezónní složky prohlubuje znalost o chování sledovaného ukazatele. V mnoha případech je tato znalost neméně důležitá než znalost trendu. Typicky je významná při plánování skladových kapacit apod.

- **Sezónně očistit časovou řadu.** Pro mnoho aplikací a analýz je výhodnější nejprve časovou řadu očistit o sezónní vlivy a pak dál pracovat s takto upravenou řadou. Srovnání vývoje mezi různými obdobími není v tomto případě vychýleno periodicky se opakujícími vlivy. Typickým případem je sezónní očišťování inflace nebo vývoje HDP.

### 5.6.1 Jednoduché metody odhadu sezónnosti

Nejjednodušší přístup k odhadu sezónní složky podle [5] vychází z předpokladu, že časová řada je tvořena trendem, sezónní a náhodnou složkou. Předpokládejme, že trend jsme již odhadli některou z metod popsaných výše, jako $\hat{T}_t$.Potom odečtením trendu od časové řady získáme odhad sezónní a náhodné složky pro **aditivní model sezónnosti**

$$S_t + E_t = y_t - \hat{T}_t. \tag{5.43}$$

Podobného výsledku dosáhneme vydělením řady trendem pro **multiplikativní model**

$$S_t E_t = \frac{y_t}{\hat{T}_t}. \tag{5.44}$$

Podobným způsobem můžeme získat nějakou kombinaci sezónní a náhodné složky i pro **smíšené modely sezónnosti**. Vliv náhodné složky můžeme zdárně potlačit průměrováním hodnot $S_t + E_t$, resp. $S_t E_t$, odpovídajících stejným sezónním obdobím. Výsledkem je pak odhad sezónní složky $\hat{S}_i$. V tomto případě jsou sezónní faktory odhadnuty neadaptivně, tj. jsou stejné pro všechny odpovídající si sezóny v celé časové řadě.

Na sezónní složku se často kladou další nároky: požadujeme, aby součet aditivních sezónních složek byl 0,

$$\hat{S}_1 + \hat{S}_2 + \ldots + \hat{S}_m = 0, \tag{5.45}$$

zatímco součet multiplikativních sezónních faktorů byl roven počtu sezón $m$

$$\hat{S}_1 + \hat{S}_2 + \ldots + \hat{S}_m = m. \tag{5.46}$$

K tomu se používá **normalizace** sezónní složky. Normalizace pro multiplikativní model má tvar

$$\hat{S}'_i = \frac{m}{\sum_{j=1}^{m} \hat{S}_j} \hat{S}_i. \tag{5.47}$$

Vyrovnané hodnoty řady pak mají pro **adaptivní modely** tvar

$$y_t = T_t + S_t \tag{5.48}$$

a pro **multiplikativní tvar**

$$y_t = T_t S_t. \tag{5.49}$$

## 5.7 Periodogram

Periodogram $l(\omega)$ časové řady $y_1, \ldots, y_n$ (nabývající reálných hodnot) se podle [1] definuje jako funkce proměnné $\omega$ tvaru:

$$l(\omega) = \frac{1}{4\pi}\left(a^2(\omega) + b^2(\omega)\right), \quad -\pi \leq \omega \leq \pi, \tag{5.50}$$

kde

$$a(\omega) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} y_t \cos(\omega t) \tag{5.51}$$

$$b(\omega) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} y_t \sin(\omega t). \tag{5.52}$$

Periodogram zavedl A. Schuster jako nástroj pro nalezení významných periodických složek v dané časové řadě.

## 5.8 Testy náhodnosti

Někdy se stane, že časová řada předložená k analýze nevykazuje při zběžné prohlídce nebo grafickém znázornění výskyt žádné systematické složky, takže se zdá, že je tvořena pouze bílým šumem. Podle [5] se však pro jistotu provádí objektivní statické testy, které tuto hypotézu potvrdí. V těchto testech se jako $H_0$ testuje, zde předložená testování jsou realizace vzájemně nezávislých stejně rozdělených náhodných veličin, které nemusí mít jako bílý šum nulovou střední hodnotu. Při zamítnutí této hypotézy zřejmě analyzovaná reziduální složka nemůže být klasickým bílým šumem.

### 5.8.1 Test založený na bodech zvratu

Bod $y_t$ je podle [5] horním bodem zvratu uvažované časové řady, když $y_{t-1} < y_t > y_{t+1}, t = 2, \ldots, n-1$. Analogicky se definuje dolní bod zvratu. Nechť $r$ označuje celkový

počet horních a dolních bodů zvratu dohromady. Lze odvodit, že:

$$E(r) = \frac{2(n-2)}{3}, \tag{5.53}$$

$$var(r) = \frac{16(n-29)}{90}. \tag{5.54}$$

Při větším $n$ hypotézu $H_0$ zamítáme, když

$$\frac{|r - 2(n-2)/3|}{\sqrt{(16n-29)/90}} \geq u(p/2). \tag{5.55}$$

### 5.8.2 Jednovýběrový Wilcoxonův test

Jde o neparametrickou obdobu testu správnosti - $H_0 : \mu = \tilde{x}$, $H_1 : \mu \neq \tilde{x}$ . Od prvků výběru se odečte správná hodnota a absolutní hodnoty rozdílů seřadíme do neklesající posloupnosti. Každé hodnotě přiřadíme pořadové číslo (pořadí). Vytvoříme sumu pořadí nezáporných prvků $S^+$ a sumu pořadí záporných prvků $S^-$. Při shodě pořadí se použije průměrné pořadí. Je-li menší číslo z dvojice $S^+$ a $S^-$ menší nebo rovno tabelované hodnotě $W_{(n,0,05)}$, nulová hypotéza o správnosti se zamítá.

# Kapitola 6

# Postup práce

V této kapitole jsou uvedeny veškeré postupy použité při zpracování dat. Dále zde jsou příklady zpracování, ukázky surových a opravených dat. Nakonec zde uvádím i postup zpracování analýzy časových řad pro jednotlivé datasety buď v projektu R nebo v programu Microsoft Excel XP Professional. K analýze byly použity tyto datasety[1] z roku 2007 (tj. od 1.1. do 31.12.):

- ALA1_pokus_opr
- ALA_Prutok_pokus_opr
- Fiedler_all
- Halenkovice_srazky_manualni_mereni[2]
- Kosiky[3] (data ze stanice v Košíkách).

## 6.1 Zpracování a oprava surových dat

Jelikož data byla naměřena na přístrojích pracujících v terénu, je logické, že se občas vyskytne nějaký výpadek. Občas může přístroj měření vynechat nebo prostě nebude fun-

[1]Jako výčet jsou uvedeny názvy opravených datasetů z jednotlivých přístrojů tak, jak jsou uloženy na přiloženém CD ve složce *Source*.

[2]Tento jediný dataset je v rozmězí sedni let, konkrétně od 1.1.2001 do 31.12.2007

[3]Dataset *Kosiky* mi byl poskytnut pro potřeby bakalářské práce panem Petrem Malinou, který vlastní soukromou amatérskou meteostanici v obci Košíky (cca 2,5 km jihozápadně od Halenkovic). Všechna jeho měření a výsledky analýz jsou k dispozici na jeho stránkách věnovaných meteorogii, *http://www.hpa1.unas.cz/*. Tímto mu znovu děkuji za poskytnutá data a za souhlas s jejich prezentací.

govat po delší dobu, například dojde baterie nebo přestane fungovat snímací čidlo. Tyto výpadky je potřeba pro další zpracování odfiltrovat.

Přístroje ukládaly ve většině případů naměřená data do formátů *.txt nebo *.csv. Tyto formáty lze bez větších problémů importovat například do programu Microsoft Excel (nebo Open Office Calc), aniž by se poškodila struktura dat.

```
20071005160000.0,1,16.6,15.3,14.5,14.6,14.4,,30.9,38.0,-6.6,-1.9,0.2,13.0
20071005161500.0,1,15.9,15.0,14.3,13.8,14.3,,30.6,38.2,-5.7,-1.3,0.2,13.0
20071005163000.0,1,16.0,15.1,14.3,13.8,14.0,,30.7,38.3,-5.5,-1.0,0.2,13.0
20071005164500.0,1,16.0,15.2,14.3,13.7,14.1,,30.8,38.5,-5.4,-1.0,0.2,13.0
20071005170000.0,1,16.0,15.3,14.5,13.8,14.1,,30.8,38.1,-5.3,-1.0,0.2,13.0
20071005171500.0,1,15.8,15.1,14.3,13.6,13.9,,30.7,38.3,-5.1,-0.8,0.2,13.0
20071005173000.0,1,15.7,15.0,14.2,13.6,13.9,,30.7,38.4,-5.1,-0.6,0.2,13.0
20071005174500.0,1,15.6,15.0,14.2,13.6,13.8,,30.6,38.4,-4.8,-0.4,0.2,13.0
20071005180000.0,1,15.5,15.0,14.2,13.6,13.8,,30.7,38.1,-4.6,-0.3,0.2,13.0
```

Obr 12.: Ukázka surových dat

Nejdříve bylo nutné naimportovat data do programu Microsoft Excel. Pro importování lze použít příkaz na kartě **Data**, záložka **Importovat externí data**, **Importovat data**. Vyberu dataset, který chci zpracovat a provedu import dat. Jako oddělovač byla v datasetech použita čárka, kterou bylo nutné nastavit při druhém kroku importu. Takto naimportovaný dataset uložím do souboru s příponou *.xls. Tento formát je primárně nastaven pro ukládání souborů programem Microsoft Excel.

