

Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra kybernetiky

Bakalářská práce

**Datová analýza dostupných dat z
obnovitelných zdrojů se
zaměřením na data z větrných
farem**

Plzeň, 2013

Michal Kubát

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

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů, jejich úplný seznam je její součástí

V Plzni dne

.....

Poděkování

Rád bych poděkoval paní Ing. Andree Zápotocké, Ph.D., vedoucí mé bakalářské práce, za odborné vedení a cenné rady a připomínky, které mi pomohly zpracovat tuto práci.

Anotace

Tato bakalářská práce pojednává o vytvoření filtrů dat a následné datové analýze. Data, která jsou v práci zpracovávána a analyzována jsou naměřené hodnoty činného výkonu vyrobené větrné energie. Hlavním cílem práce je vytvořit filtry pro čištění těchto dat.

Klíčová slova

filtr, větrná energie, korelační koeficient, střední hodnota, denní diagram

Annotation

This work is about creating data filters and subsequent data analysis. Processed and analyzed data are the measured values of active power produced wind energy. The main goal is to create filters for cleaning data.

Key words

filter, wind energy, correlation coefficient, mean value, daily chart

Obsah

1	Úvod	5
2	Použité matematické pojmy	6
2.1	Střední hodnota	6
2.2	Rozptyl a směrodatná odchylka	6
2.3	Korelace	7
2.4	Exponenciální rozdělení	8
2.5	Weibullovo rozdělení	8
2.6	Histogram	8
2.7	Lineární interpolace	9
3	Použitá data	10
4	Vytvořené filtry a čištění dat	11
4.1	Načtení dat	11
4.1.1	Načtení hodinových dat	11
4.1.2	Načtení minutových dat	11
4.2	Filtry na opravu dat	12
4.2.1	Odstranění malých hodnot	12
4.2.2	Odstranění konstantních hodnot	13
4.2.3	Odstranění extrémně velkých hodnot	13
4.3	Náhrada dat a lineární interpolace	15
4.4	Výpisy a čištění dat	17
4.4.1	Výpis souboru dat bez filtrace	17
4.4.2	Výpis souboru dat s filtrací	18
4.5	Zhodnocení	18
5	Datová analýza	20
5.1	Korelační analýza	20
5.1.1	Korelace činného výkonu s naměřeným počasím	23
5.1.2	Korelace činného výkonu s předpověďmi počasí	26
5.2	Statistická analýza	29
5.2.1	Vyrobená energie v jednotlivých rocích	29
5.2.2	Vyrobená energie v ročních obdobích	33
5.2.3	Vyrobená energie v jednotlivých měsících	35
5.2.4	Zhodnocení	36
6	Závěr	37
7	Literatura	38
8	Příloha - Histogramy	39
9	Příloha - Denní diagramy	40
10	Příloha - Algoritmy	42

1 Úvod

Hlavním úkolem této bakalářské práce je vytvořit vhodné filtry pro vyčištění problémových dat. Dalším cílem je takto zpracovaná data analyzovat. Zde použitá data se týkají naměřeného činného výkonu z větrných farem. Analýzou bude určena oblast s nejlepšími naměřenými hodnotami vyrobené elektrické energie. Dále budou pomocí korelačního koeficientu vyhodnoceny nejlepší lokality a kraje pro danou oblast měření z hlediska závislosti rychlosti větru a teplotě ovzduší na vyrobené energii. Ve statistické analýze určíme, kdy dochází k nejvyšší produkci z hlediska ročních období a měsíců a rovněž bude popsán vývoj vyráběné větrné energie během dne prostřednictvím denních diagramů.

Větrná energie vzniká působením aerodynamických sil na listy rotoru větrné elektrárny a turbína umístěná na stožáru převádí energii větru na rotační energii mechanickou, která je poté pomocí generátoru převedena na elektrickou energii.[6]

Práce je rozdělena do několika kapitol. Druhá kapitola je věnována použitým matematickým pojmům, které se z větší části týkají teorie pravděpodobnosti a základům statistiky. Třetí kapitola pojednává o datech použitých datech. Čtvrtá popisuje vytvořené datové filtry a pátá se zabývá korelační a statistickou analýzou. Tyto dvě kapitoly tvoří stěžejní část práce.

Pro vytvoření filtrů, veškeré výpočty a vykreslení grafů byl použit software *MATLAB 7.10.0 (R2010a)*.

2 Použité matematické pojmy

V následující části jsou uvedeny základní vztahy a pojmy z pravděpodobnosti a statistiky, které se vyskytují v této práci. Pokud není uvedeno jinak jsou převzaté z [1], [2] nebo [3].

