

**Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
Katedra kybernetiky**

**DATOVÁ ANALÝZA A SIMULÁTOR  
RYCHLOSTI VĚTRU Z VEŘEJNĚ  
DOSTUPNÝCH DAT**

Diplomová práce

Plzeň, 2014

Vypracoval: Bc. Martin Dlouhý  
Vedoucí práce: Ing. Andrea Zápotocká, Ph.D.

## Prohlášení

Předkládám tímto k posouzení a obhajobě diplomovou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

V Plzni dne .....

.....  
*vlastnoruční podpis*

## Poděkování

Na tomto místě bych rád poděkoval vedoucí diplomové práce *Ing. Andree Zápotocké, Ph.D.* za konzultace, cenné rady, připomínky a odborné vedení.

## **Anotace**

Tato diplomová práce zkoumá statistické vlastnosti rychlosti větru. Zásadní část práce je datová analýza rychlosti větru a porovnání předpovědí z Českého hydrometeorologického ústavu ([www.chmi.cz](http://www.chmi.cz)) a norského serveru [www.yr.no](http://www.yr.no). Jako další část byl vytvořen simulátor rychlosti větru.

## **Klíčová slova**

datová analýza, rychlost větru, autoregresní model, předpověď, stochastický systém

## **Annotation**

This thesis examines the statistical properties of wind speed. The essential part is the wind speed data analysis and comparison predictions from the Czech Hydrometeorological Institute ([www.chmi.cz](http://www.chmi.cz)) and the Norwegian server [www.yr.no](http://www.yr.no). As another part was created simulator wind speed.

## **Key words**

data analysis, wind speed, autoregression model, forecast, stochastic system

# Obsah

<b>1 Úvod</b>	<b>1</b>
<b>2 Teoretická fakta</b>	<b>2</b>
2.1 Vznik větru . . . . .	2
2.1.1 Obecná charakteristika větru . . . . .	2
2.2 Chování větru v české republice . . . . .	3
2.3 Meteostanice . . . . .	4
2.3.1 Mereorologická pozorování . . . . .	4
2.3.2 Dělení meteorologických stanic . . . . .	4
2.4 Klasifikace předpovědí počasí . . . . .	6
2.4.1 Dělení podle účelu . . . . .	6
2.4.2 Dělení podle metody zpracování . . . . .	6
2.4.3 Dělení podle prostorové platnosti . . . . .	7
2.4.4 Dělení podle doby platnosti . . . . .	7
<b>3 Datová analýza</b>	<b>8</b>
3.1 Použitá data . . . . .	8
3.2 Naměřená data z CHMI . . . . .	8
3.2.1 Rozbor rychlosti větru . . . . .	9
3.2.2 Denní diagram . . . . .	11
3.2.3 Výpočet korelací rychlosti větru . . . . .	11
3.2.4 Analýza rychlosti větru . . . . .	12
3.2.5 Četnost podle tříd . . . . .	14
3.3 Předpověď dat z CHMI . . . . .	15
3.3.1 Směrodatná odchylka chyby předpovědi . . . . .	17
3.3.2 Analýza chyby . . . . .	18
3.3.3 Korelační analýza chyby předpovědi . . . . .	20
3.4 Podrobnější pohled na vybrané lokality pro data z CHMI . . . . .	20
3.4.1 Rychlost větru . . . . .	20
3.4.2 Chyba předpovědi . . . . .	22
3.5 Naměřená data z YR . . . . .	25
3.5.1 Chyba předpovědi . . . . .	25
3.5.2 Výpočet autokorelací . . . . .	26
3.5.3 Výpočet plošné korelace pro YR . . . . .	26
3.5.4 Směrodatná odchylka chyby předpovědi YR . . . . .	27
3.5.5 Analýza chyby . . . . .	29
3.6 Podrobnější pohled na některé lokality pro data z YR . . . . .	31
3.6.1 Chyba předpovědi . . . . .	31
3.7 Porovnání norského serveru a CHMI . . . . .	32

<b>4 Simulátor</b>	<b>34</b>
4.1 Markovské procesy	34
4.1.1 Markovské řetězce	34
4.1.2 Pravděpodobnostní vektor v čase $t$	34
4.1.3 Pravděpodobnost přechodu, přechodová matice	35
4.1.4 Diskrétní čas	35
4.1.5 Přechodový graf	35
4.1.6 Klasifikace stavů a typy řetězců	35
4.2 Transformace	36
4.2.1 Metoda inverzní transformace	36
4.2.2 Jonhonsův translační systém	36
4.3 Simulace	36
4.3.1 Generátor tříd rychlosti větru	38
4.3.2 Výsledky simulace	39
<b>5 Závěr</b>	<b>42</b>
<b>Literatura</b>	<b>45</b>
<b>A Obsah Příloženého CD</b>	<b>46</b>
<b>B Pojmy z pravděpodobnosti a statistiky</b>	<b>48</b>
B.1 Základní pojmy	48
B.1.1 Náhodná veličina	48
B.1.2 Hustota pravděpodobnosti a distribuční funkce	49
B.1.3 Korelace	49
B.1.4 Směrodatná odchylka	49
B.1.5 Kvantil	49
B.1.6 Modus	50
B.2 Pravděpodobnostní rozdělení	50
B.2.1 Normální rozdělení	50
B.2.2 Weibullovo rozdělení	50
B.2.3 Laplaceovo rozdělení	50
B.2.4 Raygleyho rozdělení	50
B.2.5 Exponenciální rozdělení	51
B.3 Histogram	51
B.4 Boxplot	51
<b>C Doplnující informace</b>	<b>52</b>
C.1 Autoregresní proces (AR)	52
C.2 Euklidova norma	52

# Seznam obrázků

2.1	Pole průměrné rychlosti větru ve výšce 10m . . . . .	3
2.2	Mapa různých typů meteorologických stanic v České republice [5] . . . . .	5
3.1	Rychlost větru na jaře a na podzim . . . . .	9
3.2	Rychlost větru v létě a v zimě . . . . .	9
3.3	Průměrná rychlost větru během celého roku . . . . .	10
3.4	Porovnání nejnižší a nejvyšší měřené polohy . . . . .	10
3.5	Porovnání nejvyšší a nejnižší měřené polohy . . . . .	11
3.6	Histogram četnosti rychlosti větru s hustotami pravděpodobnosti . . . . .	12
3.7	Porovnání rychlostí větrů pro všechny lokality pomocí boxplot . . . . .	13
3.8	Četnost typů větru podle Beaufortovy stupnice . . . . .	14
3.9	Chyba předpovědi pro předpovězené minimální a maximální hodnoty . . . . .	16
3.10	Histogram četnosti chyby předpovědi s hustotami pravděpodobnosti . . . . .	18
3.11	Porovnání chyb předpovědi pro všechny lokality pomocí boxplot . . . . .	19
3.12	Histogram četnosti větru pro Plzeň v říjnu . . . . .	21
3.13	Histogram četnosti rychlosti větru pro Plzeň v jarních měsících . . . . .	22
3.14	Histogram četnosti chyb předpovědi pro říjnový měsíc . . . . .	22
3.15	Histogram četnosti chyb předpovědi pro Plzeň v jarních měsících . . . . .	23
3.16	Porovnání denních diagramů pro jednotlivé roční období pro Plzeň . . . . .	23
3.17	Porovnání dvou lokalit v ČR, kde každá je na jiné straně dosahu družice . . . . .	25
3.18	Porovnání chyb dvou předpovědí, pro lokality si blízké . . . . .	26
3.19	Chyba předpovědi z norského serveru pro různá data . . . . .	28
3.20	Histogramy četnosti chyb pro norskou předpověď . . . . .	29
3.21	Porovnání chyby CHMI a norského serveru pro stejnou lokalitu . . . . .	30
3.22	Histogram četnosti chyby předpovědi z norského serveru pro Plzeň . . . . .	31
3.23	Histogram četnosti chyb předpovědi pro podzimní měsíce . . . . .	32
3.24	Detail porovnání předpovědi CHMI a norského serveru . . . . .	33
4.1	Příklad grafu popisující Markovské řetězce . . . . .	37
4.2	Detail porovnání simulovaných a naměřených dat s hodinovou granularitou . . . . .	40
4.3	Detail porovnání průměrných rychlostí pro simulovaná a naměřená data . . . . .	40
4.4	Výsledek simulátoru a porovnání s naměřenými daty . . . . .	41

# Seznam tabulek

2.1	Stupnice . . . . .	2
3.1	Třídění měsíců podle ročních období . . . . .	9
3.2	Ukázka hodnot korelačních koeficientů za rok 2013 pro několik vybraných lokalit . . . . .	12
3.3	Třídy větru s modus z naměřených dat . . . . .	14
3.4	Ukázka hodnot směrodatných odchylek chyby předpovědi za rok 2013 pro několik vybraných lokalit . . . . .	17
3.5	Ukázka korelační analýzy chyby předpovědi udané minimální rychlostí větru pro rok 2012 . . . . .	20
3.6	Ukázka korelační analýzy chyby předpovědi udané maximální rychlostí větru pro rok 2012 . . . . .	20
3.7	Ukázka hodnot autokorelačního koeficientu pro vybrané lokality za rok 2013 . . . . .	26
3.8	Ukázka korelačních koeficientů pro chybu předpovědi pro rok 2013 . . . . .	27
3.9	Ukázka hodnot směrodatné odchylky pro předpověď za rok 2013 . . . . .	29
4.1	Tabulka četností po sobě jdoucích stavů . . . . .	38
4.2	Pravděpodobnost přechodů mezi stavy . . . . .	38
4.3	Střední doba setrvání ve stavu . . . . .	38
4.4	Parametry AR procesů predotlivé třídy větru . . . . .	39



# Kapitola 1

## Úvod

V této diplomové práci má čtenář možnost se seznámit s výsledky datové analýzy rychlosti větru a chyby předpovědi rychlosti větru. Důležité je že zpracovávaná data jsou veřejně dostupná. Naměřená data jsou získána z Českého hydrometeorologického ústavu (<http://www.chmi.cz>) a předpovědi jsou získány z Českého hydrometeorologického ústavu (<http://www.chmi.cz>) a z norského serveru <http://www.yr.no>.

Cílem této práce je porovnat předpovědi z výše zmíněných serverů a zjistit, která z předpovědí je pro Českou republiku přesnější a vytvořit matematický model, který bude reprezentovat rychlost větru v daných lokalitách.

Práce je dělena do několika kapitol. Druhá kapitola je věnována obecným teoretickým faktům o větru. Jeho vzniku, chování, okolních vlivů, které rychlost větru ovlivňují. Dále jsou zde zmíněny místa v České republice, kde jsou potřebná data sbírána.

V následující třetí kapitole je obsažena datová analýza, která je rozdělena na dvě podkapitoly pro Český hydrometeorologický ústav a pro norský server <http://www.yr.no>. Na konci této kapitoly je porovnání předpovědí mezi těmito meteorologickými servery a zhodnocení jejich předpovědí.

Ve čtvrté kapitole je popsán simulátor rychlosti větru, od jeho návrhu po prezentaci jeho výsledků.

V poslední, páté kapitole jsou shrnuty veškeré získané výsledky do závěru.

Pro přehlednost jsou v příloze popsány potřebné podklady k porozumění níže sepsaného textu.

Pro veškeré výpočty byl použit Matlab R2013a a získané obsáhlé tabulky jsou uloženy v MS Excel.

# Kapitola 2

## Teoretická fakta

### 2.1 Vznik větru

Vítr vzniká v atmosféře na základě rozdílu atmosférických tlaků v důsledku nerovnoměrného ohřívání zemského povrchu včetně oceánu. Také je ovlivněn vertikálními pohyby vzduchu při jeho zahřívání/ochlazení a rotací Země (Coriolisova síla, odstředivá síla). Další vliv má (v přízemních výškách) morfologie krajiny, rostlinstvo, vodní plochy...

Rychlost větru s výškou nad povrchem roste přibližně exponenciálně, což souvisí s odpory, víry a tření vzduchového proudu, které při povrchu vznikají. [1]

#### 2.1.1 Obecná charakteristika větru

V závislosti na rozložení atmosférického tlaku (a tedy i tlakových výší, tlakových nízí, hřebenů vysokého tlaku, brázd nízkého tlaku) se vzduch nepřetržitě přemísťuje, a to především v horizontálním směru. Toto přemísťování vzduchu se nazývá vítr. Rychlost větru se vyjadřuje v m/s nebo v km/h ( $1 \text{ m/s} = 3,6 \text{ km/h}$ ). Je měřen pomocí meteorologických stanicích ve výšce 10 m nad zemí. Pro slovní vyjádření síly větru se často používá Beaufortova stupnice (tabulka 2.1), podle které jsou rychlosti větru na základě jejich projevů rozděleny do 12 stupňů, přičemž každému stupni je přiřazen určitý název<sup>1</sup>. Veškeré údaje v této tabulce byly získány z [2].

Beaufortova stupnice	Kód	síla větru	rychlost[km/hod]	rychlost[m/s]
0,1	1	bezvětří (klidno)	0-5	0-1
2	2	slabý vítr	5-10	1-3
3	3	mírný vítr	10-20	3-6
4,5	4	čerstvý vítr	20-35	6-10
6	5	silný vítr	35-55	10-15
7,8		velmi silný vítr	55-75	15-21
9,10		vichřice	75-110	21-30
11,12		orkán	nad 110	nad 30

Tabulka 2.1: Beaufortova stupnice

V předpovědích a výstražných informacích se udává směr a rychlost větru, někdy i jeho nárazy. Směr větru udává převládající směr, odkud vítr vane (severozápadní, jižní, ...).

Rychlostí větru se rozumí průměrná rychlost větru měřená zpravidla za období 10 minut. Náraz větru je krátkodobé zvýšení rychlosti větru (po dobu alespoň 1 s, nejvýše však 20 s) alespoň o 5 m/s nad průměrnou rychlost větru, zpravidla nad hranici 12 m/s.

<sup>1</sup>Hodnota kódu je použita až v kapitole 4.3, kde je je o kódování typů větru napsáno více.

V mimotropických zeměpisných šířkách (a tedy i v Evropě) dochází často k náhlým změnám směru a rychlosti větru, oproti stálejšímu režimu větru v tropech. Průměrná rychlost větru se při zemském povrchu pohybuje většinou od 2 do 8 m/s a zřídka převyšuje 15 m/s. V tropických cyklónách, které vznikají nad tropickými oblastmi oceánů a dosáhly stádia hurikánu či tajfunu, dosahuje vítr rychlostí větších než 33 m/s (119 km/h). V tornádech na některých místech zeměkoule (zejména v USA) rychlost větru může v krátkých časových intervalech dosahovat až 100 m/s (360 km/h).

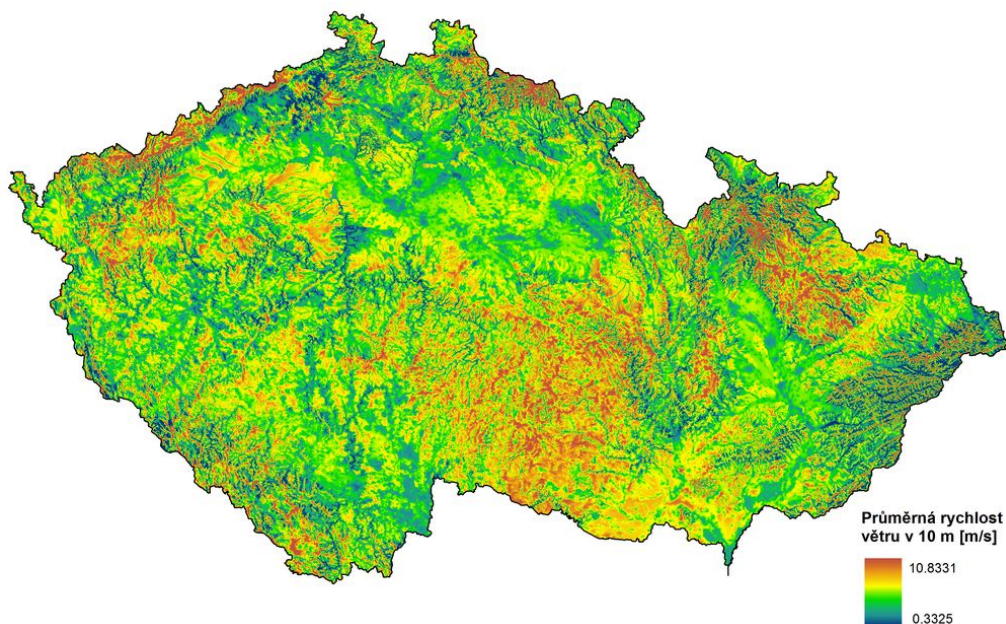
Směr a rychlost větru je ve značné míře ovlivněn členitostí terénu. Nad oceánem je vítr silnější a jeho směr poměrně stálý, nad pevninou často bývá výrazně ovlivněn reliéfem, lesním porostem či zástavbou. Rychlost větru obvykle s výškou stoupá a je nejvyšší zejména u orografických překážek (kopce, hřebeny hor apod.).

Vítr je schopen ovlivnit i teplotní poměry. Zesilováním větru se zvyšuje intenzita výparu z vodních ploch, půdy i vlhkých těles. Tím je odnímáno teplo, včetně tepla z těla člověka, čímž dochází k poklesu teploty. Je ovlivňována lidská psychika a organismus je zatížen, zejména ve spojitosti s nízkými teplotami.

V České republice se nebezpečné rychlosti větru vyskytují v zimní polovině roku nejčastěji při postupu hlubokých tlakových níží přes střední Evropu k východu, v letní polovině roku při intenzivní bouřkové činnosti. V prvním případě s přibližující se hlubokou tlakovou níží tlak vzduchu začíná silně klesat. Pokud střed tlakové níže postupuje severně od nás, otepluje se, neboť naše území se přechodně dostává do tzv. teplého sektoru této tlakové níže, který je obvykle i její největrnější částí. Čím je tlaková níže hlubší, tzn. čím větší je rozdíl mezi tlakem vzduchu na jejím okraji a uprostřed níže, tím fouká silnější vítr. Nejsilnější poryvy větru jsou obvykle spojeny s přechodem studené fronty a za ní ve studeném vzduchu v týlu tlakové níže dochází k prudkému vzestupu tlaku vzduchu. Největší rychlosti větru jsou zaznamenány na hřebenech hor, ale i na všech vyvýšených místech v republice, včetně Českomoravské vrchoviny. [3]

## 2.2 Chování větru v české republice

V České republice byla rychlost větru měřena Ústavem fyziky atmosféry AV ČR. Měření bylo prováděno 10 m nad zemí. Vypočtené průměrné rychlosti jsou graficky zobrazeny na obr. 2.1 .



Obrázek 2.1: Pole průměrné rychlosti větru ve výšce 10m

Ze získaných výsledů bylo zjištěno, že průměrná rychlost větru v ČR je 3-3,5 m/s. Tyto rychlosti se vyskytují zejména v níže položených regionech na lokalitách, které jsou otevřeny proudění vzduchu nebo na méně otevřených lokalitách, ležících ve středních polohách

Vyšší rychlosti větru se budou vyskytovat na místech exponovaných vůči převládajícím směrům větru a obecně ve vyšších polohách, pokud zároveň není rychlost větru výrazněji snížena lokálními okolnostmi (místa v údolích, lesnaté oblasti ap.). V prostoru našich vrchovin lze na vyvýšených a otevřených místech očekávat převážně rychlosti větru kolem 4 m/s ve výšce 10 m nad povrchem, ve výrazněji exponovaných polohách ve výškách nad 600 m n.m. se průměrná rychlost větru v 10 m může blížit až 5 m/s. Ještě výrazně vyšší průměrné rychlosti větru budou dosahovány na nejvyšších horských hřebenech a vrcholcích.