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 20071005160000 | 1 | 16,60 | 15,30 | 14,50 | 14,60 | 14,40 | 30,90 | 38,00 | -6,60 | -1,90 | 0,20 | 13,00 |
| 2 | 20071005161500 | 1 | 15,90 | 15,00 | 14,30 | 13,80 | 14,30 | 30,60 | 38,20 | -5,70 | -1,30 | 0,20 | 13,00 |
| 3 | 20071005163000 | 1 | 16,00 | 15,10 | 14,30 | 13,80 | 14,00 | 30,70 | 38,30 | -5,50 | -1,00 | 0,20 | 13,00 |
| 4 | 20071005164500 | 1 | 16,00 | 15,20 | 14,30 | 13,70 | 14,10 | 30,80 | 38,50 | -5,40 | -1,00 | 0,20 | 13,00 |
| 5 | 20071005170000 | 1 | 16,00 | 15,30 | 14,50 | 13,80 | 14,10 | 30,80 | 38,10 | -5,30 | -1,00 | 0,20 | 13,00 |
| 6 | 20071005171500 | 1 | 15,80 | 15,10 | 14,30 | 13,60 | 13,90 | 30,70 | 38,30 | -5,10 | -0,80 | 0,20 | 13,00 |
| 7 | 20071005173000 | 1 | 15,70 | 15,00 | 14,20 | 13,60 | 13,90 | 30,70 | 38,40 | -5,10 | -0,60 | 0,20 | 13,00 |
| 8 | 20071005174500 | 1 | 15,60 | 15,00 | 14,20 | 13,60 | 13,80 | 30,60 | 38,40 | -4,80 | -0,40 | 0,20 | 13,00 |
| 9 | 20071005180000 | 1 | 15,50 | 15,00 | 14,20 | 13,60 | 13,80 | 30,70 | 38,10 | -4,60 | -0,30 | 0,20 | 13,00 |
| 10 | 20071005181500 | 1 | 15,30 | 14,90 | 14,20 | 13,60 | 13,90 | 30,70 | 38,20 | -4,60 | -0,20 | 0,20 | 13,00 |
| 11 | 20071005183000 | 1 | 15,20 | 14,90 | 14,20 | 13,60 | 13,90 | 30,60 | 38,40 | -4,80 | -0,30 | 0,20 | 13,00 |
| 12 | 20071005184500 | 1 | 15,10 | 14,80 | 14,30 | 13,60 | 13,80 | 30,50 | 38,30 | -4,50 | 0,10 | 0,20 | 13,00 |
| 13 | 20071005190000 | 1 | 14,90 | 14,80 | 14,30 | 13,60 | 13,90 | 30,50 | 38,10 | -4,40 | 0,20 | 0,20 | 13,00 |
| 14 | 20071005191500 | 1 | 14,80 | 14,80 | 14,30 | 13,60 | 13,90 | 30,30 | 37,90 | -4,40 | 0,10 | 0,20 | 13,00 |
| 15 | 20071005193000 | 1 | 14,60 | 14,70 | 14,30 | 13,60 | 13,80 | 30,50 | 38,30 | -4,30 | 0,30 | 0,20 | 13,00 |
| 16 | 20071005194500 | 1 | 14,60 | 14,60 | 14,30 | 13,60 | 13,80 | 30,40 | 38,50 | -4,20 | 0,30 | 0,20 | 13,00 |
| 17 | 20071005200000 | 1 | 14,50 | 14,60 | 14,30 | 13,60 | 13,80 | 30,60 | 38,00 | -4,30 | 0,30 | 0,20 | 13,00 |
| 18 | 20071005201500 | 1 | 14,30 | 14,50 | 14,30 | 13,60 | 13,80 | 30,30 | 38,20 | -4,30 | 0,30 | 0,20 | 13,00 |
| 19 | 20071005203000 | 1 | 14,20 | 14,50 | 14,30 | 13,60 | 13,80 | 30,50 | 38,20 | -4,30 | 0,30 | 0,20 | 13,00 |
| 20 | 20071005204500 | 1 | 14,10 | 14,50 | 14,30 | 13,60 | 13,90 | 30,60 | 38,20 | -4,20 | 0,40 | 0,20 | 13,00 |
| 21 | 20071005210000 | 1 | 14,00 | 14,40 | 14,30 | 13,70 | 13,80 | 30,60 | 38,50 | -4,30 | 0,40 | 0,20 | 13,00 |
| 22 | 20071005211500 | 1 | 14,00 | 14,30 | 14,30 | 13,70 | 13,90 | 30,50 | 38,20 | -4,00 | 0,50 | 0,20 | 13,00 |
| 23 | 20071005213000 | 1 | 13,80 | 14,30 | 14,30 | 13,70 | 13,90 | 30,60 | 38,30 | -4,00 | 0,50 | 0,20 | 13,00 |
| 24 | 20071005214500 | 1 | 13,80 | 14,20 | 14,30 | 13,70 | 13,90 | 30,60 | 38,10 | -3,90 | 0,60 | 0,20 | 13,00 |
| 25 | 20071005220000 | 1 | 13,60 | 14,10 | 14,20 | 13,70 | 13,90 | 30,70 | 38,20 | -4,10 | 0,70 | 0,20 | 13,00 |
| 26 | 20071005221500 | 1 | 13,60 | 14,10 | 14,20 | 13,70 | 13,90 | 30,50 | 37,90 | -3,90 | 0,70 | 0,20 | 13,00 |
| 27 | 20071005223000 | 1 | 13,50 | 14,00 | 14,20 | 13,70 | 13,90 | 30,60 | 38,40 | -3,80 | 0,70 | 0,20 | 13,00 |
| 28 | 20071005224500 | 1 | 13,40 | 14,00 | 14,20 | 13,70 | 13,90 | 30,50 | 38,30 | -4,00 | 0,70 | 0,20 | 13,00 |
| 29 | 20071005230000 | 1 | 13,30 | 13,90 | 14,10 | 13,60 | 13,90 | 30,80 | 38,20 | -3,90 | 0,60 | 0,20 | 13,00 |
| 30 | 20071005231500 | 1 | 13,20 | 13,80 | 14,10 | 13,70 | 13,90 | 30,70 | 38,30 | -4,10 | 0,50 | 0,20 | 13,00 |
| 31 | 20071005233000 | 1 | 13,10 | 13,80 | 14,10 | 13,70 | 13,90 | 30,50 | 38,10 | -4,00 | 0,60 | 0,20 | 13,00 |
| 32 | 20071005234500 | 1 | 13,00 | 13,70 | 14,10 | 13,70 | 13,90 | 30,60 | 38,20 | -3,90 | 0,60 | 0,20 | 13,00 |
| 33 | 20071006000000 | 1 | 13,00 | 13,70 | 14,10 | 13,70 | 13,90 | 30,80 | 38,30 | -3,90 | 0,60 | 0,20 | 13,00 |

Obr 13.: Data naimportovaná do programu Microsoft Excel

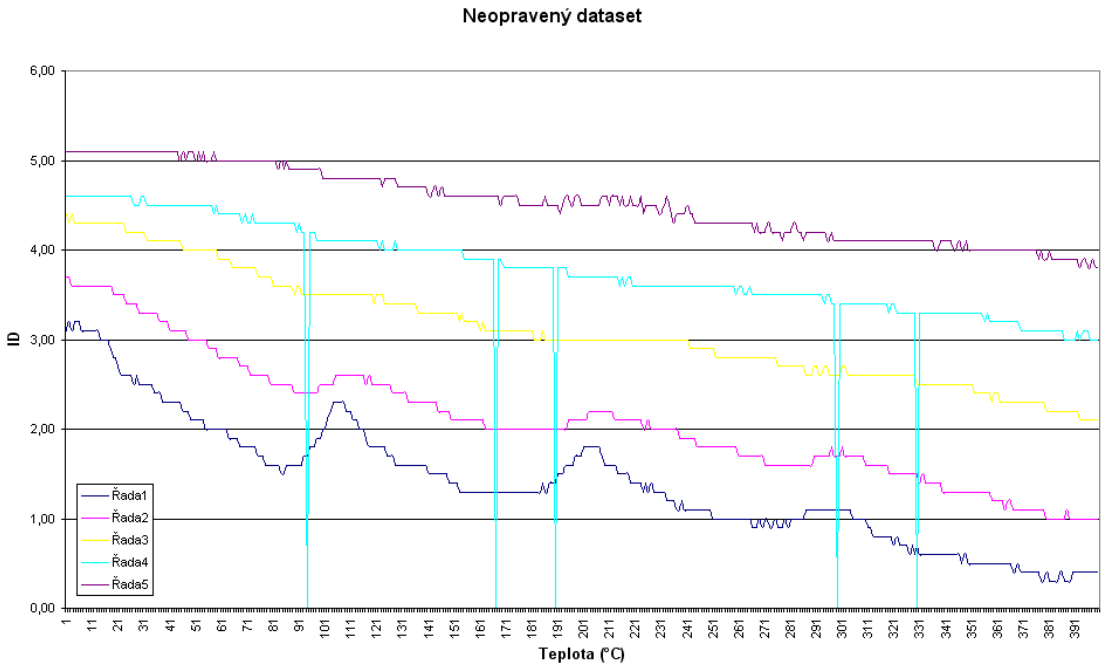
Takto naimportovaná data ovšem obsahují výše zmíněné chyby a vizualizace dat nevypadá nejlépe. Velmi dobře patrné je to z následujícího obrázku. Navíc by takové výpadky mohly výrazně ovlivnit další analýzu datasetů a výsledky by nemusely odpovídat skutečnosti.