Náhodná veličina

Jedná se o výsledek náhodného pokusu, který můžeme někdy vyjádřit jen slovně, ovšem většinou je možno jednotlivým elementárním jevům přiřadit číselné hodnoty. Výsledek těchto náhodných pokusů, bývá (po vhodné formulaci) vyjádřitelný reálným číslem. V takových případech lze považovat naměřené hodnoty za hodnoty nějaké funkce definované na množině Ω všech elementárních jevů. Zmíněná funkce se nazývá **náhodná veličina**. Jejím hlavním rysem je proměnlivost jejích hodnot při opakovaných pokusech.

Náhodné veličiny se označují velkými latinskými písmeny z konce abecedy, tedy X, Y, Z . Jejich konkrétní číselné realizace pak odpovídají malým písmenům x, y, z

2.1 Střední hodnota

Střední hodnota náhodné veličiny X , označována $E(X)$, je reálné číslo definované následovně:

- pro **diskrétní** náhodnou veličinu, která nabývá hodnot x_1, x_2, \dots s pravděpodobnostmi p_1, p_2, \dots [tj. $P(X = x_i) = p_i$] je

$$E(X) = \sum_i x_i p_i = x_1 p_1 + x_2 p_2 + \dots$$

- pro **spojitou** náhodnou veličinu, s hustotou pravděpodobnosti $f(x)$ je

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

2.2 Rozptyl a směrodatná odchylka

Rozptyl náhodné veličiny X , píšeme $D(X)$, je reálné číslo definované předpisem

$$D(X) = E(X - E(X))^2.$$

Platí rovnost (tzv. **výpočetní tvar rozptylu**)

$$D(X) = E(X^2) - E^2(X),$$

kde $E^2(X)$ je zkrácený zápis pro $(E(X))^2$,

$E(X^2) = \sum_i x_i^2 p_i$, je-li X diskrétní náhodná veličina,

resp. $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$, je-li X spojitá náhodná veličina.

Směrodatná odchylka náhodné veličiny X , označována $\sigma(X)$, je definována jako druhá odmocnina z rozptylu,

$$\sigma(X) = \sqrt{D(X)}.$$

Udává, do jaké míry se od sebe liší typické případy v souboru zkoumaných dat.

2.3 Korelace

Náhodný vektor

Je to takový vektor, jehož složkami jsou náhodné veličiny. Počet těchto veličin označme k a sledovanou k -rozměrnou náhodnou veličinu označíme

$$X = (X_1, X_2, \dots, X_k).$$

Kovariance

Vyjadřuje vzájemné vztahy mezi dvěma náhodnými veličinami X, Y a je definována předpisem

$$\text{cov}(X, Y) = E([X - E(X)] \cdot [Y - E(Y)]).$$

Korelační koeficient

Uvažujme dvourozměrný náhodný vektor a jeho složky označme X, Y . Dále předpokládejme, že veličiny X, Y nejsou konstantní, a tím pádem jejich směrodatné odchylky $\sigma(X), \sigma(Y)$ jsou kladné. Korelační koeficient $\rho(X, Y)$ je definován vztahem

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Korelační koeficient $\rho = \rho(X, Y)$ udává míru statistické lineární závislosti veličin X, Y . V případech kdy $\rho = \pm 1$ jsou náhodné veličiny X, Y přesně lineárně závislé. Znaménko mínus značí nepřímou úměrnost a plus přímou úměrnost. Je-li $\rho = 0$, náhodné veličiny jsou na sobě nezávislé tzv. nekorelované. Veličiny X, Y ovšem mohou být nekorelované a zároveň závislé.

Koeficient determinace

Druhá mocnina korelačního koeficientu se nazývá koeficient determinace, který po vynásobení číslem 100 udává procentuální míru závislosti dvou veličin.

2.4 Exponenciální rozdělení

Exponenciální rozdělení s parametrem $\delta > 0$:

X má hustotu

$$f(x) = \begin{cases} 0 & x \in (-\infty, 0] \\ \frac{1}{\delta} e^{-(x/\delta)} & x \in (0, \infty). \end{cases}$$

Píšeme: $X \sim \text{Exp}(\delta)$.

Distribun funkce :

$$F(x) = \begin{cases} 0 & x \in (-\infty, 0] \\ 1 - e^{-(x/\delta)} & x \in (0, \infty). \end{cases}$$

Výpočetem: $E(X) = \delta$, $D(X) = \delta^2$

2.5 Weibullovo rozdělení

Náhodná veličina X má Weibullovo rozdělení s parametry $c > 0$, jestliže X má distribuční funkci

$$F(x) = \begin{cases} 0 & x \in (-\infty, 0] \\ 1 - e^{-(x/\delta)^c} & x \in (0, \infty). \end{cases}$$

Píšeme pak $X \sim W(c, \delta)$.