Naopak nižší rychlosti větru lze očekávat v místech vůči proudění málo otevřených. Jedná se zejména o polohy znevýhodněné svou polohou, například místa v údolích či kotlinách a v řadě případů též v podhůří horských celků. V takových místech lze očekávat průměrné rychlosti větru mezi 2,5 a 3 m/s, v úzkých údolích a uzavřených kotlinách i méně. Rychlosti větru mohou být také snižovány výskytem rozsáhlejších lesních porostů či rozsáhlé zástavby v širším okolí lokality. Ty vedou ke zvýšení tzv. drsnosti zemského povrchu a k redukci rychlostí větru z příslušných směrů až o desítky procent. Vlivy drsnosti i orografie se navzájem sčítají, takže například v lesnatých údolích či obydlených kotlinách mohou být průměrné rychlosti větru i pod 2 m/s.[4]

## 2.3 Meteostanice

Pod pojmem meteorologická stanice se v meteorologickém slovníku skrývá tato definice:

**Definice 1.** *Místo, v němž se konají stanovená meteorologická pozorování podle dohodnutých mezinárodních nebo vnitrostátních postupů. Jedná se tedy o místo, kde se pozorují určité meteorologické prvky a veličiny v daném čase pozorování, které se kódují do předem dohodnutých zpráv a ty se posléze odesílají do vnitrostátní či mezinárodní výměny. Ze získaných dat se pak vykrešlují synoptické mapy, výškové mapy, aerologické diagramy a mnoho dalších výstupů, dále pak informace vstupují do numerických modelů a slouží jako podklady ke tvorbě předpovědí. Právě čím hustší je síť meteorologických stanic, tím přesnější předpověď numerický model vypočte[4].*

### 2.3.1 Mereorologická pozorování

Meteorologická pozorování spočívají v měření teploty vzduchu, vlhkosti vzduchu, atmosférického tlaku, délky slunečního svitu, srážek, výšky sněhové pokrývky, rychlosti a směru větru, dohlednosti, sledování oblačnosti, bouřek či dalších jevů, které mohou být různé podle specializace. K měření těchto jevů a prvků slouží celá řada přístrojů, z nichž mnohé mají speciální název, případně některé jevy určuje pozorovatel. V dnešní době, kdy světu vládnou výpočetní technologie, se však mnoho veličin přestalo odečítat lidským okem a zapisovat do deníků, jako tomu bývalo v minulosti. Jsou měřeny automaticky, automaticky se také odesílají. Daná meteorologická stanice by také měla být postavena tak, aby byla reprezentativní pro své okolí a aby byla co nejméně ovlivněna rušivými vlivy. Nevhodné je například umístit observatoř uprostřed panelového sídliště, kde je proudění větru zástavbou výrazně ovlivněno.[4]

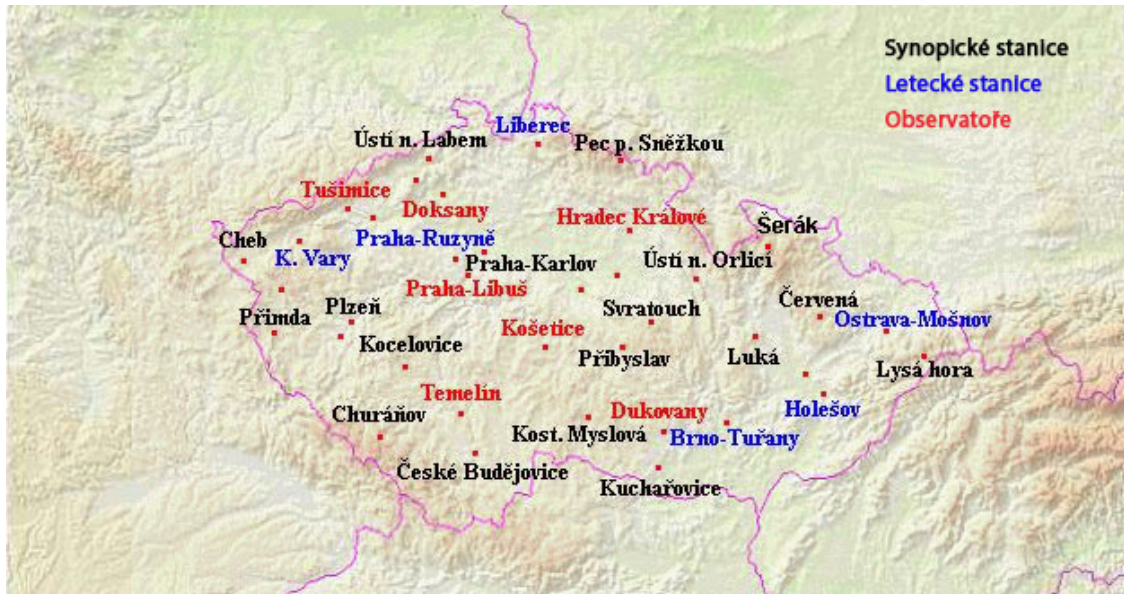
### 2.3.2 Dělení meteorologických stanic

Každá stanice může plnit úkoly různého odborného zaměření a rozsahu. Meteorologické stanice lze dělit:

1. Podle zaměření pracoviště – meteorologické stanice synoptické, klimatologické, speciální (observatoře), letecké, agrometeorologické a silniční
2. Podle umístění - pozemní, námořní a letounové
3. Podle charakteru získávání dat – přízemní a aerologické (měří ve volné atmosféře)

4. Podle stupně součinnosti s člověkem - automatické a s lidskou obsluhou.

Podrobněji si rozebereme dělení meteorologických stanic podle zaměření pracoviště a umístění stanice, které jsou nejčastější.



Obrázek 2.2: Mapa různých typů meteorologických stanic v České republice [5]

### Synoptické stanice

Synoptické stanice, kde se provádějí běžná měření, slouží pro přehled aktuálního počasí a vypracování předpovědí. Právě synoptické stanice disponují špičkovým vybavením, lidskou obsluhou a zpravidla i dlouhou dobou pozorování. V Česku je jich 18. Data jsou téměř v reálném čase k dispozici celému světu. [5]

### Observatoře

Dalšími stanicemi jsou observatoře, kde se vedle meteorologických měření zpravidla provádí další specializovaná měření nebo výzkum, například zabezpečují provoz jaderných elektráren po stránce meteorologických měření (Temelín, Dukovany), provádí solární měření (Hradec Králové), zaměřují se na agrometeorologii<sup>2</sup> a fenologii<sup>3</sup> (Doksany) nebo vypouští meteorologický balón (Prostějov, Praha Libuš). Meteorologický balón má pod sebou sondu a ta při stoupání vzhůru zaznamenává například teplotu či směr a rychlost větru (jedná se o tzv. aerologická měření). S ohledem na svoji nákladnost se tato měření provádí jen 3x denně v termínech 02:00, 08:00 a 14:00 letního času. [5]

### Letecké stanice

Důležitou roli hrají i letecké stanice, které například předávají informace potřebné k hlášení pilotům letadel o stavu i pravděpodobném vývoji počasí. Umístěny jsou na letištích (Brno - Tuřany, Praha - Ruzyně, Ostrava - Mošnov, Holešov, Karlovy Vary nebo Liberec). [5]

<sup>2</sup>Odvětví aplikované meteorologie. Zabývá se vlivem počasí a podnebí na zemědělství.

<sup>3</sup>Nauka o časovém průběhu základních životních projevů v závislosti na změnách počasí, střídání ročních období a prostředí. Předmětem zkoumání jsou obecně se opakující jevy (fenofáze) ve vývoji živých organismů – rostlin, živočichů a hub.



### Klimatologické stanice

Dále jsou důležité klimatologické stanice, jež jsou zapotřebí pro analýzy klimatu, různé typizace počasí nebo pozorují pouze jeden prvek (například srážkoměrné). Klimatologických stanic, které měří většinu důležitých meteorologických prvků, je v Česku přes 150. Nejvíce je ovšem srážkoměrných stanic. Jejich počet přesahuje 600, ale naprostá většina těchto stanic zatím nebyla automatizována.[5]

## 2.4 Klasifikace předpovědí počasí

Předpověď počasí vyjadřuje budoucí stav atmosféry nad sledovaným územím na základě znalosti jejího stavu v současnosti i v minulosti a znalosti zákonitostí, jimiž se atmosféra řídí při přechodu z jednoho stavu do druhého. Předpověď počasí je základem dalších meteorologických předpovědí, které lze rozlišovat podle:

- **účelu**, pro který jsou vydávány; rozlišují se všeobecné předpovědi počasí, speciální předpovědi (např. letecké, hydrologické, zemědělské) a výstrahy,
- **metody zpracování**, např. předpovědi synoptické, numerické, statistické,
- **prostorové platnosti**; rozlišují se předpovědi bodové, liniové, regionální a předpovědi pro celé sledované území,
- **doby platnosti**; rozlišují se předpovědi velmi krátkodobé, krátkodobé, střednědobé, dlouhodobé.

### 2.4.1 Dělení podle účelu

- **Všeobecné předpovědi** počasí jsou určeny široké veřejnosti. Bývají rozšiřovány hromadnými sdělovacími prostředky. Obsahují předpověď oblačnosti, minimálních a maximálních teplot vzduchu, směru a rychlosti větru, výskytu atmosférických srážek včetně jejich druhu, tlakové tendence a rozptylových podmínek. Většinou bývají uváděny stručnou charakteristikou a předpovědí celkové povětrnostní situace, která je příčinou počasí nad sledovaným územím a jeho změn v předpovídaném období.
- **Speciální předpovědi** jsou sestavovány podle potřeb konkrétních uživatelů z jednotlivých oborů lidské činnosti. Soustřeďují se na předpověď těch meteorologických prvků a dějů, které jsou v daném oboru zvláště důležité. Např. v leteckých předpovědích jsou to vodorovná viditelnost, výška spodní základny oblaků a vítr.
- **Výstrahy** poskytují meteorologické informace o pravděpodobném výskytu nebo dalším trvání nebezpečných povětrnostních jevů během několika nejbližších hodin v určitém místě nebo oblasti.

### 2.4.2 Dělení podle metody zpracování

- **Synoptické předpovědi** jsou sestaveny pomocí porovnávání za sebou jdoucích analyzovaných synoptických map, na kterých se sleduje vývoj tlakových útvarů a postup atmosférických front. Pomocí analogií (zkušeností z minulosti) a empirických pravidel se pak usuzuje na budoucí vývoj počasí. Protože meteorologové při sestavování těchto předpovědí využívají svých zkušeností, bývají tyto předpovědi označovány jako subjektivní.
- **Numerické předpovědi** prostorového rozložení meteorologických prvků v troposféře jsou výsledkem výpočtů výkonných počítačů. V současné době se běžně provádějí početní předpovědi atmosférického tlaku, teploty, vlhkosti, rychlosti větru, oblačnosti a srážek. Základem těchto výpočtů je pravidelná síť bodů pokrývající sledovanou oblast zemského povrchu. Vypočtené

prostorové rozložení meteorologických prvků slouží jako jeden z podkladů pro synoptickou předpověď počasí.

- **Statistické předpovědi** meteorologických prvků a jejich prostorového rozložení bývají vypracovávány metodami matematické statistiky a teorie pravděpodobnosti. Mohou být součástí předpovědí synoptických i numerických.

### 2.4.3 Dělení podle prostorové platnosti

- **Bodové předpovědi** počasí se vydávají pro města, letiště a další lokality ne větší než několik  $\text{km}^2$ .
- **Liniové předpovědi** se vydávají pro vybrané úseky dálnic a jiných komunikací.
- **Regionální předpovědi** jsou vydávány pro oblasti o rozloze do několika stovek  $\text{km}^2$ . Tyto oblasti mohou být vymezeny geograficky (např. v ČR jižní polovina území, Českomoravská vrchovina) nebo politicky (např. Morava, Slezsko).

### 2.4.4 Dělení podle doby platnosti

- **Předpovědi velmi krátkodobé** jsou předpovědi na dobu kratší než 1 den (letecké předpovědi, výstrahy), součástí je nowcasting – předpověď na 0-2 hodiny.
- **Předpovědi krátkodobé** jsou předpovědi s dobou platnosti 1-3 dny. Vydávají se několikrát denně.
- **Předpovědi střednědobé** jsou předpovědi na 4-10 dnů. Vydávají se jednou za den.
- **Předpovědi dlouhodobé** jsou předpovědi na období delší než 10 dnů. Tyto předpovědi jsou založeny na studiu cirkulačních procesů na celé polokouli, vztazích mezi oceánem a atmosférou apod. Narozdíl od ostatních předpovědí nevycházejí z aktuálního stavu počasí. Vydávají se několikrát měsíčně.

Veškeré poznatky k této kapitole byly získány z [6].

## Kapitola 3

# Datová analýza

### 3.1 Použitá data

V této práci byla použita naměřená data a hodnoty pro předpověď počasí. Důležité je, že se jedná o veřejně dostupná data.

Naměřená data jsou z CHMI ([www.chmi.cz](http://www.chmi.cz)) a předpovědi (všeobecné předpovědi) jsou jak z portálu CHMI, tak z norského portálu [www.yr.no](http://www.yr.no).

Pro naměřená data rychlosti větru se jedná o hodnoty získané hodinově během celého roku. Naměřená data mají stejný charakter s lokalitami z obrázku 2.2 a jsou získávány každou hodinu během celého dne.

Získané předpovědi z CHMI byly udány jednou denně, ale nejsou pro lokality, nýbrž pro kraje. Předpověď byla udávána na den, noc a na další den. Pod pojmem den je v této práci chápán časový interval v rozmezí 7:00-19:00. V předpovědi rychlosti větru byly získány dvě hodnoty udávající minimální a maximální rychlost větru.

Předpověď norské družice je udávána stejně jako naměřená data z CHMI v hodinových nebo tří hodinových řezech během celého dne. Předpovědi jsou udávány pro lokality, které jsou na obrázku 2.2. Hodinové předpovědi jsou udávány pro takové lokality, které jsou v dosahu norské družice. Pro ostatní lokality je předpověď udávána v rozmezí tří hodin. Dosah družice nejprve dělil ČR na dvě poloviny. Od Prahy na západ byla předpověď udávána hodinově, jelikož zde měla družice dosah. Od Prahy na východ byla předpověď udávána v tří hodinovém intervalu. Ale během roku 2013 se dosah družice změnil a hodinová předpověď je udávána jen pro malou část od Prahy na sever. Pro větší část ČR jsou nyní předpovědi udávány v rozmezí tří hodin. Pro představu pomyslná hranice dosahu družice je na přímce Praha – Pardubice.

Naměřená data z CHMI byla získávána od roku 2011 do současnosti. Předpovědi z CHMI byla zislávána od roku 2011 do roku 2013. Pro rok 2014 nejsou poskytována žádná další data ke stažení.

Předpovědi z norského serveru jsou stahována od druhé poloviny roku 2012 do současnosti. Předpovědi rychlosti větru jsou udány pouze jednou hodnotou.

### 3.2 Naměřená data z CHMI

Jako první se podíváme na data získaná z CHMI.

V úvodních kapitolách se podíváme na rychlosti větru v celé ČR, po té se zaměříme na jednotlivé lokality. Podíváme se také, na trendy rychlosti větru během dne a jako poslední práci s naměřenými daty se podíváme na korelační analýzu.



### 3.2.1 Rozbor rychlosti větru

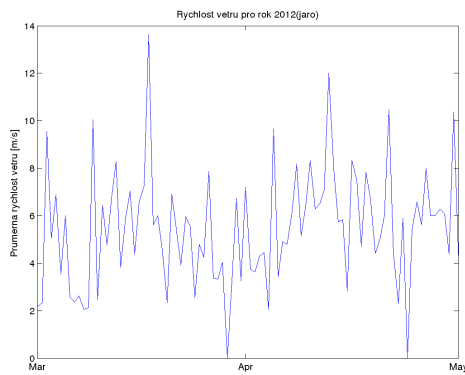
Vycházejme z předpokladů, které byly zaznamenány v kapitole 2.1.

V zimních měsících je průměrná rychlost větru vyšší. Za druhé se podíváme na průměrnou rychlost v horských oblastech. Ta by měla být vyšší než v jiných místech ČR.

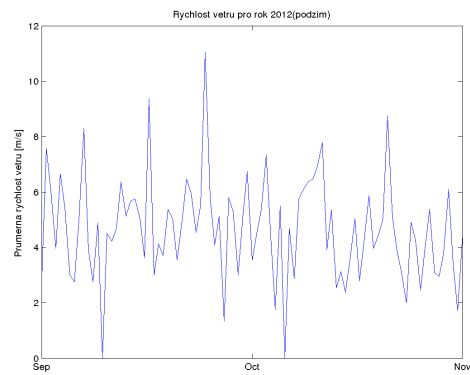
Abychom mohli tuto analýzu provést, nadefinujme si roční období, které je tříděno jenom vzhledem k měsícům. Zanedbáváme astronomické data ročních období.

<b>Jaro</b>	Březen, Duben, Květen
<b>Léto</b>	Červen, Červenec, Srpen
<b>Podzim</b>	Září, Říjen, Listopad
<b>Zima</b>	Leden, Únor, Prosinec

Tabulka 3.1: Třídění měsíců podle ročních období



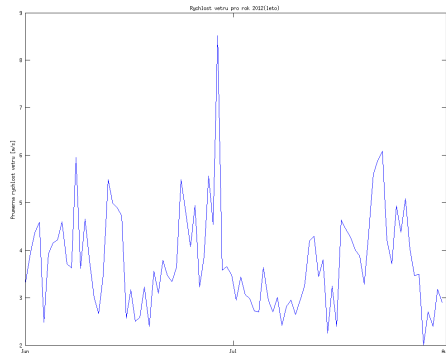
(a) Jaro



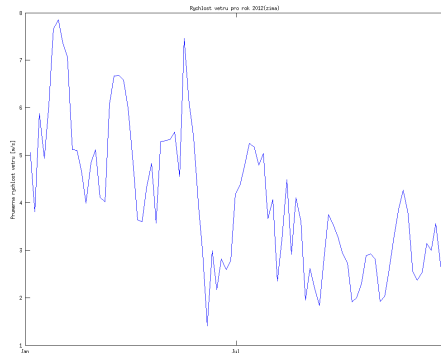
(b) Podzim

Obrázek 3.1: Rychlost větru na jaře a na podzim

Na obr. 3.1 lze vidět, že rychlost větru je skoro totožná jak v jarních, tak v podzimních měsících. Jako další byly porovnány letní a zimní měsíce.



(a) Léto



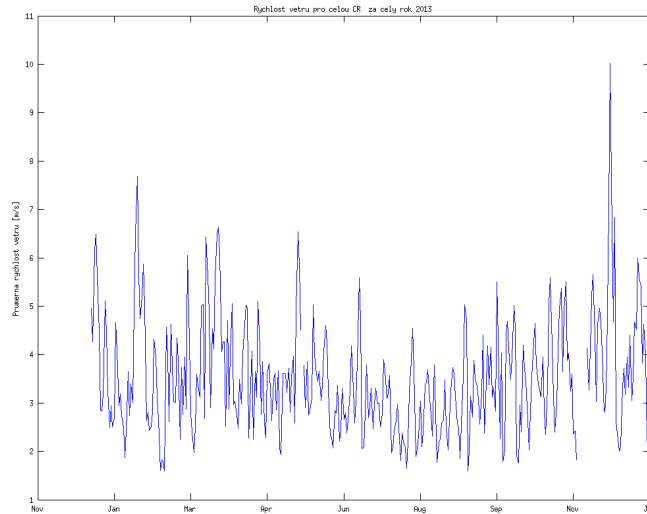
(b) Zima

Obrázek 3.2: Rychlost větru v létě a v zimě

Na obrázku 3.2 jsou vidět jasné rozdíly rychlostí větru.