Následovala vizualizace dat a zjištění chyb. Vyskytující se chyby byly dvojího typu:

- dlouhodobé výpadky
- náhodné výpadky
- výpadky během stahování dat a výměny baterie

Dlouhodobé výpadky po dohodě s vedoucím práce nejsou odstraněny, byly ponechány jako nulové hodnoty. Tyto výpadky tedy nemají žádný další vliv na následné zpracování datasetu a výsledná data tedy nejsou zkreslena.

Náhodné výpadky spočívají v nezapočítání jedné naměřené hodnoty, která je ve většině případů nulová. Tyto výpadky se objevují v nepravidelných intervalech na přístrojích ALA a ALA1. Jejich příčina je neznámá. Vyskytují se téměř vždy v hloubce 35 cm. Jediný výpadek mimo tuto hloubku byl pozorován jen v jediném případě a to v hloubce 20 cm. Tyto chyby bylo nutné ošetřit, protože jinak by mohlo dojít ke značnému zkreslení údajů, které lépe dokazuje následující tabulka.




Obr 14.: Neopravená data vizualizována v grafu

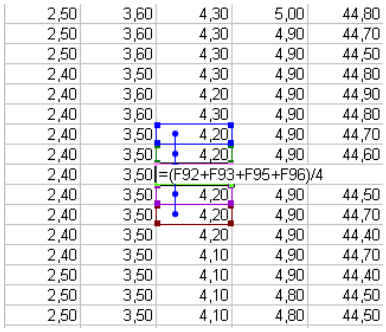
Výpadky během stahování dat do PC jsou zanedbatelné, protože se jedná asi jen o 6 drobných výpadků. To je způsobeno přerušením měřícího cyklu během odesílání dat do PC. Podobně to

platí pro výpadky zaviněné výměnou baterie. Během výměny nemohl pochopitelně přístroj měřit, proto došlo k dočasnému výpadku měření.

**Tabulka 6.1: Příklad zkreslení neopravených dat**

| | **5:00** | **5:15** | **5:30** | **5:45** | **Průměr** |
|---|---|---|---|---|---|
| Neopravená data | 10,30 | 10,30 | 0,00 | 10,30 | 7,725 |
| Opravená data | 10,30 | 10,30 | 10,30 | 10,30 | 10,30 |

K odstranění takovýchto chyb byl použit další program od společnosti Microsoft, Visual Basic 6.0. Tento program umožňuje poměrně dobrou práci s ostatními produkty firmy Microsoft. Proto nebylo obtížné pomocí vlastní naprogramované aplikace chyby v datasetech najít a odstranit. Program automaticky kontroloval celý dataset. Jakmile zjistil mezi 2 sousedními hodnotami rozdíl větší, než určitá hranice, dopočetl hodnotu aritmetickým průměrem 2 nejbližších hodnot z každé strany, viz. obr. 15. Hranice byly odlišné pro jednotlivé případy. Pro teplotu půdy byla 3°C, teplotu vzduchu 2°C a pro vlhkost půdy 5%.

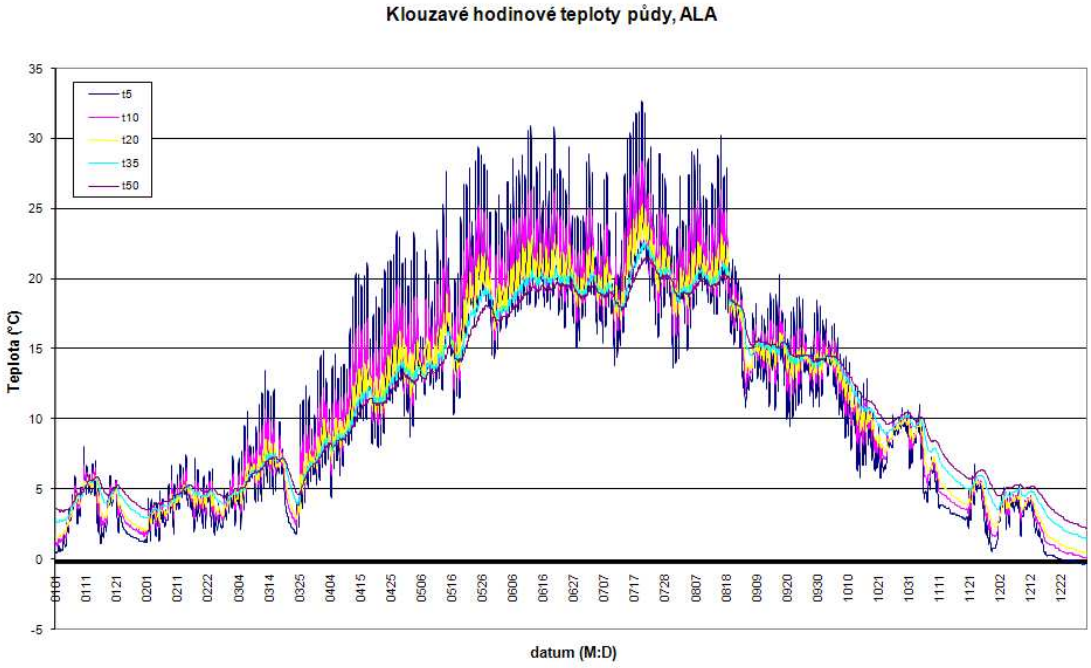



Obr 15.: Neopravená data vizualizována v grafu

Finálně opravený dataset vypadal oproti surovému stavu úhledně a při vizualizaci do grafů již byly všechny sledované časové řady velmi dobře patrné. Příklad opraveného grafu je možné vidět na obr. 16. Takto opravené časové řady byly dále zpracovány.

Každý dataset byl pomocí metody klouzavých průměrů (viz. oddíl 5.5 na straně 29) shlazen do hodinových, denních a měsíčních průměrů. Všechny shlazené datasety jsou samozřejmě přiloženy na CD se všemi přílohami.

Pro analýzu časových řad byly podle oddílu 4.4.1 na straně 15 použity denní klouzavé průměry, protože již byly očištěny od denních amplitud (byla shlazena denní složka časové řady). Tyto datasety jsou tedy nejlepší ke studiu sezónních jevů na sledované lokalitě. Dlouhodobé výpadky v datasetech Fiedler a ALA - průtok byly v datasetech vynechány, aby nezkreslovaly správně naměřená data při následném zpracování. Navíc nás téměř nelimituje HW ani SW složka, protože dataset obsahuje ”jen” 365 záznamů.




Obr 16.: Opravená data vizualizována v grafu, stanice ALA

## 6.2 Vlastní analýza dat

### 6.2.1 Statistická analýza

Ačkoli teorie k této látce je uvedena už na začátku celé práce, tak tato problematika musela být řešena až v závěru práce, kdy bylo jisté, že se již nebude v datasetech nic měnit. Navíc je lepší, když do této analýzy vstupují data již opravená. Předejde se tím jistému zkreslení při výpočtu jednotlivých statistických charakteristik, které by jistě byly ovlivněny výpadky. K výpočtu jednotlivých charakteristik byly použity oba dva programy, jak projekt R, tak Microsoft Excel.

Práce v programu od firmy Microsoft, v aplikaci Excel, byla velmi jednoduchá. Stačilo nainstalovat doplňkovou sadu **Analytické nástroje**. Ta obsahuje mimo jiné modul **Popisná statistika**, který vypočte základní uvedené charakteristiky automaticky. Pro výpočet všech charakteristik je nutností zatrhnout kolonku **Celkový přehled**. Tímto se liší od projektu R, který přistupuje ke každé sledované statistice jednotlivě. Výsledek získaný pomocí programu Microsoft Excel je možné vidět na následujícím příkladu.

```
Stř. hodnota: 11,32762689
Chyba stř. hodnoty: 0,406151283
Medián: 10,08125
```

```
Modus: 22,440625
Směr. odchylka: 7,630889815
Rozptyl výběru: 58,23047937
Špičatost: -1,349653559
Šikmost: 0,190306904
Rozdíl max-min: 26,66145834
Minimum: -0,347916667
Maximum: 26,31354167
Součet: 3998,652292
Počet: 353
Největší (1): 26,31354167
Nejmenší (1): -0,347916667
```

Projekt R takto najednou všechny statistiky bohužel nevypočte. Zato nabízí mnoho jiných nástrojů, které jsou na pokročilejší úrovni než aplikace Excel. Navíc lze jednotlivé příkazy vložit do funkce a stejnou funkci lze použit kdykoli jindy při jiné práci. Dále projekt R, na rozdíl od aplikace Excel, automaticky nezaokrouhluje hodnoty. Tím je zajištěna vyšší přesnost výsledků, než při použití modulu Popisná statistika. Pro výpočty sledovaných charakteristik byly použity tyto příkazy:

Aritmetický průměr: **mean(ALA[,1])**
Směrodatná odchylka: **sd(ALA[,1])**
Rozptyl: **var(ALA[,1])**
Korelační koeficient: **cor(ALA[,1], ALA[,2])**[4]
Křivost: **kurtosis(ALA[,1])**[5]
Špičatost: **skewness(ALA[,1])**[6]

Veškeré takto spočtené charakteristiky jsou uvedeny v tabulkách v přílohách.