Při volbě $c = 1$, dostaneme exponenciální rozdělení.

Pro $c < 1$ je hustota pravděpodobnosti klesající na $(0, +\infty)$.

Pro $c > 1$ má hustota pravděpodobnosti na $(0, +\infty)$ tvar nesymetrického zvonu podobně jako u logaritmicke-normálního rozdělení.

Střední hodnotu a rozptyl této veličiny lze vyjádřit pomocí funkce gama

$$E(X) = \Gamma\left(\frac{1}{c} + 1\right) \delta$$

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

2.6 Histogram

Jedná se o grafické vyjádření výskytu sledované veličiny v daném intervalu (třídě). Zobrazuje se pomocí něj rozložení hodnot pro soubory velkého rozsahu.

2.7 Lineární interpolace

Jedná se o metodu prokládání křivek použitím lineárních mnohočlenů. Jsou dány dva body se souřadnicemi (x_0, y_0) a (x_1, y_1) . Lineární interpolace je přímka mezi těmito dvěma body. Platí rovnice:

$$\frac{y - y_0}{y_1 - y_0} = \frac{x - x_0}{x_1 - x_0}$$

Vyřešením této rovnice pro y dostaneme vzorec pro lineární interpolaci:

$$y = y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0}$$

Definice převzata z [4].

3 Použitá data

Použitá data pro zpracování se skládají z různých období roků 2011 až 2012.

Naměřený činný výkon

Činný výkon je měřen v rámci:

1. Oblastí

Jedná se o minutová data vyrobené větrné energie v MW. Výkon je měřen v oblastech s názvy E.ON. Východ, SME - ČEZ Morava, SCE - ČEZ Sever, STE - ČEZ Střed a VCE - ČEZ Východ. Přičemž v analýze je největší důraz kladen na data E.ON. Východ, která jsou nejkvalitnější.

2. Celé České Republiky V tomto případě se jedná o hodinové průměry vyrobené větrné energie MWhod/hod.

Naměřené počasí

Jde o hodinová data počasí v České republice rozdělené do 39 lokalit. Ke každé lokalitě je rovněž uveden kraj, do kterého spadá. Data z tohoto měření použítá pro analýzu obsahují informace o rychlosti větru, směru větru, teplotě ovzduší a oblačnosti. Oblačnost vyjadřuje stupeň pokrytí oblohy oblačností v osminách.

Předpověď počasí

Jedná se o denní předpověď počasí pro všech 14 krajů České Republiky. Tedy pro každý den v roce data obsahují předpověď na daný den, noc a den následující. Data použítá pro analýzu opět obsahují informace o rychlosti větru, směru větru, teplotě ovzduší a oblačnosti.

4 Vytvořené filtry a čištění dat

4.1 Načtení dat

Veškeré skripty jsou napsány a používány v softwaru *MATLAB*. Soubor ve formátu *.csv* načteme pomocí funkce *importdata()*. Jedná se o tabulkový soubor, kde první sloupec obsahuje časový údaj rozdělený po minutách a zbylých 14 sloupců určuje naměřenou hodnotu činného výkonu v daný čas pro konkrétní lokalitu. Data tedy rozdělíme do struktur na matici hodnot s typickým názvem *signals.values* a časový vektor *time*, kde dané datum převedeme použitím funkce *datenum()* do číselného kódu. Data zpracováváme pro každý rok zvlášť.

4.1.1 Načtení hodinových dat

Jde o hodinová data měřená pro celou Českou Republiku. Přesněji řečeno se jedná o hodinové průměry. Z důvodu zachování ekvidistance jsou chybějící data nahrazena NaN (Not-a-Number), se kterými *MATLAB* dokáže pracovat.

4.1.2 Načtení minutových dat

Naměřená minutová data pro oblasti České Republiky získáme ze zdrojového souboru, podobně jako u hodinových dat. Ovšem v tomto případě vytváříme matici s devíti sloupci. Názvy těchto devíti oblastí ve kterých jsou k dispozici naměřené hodnoty jsou uvedeny v sekci použitá data. Sloupce vytvořené matice, pro které data z odpovídajících oblastí nejsou dostupná opět obsahují hodnoty NaN.