Obrázky 3.1 a 3.2 jsou potvrzovány praxí. Zamysleme se jak se chová vítr v jednotlivých ročních obdobích a zjistíme, že realita správnost získaných grafů potvrzuje.

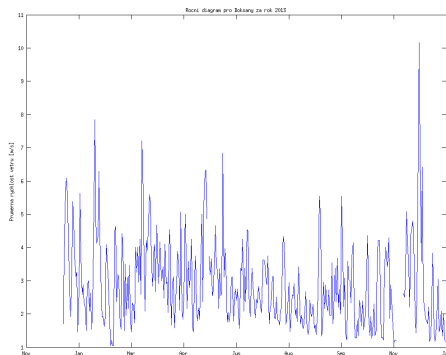
Jako poslední se podíváme na rychlost větru během celého roku



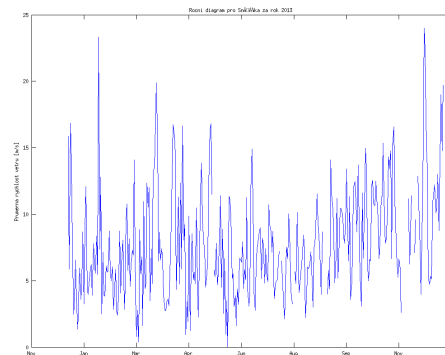
Obrázek 3.3: Průměrná rychlost větru během celého roku

I zde, na obrázku 3.3, jsou potvrzena všechna tvrzení, která jsme si postupnou analýzou grafů definovali.

Nyní se podíváme na další námi zadaný předpoklad, který se týkal závislosti rychlosti větru v nadmořské výšce. Pro názornost porovnejme mezi sebou nejnižší měřené místo (Doksany - 158 m.n.m) a nejvyšší měřené místo (Sněžka - 1602m.n.m) v ČR.



(a) Doksany



(b) Sněžka

Obrázek 3.4: Porovnání nejnižší a nejvyšší měřené polohy

Na obrázku 3.4 vidíme, že v Doksanech se pohybuje průměrná rychlost větru okolo 3,5 m/s. Oproti tomu na Sněžce je průměrná rychlost větru dvojnásobná.

Tímto obrázkem můžeme zcela potvrdit tvrzení, že v horských oblastech je průměrná rychlost vyšší než v nížinných oblastech. Z geologického hlediska toto tvrzení také můžeme potvrdit, jelikož nížinné lokality jsou v údolích, a tím pádem jsou před prudkými větry chráněny.

Bohužel vzhledem k různým nadmořským výškám, zde nemůžeme použít stejná měřítka pro y-ovou osu.

### 3.2.2 Denní diagram

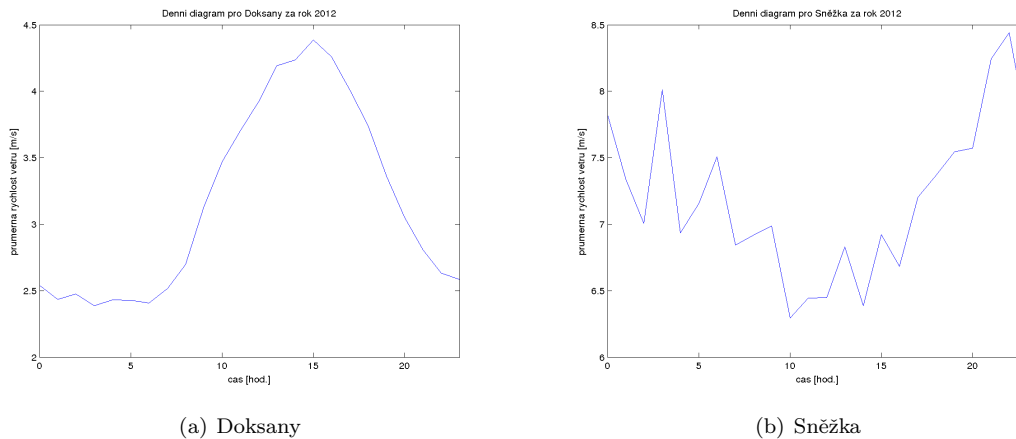
Pomocí denního diagramu jsme schopni odhadnout trend průběžné rychlosti větru během dne.

Očekáváme, že ve vysokohorských oblastech by měl mít trend tvar písmene „U“ a v níže položených lokalitách by měl být zcela inverzní. V této části jsme očekávali, že v ČR budou mít spíše vysokohorský trend denního diagramu. Domníváme se, že abychom měli i nížinný trend, museli bychom mít měřené lokality mít nižší nadmořskou výšku než 158 m.n.m. Pokusme se následující analýzou naše předpoklady potvrdit, či vyvrátit.

Dení diagram jsme vypočítali podle následujícího vztahu :

$$DD(t) = \frac{1}{PD} \sum_{i=1}^{PD} x[(i-1) - 24 + t], \quad (3.1)$$

kde  $PD$  je počet dní,  $x$  jsou naměřená data a  $t = 1, 2, \dots, 24$  Podívejme se opět na Doksany a Sněžku, které jsme použili již v kapitole 3.2.1.



Obrázek 3.5: Porovnání nejvyšší a nejnižší měřené polohy

Z obrázkem 3.5 vyplývá, že jsme se v očekávání, které je popsáno výše, mýlili. I v ČR jsou lokality, které mají jasně viditelný nížinný trend. Dále vidíme, že poloha denního diagramu je ovlivněna nadmořskou výškou. S rostoucí nadmořskou výškou roste i poloha denního diagramu vůči y-ové ose.

### 3.2.3 Výpočet korelací rychlosti větru

Ze známých naměřených dat byla vypočtena korelační matice, ve které jsou spočteny vzájemné korelace mezi všemi lokalitami. Pro výpočet byl použit vztah viz B.1.3.

Není-li v práci uvedeno jinak, myslíme tím plošnou korelací mezi lokalitami.

Výsledná matice je symetrická a uložena to tabulky pomocí programového prostředí MS Excel (*korelace\_lokace.xls*).

Na diagonále jsou 1, což značí přímou závislost na dané lokaci. Při pohledu na mezi sebou nejbližší lokality (Cheb, Lysá hora), je korelace blízká hodnotě 0. Z čehož vyplývá, že se rychlost větru v těchto dvou lokalitách neovlivňuje.

Pokud se podíváme na lokality, které jsou blízko u sebe (Praha–Ruzyně, Praha–Karlovy, Praha–Libuň, Praha–Kbely), je hodnota korelačního koeficientu blízká 0.8, což značí, že tyto lokality jsou mezi sebou závislé.

Jako příklad se můžeme podívat na vybrané hodnoty, které jsou uvedeny v tabulce 3.2.

	Cheb	Karlovy Vary	Přímda	Tušimice	Plzeň – Míkulka	Churáňov
Cheb	1,0000	0,6957	0,6054	0,5417	0,6205	0,4828
Karlovy Vary	0,6957	1,0000	0,6420	0,5958	0,6501	0,4949
Přímda	0,6054	0,6420	1,0000	0,4384	0,6179	0,6252
Tušimice	0,5417	0,5958	0,4384	1,0000	0,5737	0,4235
Plzeň – Míkulka	0,6205	0,6501	0,6179	0,5737	1,0000	0,5520
Churáňov	0,4828	0,4949	0,6252	0,4235	0,5520	1,0000

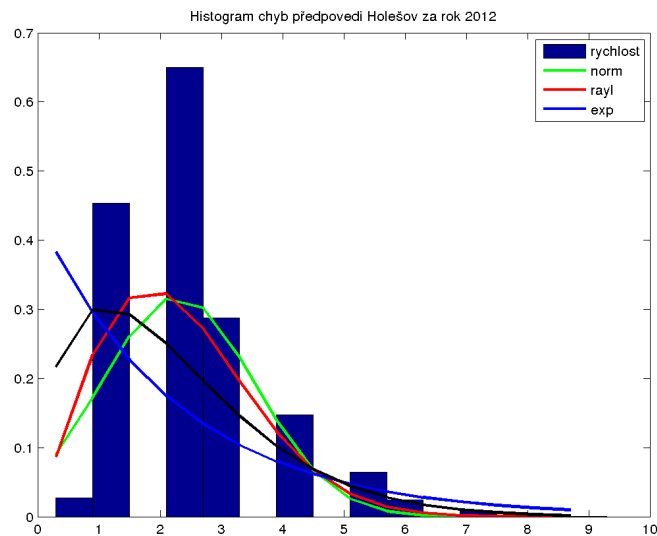
Tabulka 3.2: Ukázka hodnot korelačních koeficientů za rok 2013 pro několik vybraných lokalit

### 3.2.4 Analýza rychlosti větru

Pokusme se zjistit, které pravděpodobnostní rozložení nejlépe odpovídá námi získaným histogramům v různých lokalitách. Naměřené hodnoty jsou prokládány čtyřmi rozděleními :

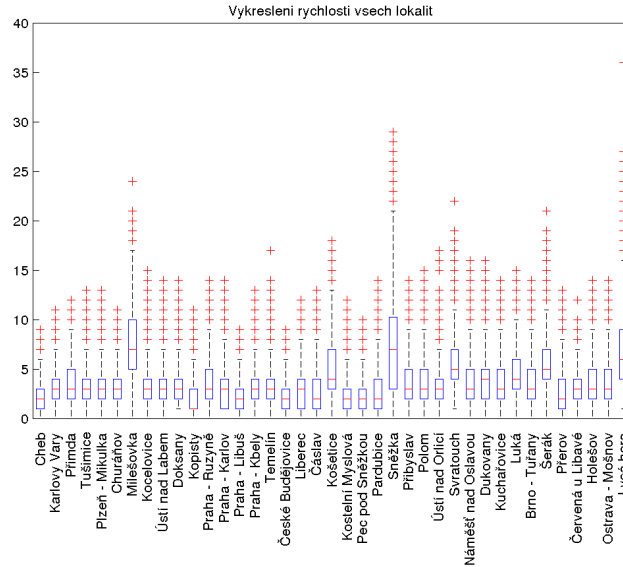
- Normální rozdělení,
- Exponenciální rozdělení,
- Weibullovo rozdělení,
- Raygleyho rozdělení.

Po prozkoumání histogramů, jsme zjistili, že Weibullovo a Raygleyho rozdělení reprezentují nejlépe chování větru. Pro důkladnější analýzu jsme vypočetli Euklidovu normu (viz C.2), pomocí které jsme zjistili, že nejlépe reprezentuje chování větru v ČR Raygleyho rozdělení. Pro ukázkou je zde uveden obrázek 3.6.



Obrázek 3.6: Histogram četnosti rychlosti větru s hustotami pravděpodobnosti

Pro názornost byl vykreslen boxplot (viz B.4) pro všechny měřené lokality a pro kraje. Podívejme na graf pro celou ČR (obrázek 3.7)



Obrázek 3.7: Porovnání rychlostí větrů pro všechny lokality pomocí boxplot

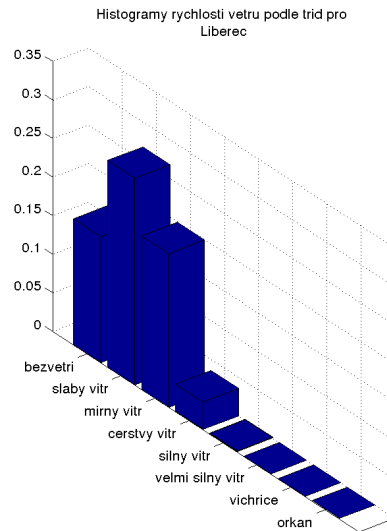
Obrázek 3.7 nám říká, že průměrná rychlost větru se ve všech lokalitách pohybuje přibližně kolem stejných hodnot. Vyjímka jsou horské lokality (Milešovka, Sněžka, Lysá hora), kde je průměrná rychlost vyšší. Pokud se ale podíváme jenom na tyto tři horské lokality, vidíme, že i zde je průměrná rychlost větru stejná. Jenom na Sněžce je větší variabilita rychlosti větru.

Tyto lokality byly ještě vykresleny do grafů vzlašť podle krajů. Jsou uloženy na příloženém CD/DVD.

### 3.2.5 Četnost podle tříd

Rychlosti větru v jednotlivých lokalitách, jsme vyřídili rychlosti podle Beaufortovy stupnice (tabulka 2.1). Získali jsme histogramy, ze kterých jsme mohli učinit několik závěrů.

Převážně v celé ČR je vítr v rozmezí bezvětří až čerstvý vítr. Nejvíce se vyskytuje slabý vítr. V horských oblastech je převážně čerstvý vítr. Velmi zajímavé výsledky jsme získali pro Pec pod Sněžkou, kde převládá bezvětří. To může být zapříčiněno buď špatným měřením nebo okolním reliéfem. Pro názornost je na obrázku 3.8 uveden příklad získaného histogramu.



Obrázek 3.8: Četnost typů větru podle Beaufortovy stupnice

Po prozkoumání získaných grafů můžeme říci, že potvrzují správnost obrázku 2.1.

Dále jsme pro jednotlivé třídy větru spočetli medián (více viz. B.1.5). Získané výsledky jsou prezentovány v tabulce 3.3.

síla větru	rychlost [m/s]	modus [m/s]
bezvětří (klidno)	0-1	1
slabý vítr	1-3	3
mírný vítr	3-6	5
čerstvý vítr	6-10	8
silný vítr	10-15	12
velmi silný vítr	15-21	17
vichřice	21-30	24
orkán	nad 30	36

Tabulka 3.3: Třídy větru s modus z naměřených dat

### 3.3 Předpověď dat z CHMI

Jako další se podíváme na chybu předpovědi. Předpověď rychlosti větru je udávána dvěma hodnotami (minimální a maximální rychlost větru).

Dalším problémem byla samotná předpověď, která je udávána pouze pro kraje, ale naměřená data jsou pro lokality. Proto zde bylo velmi důležité kontrolovat, zda-li zkoumaná lokalita je porovnávána s příslušným krajem do kterého spadá.

Chyba byla vypočtena dle následujícího vztahu

$$e = x - \hat{x}, \quad (3.2)$$

kde  $x$  jsou průměrné rychlosti a  $\hat{x}$  jsou předpovězené hodnoty.

Předpověď obsahovala následující údaje:

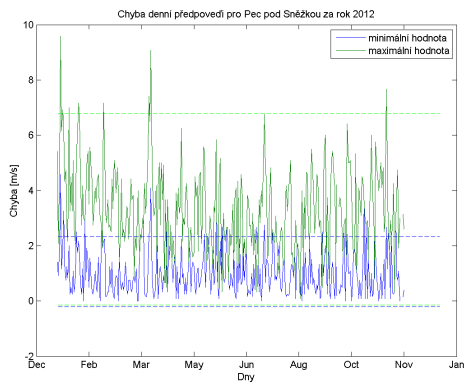
- předpověď na den,
- předpověď na noc,
- předpověď na další den.

Pod pojmem den, jak již bylo řečeno v kapitole 3.1, den začíná v 7:00, noc začíná v 19:00. Ohledně předpovědi na další den, zde bereme stejné hodinové rozmezí.

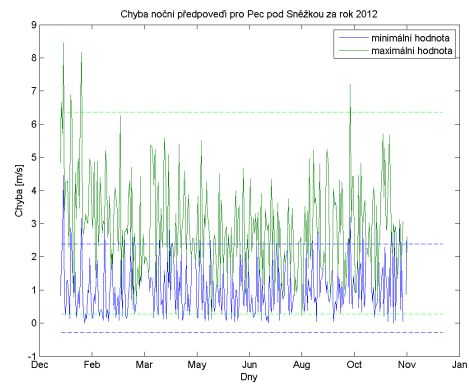
Následující grafy byly získány tak, že jsme ze získaných dat vyseparovali potřebné údaje o rychlosti větru v rozmezí požadovaného časového intervalu. Tyto hodnoty jsme pak odečetli od maximální a minimální předpovězené rychlosti větru.

Z obrázku 3.9 můžeme udělat několik závěrů.

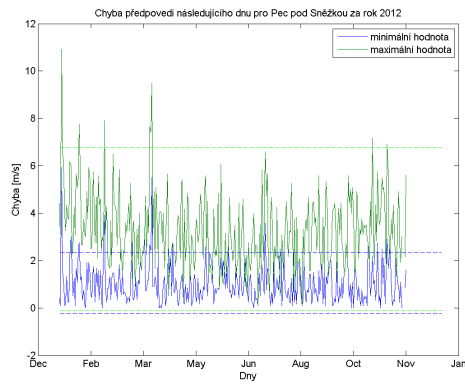
Hodnota chyby pro minimální rychlost větru je nižší, než pro maximální hodnoty. V zimních měsících je velikost chyby vyšší. Velikost chyby pro den, noc a následující den je přibližně stejná.



(a) Chyba předpovědi na den



(b) Chyba předpovědi na noc



(c) Chyba předpovědi na další den

Obrázek 3.9: Chyba předpovědí pro předpovězené minimální a maximální hodnoty



### 3.3.1 Směrodatná odchylka chyby předpovědi

Pro výpočet je použit vztah, který je uveden v B.1.4

Veškerá data jsou opět uložena v tabulce pomocí MS Excel pod názvem *STD\_lokace.xls*.

Pro výpočet chyby byl použit vztah 3.2.

Po prozkoumání těchto dat můžeme usoudit, že směrodatná odchylka předpovědi pro minimální hodnoty rychlost větru je menší než pro maximální hodnoty. Dále jsme zjistili, že vypočtené hodnoty odchylky jsou pro den (popřípadě noc) nižší než pro předpověď na další den. Tento poznatek platí jak pro minimální, tak pro maximální hodnoty předpovědi.