### 6.2.2 Analýza časových řad v praxi

Pomocí metody klouzavých průměrů byly "shlazeny" veškeré časové řady použité v práci. K dispozici totiž byly údaje za každých patnáct minut po dobu jednoho roku, což je dohromady přes 40 000 záznamů. Toto opravdu velké množství údajů by nemělo smysl z hlediska analýzy časových řad analyzovat, protože hustota údajů je příliš vysoká. Proto bylo provedeno troje shlazení:

[4]Vyžaduje vždy 2 hodnoty, slouží pro porovnání vlivu jedné sledované hodnoty na druhou.
[5]Pro výpočet křivosti je nutné mít nainstalovaný balík CAR.
[6]Pro výpočet špičatosti je nutné mít nainstalovaný balík CAR.

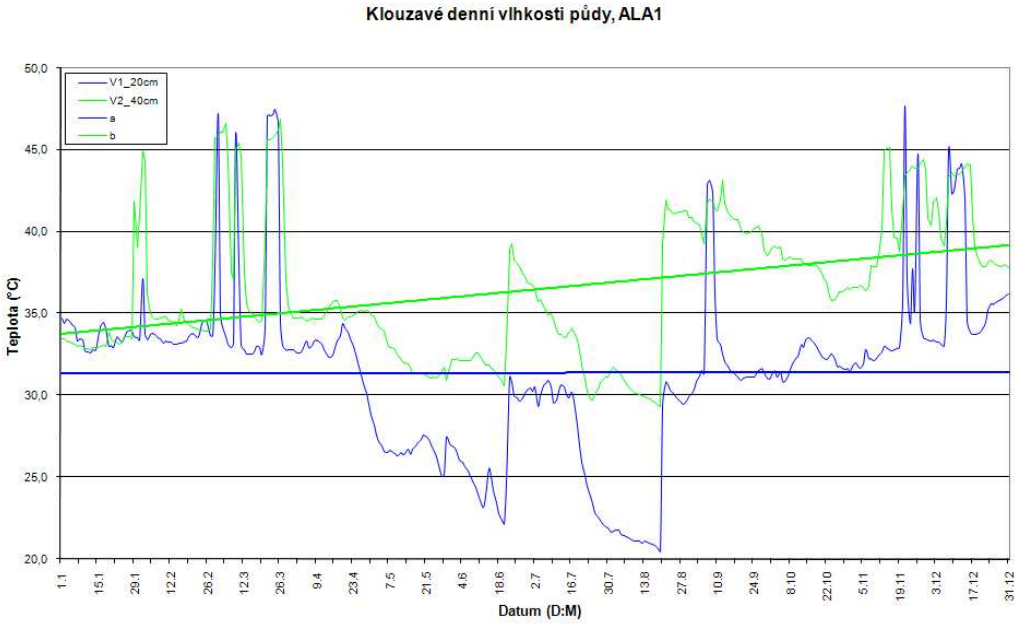
- hodinové klouzavé průměry,
- denní klouzavé průměry,
- měsíční klouzavé průměry.

Jako ideální se jeví denní klouzavé průměry, které již nezachovávají denní periodicitu, ale zachovávají sezónnost, která je z datasetu velice dobře patrná. Veškerá další analýza proto proběhne na datasetech s denními klouzavými průměry.

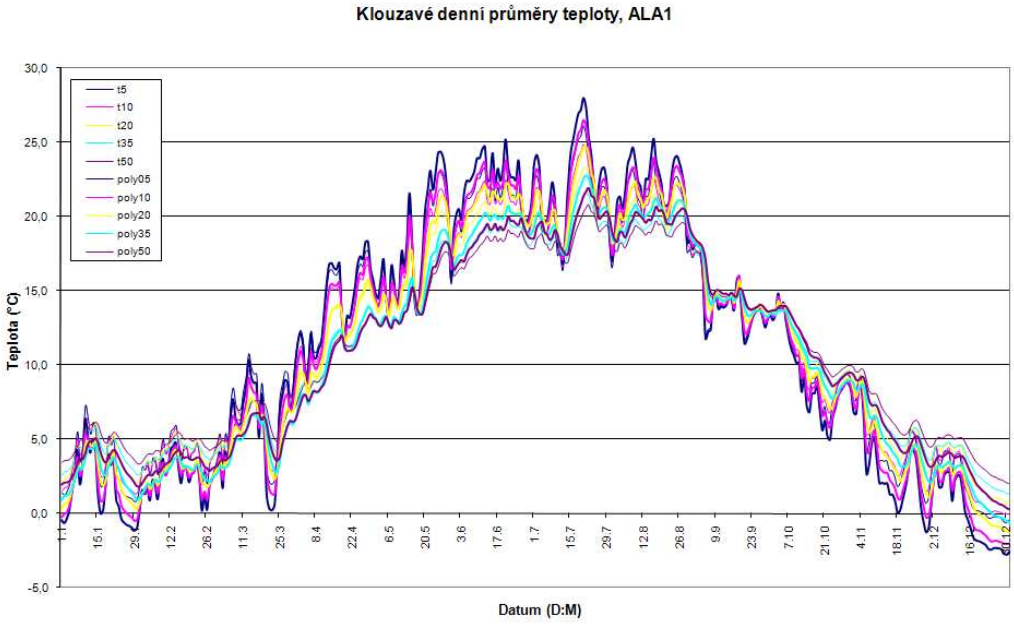
Stejně jako předchozí krok, statistická analýza, byla i tato analýza zpracována v obou programech. Ovšem s tím rozdílem, že základní výpočty, tj. určení parametrů pro lineární a polynomickou funkci, byly provedeny pomocí aplikace Microsoft Excel. Dále zde byly provedeny i odečty vypočtených hodnot od naměřených, což vedlo ke vzniku reziduální složky. Ta byla exportována do formátu *.txt a dále testována v projektu R.

1. **Výpočet parametrů křivky**
   V tomto kroku jsem jako začátečník s časovými řadami byl nucen využít radu vedoucího práce. Nevěděl jsem, která křivka by nejlépe vystihovala vlhkosti a teploty. Po domluvě s Mgr. Tučkem byl nakonec pro vlhkosti půdy použit lineární trend, pro teploty půdy a vzduchu polynom IV. řádu. Hodnoty parametrů těchto křivek byly vypočteny podle vzorců uvedených v oddílech 3.4.2 a 3.4.3.

2. **Konstrukce křivky**
   Z vypočtených parametrů bylo nutné sestrojit hledanou křivku. Tou byla následně data proložena.