Při pohledu na tyto data vidíme, že většina hodnot se pohybuje v rozmezí 1,5 - 2,1 m/s u předpovědi pro minimální rychlost větru. Horší jsou hodnoty pro výše položené lokality, např. Sněžka, Šerák. Zde se hodnoty odchylky pohybují přes 3. Zajímavé jsou hodnoty u Šeráku. Ty jsou naprosto obrácené, než u zbytku lokalit. Vypočtená směrodatná odchylka je větší pro minimální hodnoty než pro maximální. Naprosto největší hodnoty jsou, dle očekávání, na Sněžce, které se pohybují v rozmezí 5-6 m/s.

lokality	STD - den		STD-noc		STD - další den	
	min. hod.	max. hod.	min. hod.	max. hod.	min. hod.	max. hod.
<b>Cheb</b>	1,1000	2,7295	1,2688	2,44519	1,3508	2,7295
<b>Karlovy Vary</b>	1,6320	2,0208	1,5466	1,8808	1,87055	2,02080
<b>Přimda</b>	2,1781	1,5803	1,9422	1,8889	2,03874	1,5803
<b>Plzeň – Mikulka</b>	1,8273	1,6371	1,2660	2,3823	1,5859	1,6371
<b>Sněžka</b>	5,6191	4,07051	5,7629	4,3216	5,6370	4,0705
<b>Šerák</b>	3,3925	1,9310	3,7907	2,06035	3,5897	1,9310

Tabulka 3.4: Ukázka hodnot směrodatných odchylek chyby předpovědi za rok 2013 pro několik vybraných lokalit

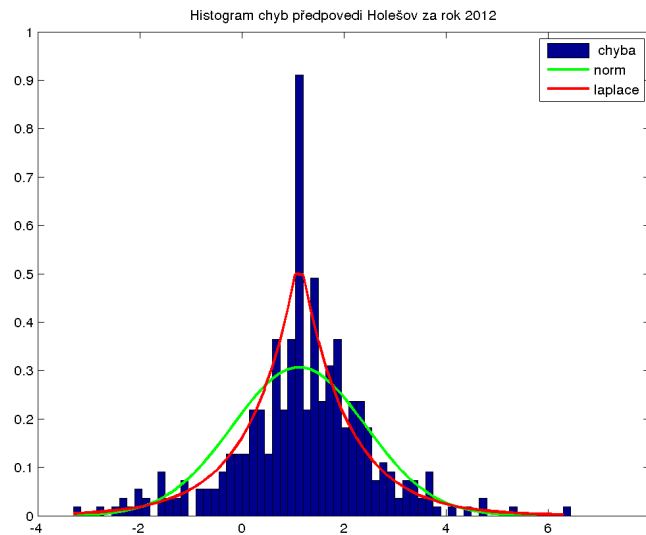
### 3.3.2 Analýza chyby

Jako další se podíváme na histogramy chyb. Pro výpočet chyby byl použit vztah 3.2. Výsledné histogramy jsme tentokrát prokládali dvěma rozloženími :

- Normálním rozdělením,
- Laplaceovým rozdělením.

Více o těchto rozloženích je uvedeno v B.2.1 a B.2.3.

Výsledný ukázkový histogram můžeme vidět na obrázku 3.10.

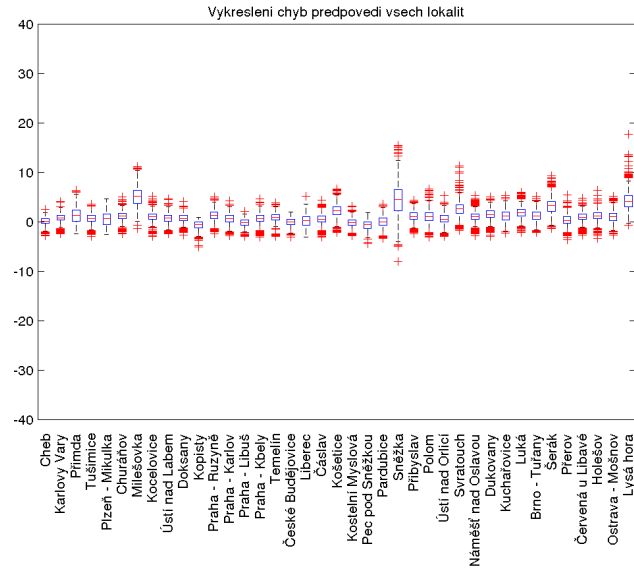


Obrázek 3.10: Histogram četnosti chyby předpovědi s hustotami pravděpodobnosti

Připomeňme si, že jak bylo již uvedeno výše, předpovědi z CHMI jsou uváděny pro celé kraje a jejich granularita je pouze jedno číslo na celý den.

I zde pro nejasný výsledek byla použita Euklidova norma. Po zhodnocení výsledků jsme zjistili, že nejlépe chování chyby vyhovuje Laplaceovo rozdělení.

Pro získaná data, jsme si vykreslili boxplot, který je na obrázku 3.11



Obrázek 3.11: Porovnání chyb předpovědí pro všechny lokality pomocí boxplot

Z obrázku 3.11 můžeme zjistit, že např. v Chebu a Kopistech je hodnota chyby předpovědi nejmenší, a tudíž je zde nejpřesnější předpověď počasí, alespoň co se týká rychlosti větru. Opět vidíme vyčnívající horské oblasti, pro které je těžké rychlost větru předvídat a podle toho i vypadá hodnota chyby. Sněžka v tomto grafu opět dominuje svou polohou kvartilů.

### 3.3.3 Korelační analýza chyby předpovědi

V této kapitole jsme se zaměřili na korelační analýzu chyb předpovědi.

Z důvodu, že je předpověď z CHMI udávána pro kraje, je i tato analýza uváděna pro jednotlivé kraje.

Získané výsledky jsou uloženy do tabulek v programovém prostředí MS Excel pod názvem *korelace\_lokaci\_predp\_-,typ*.xls. Kde „typ“ značí, na jaký časový úsek je daná předpověď udána (den, noc, další den).

V této části bylo důležité si vytřídit dané lokality pro jednotlivé kraje. Z vytříděných hodnot jsme vypočetli průměrné hodnoty, které byly nadále použity pro výpočet podle vztahu B.7. Pro výpočet chyby byl použit vztah 3.2, který je uveden v kapitole 3.3.

Po prozkoumání získaných korelačních koeficientů jsme zjistili, že hodnoty koeficientů si jsou podobné pro všechny typy předpovědí. Hodnoty korelačních koeficientů jsou pro minimální hodnoty nižší než pro maximální hodnoty předpovědi. Hodnota koeficientu se s rostoucí vzdáleností mezi jednotlivými kraji snižuje, i když málo.

Pro ukázkou jsou zde uvedeny dvě tabulky pro minimální hodnoty předpovědi (tabulka 3.5) a maximální hodnoty předpovědi (tabulka 3.6)

	Karlovarský	Plzeňský	Ústecký	Středočeský	Praha
Karlovarský	1,0000	0,8702	0,9324	0,9353	0,9160
Plzeňský	0,8702	1,0000	0,8661	0,8534	0,7870
Ústecký	0,9324	0,8661	1,0000	0,9112	0,8362
Středočeský	0,9353	0,8534	0,9112	1,0000	0,9366
Praha	0,9160	0,7870	0,8362	0,9366	1,0000

Tabulka 3.5: Ukázka korelační analýzy chyby předpovědi udané minimální rychlostí větru pro rok 2012

	Karlovarský	Plzeňský	Ústecký	Středočeský	Praha
Karlovarský	1,0000	0,9704	0,9870	0,9869	0,9841
Plzeňský	0,9704	1,0000	0,9644	0,9618	0,9556
Ústecký	0,9870	0,9644	1,0000	0,9865	0,9813
Středočeský	0,9869	0,9618	0,9865	1,0000	0,9914
Praha	0,9841	0,9556	0,9813	0,9914	1,0000

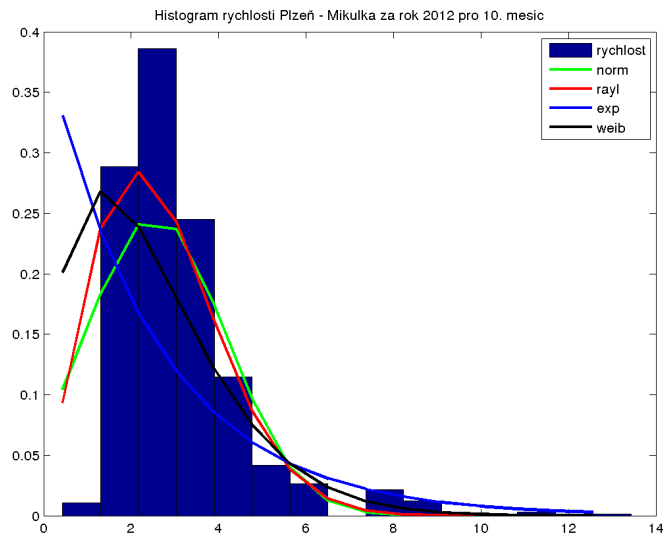
Tabulka 3.6: Ukázka korelační analýzy chyby předpovědi udané maximální rychlostí větru pro rok 2012

## 3.4 Podrobnější pohled na vybrané lokality pro data z CHMI

Pro názornost byly vybrány čtyři lokality z ČR (Plzeň–Mikulka, Temelín, Tušimice, Ostrava–Mošnov), které dostatečně a názorně reprezentují chování větru v ČR. Zaměříme se zde pouze na lokalitu Plzně, jejíž rozbor je v kapitolách 3.4.1 a 3.4.2. Zbylé lokality jsou na příloženém CD/DVD.

### 3.4.1 Rychlost větru

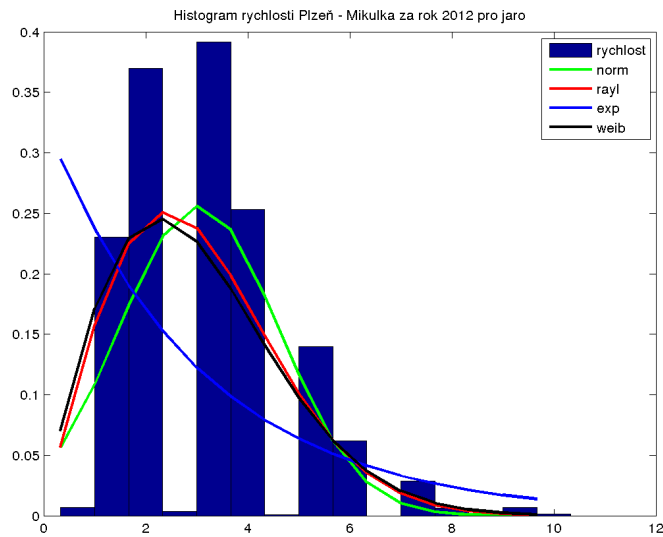
Podívejme se na rychlosti větru z hlediska měsíců a ročních období. Po prozkoumání dat, bylo zjištěno, že rychlost větru se s postupujícími měsíci nijak závažně nemění. Jedinými dvěma měsíci, kde si můžeme říci, že je vítr silnější jsou září a listopad. I zde byly histogamy prokládány již výše zmíněnými pravděpodobnostními rozděleními. Získané výsledky jenom potvrzují již známý fakt o Raygleyho rozdělení. Jako ukázkový příklad se můžeme podívat na obrázek 3.12.



Obrázek 3.12: Histogram četnosti větru pro Plzeň v říjnu

Nyní se podívejme na rychlost větru vzhledem k ročním obdobím. Při dělení měsíců na roční období jsme opět zanedbávali astronomické data a řídili jsme se tabulkou 3.2.1.

Zjistili jsme, že rychlost větru na jaře a na podzim jsou si velmi blízké, což potvrzuje opět již zjištěný fakt z kapitoly 3.2.1. Ale pravým opakem jsou letní a zimní měsíce, kde podle získaných histogramů více fouká v létě než v zimě. Podívejme se pro ukázkou na jarní měsíce, viz obrázek 3.13.

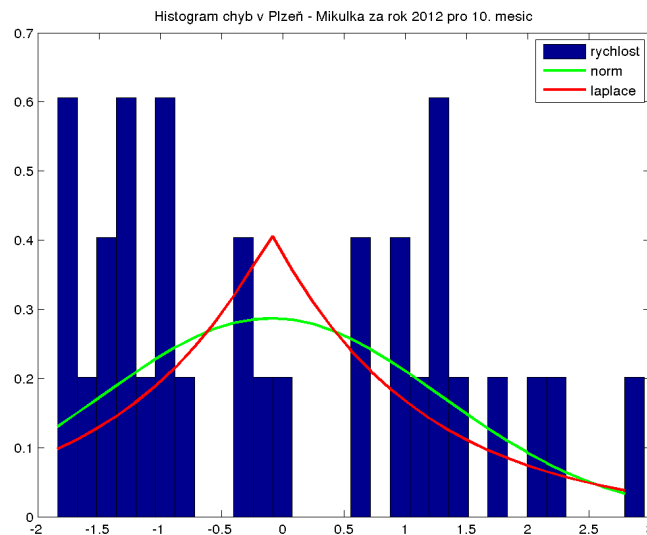


Obrázek 3.13: Histogram četnosti rychlosti větru pro Plzeň v jarních měsících

### 3.4.2 Chyba předpovědi

Jelikož nás zajímá, jak vypadá předpověď pro Plzeň, podívejme se v této kapitole na chyby předpovědi. Chyba bylo opět počítána pomocí vztahu 3.2.

Podívejme se opět na říjnový měsíc.

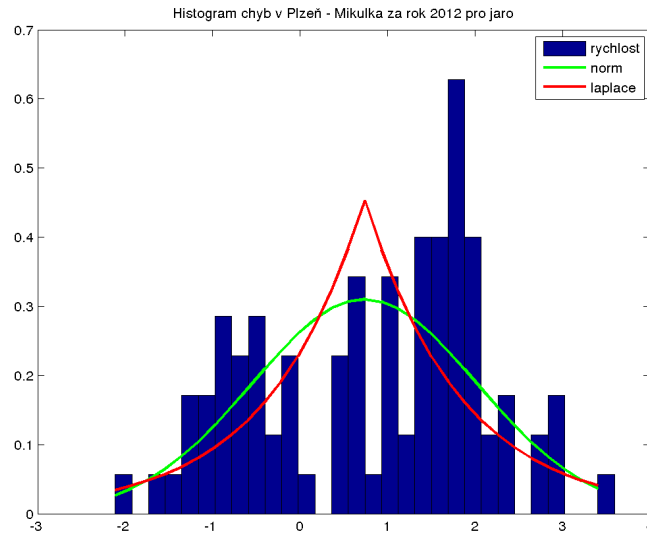


Obrázek 3.14: Histogram četnosti chyb předpovědi pro říjnový měsíc

Z obrázku 3.14 nemůžeme vyčíst, zda-li se pro popis chyby lépe hodí Normální a nebo Laplaceovo rozdělení. Tento závěr můžeme udělat pro všechny získané histogramy.

Pokusme se zjistit, jestli jsme schopni něco zjistit alespoň z ročních období.

Na obrázku 3.15 vidíme, že proložení hustotami pravděpodobnosti už je lepší, ale stále nemůžeme



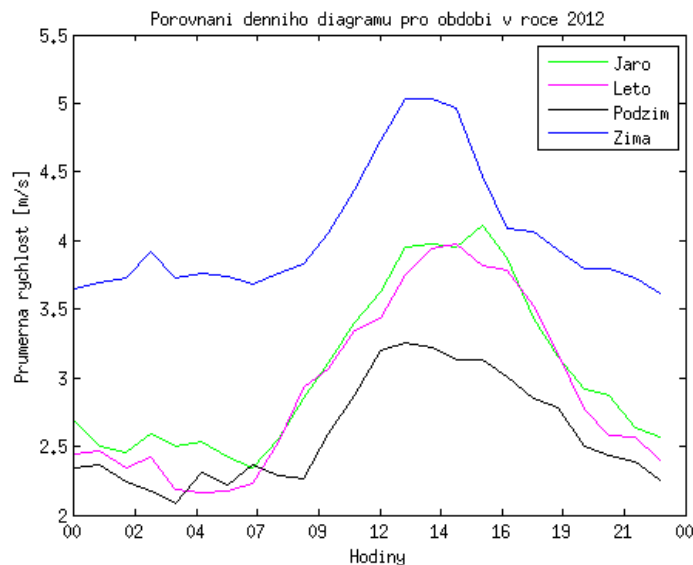
Obrázek 3.15: Histogram četnosti chyb předpovědi pro Plzeň v jarních měsících

nic usoudit. Opět jsme bohužel získali podobné grafy pro všechna roční období. Nejhůře je na tom histogram pro zimní měsíce.

Jako poslední nás zajímala hodinová data a jejich četnost z hlediska denních diagramů. Tuto analýzu jsme provedli pouze pro lokalitu Plzeň. Analýza byla provedena z hlediska ročních období a tříd větru. Výsledná data byla uložena do datové struktury *Plzeň - Mikulka\_DD.mat*.

Pokud se podíváme do již zmíněné datové struktury na vyříděné rychlosti s ohledem na roční období, nemůžeme denní diagramy vykreslit, jelikož požadujeme pro jednu hodinu denního diagramu minimálně deset údajů o rychlosti větru. To splňuje pouze slabý a mírný vítr. Tento poznatek potvrzuje získaný obrázek 3.8.

Podívejme se na obrázek 3.16, který denní diagramy během jednotlivých ročních období.



Obrázek 3.16: Porovnání denních diagramů pro jednotlivé roční období pro Plzeň

Z obrázku 3.16 můžeme vyčíst, že v zimních měsících je rychlost větru během dne vyšší. Překvapivé jsou denní diagramy jara a léta, které jsou prakticky totožné. Většinou jsou jarní rychlosti větru podobné podzimním, které jsou v tomto případě nejnižší.



### 3.5 Naměřená data z YR

Po zpracovaných datech z CHMI se nyní podíváme na norskou předpověď, která je udávána na [www.yr.no](http://www.yr.no). Předpověď z norské družice je hodinová pro část ČR směrem od Prahy na západ. Předpověď pro druhou část ČR je udávána co tři hodiny. Bohužel však na pomezí roku 2013 a 2014 se opět dosah družice změnil. Nyní je i lokalita Plzně v části, kde se udává předpověď jednou za tři hodiny. Hodinové předpovědi jsou nadále udávány pro severozápadní a severní část ČR.

Předpovědi jsou ukládány v datové struktuře do sloupců, kde první sloupec značí nejbližší předpověď pro danou lokalitu. S rostoucím počtem sloupců se aktuálnost předpovědi snižuje.

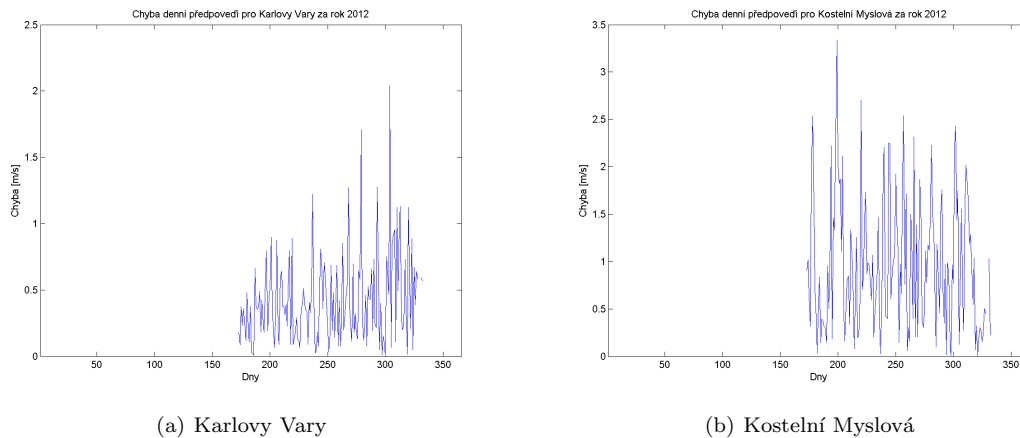
Pro porovnání chyb předpovědi používáme první sloupec (nejbližší možná předpověď) z dostupných dat.

Pokud se vyskytují v textu předpovědi z jiných sloupců, mají pouze informační charakter.

#### 3.5.1 Chyba předpovědi

I zde se podíváme na chybu předpovědi, jak tomu bylo v kapitole 3.3 a pokusme se porovnat oblasti, kde družice dosáhne s oblastmi, kde jsou předpovědi spíše odhadovány.

Pro naše porovnání se podíváme na dvě lokality. Na Karlovy Vary (606 m.n.m) a Kostelní Myslovou (659 m.n.m)<sup>1</sup>.