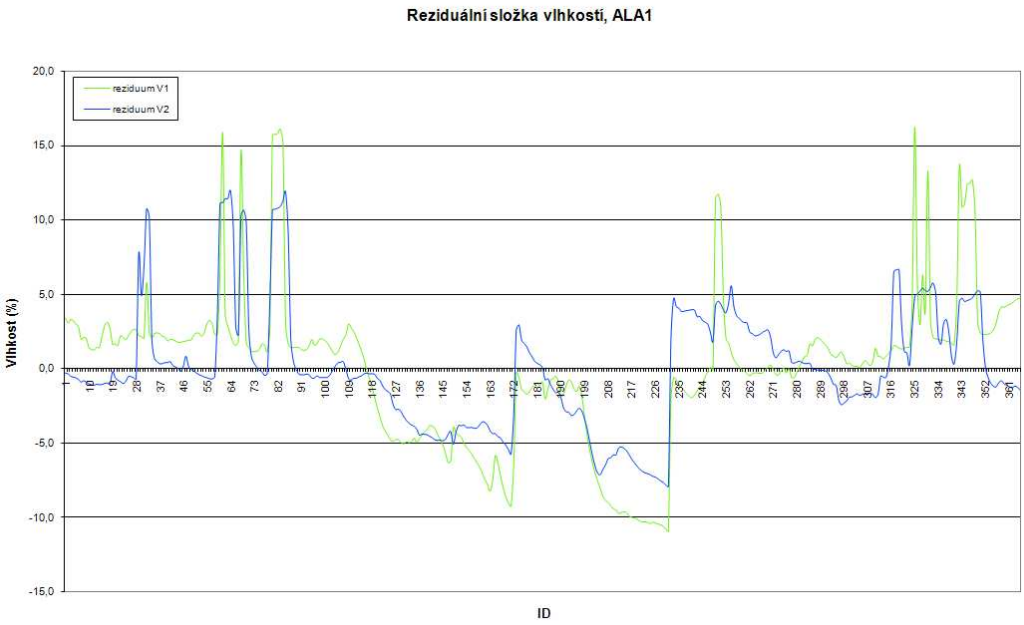

Obr 17.: Lineární trend proložený vlhkostmi na přístroji ALA1



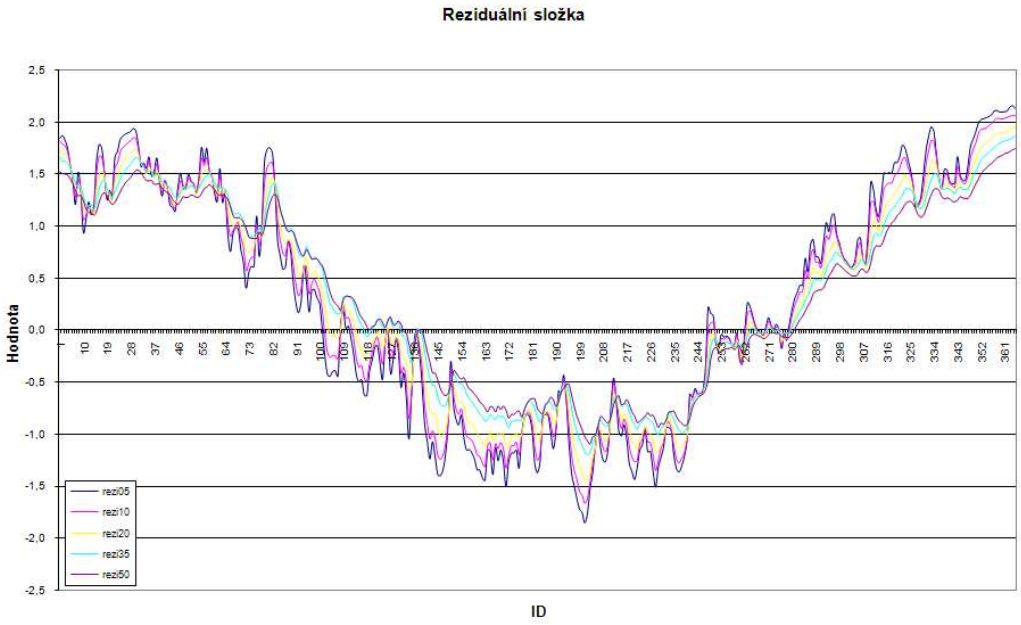

Obr 18.: Polynomiální trend proložený teplotami půdy na přístroji ALA1

3. **Získání rezidua**

   Získání reziduální složky sledované řady vzniklo odečtením hodnot použitých křivek od naměřených hodnot. Následně byla rezidua exportována do formátu *.txt a načtena do programu R.




Obr 19.: Lineární trend odečtený od vlhkostí, ALA1




Obr 20.: Polynomiální trend odečtený od teplot půdy, ALA1

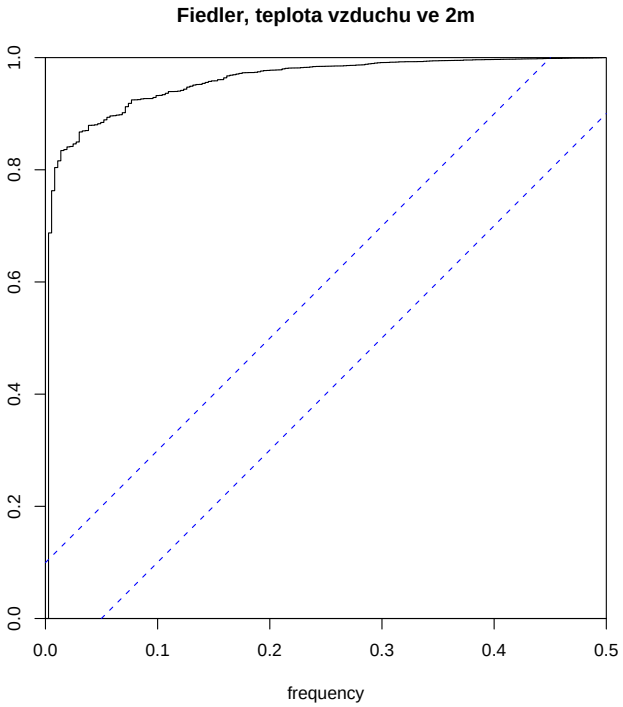
4. **Hledání period**

   Abychom se přesvědčili, že jsme časovou řadu opravdu očistili od všech period, tj. že jsme ji proložili vhodnou křivkou, je nutné provést test periodicity. Tyto testy odhalí všechny podstatné periody, které by mohly v reziduální složce zůstat. Právě tyto periody jsou velmi podstatné například v ekonomii, např. Marshallův cyklus opakující se jednou za desítky let.

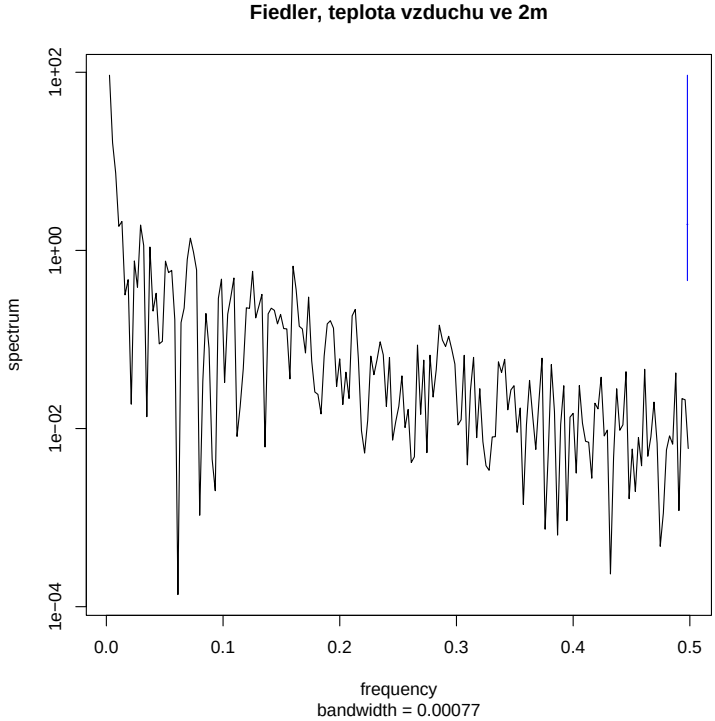
   ```
   cpgram
   Plot Cumulative Periodogram
   spec.pgram
   Estimate Spectral Density of a Time Series by a Smoothed Periodogram
   ```

   Aplikací příkazu **cpgram** vyšly periody vyjádřené kumulativně. Pomocí druhého příkazu, **spec.pgram**, byly periody vypočteny pomocí rychlé Fourierovy transformace. Rozdíl mezi výstupy obou příkazů je velmi dobře patrný z následujících obrázků.

   Zde jsou uvedeny jen tyto dva obrázky, protože ostatní grafické výstupy vypadají téměř stejně. Veškeré grafické výstupy jsou k dispozici na CD s přílohami a datasety.




Obr 21.: Stanice Fiedler, teplota vzduchu ve 2m, příkaz cpgram




Obr 22.: Stanice Fiedler, teplota vzduchu ve 2m, příkaz spec.pgram

5. **Test period**

Tento krok následuje po hledání period. Následuje po odhalení a očištění periody. Může

se stát, že testy uvedené výše periodu nerozpoznají a proto je nutný další test. Přesně vzato, v tomto testu se jako nulová hypotéza $H_0$ testuje, zda předložená pozorování jsou realizace vzájemně nezávislých stejně rozdělených náhodných veličin, kde nemusí mít bílý šum nulovou střední hodnotu. Při nezamítnutí této nulové hypotézy se opravdu jedná o bílý šum. To provedeme pomocí projektu R pomocí Wilcoxonova testu (jednovýběrový, znaménkový) a určíme body zvratu. Tyto výpočty se provádějí za pomoci programu, který vše vypočítá za nás. Podrobně, pro ruční počítání, je postup uveden v [5], str. 94-99.

```
#Dataset ALA:
ALA05 = ALA[,1]
ALA10 = ALA[,2]
wilcox.test(ALA05, ALA10, paired = TRUE, alternative = "greater")
wilcox.test(ALA10 - ALA05, alternative = "less",
exact = FALSE, correct = FALSE)

#Dataset ALA1:
ALA105 = ALA1[,1]
ALA110 = ALA1[,2]
wilcox.test(ALA105, ALA110, paired = TRUE, alternative = "greater")
wilcox.test(ALA110 - ALA105, alternative = "less",
exact = FALSE, correct = FALSE)

#Dataset Fiedler:
F10 = Fiedler[,1]
F60 = Fiedler[,2]
wilcox.test(F10, F60, paired = TRUE, alternative = "greater")
wilcox.test(F60 - F10, alternative = "less",
exact = FALSE, correct = FALSE)
```

Tento test vyšel ve všech případech s výsledkem, že **nelze zamítnout** $H_0$. Tímto lze považovat zbývající část časové řady za reziduální složku a časová řada je plně očištěna od veškerých period.