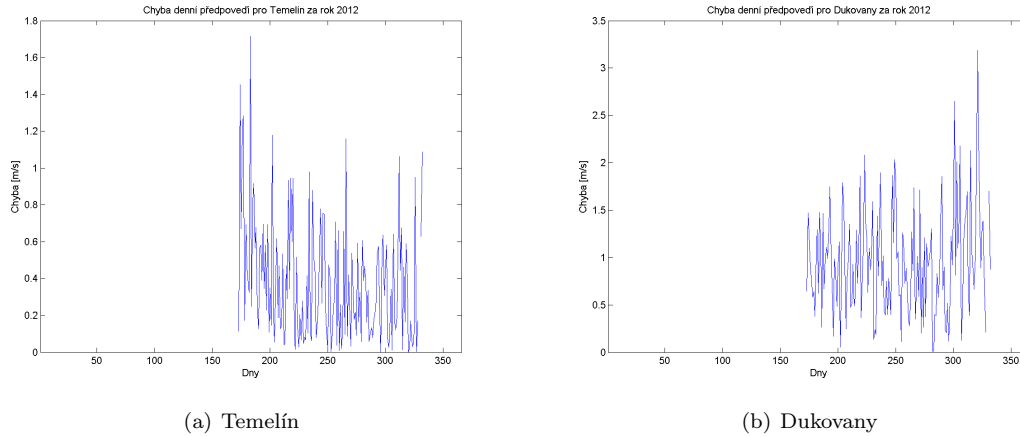


Obrázek 3.17: Porovnání dvou lokalit v ČR, kde každá je na jiné straně dosahu družice

Na obrázku 3.17 už na první pohled vidíme, že v Kostelní Myslové je chyba větší, a to zhruba dvojnásobně.

<sup>1</sup>Bohužel nebyly nalezeny lokality a měli by podobnou nadmořskou výšku. Tento výběr byl nejlepší

Zkusme ještě porovnat lokality, které jsou si blízké, ale jedna je v dosahu družice, zatímco druhá nikoliv. Pro tuto analýzu nám poslouží Temelín (503 m.n.m) a Dukovany (400 m.n.m)



Obrázek 3.18: Porovnání chyb dvou předpovědí, pro lokality si blízké

Na obrázku 3.18 vidíme, že i přes vyšší nadmořskou výšku má Temelín stále lepší předpověď než Dukovany.

### 3.5.2 Výpočet autokorelací

Podívejme se na vypočtené korelační koeficienty. Hodnoty pro rok 2012 jdou výrazně nižší, než pro rok 2013. To může být způsobeno tím, že jsme začali sbírat data teprve v polovině roku 2012. Podívejme se proto důkladněji na rok 2013, který nám z hlediska úplnosti dat řekne více. Pro horské oblasti jako jsou již už mnohokrát zmiňovaná Sněžka a Lysá hora je korelační koeficient blízký hodnotě 0.2. Můžeme tedy říci, že pro tyto lokality předpověď není zcela přesná. Zajímavé výsledky jsme získali z pražských lokalit, kde nejvyšší koeficient je na Ruzyni, který je přibližně 0.7. Oproti tomu ostatní pražské lokality (Karlov, Libuš, Kbely) mají hodnoty pouze okolo 0.4. Takový rozdíl hodnot mohl vzniknout tím, že na Ruzyni, mají vlastní, možná i přesnější, meteostanici kvůli letovému provozu.

Výsledná data jsou uložena to tabulky v MS Excel pod názvem *Korelace\_norsko.xls*.

Na ukázkou několika hodnot se můžeme podívat do tabulky 3.7

lokality	korelační koeficient
<b>Cheb</b>	0,7263
<b>Karlovy Vary</b>	0,7317
<b>Přimda</b>	0,7501
<b>Tušimice</b>	0,6632
<b>Plzeň – Mikulka</b>	0,6995

Tabulka 3.7: Ukázka hodnot autokorelačního koeficientu pro vybrané lokality za rok 2013

### 3.5.3 Výpočet plošné korelace pro YR

Jako tomu bylo již v analýze CHMI (kapitola 3.2.3), i zde jsme počítali korelace. Vztah pro výpočet korelačního koeficientu je uveden v kapitole B.1.3.

V tabulce 3.7 jsme se zaměřili na koeficienty korelačních koeficientů chyby předpovědi.

Zjistili jsme, že s rostoucí vzdáleností mezi jednotlivými lokalitami hodnota koeficientu stejná nebo mírně klesající. Divné je, že pro horské oblasti získáváme pro některé lokality záporné korelace.

Při pohledu na koeficienty vypočtených z dat z roku 2012, jsme zjistili, že výsledky jsou neuspokojivé. Veškeré hodnoty koeficientů jsou kolem 0,5. To ovšem neplatí pro rok 2013 kde jsou hodnoty zcela jiné. Neuspokojivé korelace přisuzujeme nedostatečným datům z roku 2012, které začaly být získávány od poloviny již zmíněného roku.

	<b>Cheb</b>	<b>Karlovy Vary</b>	<b>Přímda</b>	<b>Tušimice</b>	<b>Plzeň – Míkulka</b>
<b>Cheb</b>	0,9391	0,9178	0,8941	0,8776	0,9116
<b>Karlovy Vary</b>	0,9178	0,9215	0,8995	0,8826	0,9155
<b>Přímda</b>	0,8941	0,8995	0,9159	0,8753	0,9115
<b>Tušimice</b>	0,8776	0,8826	0,8753	0,8926	0,8956
<b>Plzeň – Míkulka</b>	0,9116	0,9155	0,9115	0,8956	0,9157

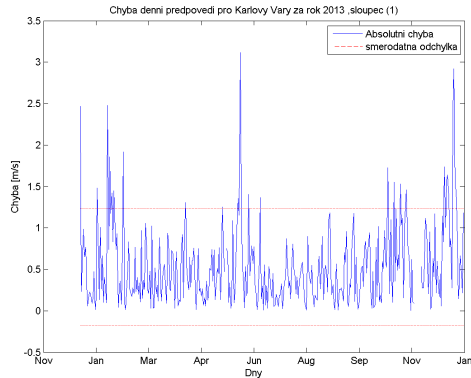
Tabulka 3.8: Ukázka korelačních koeficientů pro chybu předpovědi pro rok 2013

### 3.5.4 Směrodatná odchylka chyby předpovědi YR

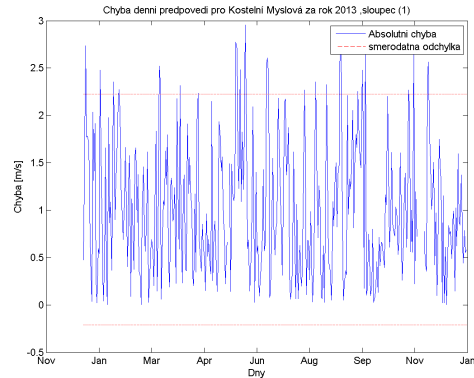
Jak tomu bylo již v kapitole 3.3.1, podívejme se na hodnoty směrodatné odchylky.

Zaměříme se na stejné lokality, jako tomu bylo v kapitole 3.5.1.

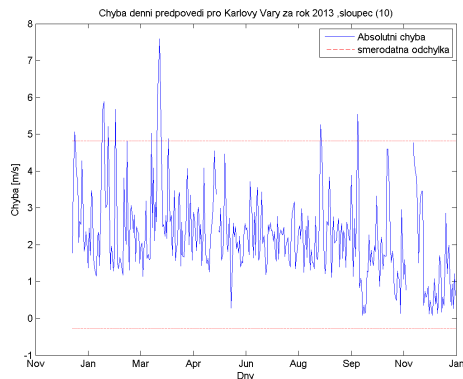
Připomeňme si, že předpovědi ze serveru [www.yr.no](http://www.yr.no) jsou uloženy do několika sloupců, přičemž první sloupec značí nejbližší možnou předpověď. Proto jsou v popisu jednotlivých obrázků napsány sloupce. Pokusme se zaměřit i na to jak je předpověď přesná i z tohoto hlediska.



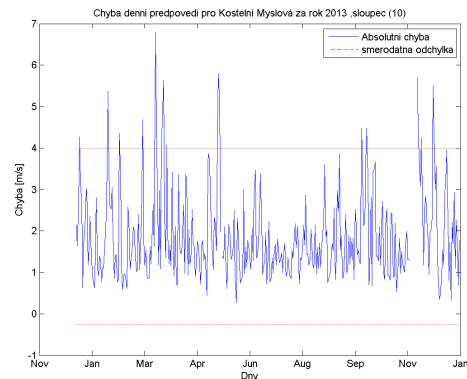
(a) Karlovy Vary – sloupec 1



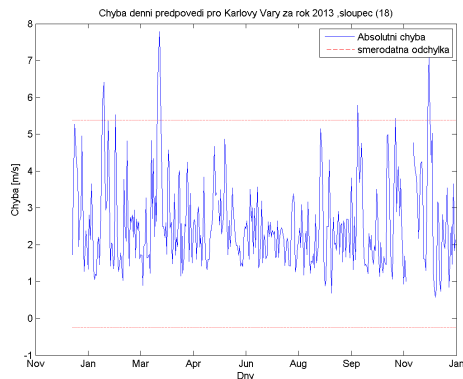
(b) Kostelní Myslová – sloupec 1



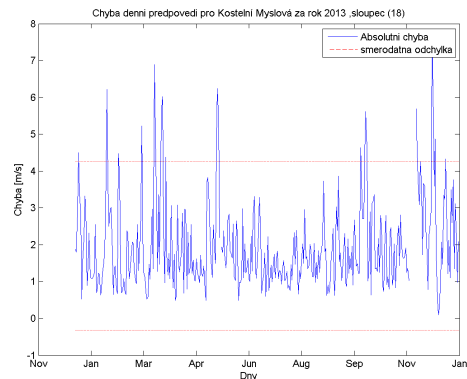
(c) Karlovy Vary – sloupec 10



(d) Kostelní Myslová – sloupec 10



(e) Karlovy Vary – sloupec 18



(f) Kostelní Myslová – sloupec 18

Obrázek 3.19: Chyba předpovědi z norského serveru pro různá data

Z obrázku 3.19 můžeme vidět, že na prvním porovnávaném grafu jsou hranice směrodatné odchylky užší v Karlových Varech, to je způsobeno nejspíše polohou Karlových Varů. Očekávali jsme, že pás mezi směrodatnými odchylkami bude užší v případě Karlových Varů, ale pás směrodatné odchylky pro desátý a osmnáctý sloupec, je pro Kostelní Myslovou dokonce užší. Na druhou stranu se jedná o hodnotu přibližně 1 m/s. Různá šířka pásů je nejspíše zapříčiněna členitostí terénu a svůj vliv zde má i různá nadmořská výška.

Po prozkoumání termínů předpovědí, jsme zjistili, že desátý sloupec pro Karlovy vary reprezen-

tuje předpověď pět dní starou. Předpověď z osmnáctého sloupce je stará šest dní. Stejný poznatek platí i pro Kostelní Myslovou.

Pro ukázkou jsou některé hodnoty zapsány v tabulce 3.9.

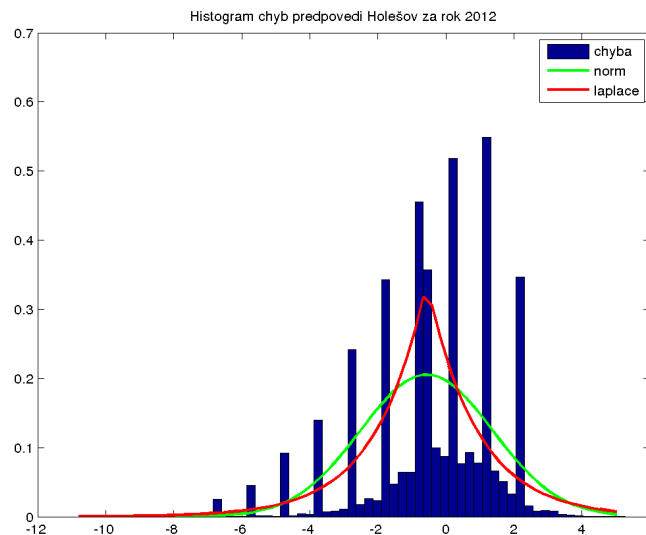
lokality	STD - sloupec 1	STD - sloupec 10	STD - sloupec 18
<b>Cheb</b>	1,2649	1,8585	2,0126
<b>Karlovy Vary</b>	0,7053	2,5421	2,8166
<b>Přimda</b>	0,8561	3,2222	3,6370
<b>Sněžka</b>	5,9825	8,0624	8,1834
<b>Šerák</b>	3,3456	5,6866	5,8338

Tabulka 3.9: Ukázka hodnot směrodatné odchylky pro předpověď za rok 2013

### 3.5.5 Analýza chyby

Jak tomu již bylo v kapitole 3.3.2. Bylo použito i stejné pravděpodobnostní rozdělení.

Podívejme se pro histogramy četnosti chyb předpovědí pro Holešov jako tomu bylo ve výše zmíněné kapitole na obrázku 3.10 a pokusme se je porovnat.

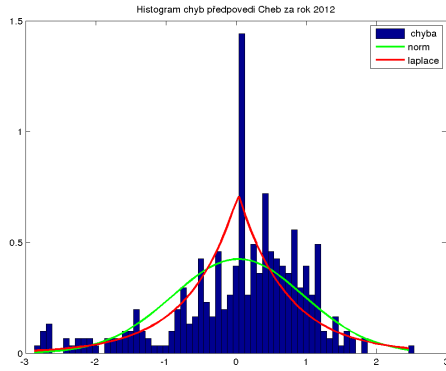


Obrázek 3.20: Histogramy četnosti chyb pro norskou předpověď

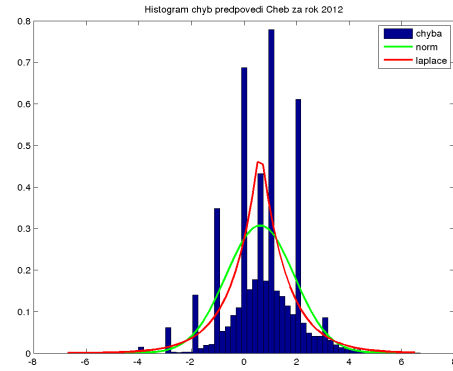
Na obrázku 3.20 vidíme, že se ve stejných intervalech objevují výrazně vyšší hodnoty.

Porovnejme výsledky chyb předpovědí z obrázků 3.10 a 3.20. Vidíme, že histogramy všech rozložení jsou na obou obrázcích posunuty mírně doleva. Při zanedbání chyby předpovědi norského serveru, vidíme že chyby jsou nižší. Možná kdybychom tyto hodnoty nějak zprůměrovali, získali bysme podobné grafy.

Pravděpodobně je tento problém zapříčiněn tím, že Holešov je mimo dosah družice. Podívejme se proto např. na Cheb.



(a) CHMI chyba předpovědi



(b) YR chyba předpovědi

Obrázek 3.21: Porovnání chyby CHMI a norského serveru pro stejnou lokalitu

Na obrázku 3.21 lze vidět, že vysoké opakující se hodnoty nejsou jenom v místech, kde není dosah družice, nýbrž všude. Bohužel po zkoumání vstupních dat jsme nezjistili, čím jsou tyto vysoké hodnoty zapříčiněny.

Samozřejmě nás také zajímalo, které pravděpodobnostní rozložení lépe reprezentuje chování chyby předpovědi. Opět jsme použili porovnávání pomocí Euklidovy normy (C.2). Zjistili jsme, že i zde chybu předpovědi lépe reprezentuje Laplaceovo rozdělení.

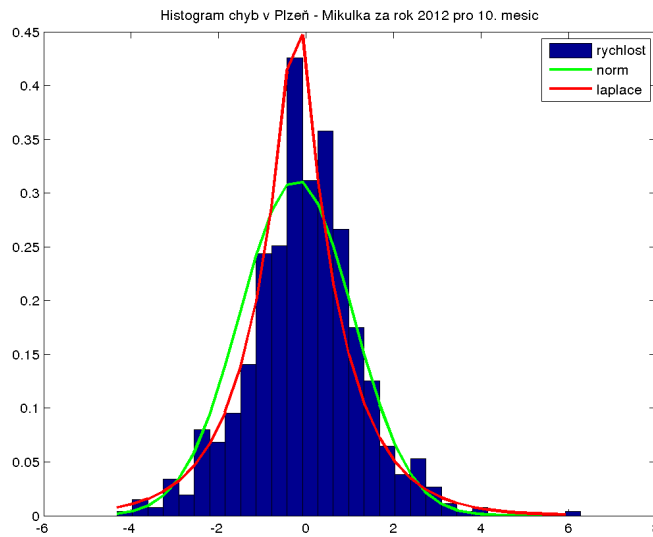
## 3.6 Podrobnější pohled na některé lokality pro data z YR

Opět jsme se zaměřili na stejné lokality jako tomu bylo v kapitole 3.4.

### 3.6.1 Chyba předpovědi

I zde se zaměříme na lokalitu Plzeň. Bude provedena analýza jak z hlediska měsíců, tak ročních období.

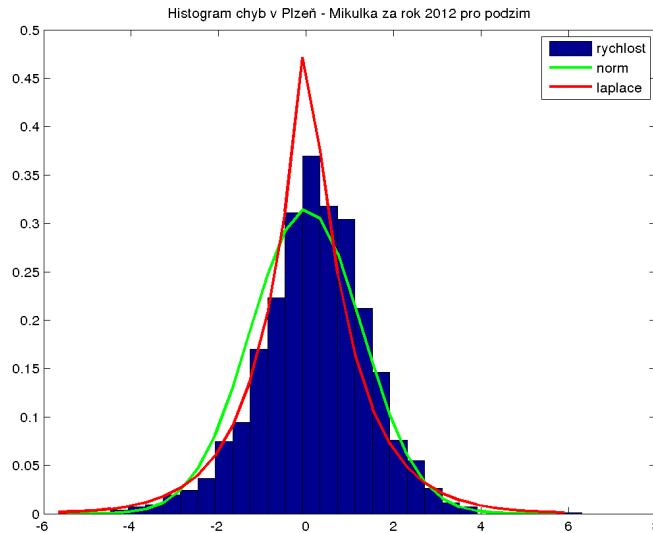
Jako první se podíváme na chybu předpovědi říjnového měsíce, která je na obrázku 3.22.



Obrázek 3.22: Histogram četnosti chyby předpovědi z norského serveru pro Plzeň

Tento graf můžeme porovnat s 3.14. Oproti grafu získanému z CHMI vidíme, že pravděpodobnostní rozložení chyby lépe reprezentuje. To je zapříčiněno tím, že z norského serveru se předpověď aktualizována pro každou hodinu. Z tohoto hlediska můžeme říci, že je předpověď norského serveru lepší.

Pokusme se ještě podívat na roční období. Ale bohužel jelikož jsme data z norského serveru začali stahovat až od poloviny roku 2012, nemáme histogram pro jarní měsíce. Porovnejme proto podzimní měsíce (3.23), jelikož podzim a jaro mají podobný průběh ohledně rychlosti větru.