## 6.3 Výsledky analýzy časových řad

Výsledky analýzy časových řad pro mě osobně byly zklamáním, protože nebyla nalezena jediná významná perioda. Nejvýznamnější nalezená perioda byla zjištěna u srážek a měla hodnotu 0,1, takže ze statistického hlediska se dá považovat za reziduum, nikoli za periodu. Ani po testu sezónnosti nabyla objevena perioda, dokonce i testy náhodnoti prokázaly, že se po očištění

dat od trendu jedná o reziduální složku. Ovšem tyto výsledky se daly očekávat, když jsem zpracovával data za jediný rok (mimo manuální měření, zde byla data za 7 let). Kdybych zpracovával data na hodinové úrovni, tak by pravděpodobně byla objevena denní perioda. Na druhou stranu by byly zapotřebí lepší modely na data a to by kladlo větší nároky na programové i technické zařízení. Navíc by kvůli příliš vysoké hustotě měření mohly být tyto periody zanedbány.

Jedinou zajímavostí bylo zjištění klesajícího trendu vlhkostí půdy v hloubkách 20 i 40 cm na stanici ALA. Vzhledem k tomu, že se zde nachází aktivní sesuv, tak byl očekáván rostoucí trend vlhkosti, který byl zjištěn u vlhkostí na přístroji ALA1.

Na závěr byla data vzniklá opravou a shlazením do denních hodnot porovnána s daty z Atlasu podnebí Česka a s daty z amatérské meteostanice v Košíkách.

## 6.4 Srovnání

Na závěr práce byly porovnány zpracovaná data s daty z Košíků a z Atlasu podnebí Česka, která mi laskavě pro účely bakalářské práce poskytnuty Českým Hydrometeorologickým Ústavem. Srovnání s daty z Košíků probíhalo na denní a měsíční úrovni, srovnání s Atlasem jen na měsíční úrovni.

### 6.4.1 Srážky

Jako první uvádím srovnání srážek. Atlas podnebí uvádí dvoje charakteristiky, sněhové a dešťové srážky. Co se týče sněhových srážek, Atlas uvádí průměrně 43 dní ročně se sněžením. V tomto ohledu se statistiky liší téměř dvojnásobně, kde v Košíkách bylo 22 dní se sněžením, v Halenkovicích jen 19. Další srovnání je vidět v následující tabulce. Data u Atlasu podnebí jsou ovšem vypočtena jako dlouhodobý průměr. Dále mohlo dojít k jistému zkreslení dat z Atlasu podnebí vlivem interpolace, kdy sice byly pro výpočet použity dva výškové modely ČR, ovšem se mi nepodařilo zjistit, jak byly tyto modely přesné a s jakou odchylkou pracovaly.

Data pro srovnání pochází primárně ze stanice Fiedler, pro celkové statistiky byl navíc použit dataset manuálního měření.

Při bližším porovnání se stanicí v Košíkách jsou patrné místní rozdíly, ale data se vesměs shodují, což se teoreticky dalo očekávat. Data se liší jen minimálně a to zejména při náhlých skocích hodnot. Tyto skoky vznikly velkou intenzitou srážek, z čehož usuzuji, že se pravděpodobně jednalo o srážky bouřkové. U bouřek je možné, že zasáhly jen jednu lokalitu a druhé se vyhnou. Patrně proto je rozdíl mezi roční sumou srážek z Halenkovic a Košíků celkem vysoký, 127 mm. Grafická interpretace srovnání srážek je vyjádřená v příloze.

**Tabulka 6.2: Srovnání srážek za rok 2007**

| Sledovaná statistika | Atlas podnebí | Košíky | Halenkovice |
|---|---|---|---|
| Roční úhrn srážek | 628 | 887 | 760 |
| Srážkové dny s úhrnem > 0,1 mm | 132 | 135 | 173 |
| Srážkové dny s úhrnem > 1 mm | 95 | 109 | 93 |
| Srážkové dny s úhrnem > 5 mm | 36 | 46 | 34 |
| Srážkové dny s úhrnem > 10 mm | 17 | 18 | 17 |
| Maximum denních úhrnů srážek | 37 | 95,9 | 98,2 |
| Datum prvního sněžení | 10.-20.11. | 20.11. | 20.11. |
| Datum posledního sněžení | 31.3.-10.4. | 21.3. | 20.3. |
| Průměrný počet dní se sněžením | 43 | 22 | 19 |
| Počet dní se sněžením v prosinci | 11 | 9 | 7 |
| Počet dní se sněžením v lednu | 13 | 5 | 6 |
| Počet dní se sněžením v únoru | 11 | 4 | 3 |
| Počet dní se sněžením v březnu | 8 | 4 | 3 |

Z tabulkového srovnání je velmi dobře patrné, že tento rok byl oproti dlouhodobému průměru vlhčí, kdy v Halenkovicích spadlo o 126 mm srážek více, než je dlouhodobý průměr. To lze pravděpodobně přičíst slabé zimě, kdy počet dní se sněžením nedosahoval ani polovině dlouhodobého průměru. Proto lze předpokládat, že srážky spadly na zem ve formě deště, nikoli sněhu. Tyto závěry dále potvrzují hodnoty získané zpracováním teplot vzduchu, které budou uvedeny v dalším odstavci.

Následuje srovnání dlouhodobých měření, a to od 1.1.2001 do 31.12.2007. Toto porovnání bylo možné provést pouze se stanicí v Košíkách, protože Český hydrometorologický úřad odmítnul poskytnout data ze své profesionální stanice v Otrokovicích. Měsíční sumy srážek za jednotlivé roky jsou uvedeny v následujících tabulkách. Srovnání probíhalo jen na měsíční úrovni. Graficky vizualizovaná data jsou k dispozici v digitální podobě v příloze.

Z tabulek je zřejmé, že nejsušším rokem vůbec byl rok 2003, kdy byly naměřeny minimální hodnoty srážek, 480,3 mm. Naopak nejvlhčím rokem se stal rok 2007 se 759,8 mm srážek. Zřejmě díky mírné zimě a nízkému počtu dní se sněžením, proto srážky spadly na zem ve formě deště, nikoli sněhu.

**Tabulka 6.3: Měsíční hodnoty srážek, Halenkovice**

| Sledovaný rok | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|
| **leden** | 44,6 | 10,0 | 40,5 | 42,1 | 9,7 | 53,7 | 53,6 |
| **únor** | 8,0 | 33,9 | 4,5 | 38,9 | 33,9 | 58,1 | 22,9 |
| **březen** | 50,3 | 15,4 | 6,6 | 63,8 | 15,4 | 58,6 | 89,2 |
| **duben** | 41,6 | 32,3 | 35,7 | 21,4 | 32,3 | 77,7 | 6,3 |
| **květen** | 65,3 | 28,5 | 43,6 | 30,4 | 28,5 | 101,5 | 57,1 |
| **červen** | 50,7 | 94,6 | 28,8 | 100,5 | 94,6 | 64,9 | 114,7 |
| **červenec** | 139,2 | 86,5 | 146,0 | 44,8 | 86,5 | 4,3 | 51,0 |
| **srpen** | 38,0 | 70,1 | 13,1 | 29,7 | 70,1 | 107,2 | 120,2 |
| **září** | 138,3 | 39,8 | 29,9 | 51,7 | 39,8 | 12,0 | 133,9 |
| **říjen** | 10,5 | 87,9 | 48,9 | 54,3 | 87,9 | 22,1 | 34,2 |
| **listopad** | 20,0 | 44,8 | 36,3 | 62,2 | 44,8 | 43,0 | 44,1 |
| **prosinec** | 38,1 | 32,9 | 46,4 | 23,8 | 61,4 | 27,3 | 32,6 |
| **roční suma** | 644,6 | 576,7 | 480,3 | 563,6 | 604,9 | 630,4 | 759,8 |

**Tabulka 6.4: Měsíční hodnoty srážek, Košíky**

| Sledovaný rok | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|
| **leden** | 46,3 | 10,1 | 41,3 | 51,3 | 32,7 | 52,6 | 59,9 |
| **únor** | 12,5 | 42,0 | 5,2 | 54,4 | 87,4 | 60,4 | 25,0 |
| **březen** | 62,9 | 20,4 | 13,7 | 93,1 | 19,7 | 73,1 | 83,7 |
| **duben** | 65,9 | 29,1 | 43,2 | 24,0 | 85,8 | 95,3 | 7,2 |
| **květen** | 61,8 | 30,2 | 46,3 | 34,5 | 85,0 | 105,6 | 62,0 |
| **červen** | 48,3 | 108,0 | 59,1 | 128,9 | 51,6 | 68,1 | 155,7 |
| **červenec** | 120,0 | 108,3 | 143,8 | 46,1 | 119,2 | 11,9 | 61,8 |
| **srpen** | 43,4 | 65,5 | 17,2 | 35,1 | 68,8 | 111,5 | 132,4 |
| **září** | 156,0 | 44,5 | 27,3 | 45,4 | 13,7 | 18,7 | 162,8 |
| **říjen** | 8,8 | 97,1 | 53,8 | 77,8 | 10,1 | 23,1 | 47,9 |
| **listopad** | 18,2 | 46,3 | 38,2 | 46,7 | 54,9 | 44,1 | 51,9 |
| **prosinec** | 41,4 | 35,7 | 53,6 | 25,8 | 122,6 | 29,0 | 36,7 |
| **roční suma** | 685,5 | 637,2 | 542,7 | 663,1 | 751,5 | 693,4 | 887,0 |

### 6.4.2 Teplota vzduchu

Jako druhé v pořadí následuje srovnání teplot. I tato data pochází ze stanice Fiedler. Na rozdíl od srážek ale měření pokračovalo i v lednu a v únoru, takže k dispozici byla data téměř za celý rok. V následující tabulce jsou uvedeny průměrné měsíční teploty vzduchu, opět v porovnání s Atlasem podnebí a s Košíky.

**Tabulka 6.5: Průměrné měsíční teploty vzduchu v roce 2007**

| Měsíc | Halenkovice | Košíky | Atlas podnebí |
|---|---|---|---|
| leden | 3,3 | 2,9 | -1,3 |
| únor | 3,2 | 3,4 | -0,7 |
| březen | 6,3 | 6,0 | 3,2 |
| duben | 12,1 | 11,1 | 8,4 |
| květen | 16,2 | 15,7 | 13,6 |
| červen | 20,0 | 19,3 | 16,1 |
| čevrenec | 20,8 | 19,8 | 18,8 |
| srpen | 20,1 | 19,6 | 17,6 |
| září | 12,8 | 12,0 | 13,2 |
| říjen | 8,4 | 7,9 | 8,8 |
| listopad | 2,4 | 2,4 | 3,5 |
| prosinec | -0,8 | -1,0 | -0,3 |

Tabulka uvedená výše jen potvrzuje závěr uvedený uvedený u srážek, a to že rok 2007 byl oproti dlouhodobému průměru výrazně teplejší. Tento závěr potvrzují jak hodnoty ze srážkoměru, tak naměřené hodnoty teplot. V roce 2007 byly v lokalitě Košíky naměřeny tři ze čtyř nejvyšších naměřených teplot, což tento závěr také potvrzuje.

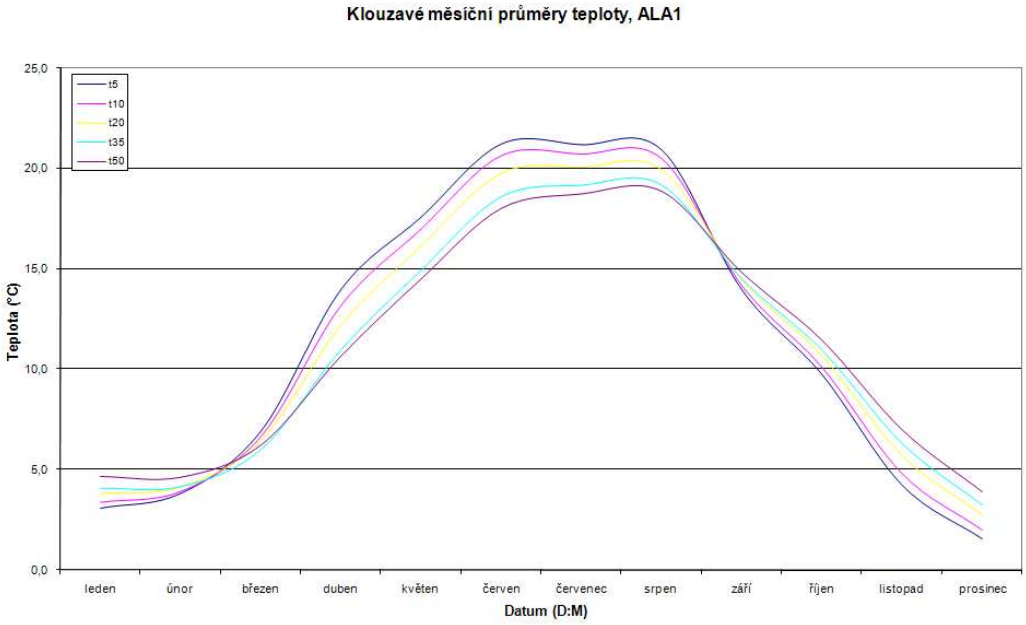
### 6.4.3 Teplota půdy

Pro zpracování teplot půdy jsem měl sice nejvíce dat, ale na druhou stranu téměř žádné možnosti srovnání, protože v Atlase podnebí se této problematice věnují velice zběžně a to jen na pěti stránkách. Ze zpracovaných dat jsou ale velmi dobře patrné důležité lomové oblasti, a to hlavně nástup jara a podzimu. Velmi dobře patrné je to z následujícího obrázku, kdy jaro nastupuje na začátku března a podzim na konci srpna. Tento přechod je zřetelně patrný jak na stanici ALA (s průtokoměrem), tak na stanici ALA1.

**Tabulka 6.6: Extrémní teploty v obci Košíky, maximální teploty**

| Pořadí | Den | Teplota (ve °C) |
|---|---|---|
| 1. | 20.7.2007 | 39,3 |
| 2. | 17.7.2007 | 38,3 |
| 3. | 28.8.2003 | 37,9 |
| 4. | 01.8.2007 | 36,4 |
| 5. | 21.7.2006 | 36,3 |

*Zdroj: http://hpa1.unas.cz/extrem.php, on-line 27.2.2008*



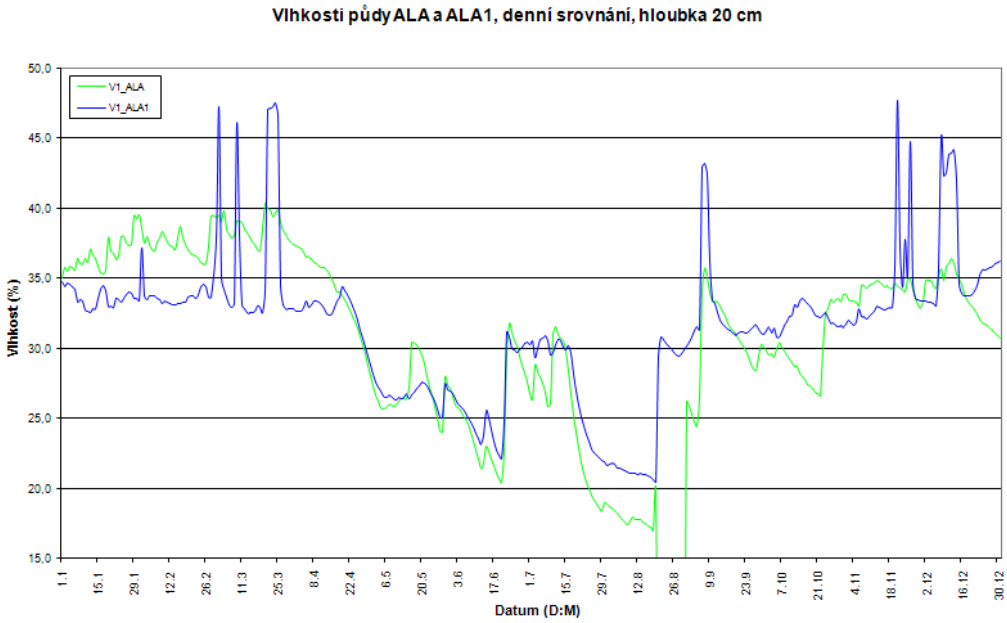

Obr 23.: Stanice ALA1, roční chod teplot půdy

### 6.4.4 Vlhkosti půdy

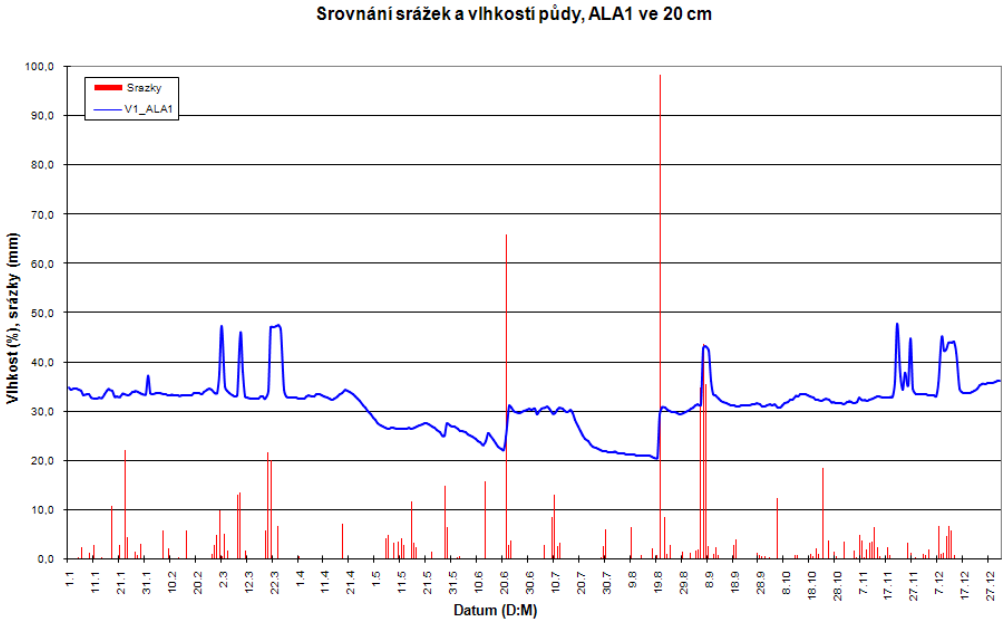
Této kapitole se Atlas podnebí bohužel nevěnuje vůbec, ani pan Malina na své stanici v Košíkách nemá možnost měření vlhkosti půdy. Proto jsem tato data porovnával pouze se srážkami, abych zkusil zjistit vzájemnou korelaci. Ze vzájemné vizualizace dat z obu přístrojů je možné vidět jisté nesrovnalosti, způsobené pravděpodobně umístěním přístrojů. Ačkoli je studované území poměrně malé, je patrný rozdíl mezi patou a střední částí svahu.