Obrázek 3.23: Histogram četnosti chyb předpovědi pro podzimní měsíce

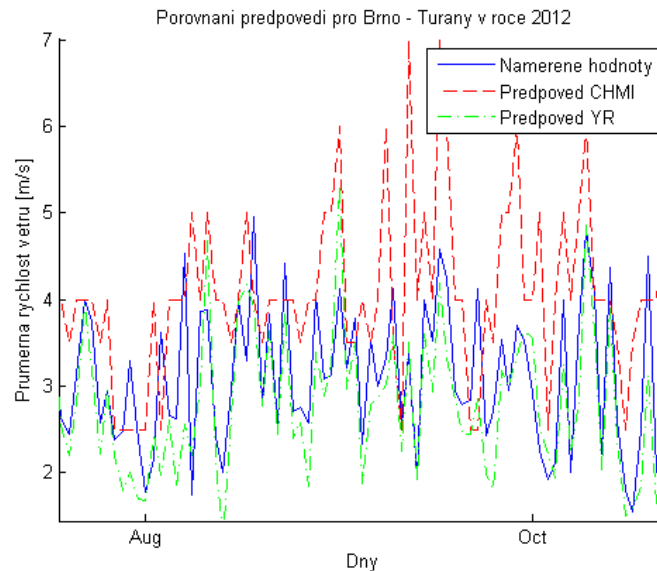
Obrázek 3.23 můžeme opět porovnat s obrázkem 3.15, který byl pořízen ze získaných dat z CHMI. Vidíme, že i zde jsou histogramy přehlednější a více nám o chybě předpovědi řeknou, než u již zmíněného obrázku pro Plzeň z CHMI.

### 3.7 Porovnání norského serveru a CHMI

Pokusme se v této kapitole shrnout vše, co jsme v kapitolách 3.2 a 3.5 zjistili a porovnejme předpovědi jednotlivých serverů mezi sebou.

Jako první se podívejme na předpovědi z obou serverů. Je nutno podotknout, že z norské družice jsme brali pouze nejbližší předpověď, tzn. první sloupec. Co se týká předpovědi z CHMI, zde jak víme již z kapitoly 3.3, je předpověď udávána v minimální a maximální hodnotě rychlosti větru, tyto hodnoty jsme průměrovali a vykreslili do grafu 3.24. Zde, v této práci, je pro přehlednost vykreslen jenom náhodný detail celkového porovnání předpovědí z obou serverů.





Obrázek 3.24: Detail porovnání předpovědi CHMI a norského serveru

Z obrázku 3.24 můžeme vyčíst, že ani jedna předpověď není zcela dokonalá, ale ani o jedné z nich nemůžeme říci, že je lepší. To ani v místech, kde norská družice nemá dosah.

Jako další se zaměříme na získané histogramy v kapitolách 3.2.4, 3.3.2, 3.4, 3.5.5 a 3.6.

Pro reprezentaci chování větru jsme pomocí Euklidovy normy (kapitola C.2) zjistili, že je nejlepší Raygleyho rozdělení. Co se týče chyby, zde jsme zjistili, že se pro popis hodí Laplaceovo rozdělení.

V kapitolách 3.4 a 3.6 jsme se zaměřili na vybrané lokality. Po porovnání obrázků pro říjen (obrázek 3.14 a 3.22) můžeme říci, že zde je lepší předpověď z norského serveru. A to z důvodu granularity dat, která je zde hodinová. Stejný závěr můžeme udělat i pro roční období, které jsou na obrázcích 3.15 a 3.23.

Ale tím neříkáme, že norská předpověď je lepší než předpověď uvedená CHMI. Musíme brát v potaz to, že předpověď z CHMI pokrývá celou Českou republiku, ikdyž jenom po krajích a předpověď je pouze jedna na celý den.

Po této datové analýze můžeme říci, že obě předpovědi mají své výhody a nevýhody. Pokud tohle vše vezmeme v potaz, můžeme říci, že ve výsledku jsou si obě předpovědi rovny.

# Kapitola 4

## Simulátor

V této kapitole se zaměříme na požadovanou simulaci rychlosti větru v České republice. K tomu si budeme muset vytrýbit několik pojmů a teoretických poznatků. Ty si vyvětlíme v několika nadcházejících kapitolách než se dostaneme k samotné simulaci.

Poznatky byly čerpány ze [7], [8], [9] a [10]

### 4.1 Markovské procesy

#### 4.1.1 Markovské řetězce

Je to náhodný proces s diskretní množinou stavů, diskretním časem a takový, že pravděpodobnost  $\pi_i(t)$ , že v čase  $t$  bude proces ve stavu  $i$ , je stochasticky závislá pouze na stavu v časovém okamžiku  $t - 1$ . Markovské řetězce můžeme definovat následující způsobem :

- $V = \{v_1, v_2, \dots, v_r\}$  je konečná množina hodnot stavu  $y$ .
- Počáteční stav řetězce je popsán vektorem pravděpodobnosti hodnot počátečního stavu označeného jako  $\alpha$ .
- Stav systému se mění – systém přechází náhodně z jednoho stavu do druhého. Změna jednoho stavu do druhého se nazývá událost.
- Je-li systém ve stavu  $v_i$ , pak pravděpodobnost, že následující stav po uplynutí času  $\Delta$  bude  $v_j$  je  $\pi_{i,j}(t, \Delta)$ . Pravděpodobnost  $v_j$  je  $\pi_{i,j}(t, \Delta)$  bude nezávislá na předchozích stavech systému a pro homogenní přetěžce závisí pouze na intervalu  $\Delta$ .

#### 4.1.2 Pravděpodobnostní vektor v čase $t$

Znalost o tom, ve kterém stavu se nacházel, nachází a nebo bude nacházet markovský řetězec v čase  $t$  budeme vyjadřovat jako pravděpodobnostní vektor

$$\pi(t) = [\pi_1(t), \pi_2(t), \dots, \pi_n(t)] \quad (4.1)$$

Tento vektor, popisující rozdělení pravděpodobnosti stavů v čase  $t$ , obsahuje veškeré informace o stavu řetězce a je využíván pro popis chování systému.

Pravděpodobnostní vektor má vždy všechny složky nezáporné a jejich součet je roven jedné. Pravděpodobnostní vektor budeme pokládat za vektor řádkový ( kvůli pohodlnému násobení tzv. přechodovou maticí).

### 4.1.3 Pravděpodobnost přechodu, přechodová matice

Pravděpodobnost přechodu hodnoty atributu  $y(t_k) = v_i$  na hodnotu  $y(t_k + 1) = v_j$  je závislá na velikosti kroku  $\Delta$ . Označíme-li  $P(y(t_{k+1}) = v_j | y(t_k) = v_i)$  jako  $p_{i,j}(\Delta)$  můžeme matici přechodu hodnot stavu této posloupnosti do následujícího okamžiku popsat maticí :

$$P(\Delta) = \begin{bmatrix} p_{1,1}(\Delta) & p_{1,2}(\Delta) & \cdots & p_{1,n}(\Delta) \\ p_{2,1}(\Delta) & p_{2,2}(\Delta) & \cdots & p_{2,n}(\Delta) \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1}(\Delta) & p_{n,2}(\Delta) & \cdots & p_{n,n}(\Delta) \end{bmatrix} \quad (4.2)$$

Rozdělení pravděpodobnosti stavů v čase  $t + \Delta$  je potom :

$$\pi(t + \Delta) = \pi(t) + P(\Delta). \quad (4.3)$$

Obecně potom platí

$$\pi(t + n \cdot \Delta) = \pi(t) + P(\Delta)^n. \quad (4.4)$$

### 4.1.4 Diskrétní čas

Předpoklad, že v markovském řetězci je čas diskrétní, znamená, že nás stavy procesu zajímají pouze v okamžicích, které tvoří (potenciálně nekonečnou) rostoucí posloupnost. Mezi těmito okamžiky se v markovském řetězci nic neděje, čas mezi těmito okamžiky v modelu neexistuje.

Zejména to znamená, že pro okamžik  $t$  má smysl hovořit o následujícím časovém okamžiku – budeme jej značit jako okamžik  $t + 1$ . Pro spojitý čas toto neplatí. Ve spojitém čase mezi každými dvěma různými okamžiky  $t_1, t_2$  leží nekonečně mnoho okamžiků  $t$  takových, že  $t_1 < t < t_2$ .

Při praktickém modelování nějakého děje pomocí markovského řetězce jsou okamžiky  $t_i$  zpravidla odvozeny od výskytu nějaké události. Často touto událostí bývá změna stavu řetězce, to odpovídá situaci, kdy nás zajímají jen změny stavů a nikoli doby, za jak dlouho ke změně stavu dochází. Okamžiky  $t_i$  však mohou být odvozeny i od jiných událostí, než pouze od změn stavu.

Jiný, rovněž častý, způsob modelování spočívá v tom, že stavy modelovaného děje sledujeme se zcela pravidelným časovým krokem, např. každých 5 milisekund nebo každého prvního v měsíci. To pak znamená, že ve skutečném ději (v ději, který modelujeme) může během časového kroku dojít i k několika změnám stavu, ale model (markovský řetězec) tyto změny nedokáže zachytit. Tento způsob modelování však na rozdíl od předchozího umožňuje modelovat dobu, která uplyne mezi změnami stavů nebo než je dosaženo nějakého cílového stavu.

### 4.1.5 Přechodový graf

Je orientovaný graf, jehož vrcholy odpovídají stavům markovského řetězce a orientované hrany odpovídají možným přímým změnám stavů, tj. změnám, které mají nenulovou pravděpodobnost. Každá hrana z vrcholu (stavu)  $i$  do vrcholu (stavu)  $j$  je ohodnocena pravděpodobností přechodu  $\pi_{i,j}(t, \Delta) > 0$ . Počet hran je tedy roven počtu nenulových prvků přechodové matice. Z každého vrcholu vychází alespoň jedna hrana a součet ohodnocení všech hran, které z vrcholu vycházejí, je roven jedné.

Graf markovského řetězce je důležitý nejen pro svou názornost, ale zejména proto, že mnoho důležitých vlastností markovského řetězce lze snadno odvodit z vlastností jeho grafu, zejména z jeho rozkladu na silně souvislé komponenty.

Příklad znázorněného grafu můžeme vidět na obrázku 4.1, který popisuje i námi navrhovaný simulátor.

### 4.1.6 Klasifikace stavů a typy řetězců

Většinu těchto pojmů definujeme pomocí vlastností přechodového grafu, neboť tyto vlastnosti lze v grafu poměrně snadno ověřit.

### Stochasticky uzavřená množina stavů

Je taková množina, ze které nevychází žádná hrana ven. Jinak řečeno, pravděpodobnost, že markovský řetězec opustí stochasticky uzavřenou množinu, je nulová.

### Ergodická množina stavů

Je stochasticky uzavřená množina, která neobsahuje menší stochasticky uzavřenou množinu.

Množina stavů je ergodická právě tehdy, když jí odpovídající podgraf přechodového grafu je silně souvislou komponentu, ze které nevychází ven žádná hrana.

Ergodický stav je stav, který je prvkem nějaké ergodické množiny. Každý markovský řetězec má alespoň jednu ergodickou množinu.

## 4.2 Transformace

### 4.2.1 Metoda inverzní transformace

Metoda inverzní transformace je založena na jednoduchém principu. Nechť  $U$  je náhodná veličina, která na rovnoměrné rozdělení intervalu  $(0, 1)$  a nechť  $F(X)$  je libovolná distribuční funkce. Potom náhodná veličina  $X = F^{-1}(U)$  má rozdělení s distribuční funkcí  $F(X)$ .

Metoda inverzní transformace je proto obzvláště vhodná pro generování náhodných veličin, u kterých umíme vyjádřit inverzní funkci explicitně. Typickým příkladem jsou Exponenciální či Wiebullovo rozdělení.

Metoda inverzní transformace je použitelná i pro případy, kdy nemůžeme najít analytické vyjádření inverzní funkce  $F^{-1}$ . V tomto případě se v rámci generování používá numerického řešení rovnice  $F(X) = U$ .

Výhodou této metody je její jednoduchost a rychlost.

### 4.2.2 Jonhonsův translační systém

V případě, že modelujeme data s neznámým pravděpodobnostním rozdělením je jednou z možností použít Johnsonova translačního systému pro náhodnou veličinu  $X$  :

$$F_X(x) = \Phi \left\{ \gamma + \delta f \left[ \frac{x - \xi}{\lambda} \right] \right\}, \quad (4.5)$$

kde  $\delta$  a  $\gamma$  jsou parametry definující tvar rozdělení,  $\xi$  je pohybový parametr,  $\lambda$  je parametr rozptylu dat,  $\Phi(\cdot)$  je distribuční funkce normovaného normálního rozdělení a  $f(\cdot)$  je jedna z transformací :

$$f(y) = \begin{cases} \log(y) & \text{pro logaritnicko-normální typy rozdělení} \\ \log(y^2 + \sqrt{y^2 + 1}) & \text{pro neomezené typy rozdělení} \\ \log\left(\frac{y}{1-y}\right) & \text{pro omezené typy rozdělení} \\ y & \text{pro normální typy rozdělení} \end{cases} \quad (4.6)$$

Poskytuje dobré znázornění pro jednovrcholová rozdělení a umožňuje získat různé tvary pravděpodobnostních rozdělení.

## 4.3 Simulace

K popisu modelu byly použity stochastické modely diskrétní v úrovni a spojitě v čase. Jednotlivé stavy odpovídají jedné ze tříd rychlosti větru – bezvětří, slabý vítr, mírný vítr, čerstvý vítr, silný vítr. Pro další výpočty vynecháme třídy velmi silný vítr, vichřice a orkán kvůli nedostatku dat.

Simulace byla prováděna v několika krocích. Pojdme si tyto kroky vypsát a postupně si ke každému z nich něco napíšeme.

1. Odstraníme sezónnost z naměřených dat rychlosti větru s využitím denních diagramů Pro jednotlivá období. Tyto denní diagramy jsou odečteny od naměřených dat.
2. Zmíněnou odchylku od denního diagramu nyní zpracujeme pomocí transformace o které jsme se zmínili v kapitole 4.2.

Tím získáme normalizovaná data. Pro transformaci byl použit vztah

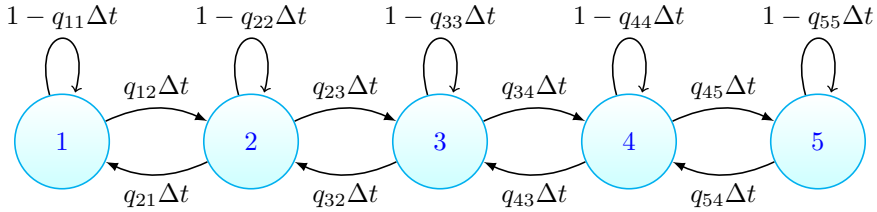
$$y_v = \Phi^{-1}(F_v(E_{data})) \stackrel{def}{=} Z_v(E_{data}), \quad (4.7)$$

kde  $\Phi(\cdot)$  je distribuční funkce normovaného normálního rozdělení,  $F_v$  je distribuční funkce rychlosti větru a  $E_{data}$  jsou naměřená data. Jelikož neznáme distribuční funkci  $F_v$ , použijeme empirickou distribuční funkci z naměřených dat a určíme transformaci  $Z_v$  v měřených bodech. Po té body proložíme po částech lineární funkcí. Výhoda toho postupu je taková, že získáme rostoucí funkci, která má i svou inverzní funkci.

Pokud známe transformační funkci, je možné přepočítat rychlost větru na modelovanou rychlost větru podle vztahu

$$y_v(t) = Z_v(E_{data}(t)) \quad (4.8)$$

3. Normovaná data rozdělíme do tříd větru.
4. Z již výše zmíněných modelovaných hodnot nyní budeme indentifikovat parametry AR procesu modelované veličiny za pomoci AR procesu.



Obrázek 4.1: Příklad grafu popisující Markovské řetězce

Poznamenejme, že čísla uvnitř kruhů odpovídají kódům typů větru v tabulce 2.1.

5. Po těchto krocích nám už jenom zbývá si určit matice četnosti po sobě jdoucích stavů (tabulka 4.1), matici pravděpodobností přechodů mezi stavy (tabulka 4.2) a matici setrvání ve stavu (tabulka 4.3). Pak bylo potřeba jenom rozhodovat, do jakého stavu se markovský řetězec přesune pro následující časový krok. To bylo provedeno pomocí matice pravděpodobnosti přechodů, která je uvedena v tabulce 4.2.
6. Pro doplnění bylo použito vykresleno několik realizací s reálnými daty.

Poznamenejme k bodu 4, že tento krok byl velice důležitý, jelikož víme, že budeme potřebovat minimálně 5 AR procesů v markovském řetězci, ale pokud budou parametry AR procesu pro období různé, budeme muset definovat 20 AR procesů (5 tříd větru  $\times$  4 roční období). Proto pro nás bylo stěžejní zjistit parametry vektoru A a B a navzájem je porovnat (více o parametrech A a B viz. C.1).

Po prozkoumání parametrů sezónnosti, jsme usoudili, že hodnoty vektoru A a B jsou si dostatečně blízko na to, abysme ze simulátoru vyřadili sezónnost. Proto náš výsledný simulátor bude mít 5 AR procesů. Pro každou třídu větru jeden.

Dále jsme se se pokusili porovnat parametry AR procesu pro třídy a období. Náš prvotní předpoklad byl, že hodnoty těchto parametrů budou při porovnání různé. Tento předpoklad se také po prozkoumání potvrdil.

### 4.3.1 Generátor tříd rychlosti větru

Tento generátor je tvořen stochastickým modelem diskrétním v úrovni. Stavby modelu jsou stejné jako třídy meteorologických dat. Z naměřených rychlostí větru byly vypočteny četnosti dvojic po sobě následujících stavů, které jsou uvedeny v tabulce 4.1. Proměnná  $x(t)$  značí zkoumaná data v čase  $t$  a  $x(t+1)$  tu samou veličinu v čase  $t+1$ .

Jelikož jsme se v simulátoru zaměřili na lokalitu Plzně, jsou i uvedené tabulky platné pouze pro oblast Plzně.

		x(t+1)				
		bezvětří	slabý vítr	mírný vítr	čerstvý vítr	silný vítr
x(t)	bezvětří	2490	1612	52	3	1
	slabý vítr	1624	8163	1603	27	3
	mírný vítr	47	1608	4070	338	5
	čerstvý vítr	1	19	352	421	8
	silný vítr	0	2	1	11	11

Tabulka 4.1: Tabulka četností po sobě jdoucích stavů

Zde uvedené modely budou použity pro simulace, potom je vhodné používat spojitý Markovský řetězec. Matice četnosti po sobě jdoucích stavů byla přepočtena na matici pravděpodobnosti přechodů mezi stavy pomocí vztahu

$$p(x) = \frac{x_{i,j}}{\sum_{j=1}^n x_{i,j}}, \quad (4.9)$$

kde  $x$  je daný prvek v matici s pozicí  $i$  a  $j$ ,  $n$  je počet tříd větru (v naše případě 5). Matice pravděpodobností přechodu mezi stavy je udávána s krokem jedné hodiny.

		x(t+1)				
		bezvětří	slabý vítr	mírný vítr	čerstvý vítr	silný vítr
x(t)	bezvětří	-	0,9664	0,0312	0,0018	$5,9952 \cdot 10^{-4}$
	slabý vítr	0,4986	-	0,4922	0,0083	$9,2109 \cdot 10^{-4}$
	mírný vítr	0,0235	0,8048	-	0,1692	0,0025
	čerstvý vítr	0,0026	0,0500	0,9263	-	0,0211
	silný vítr	0,000	0,1429	0,0714	0,7857	-

Tabulka 4.2: Pravděpodobnost přechodů mezi stavy

a střední dobu setrvání ve stavu

bezvětří	slabý vítr	mírný vítr	čerstvý vítr	silný vítr
2,4928	3,5063	3,0370	2,1079	1,7857

Tabulka 4.3: Střední doba setrvání ve stavu

V tabulce 4.2 vidíme, že je možnost, aby například mírný vítr přešel do bezvětří. To je nejspíše zapříčiněno tím, že data jsou sbírána jednou za hodinu a během té hodiny se několikrát změní typ větru. V praxi není možné, aby rychlost větru měla charakter skokové funkce. Proto při určování následujícího stavu v simulaci budeme brát v potaz nejbližší typ větru. Z toho vyplývá, že se náš příklad s mírným větrem nastat nemůže.