Z níže uvedených obrázků je velmi dobře patrné, jaký vliv má intenzita srážek na vlhkost půdy. Při náhlých a intenzivních srážkách dochází ke skokovému nárůstu vlhkosti půdy na obou přístrojích. Z Obr. 24 je také patrné, že se vlhkost půdy mění v závislosti na umístění přístroje.

Větší výkyvy hodnot během srážek s velkou intenzitou lze vidět v březnu, na začátku září a od konce listopadu do poloviny prosince.




Obr 24.: Stanice ALA a ALA1, denní srovnání vlhkostí v hloubce 20cm




Obr 25.: Korelace srážek ze stanice Fiedler a vlhkosti půdy ve 20cm ze stanice ALA1

## 6.5 Závěrečné shrnutí výsledků

V tomto shrnutí jsou sesumarizovány veškeré poznatky plynoucí z vypracování bakalářské práce. Výsledky popisné statistiky jsou uvedeny v tabulkách v příloze, zvlášť pro každý dataset.

Dále jsou v příloze uvedeny parametry polynomů, které byly použity pro proložení jednotlivých časových řad. Nyní již ale k časovým řadám. Při analýze všech datasetů se nepodařilo najít ani jednu významnou periodu, dá se tedy říct, že všechny datasety obsahují jen trend a reziduální složku. Při porovnávání dat za jediný rok se ovšem tyto výsledky daly očekávat. Při porovnávání naměřených hodnot s hodnotami z Košíků a z Atlasu podnebí bylo zjištěno, že rok 2007 byl teplotně nadprůměrný. Jednalo se zejména o velmi teplou zimu a velké množství dešťových srážek na úkor sněhových. To mohlo nepříznivě ovlivnit aktivitu sesuvu.

# Kapitola 7

# Diskuze

Sporným bodem v práci jsou dlouhodobé výpadky. Při analýze časových řad se v zásadě postupuje dvěma způsoby, ale ani o jednom z nich nelze říci, že je ideální. Prvním přístupem je vynechání dat a nahrazení výpadku nulovými hodnotami. Tato metoda zamezí použití dat, která prostě nejsou brána v úvahu a tudíž vypadnou z analýzy i data, která tam jsou a řada se stává nespojitou. Druhým případem je predikce dat, například lineárním trendem. Tato metoda může, ale nemusí, být užitečnější, protože může výrazně zkreslit data a analýza se stane nepřesnou.

Pro teploty půd byl nejvhodnějším trendem polynom. Ovšem o stupni polynomu lze vést delší diskuzi. Z matematického hlediska je IV. stupeň příliš vysoký, teoreticky by šlo využít II., maximálně III. stupně. Tím by ovšem mohlo dojít k výraznějšímu odchýlení od křivky, i když by odchylky byly v řádech setin. Pro delší období, např. za 3 roky, by byl určitě vhdonější polynom III. stupně. Při testu stupně polynomu byly výsledky pro IV. stupeň 97,45%, pro III. stupeň 91,12%.

Výsledný klesající trend vlhkosti půdy ve hloubce 20 i 40 cm na stanici ALA byl jedním z velkých překvapení, všeobecně byl očekáván trend rostoucí, vzhledem k aktivitě sesuvu.

Porovnání s Atlasem podnebí a měřením z roku 2007 sice jednoznačně ukazuje na nadprůměrné hodnoty, zejména v prvních 9 měsících, ale data v Atlase jsou vypočtena jako dlouhodobý průměr.

# Kapitola 8

# Závěr

Bakalářská práce se zabývá sesuvem v Halenkovicích, okres Zlín, který byl aktivován po katastrofálních dešťových srážkách v roce 1997. Cílem práce bylo provést základní statistické charakteristiky naměřených dat a zpracovat tato data pomocí analýzy časových řad.

Práce je členěna do několika částí. Úvodní část obsahuje jednoduchou charakteristiku zájmového území a přístojů, ze kterých data pochází. Druhou částí jsou teoretické základy nutné pro praktické vypracování. Tyto základy jsou poměrně obsáhlé, proto jsou členěny do tří kapitol: Popisná statistika, Časové řady a Základní přístupy k analýze časových řad. Třetí, nejdůležitější část práce, je věnována praktickému zpracování analýzy a prezentaci výsledků, zejména srovnání dat s Atlasem Podnebí a amatérskou stanicí v Košíkách.

Asi největším problémem bylo nastudování literatury, protože i když je tato problematika součástí státní zkoušky, ve výuce se na něj prostě pozapomnělo. Zabývám se konkrétně statistickými charakteristikami datasetu a dekompozicí časových řad. Čím hlouběji jsem se nořil do tajů statistiky, tím více jsem musel zasahovat do datasetů, což vyústilo v jejich kompletní opravu a shlazení. Po shlazení následovaly analýzy, proložení trendem, očištění od trendu a test period. Cílem analýzy je nalézt náhodnou složku časové řady.

Veškerá data jsou vizualizována v grafech a tabulkách, takže jsou přehledně nachystána pro další případné zpracování. Veškeré testy poukazují, že časové řady očištěné od trendu nevykazují další složku. Po testech reziduální složky, které také nic neobjevily, lze říct, že se jedná o bílý šum. Tento výsledek se vzhledem k povaze dat (měření za jeden rok, denní průměry) dal předem očekávat. Tato práce má ale velký potenciál hlavně do budoucna, protože při případném zpracování dat v rámci magisterské práce by již byla zpracovávána data za tři roky a výsledky by byly určitě matematicky zajímavější.

Na závěr této práce bych ještě rád zmínil několik osob, které se většinou úzce podílely na vzniku práce, jmenovitě to jsou:

- Pavel Žárský, vedoucí Aerologického oddělení ČHMÚ

- Petr Malina, amatérský meterolog z Košíků
- Mgr. Pavel Tuček, vedoucí práce
- RNDr. Michal Bíl Ph.D., konzultant a poskytovatel dat

Všem výše zmíněným osobám tímto děkuji za spolupráci. Ještě přikládám omlouvu pro pana Mgr. Miloslava Jančíka, kterého jsem při hledání vedoucího své práce vyrušoval od jeho vlastní práce. Dále bych rád zmínil další použitou literaturu, díky které mohla být tato práce napsána. Jedná se o [7], [8], [9] a [10], díky nimž jsem se ponořil do hloubek LaTeXu, což mi velice usnadnilo psaní matematické části této práce.

# Summary

My bachelor work deals with the processing of data from the model locality Halenkovice through analysis of time series. Data are processed from the crude state, this is how they were measured with device and sent to the server, possibly as downloaded straight from the device. At first I had to clean the data from these station from random blackouts of measurement. This was done through my own programme, which I wrote in the programming language Microsoft Visual Basic 6.0. Longterm blackouts staid as null. This data was for the analysis of time series still inappropriate, because a fixed dataset had values from 10 000 to 40 000 records. That´s why I had to use the sliding averages and smooth data on hourly, daily and monthly averages. For the analysis of time series were the most suitable daily averages, beacause they had been cleaned from day period and dataset had contained acceptable 365 values. Monthly average served for further comparison with reference data from the Atlas podnebí Česka and the amateur meteostation in the village named Košíky.

# Literatura


[1] ANDĚL, Jiří: *Statistická analýza časových řad.* 1. vydání, SNTL 1976, Praha, 271 s.

[2] ANDĚL, Jiří: *Matematická statistika.* SNTL 1985, Praha.

[3] ANDĚL, Jiří: *Základy matematické statistiky.* Preprint, Univerzita Karlova, Matematicko-fyzikální fakulta, Praha, 2002.

[4] ATLAS PODNEBÍ ČESKA: *Atlas podnebí Česka.* 1. vyd., Praha 2007, Olomouc 2007, 255 s, ISBN 978-80-86690-26-1 (ČHMÚ), ISBN 978-80-244-1626-7 (UP).

[5] CIPRA, Tomáš: *Analýza časových řad s aplikacemi v ekonomii.* SNTL 1986, Praha, 248 s.

[6] KVASNIČKA, Michal, VAŠÍČEK, Osvald: *Úvod do analýzy časových řad.* Masarykova univerzita 2001, Brno, 173 s.

[7] OLŠÁK, Petr: *Typografický systém TEX.* CsTUG 1995, ISBN 80-901950-0-8.

[8] OLŠÁK, Petr: *TEXbook naruby.* Konvoj 1997, ISBN 80-85615-64-9.

[9] PARTL, Hubert, SCHLEGL, Elisabeth, HYNA, Irene, SÝKORA, Petr: *LATEX Stručný popis.* Je součástí distribuce CSTEXu jako soubor uvodlat.zip.

[10] RYBIČKA, Jiří: *LATEX pro začátecníky.* 2. vydání, Konvoj 1999, ISBN 80-85615-77-0 (brož.), ISBN 80-85615-74-6 (váz.).

[12] Meteostanice v Košíkách, dostupné z: *http://www.hpa1.unas.cz* (online 16. 3. 2008).