### 4.3.2 Výsledky simulace

Pro simulaci jsme použili naměřená data z CHMI od roku 2011 do roku 2013.

Podívejme se v této kapitole na získané výsledky ze simulátoru pro lokalitu Plzně. Z časové náročnosti některých výpočtů, se opět zaměříme pouze na lokalitu Plzně.

V kapitole 4.3 jsme si v bodě pět řekli, že je pro naši práci důležité se podívat na parametry jednotlivých AR procesů. Podívejme se na parametry jednotlivých tříd větru :

	kód	$\mu$	$\sigma$	<b>A</b>	<b>FPE</b>
<b>bezvětří</b>	1	-0,6652	0,8101	$1 - 0,4935 \cdot q^{-1} - 0,2726 \cdot q^{-2}$	0,5460
<b>slabý vítr</b>	2	-0,1604	0,8767	$1 - 0,4850 \cdot q^{-1} - 0,1941 \cdot q^{-2}$	0,4819
<b>mírný vítr</b>	3	0,5843	0,8566	$1 - 0,5364 \cdot q^{-1} - 0,2444 \cdot q^{-2}$	0,4942
<b>čerstvý vítr</b>	4	1,3853	0,8302	$1 - 0,5980 \cdot q^{-1} - 0,3113 \cdot q^{-2}$	0,5290
<b>silný vítr</b>	5	1,9067	1,1936	$1 - 0,7296 \cdot q^{-1} - 0,1599 \cdot q^{-2}$	1,6902

Tabulka 4.4: Parametry AR procesů přednotlivé třídy větru

Na tabulce 4.4 jasně vidíme, že střední hodnota ( $\mu$ ) má jasně rostoucí tendenci. Hodnota  $\sigma$  je pro všechny hodnoty skoro stejná. Totéž můžeme říci i o hodnotách polynomu A u AR procesu (více viz. C.1), výjimkou je pouze polynom a u silného větru jehož hodnoty jsou vyšší. U posledního řádku, nazvaném FPE (Final Prediction Error), tato chyba je vztažena k varianci chyby modelu. Vidíme, že tato chyba je relativně stejná, až opět na poslední řádek.

Z tabulky také vidíme, že jsme pro simulátor použili AR proces druhého stupně. Původně jsme očekávali že druhý, maximálně čtvrtý řád bude optimální pro AR model. Po naprogramování a vyzkoušení různých řádů, jsme zjistili, že nejlepší výsledky získáváme pro druhý řád.

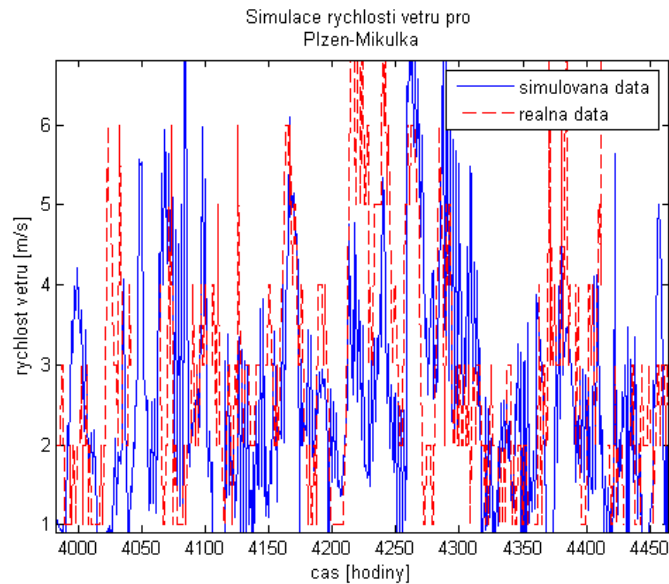
Ze simulátoru jsou získávána data hodinově, jako tomu bylo u CHMI. Pro naši simulaci a získání hodnot použijme maximální čas rovný 8760 hodinám.

Jako počáteční podmínky jsou použity :

- kód stavu, který je uveden v tabulce 2.1,
- náhodné koeficienty AR procesu,
- doba setrvání ve stavu, jejichž příklad je uveden v tabulce 4.3,
- pravděpodobnost přechodu mezi stavy, která je uvedena v tabulce 4.2

Podívejme se na získané grafické výsledky simulace.

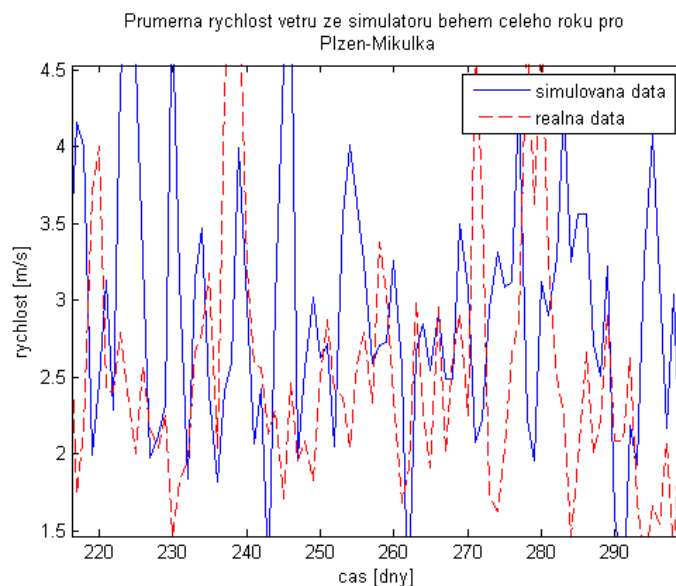
Nejdříve se podívejme na porovnání naměřených dat a jedné realizace simulovaných dat vzhledem k hodinové granularitě.



Obrázek 4.2: Detail porovnání simulovaných a naměřených dat s hodinovou granularitou

Na obrázku 4.2 vidíme, že realizace dat dokonale neodpovídají naměřeným datům, ale ani nezískáváme ze simulátoru náhodná čísla. Jsou úseky, kde simulátor se mimo realitu více, ale jsou i místa, kde simulátor relativně dobře odhadne chování větru v Plzni.

Jelikož by pro některé mohl být graf s hodinovými daty nepřehledný, byl také vytvořen graf průměrných rychlostí během celého roku. Podobný graf jsme už viděli v kapitole 3.2.1.

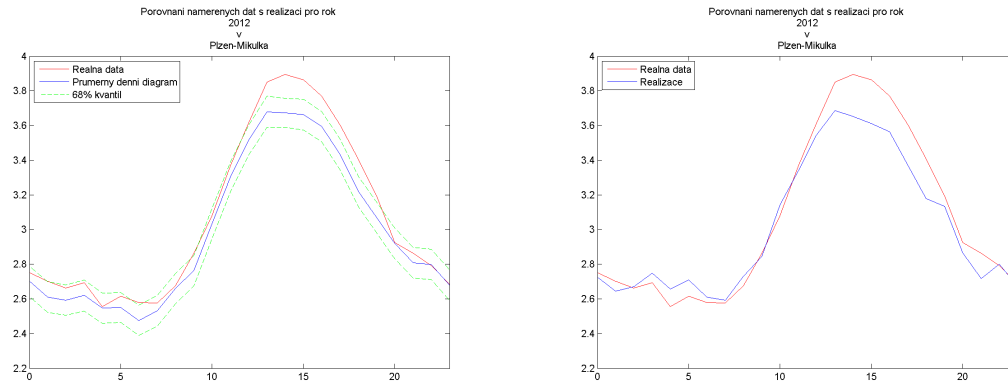


Obrázek 4.3: Detail porovnání průměrných rychlostí pro simulovaná a naměřená data



Obrázek 4.3 nám ukazuje detail porovnání průměrných rychlostí realizovaných simulátorem s reálnými naměřenými daty. I zde jasně vidíme, že simulátor není jenom generátor náhodných čísel, ale že i zde jsou úseky, které vyhovují realitě. Tento obrázek jenom potvrzuje tvrzení, která jsme udělali u obrázku 4.2.

Nakonec se podíváme na denní diagram.



(a) Porovnání denních diagramů pro tisíc simulací

(b) Porovnání denních diagramů pro simulovaná a naměřená data

Obrázek 4.4: Výsledek simulátoru a porovnání s naměřenými daty

Na obrázku 4.4(a) vidíme, že hraniční linie 85%-ního kvantilu nám ukazují pás přípustných hodnot pro tisíc realizací. Simulace byla pokaždé spuštěna s jinou náhodou počáteční podmínkou.

Z obrázku 4.4(b) vidíme, že denní diagramy jsou si velice podobné. Vykreslený denní diagram realizace byl náhodně vybrán z tisíce realizací pomocí generátoru náhodných čísel.

# Kapitola 5

## Závěr

Jelikož je celá práce rozdělena na dvě části, utvořme i zde závěr pro každou část zvlášť.

V datové analýze jsme se z počátku zabývali obecných chování větru. Na obrázcích 3.2 a 3.1 jsme viděli rozdíly mezi jednotlivými ročními období. Zamysleme se jak se chová vítr v jednotlivých ročních období a zjistíme, že realita potvrzuje správnost získaných grafů. Obrázek 3.3 je jenom další důkaz, jak se vítr chová během jednotlivých ročních období.

Po analýze ročních období jsme se zaměřili, na rozdíly rychlosti větru z různých nadmořských výškách. Pro tento účel jsme vybrali dvě lokality s naprosto různými nadmořskými výškami. Obrázek 3.4 nám ukazuje, že naše očekávání ohledně rychlosti větru v závislosti na nadmořské výšce byla naprosto oprávněná. Tento obrázek zcela potvrzuje tvrzení, že v horských oblastech je průměrná rychlost vyšší než v nížinných oblastech.

V následující kapitole 3.2.2 jsme zůstali u rozdílných nadmořských výšek, ale porovnávali jsme denní diagramy. Naše očekávání bylo takové, že více lokalit v České republice bude mít spíše trend převráceného písmene „U“, tedy trend horských oblastí. Po prozkoumání obrázku 3.5 jsme zjistili, že jsme se mýlili. V České republice jsou i lokality, které mají typicky nížinné denní diagramy, kterých je více.

V kapitole 3.3, zabývající se chybou předpovědi jsme zjistili, že u Českého hydrometeorologického ústavu je předpověď udávána pro kraje a jednou pro celý den. Předpovídaná hodnota je udávána jako minimální a maximální rychlost větru. Po prozkoumání obrázku 3.9 jsme zjistili, že chyba předpovědi pro předpověď minimální rychlosti větru je chyba nižší než pro předpovězené maximální rychlosti větru. Dále jsme zjistili, že v zimních měsících je velikost chyby vyšší než v jiných ročních období. Velikost chyby pro den, noc a následující den je přibližně stejná.

Z výpočtu korelací (kapitola 3.2.3) jsme si potvrdili předpoklad, že rychlost větru mezi dvěma vzdálenými lokalitami není navzájem ovlivňována. Jejich korelační koeficient se blíží hodnotě nula. Naopak rychlost větru je ovlivňována u lokalit sobě blízkých. Krásným příkladem tomu je Praha.

U směrodatných odchylek v kapitole 3.3.1 se nám jenom potvrdil fakt, který jsme získali už u chyby předpovědi. Směrodatné odchylky pro minimální rychlosti větru jsou nižší než pro maximální hodnoty. Dále jsme zjistili, že směrodatná odchylka je pro den a noc přibližně stejná. Pro předpověď na další den je avšak vyšší. Při pohledu na tyto data vidíme, že většina hodnot se pohybuje v rozmezí 1,5 - 2,1. Horší jsou hodnoty pro výše položené lokality, např. Sněžka, Šerák. Zde se odchylky pohybují přes hodnotu 3. Zajímavé jsou hodnoty u Šeráku. Ty jsou naprosto obrácené, než u zbytku lokalit. Vypočtená směrodatná odchylka je větší pro minimální hodnoty než pro maximální. Naprosto největší hodnoty jsou, dle očekávání, na Sněžce, které se pohybují mezi hodnotami 5-6.

Zajímavou částí bylo zkoumání rychlosti větru pomocí histogramů, které bylo provedeno v kapitolách 3.2.4 a 3.3.2. Zde jsme zjistili Raygleyho rozdělení reprezentuje nejlépe chování větru. Při pohledu na obrázek 3.7 na první pohled vidíme, že průměrná rychlost větru se skoro ve všech lokalitách pohybuje přibližně kolem stejných hodnot. Vyjímkou jsou horské lokality (Milešovka, Sněžka, Lysá hora), kde je průměrná rychlost vyšší. Pokud se ale podíváme jenom na tyto tři horské lokality, vidíme, že i zde je průměrná rychlost větru stejná. Ale jenom na Sněžce je větší

variabilita rychlosti větru. Pro chybu předpovědi jsme zjistili, že nejlépe chování chyby vyhovuje Laplaceovo rozdělení. Pro chyby předpovědí byl také vykreslen boxplot, viz obrázek 3.11. Zde lze vidět, že např. v Chebu a Kopistech je hodnota chyby předpovědi nejmenší, a tudíž je zde nejpřesnější předpověď počasí, alespoň co se týká rychlosti větru. Opět vidíme vyčnívající horské oblasti, pro které je těžké rychlost větru předvídat a podle toho i vypadá hodnota chyby. Sněžka v tomto grafu opět dominuje svou polou kvartilů. Po prozkoumání dat, bylo zjištěno, že rychlost větru se s postupujícími měsíci nijak závažně nemění.

Vykreslili jsme také histogram, který znázorňuje četnost tříd větru podle Beaufortovy stupnice. Tyto histogramy potvrzují správnost obrázku 2.1.

Jako poslední část datové analýzy pro data z Českého hydrometeorologického ústavu (kapitola 3.4), jsme se zaměřili jenom na vybrané lokality České republiky. I zde byly vykresleny histogramy. Jejich výsledky pouze potvrdili tvrzení o Raygleyho rozdělení. Jako ukázkové příklady byl uveden obrázek 3.12. Pokud jsme porovnávali rychlosti větru během jednotlivých ročních období, zjistili jsme že rychlost větru na jaře a na podzim jsou si velmi blízké, což potvrzuje opět již zjištěný fakt z kapitoly 3.2.1. Ale pravým opakem jsou letní a zimní měsíce, kde podle získaných histogramů více fouká v létě než v zimě. Pro chybu během jednotlivých měsíců nejsme schopni přesně určit kvůli nedostatku dat. Ostatně podívejme se na obrázek 3.13. Pokud se pokusíme na tento problém podívat z hlediska ročních období, nebude výsledek o nic lepší. Můžeme si sice udělat alespoň mlhavou představu o chybě předpovědi, ale i zde vzhledem k nedostatku dat nemůžeme přesně určit, které rozdělení reprezentuje chybu předpovědi (obrázek 3.14). Ke konci datové analýzy jsme pro porovnání vykreslili na obrázek 3.16, ze kterého jsme zjistili, že v zimních měsících je rychlost větru během dne vyšší. Překvapivé jsou denní diagramy jara a léta, které jsou prakticky totožné. Většinou jsou jarní rychlosti větru podobné podzimním, které jsou v tomto případě nejnižší.

Po datové analýze dat z Českého hydrometeorologického ústavu, jsme se zaměřili na předpovědi z norského serveru [www.yr.no](http://www.yr.no). Jako první jsme se pokusili porovnat v kapitole 3.5.1 dvě lokality, ale každá byla v jiné části republiky podle dostahu norské družice. Na obrázku 3.17 už na první pohled vidíme, že v Kostelní Myslové je chyba větší, a to zhruba dvojnásobně.

Z korelační analýzy, která je uvedena v kapitolách 3.5.2 a 3.5.3, jsme zjistili, že hodnoty plošných korelací pro rok 2012 jdou výrazně nižší, než pro rok 2013. To může být způsobeno tím, že jsme začali sbírat data teprve v polovině roku 2012. Podívejme se proto důkladněji na rok 2013, který nám z hlediska úplnosti dat řekne více.

Zjistili jsme, že s rostoucí vzdáleností mezi jednotlivými lokalitami hodnota koeficientu stejná nebo mírně klesající. Divné je, že pro horské oblasti získáváme pro některé lokality záporné korelace.

Pro autokorelace platí následující. Pro horské oblasti jako jsou již už mnohokrát zmiňovaná Sněžka a Lysá hora je korelační koeficient blízký hodnotě 0.2. Můžeme tedy říci, že pro tyto lokality předpověď není zcela přesná. Zajímavé výsledky jsme získaly z pražských lokalit, kde nejvyšší koeficient je na Ruzyni, který je přibližně 0.7. Oproti tomu ostatní pražské lokality (Karlovy, Libuň, Kbely) mají hodnoty pouze okolo 0.4. Takový rozdíl hodnot mohl vzniknout tím, že na Ruzyni, mají vlastní, možná i přesnější, meteostanici kvůli letovému provozu.

Při pohledu na směrodatné odchylky (kapitola 3.5.4), viz obrázek 3.19, můžeme vidět, že na prvním porovnávaném grafu jsou hranice směrodatné odchylky užší v Karlových Varech, to je způsobeno nejspíše polohou Karlových Varů. Očekávali jsme, že pás mezi směrodatnými odchylkami bude užší v případě Karlových Varů, ale pás směrodatné odchylky pro desátý a osmnáctý sloupec, je pro Kostelní Myslovou dokonce užší. Na druhou stranu se jedná o hodnotu přibližně 1 m/s. Různá šířka pásů je nejspíše zapříčiněna členitostí terénu a svůj vliv zde má i různá nadmořská výška.

V neposlední řadě, i zde jsme provedli analýzu chyby předpovědi pomocí histogramů, které jsou uvedeny v kapitole 3.5.5. Bohužel jsme nezjistili proč v pravidelných interval vznikají skoky, které můžeme vidět na obrázcích 3.20. Porovnejme výsledky chyb předpovědí z obrázků 3.10 a 3.20. Při zanedbání chyby předpovědi norského serveru, vidíme že chyby jsou nižší. Možná kdybychom tyto hodnoty nějak zprůměrovali, získali bychom podobné grafy.

Na závěr datové analýzy jsme se zaměřili v kapitole 3.6 opět na některé vybrané lokality

a v práci jsme po pozdější porovnání použili data z Plzně. Pro chybu předpovědi, která je na obrázku 3.22, jsme použili již známé Normální a Laplaceovo rozdělení a zkoumali jsme, které z rozdělení lépe reprezentuje chybu předpovědi. I zde to bylo Laplaceovo rozdělení. Při pohledu na obrázky 3.15 a 3.23, vidíme, že grafy pro roční období z norského serveru nám řeknou více než grafy získané z dat Českého hydrometeorologického ústavu. To je nejspíše zapříčiněno granularitou dat.

Závěrem nemůžeme zcela usoudit, který ze dvou zkoumaných serverů poskytující předpovědi je lepší. Ani jedna předpověď není zcela dokonalá, ale ani o jedné z nich nemůžeme říci, a to ani v místech, kde norská družice nemá dosah, že jedna z předpovědí je lepší. To potvrzuje i srovnávací detail na obrázku 3.24. Ze získaných histogramů v kapitolách 3.3.2, 3.5.5 a 3.6, se oba servery shodují a chybu předpovědi nejlépe reprezentuje Laplaceovo rozdělení. V kapitolách 3.4 a 3.6 jsme se zaměřili na vybrané lokality. Po porovnání obrázků pro říjen (obrázek 3.14 a 3.22) můžeme říci, že zde je lepší předpověď z norského serveru. A to z důvodu granularity dat, která je zde hodinová. Stejný závěr můžeme udělat i pro roční období, které jsou na obrázcích 3.15 a 3.23.

Ale tím neříkáme, že norská předpověď je lepší než předpověď uvedená CHMI. Musíme brát v potaz to, že předpověď z CHMI pokrývá celou Českou republiku, ikdyž jenom po krajích a předpověď je pouze jedna na celý den.

Po této datové analýze můžeme říci, že obě předpovědi mají své výhody a nevýhody. Pokud tohle vše vezmeme v potaz, můžeme říci, že ve výsledku jsou si obě předpovědi rovny.

V poslední části této práce byl vytvořen simulátor rychlosti větru. Výsledné vygenerované výsledky jsme porovnali s reálnými naměřenými daty získanými z Českého hydrometeorologického ústavu. Zjistili jsme, že stačí mít Markovský řetězec s pouze pěti AR procesy (viz obrázek 4.1), jelikož jsme zjistili, že parametry AR procesů pro jednotlivá roční období jsou si velice blízké. Po naprogramování a porovnání dat, jsme usoudili, že výsledky jsou vyhovující a v mnoha ohledech jsou výsledná data ze simulátoru velice podobná získaným reálným datům. K tomuto závěru jsme dospěli po prozkoumání obrázků 4.3, 4.2 a 4.4(b). Nepřesnost simulace přisuzujeme našemu pojetí přechodové matice a měření reálných dat. Pokud bychom brali přechodovou matici tak jaká je, mohli by získaná data reprezentovat naměřená data lépe. Na druhou stranu ale musíme brát v úvahu to, že vítr se nikdy nemění skokově a reálná data jsou získávána jenom jednou za hodinu a z námi použitých dat není možné poznat jak se vítr chová v mezicyklu měření.

# Literatura

- [1] Využití energie větru, Červenec 2013. <http://www.transformacni-technologie.cz/vyuziti-energie-vetru.html>.
- [2] P. Janeček, A. Zápotocká, and P. Skála. Datová reprezentace meteorologických dat modelů OZE z otevřených zdrojů. Výzkumná zpráva, ZČU, 2011.
- [3] Sivs - kód ii. vítr. <http://portal.chmi.cz/files/portal/docs/meteo/om/sivs/vitr.html>.
- [4] D. Hanslian. Větrné podmínky v České republice ve výšce 10 m nad povrchem i, Duben 2013. <http://oze.tzb-info.cz/vetrna-energie/9770-vetrne-podminky-v-ceske-republice-ve-vysce-10-m-nad-povrchem-i>.
- [5] Meteorologické stanice - dělení a význam, Srpen 2013. <http://www.in-pocasi.cz/clanky/teorie/meteorologicke-stance-rozdeleni/>.
- [6] Klasifikace předpovědí počasí. <http://www.meteocentrum.cz/encyklopedie/klasifikace-predpovedi-pocasi.php>.
- [7] J Demel. Operační výzkum, Prosinec 2002. <http://kix.fsv.cvut.cz/demel/ped/mo20/OV02.pdf>.
- [8] P. Janeček, J. Mošna, and E. Janeček. Metody a algoritmy modelování neurčitosti se zaměřením na energetické zdroje. Výzkumná zpráva, ZČU, Prosinec 2011.
- [9] I. Hochmannová. Matematický model výroby větrné energie. Diplomová práce, ZČU, Srpen 2010.
- [10] E. Janeček, R. Janeček, and A. Zápotocká. Modely, metody a algoritmy pro stanovování velikostí rozsahu podpurných služeb při nových oze zdrojích. Výzkumná zpráva, ZČU, 13.
- [11] J. Reif and Z. Kobeda. Úvod do pravděpodobnosti a spolehlivosti. Technical report, ZČU, Listopad 2004.
- [12] Náhodná veličina a rozdělení pravděpodobnosti. <http://new.euromise.org/czech/tajne/ucebnice/html/html/node6.html>.
- [13] V.Čermák. Diskrétní a spojitá rozdělení. vzorce, grafy, tabulky. Technical report, VŠE, 1993.
- [14] Boxplot, Leden 2014. <http://cs.wikipedia.org/wiki/Boxplot>.
- [15] L. Průcha. <ftp://math.feld.cvut.cz/pub/prucha/ubmi/predn/u11.pdf>.
- [16] Korelace, Říjen 2013. <http://cs.wikipedia.org/wiki/Korelace>.
- [17] L.Tesková. Lineární algebra. Technical report, ZČU, Říjen 2005.
- [18] M. Šimandl. Adaptivní řízení a zpracování signálů. Technical report, ZČU, Plzeň, 2004.

# Příloha A

## Obsah Příloženého CD

- **Získané boxploty**

*Datová analýza \img \box \CR \„rok“ \kraje\_bplot*

- **Porovnání naměřených hodnot s předpovědí CHMI a YR**

*Datová analýza \CR\_YR \„rok“*

- **Denní diagramy**

*Datová analýza \img \denni\_diagram \CR \„rok“*

- **Histogramy**

Data získaná z CHMI

*Datová analýza \img \ histogram \CR \„rok“ \chyba*

*Datová analýza \img \ histogram \CR \„rok“ \rychlost*

*Datová analýza \img \ histogram \CR \„rok“ \stupnice*

*Datová analýza \img \ histogram \CR \„rok“ \lokalita \„vybrane lokality“*

*Datová analýza \img \ histogram \CR \„rok“ \lokalita „vybrane lokality“ \chyba \mesice*

*Datová analýza \img \ histogram \CR \„rok“ \lokalita \„vybrane lokality“ \chyba \obdobi*

*Datová analýza \img \ histogram \CR \„rok“ \lokalita „vybrane lokality“ \rychlost \mesice*

*Datová analýza \img \ histogram \CR \„rok“ \lokalita \„vybrane lokality“ \rychlost \obdobi*

Data získaná z YR

*Datová analýza \img \ histogram \YR \„rok“ \chyba*

*Datová analýza \img \ histogram \YR \„rok“ \lokalita \„vybrane lokality“*

*Datová analýza \img \ histogram \YR \„rok“ \lokalita „vybrane lokality“ \chyba \mesice*

*Datová analýza \img \ histogram \YR \„rok“ \lokalita \„vybrane lokality“ \chyba \obdobi*

- **Chyby**

Data získaná z CHMI

*Datová analýza \img \chyba \CR \„rok“*

Data získaná z YR

*Datová analýza \img \chyba \YR \„rok“ \error*

*Datová analýza \img \chyba \YR \„rok“ \STD*

- **Rychlost větru**

*Datová analýza\img\rychlost \CR „rok“ \cely\_lokace*

*Datová analýza\img\rychlost \CR „rok“ \jaro\_lokace*

*Datová analýza\img\rychlost \CR „rok“ \podzim\_lokace*

*Datová analýza\img\rychlost \CR „rok“ \leto\_lokace*

*Datová analýza\img\rychlost \CR „rok“ \zima\_lokace*

*Datová analýza\img\rychlost \CR „rok“ \CR*

*Datová analýza\img\rychlost \CR \porovnaní „rok“*

*Datová analýza\img\rychlost \CR „rok“ \YR \cely\_lokace*

- **Data uložená v tabulkách a .mat soubory**

*Datová analýza\vystup*

Pod pojmem rok bereme data získaná v daném roce.

Pod pojmem „vybrané lokality“ jsou chápány lokality, které jsou uvedeny v kapitolách 3.4 a 3.6.

# Příloha B

## Pojmy z pravděpodobnosti a statistiky

V této kapitole jsou uvedeny všechny pojmy a vztahy, které jsou v tomto textu hojně používány. Většina teoretických podkladů byla čerpána z [11],[12], [13], [14], [15], [16] a [13].

### B.1 Základní pojmy

#### B.1.1 Náhodná veličina

Většina náhodných pokusů a pozorování prováděných v biologii a medicíně má výsledek vyjádřený reálným číslem. Tato čísla vytvářejí hodnoty reálné náhodné veličiny. Náhodné veličiny označujeme zpravidla velkými písmeny z konce latinské abecedy, například  $X$ ,  $Y$ ,  $Z$ , a jejich hodnoty odpovídajícími malými písmeny, například  $x$ ,  $y$ ,  $z$ . K základním charakteristikám náhodné veličiny řadíme průměr ( $\mu, E(X)$ ) a rozptyl ( $\sigma^2, D(X)$ ) - viz dále. Jak často určité hodnoty náhodné veličiny nastávají, je exaktně matematicky popsáno pomocí rozdělení pravděpodobnosti. V praxi se zpravidla setkáváme s náhodnými veličinami dvojího typu - diskrétními a spojitými náhodnými veličinami.

#### Diskrétní

Diskrétní náhodná veličina  $X$  může nabývat pouze konečného počtu hodnot, můžeme říci, že každé hodnotě  $x_i$  je přiřazena pravděpodobnost  $P(X = x_i) > 0$ . Platí, že součet všech těchto pravděpodobností je roven jedné. Pro diskrétní náhodnou veličinu spočteme průměr

$$E(X) = \sum_{i=1}^k x_i \cdot P(X = x_i) \quad (\text{B.1})$$

a rozptyl

$$D(X) = \sum_{i=1}^k (x_i - \mu)^2 P(X = x_i). \quad (\text{B.2})$$

#### Spojitá

Spojitá náhodná veličina  $X$  může nabývat všech hodnot  $x$  z určitého intervalu. Její pravděpodobnostní rozdělení je popsáno pomocí nezáporné funkce  $f(x)$ , která se nazývá hustota rozdělení. Pomocí hustoty můžeme počítat hodnoty průměru

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (\text{B.3})$$



a rozptylu

$$D(X) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx. \quad (\text{B.4})$$

### B.1.2 Hustota pravděpodobnosti a distribuční funkce

Náhodná veličina  $X$  má rozdělení spojitého typu, nexistuje-li nezáporná funkce  $f(x)$  taková, že pro všechna reálná  $x$  můžeme vyjádřit distribuční funkci ve tvaru

$$F(x) = \int_{-\infty}^{\infty} f(x) dx. \quad (\text{B.5})$$

Funkce  $f(x)$  se nazývá hustota pravděpodobnosti náhodné veličiny  $X$ , jestliže distribuční funkce  $F(x)$  je spojitá pro všechna reálná  $x$ , ve kterých existuje derivace

$$f(x) = \frac{dF(x)}{dx} \quad (\text{B.6})$$

### B.1.3 Korelace

Korelace vyjadřuje vzájemný vztah mezi dvěma procesy. Pokud se jedna z nich mění, mění se korelativně i druhá a naopak. Pokud se mezi dvěma procesy ukáže korelace, je pravděpodobné, že na sobě závisejí, nelze z toho však ještě usoudit, že by jeden z nich musel být příčinou a druhý následkem. To samotná korelace nedovoluje rozhodnout. Můžeme určit pouze korelační koeficient mezi veličinami  $x$  a  $y$

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_Y}, \quad (\text{B.7})$$

který může nabývat hodnot v intervalu  $\langle -1; 1 \rangle$ .

Pro výpočet korelace je potřeba znát hodnotu kovariance  $\text{cov}(X,Y)$ , která je dána vztahem

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y) \quad (\text{B.8})$$

### B.1.4 Směrodatná odchylka

Směrodatná odchylka náhodné veličiny  $X$ , označovaná  $\sigma(X)$ , je definovaná jako :

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - E(X))^2}, \quad (\text{B.9})$$

kde  $N$  je počet vzorků,  $E(X)$  je střední hodnota zkoumaných dat,  $x_i$  jsou zkoumaná data. Směrodatná odchylka je nejužívanější míra variability.

### B.1.5 Kvantil

Často potřebujeme k dané hodnotě pravděpodobnosti  $p \in (0;1)$  nalézt reálné číslo  $x_p$ , takové, že náhodná veličina  $X$  s pravděpodobností  $p$  nabude hodnotu z intervalu  $(-\infty; x_p]$ . Pro spojitě náhodné veličiny to znamená, že obsah plochy mezi osou  $x$  a grafem hustoty  $f(x)$  pro  $x \in (-\infty; x_p]$  je roven  $p$ . Toto vede k zavedení pojmu kvantilu.

Nechť  $X$  je spojitá náhodná veličina,  $p \in (0;1)$ .

Číslo  $x_p$  se nazývá 100%-ní kvantil veličiny  $X$  jestliže platí:

$$P(X \leq x_p) = p \quad (\text{B.10})$$

Pro některé kvantily se používají speciální názvy :

- kvantil  $x_{0,5}$  se nazývá medián (medián je tedy 50%-ní kvantil),
- kvantil  $x_{0,25}$  se nazývá dolní kvartil,
- kvantil  $x_{0,75}$  se nazývá horní kvartil.

### B.1.6 Modus

Modus  $\hat{x}$  náhodné veličiny je hodnota s největší relativní četností, tj. jedná se o hodnotu, která se v souboru dat vyskytuje nejčastěji.

## B.2 Pravděpodobnostní rozdělení

Podle tvaru rozeznáváme symetrické a asymetrické pravděpodobnostní rozdělení. Asymetrické rozdělení je buď levostranně nebo pravostranně asymetrické. Pravostranně asymetrické rozdělení (má delší pravý chvost). Pravděpodobnostní rozdělení jsou většinou jednovrcholová, ale mohou být i dvouvrcholová nebo obecně vícevrcholová. Zpravidla je vícevrcholovost způsobena nehomogenitou dat.

### B.2.1 Normální rozdělení

Normální rozdělení, někdy nazýváno také Gaussovo rozdělení, má zásadní roli ve statistice.

Náhodná veličina  $X$  má Normální rozdělení s parametry  $-\infty < \mu < \infty$  a  $-\infty < \sigma^2 < \infty$ , pokud její hustota pravděpodobnosti odpovídá

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{B.11})$$

Toto rozdělení je značeno  $\mathcal{N}(\mu, \sigma)$ , kde  $\mu$  značí střední hodnotu a  $\sigma^2$  rozptyl. Toto rozložení je symetrické podle  $x = \mu$  (platí také pro Laplaceovo rozdělení v kapitole B.2.3)

### B.2.2 Weibullovo rozdělení

Náhodná veličina  $X$  má Weibullovo rozdělení s parametry  $\delta; c > 0$ , pokud její hustota pravděpodobnosti odpovídá

$$f(x) = \begin{cases} \frac{c \cdot x^{c-1}}{\delta^c} e^{-\left(\frac{x}{\delta}\right)^c} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases} \quad (\text{B.12})$$

Pro střední hodnotu a rozptyl platí následující vztahy.

$$E(X) = \delta \Gamma\left(\frac{1}{c} + 1\right), \quad (\text{B.13})$$

$$D(X) = \delta^2 \left[ \Gamma\left(\frac{2}{c} + 1\right) - \Gamma^2\left(\frac{1}{c} + 1\right) \right]. \quad (\text{B.14})$$

### B.2.3 Laplaceovo rozdělení

Náhodná veličina  $X$  má Laplaceovo rozdělení s parametrem  $a > 0$ , pokud její hustota pravděpodobnosti odpovídá

$$f(x) = \frac{1}{2a} e^{-\frac{|x-\mu|}{a}}, \quad (\text{B.15})$$

kde  $x \in \mathbf{R}$ .

Pro střední hodnotu a rozptyl platí následující vztahy

$$E(X) = \mu, \quad (\text{B.16})$$

$$D(X) = 2a^2. \quad (\text{B.17})$$

### B.2.4 Raygleyho rozdělení

Náhodná veličina  $X$  má Raygleyho rozdělení s parametrem  $b > 0$ , pokud její hustota pravděpodobnosti odpovídá

$$f(x) = \begin{cases} \frac{x}{b^2} e^{-\frac{x^2}{2b^2}} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases} \quad (\text{B.18})$$

Pro střední hodnotu a rozptyl platí následující vztahy

$$E(X) = b\sqrt{\frac{\pi}{2}}, \quad (\text{B.19})$$

$$D(X) = b^2 \left(2 - \frac{\pi}{2}\right). \quad (\text{B.20})$$

Raygleyho rozdělení je v podstatě zjednodušené Webullovo rozdělení, pokud dosadíme za  $c = 2$  a  $\delta = \sqrt{2b}$

## B.2.5 Exponenciální rozdělení

Náhodná veličina  $X$  má Exponenciální rozložení s parametrem  $\lambda > 0$ , pokud její hustota pravděpodobnosti odpovídá

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases} \quad (\text{B.21})$$

Pro Střední hodnotu a rozptyl platí

$$E(X) = \frac{1}{\lambda}, \quad (\text{B.22})$$

$$D(X) = \frac{1}{\lambda^2}. \quad (\text{B.23})$$

Exponenciální rozdělení je používáno při prezentacích životnosti zařízení, která nepodléhají opotřebení nebo doby mezi výskytem dvou událostí.

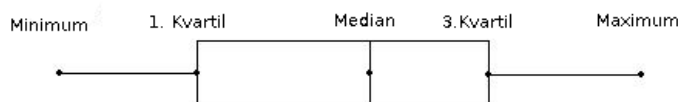
## B.3 Histogram

Histogram je graf kdy na vodorovnou osu znázorníme třídy a na svislou osu četnosti či relativní četnosti. Často se používá ve tvaru, kdy se hodnota odpovídající třídě znázorní jako sloupec s intervalem třídy jako základnou a výška je dána četností.

## B.4 Boxplot

V deskriptivní statistice je boxplot neboli krabicový graf či krabicový diagram jeden ze způsobů grafické vizualizace numerických dat pomocí jejich kvartilů. Střední “krabicová” část diagramu jen shora ohraničena 3. kvartilem, zesponu 1. kvartilem a mezi nimi se nachází linie vymezující medián. Boxploty mohou obsahovat také linie vycházející ze střední části diagramu kolmo nahoru a dolů, tzv. fousky, vyjadřující variabilitu dat pod prvním a nad třetím kvartilem. Odlehle hodnoty, tzv. outliery, pak mohou být vykresleny jako jednotlivé body.

Boxploty zobrazují rozdíly mezi datovými soubory bez jakýchkoli předpokladů normálního rozdělení dat, jsou tedy neparametrické. Rozteče mezi jednotlivými prvky střední části diagramu indikují stupeň disperze (rozptylu) a šikmosti dat. Kromě bodů samotných umožňují také vizuálně odhadnout rozmezí mezi kvartily, rozsah dat, aritmetický průměr a vážený průměr.



## Příloha C

# Doplňující informace

Informace byly čerpány z [17] a [18].

### C.1 Autoregresní proces (AR)

Autoregresní model řádu  $na$  můžeme reprezentovat následujícím vztahem.

$$A(q^{-1})y(t) = e(t), \quad (\text{C.1})$$

kde  $y(t)$  je výstup,  $e(t)$  je bílý šum definovaný jako  $\mathcal{N}(0, \sqrt{D(X)})$ ,  $q$  je zpětný posun. Polynom  $A(q^{-1})$  má následující tvar

$$A(q^{-1}) = 1 + a_1 \cdot q^{-1} + \dots + a_{na} \cdot q^{-na} \quad (\text{C.2})$$

### C.2 Euklidova norma

Nechť  $x = (x_1, \dots, x_n)$  in  $\mathcal{R}^n$ , potom Euklidova norma prvku je dána vztahem

$$\|x\| = \sqrt{(x, x)} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\sum_{i=1}^n x_i^2} \quad (\text{C.3})$$