4.2 Filtry na opravu dat

Filtry na opravu dat jsou pro hodinová data a minutová data v podstatě stejné. Z toho důvodu uvádím pouze ty pro minutová data, které jsou mírně složitější a filtry hodinových dat jsou v příloze. V následujícím textu jsou vždy popsány rozdíly mezi oběma typy. Výsledky práce každého filtru jsou uloženy do textových souborů, které jsou na přiloženém CD. Filtry ve výpočtech dále používají instalovaný výkon, který je pro jednotlivé lokality určen v následující tabulce 1.

Oblast	Instalovaný výkon [MW]
E.ON. Východ	21
SME - ČEZ Morava	43
SCE - ČEZ Sever	93
STE - ČEZ Střed	6
VCE - ČEZ Východ	21
ČR	93.05

Tabulka 1: Instalovaný výkon

První úpravou naměřených dat, která z důvodu jednoduchosti nemá samostatnou kapitolu je náhrada všech záporných hodnot nulou. Tento filtr je pro oba typy dat totožný.

4.2.1 Odstranění malých hodnot

Vstupem filtru pro odstranění malých hodnot jsou naměřená data a instalovaný výkon, který je pro jednotlivé oblasti vyčíslen v tabulce 1. Výstupem pak jsou opravená naměřená data a textový soubor s provedenými datovými změnami. Prvním parametrem filtru je mez minimální hodnoty, která z důvodu univerzálnosti algoritmu odpovídá jednomu procentu instalovaného výkonu pro danou oblast. Filtr tedy pracuje s normovanými hodnotami naměřeného výkonu. Druhým parametrem je minimální doba výskytu příliš malé hodnoty. Tato požadovaná doba je 5 po sobě následujících minut.

Hodnoty naměřeného výkonu vyhodnocené filtrem jako nežádoucí nejsou ze souboru dat odstraněny přímo z důvodu zachování ekvidistance. Jsou pouze nahrazeny hodnotou NaN, která v MATLABu označuje Not-a-Number. Rozměr vektoru dat tedy zůstane nezměněn a je možné s ním dále pracovat.

Filtr vytvořený pro hodinová data (viz Algoritmus 6 v příloze) je podstatně jednodušší. V tomto případě pracujeme s každou hodnotou, která je menší než požadované minimum. Rozdíl je rovněž v tom, že ji nenahrazujeme hodnotou NaN, nýbrž nulou.

Algoritmus 1: Filtr na odstranění malých hodnot

```
1  if instal_vykon(i)>0,
2      pom = pom/instal_vykon(i)*100;
3      index0 = find(pom(:,1) < 1 & pom(:,2) < 1 &...
4                  pom(:,3) < 1 & pom(:,4) < 1 & pom(:,5) < 1);
5      VTE_oblasti.signals.values(index0,i) = NaN;
6
7      indexNul = find((isnan(VTE_oblasti.signals.values(:,i))...
8                      ==1)&(isnan(old(:,i))==0));
9      opravovane_hodnoty = old(indexNul,i);
10     y = [indexNul'; opravovane_hodnoty'];
11     nazev = VTE_oblasti.hodnoty{i};
12     file = ['vypisy_oprav\' zpracovane_VTE_oblasti...
13            \'_ int2str(rok) \'_Odstranění malých hodnot_\'...
14            datestr(now, 'yyyy-mm-dd_HH-MM') \'.txt'];
15     fid = fopen(file, 'a');
16     fprintf(fid, '\n');
17     fprintf(fid, nazev);
18     fprintf(fid, '\n');
19     fprintf(fid, '%6.0f %10.4f \n', y);
20     status = fclose(fid);
21 end
```

4.2.2 Odstranění konstantních hodnot

V případě toho to filtru jsou vstupem pouze naměřená data. Výstupem je jejich opravená verze a textový soubor s provedenými změnami. Parametrem filtru je minimální doba trvání konstantní hodnoty, která činí 5 minut. V uvedeném algoritmu je ošetřen problém, který by nastal, kdyby posledních pět hodnot vektoru *index* nebylo nulových. Konstantní hodnoty naměřených výkonů jsou opět přepsány na hodnotu NaN.

Algoritmus filtru pro hodinová data (viz Algoritmus 7 v příloze) se liší od minutového pouze tím, že jsou hodnotou NaN nahrazovány tři po sobě následující hodiny s konstantní hodnotou.

4.2.3 Odstranění extrémně velkých hodnot

Poslední filtr odstraňuje veškeré nesmyslně velké hodnoty. Vstupem jsou naměřená data a instalovaný výkon. Využívá se zde opět normování pomocí instalovaného výkonu. Parametrem je zde mez maximální hodnoty. Jako extrémně vysoké jsou klasifikovány takové hodnoty, které dvojnásobně překračují daný instalovaný výkon. Je tedy zaručena univerzálnost použití. Výstupem jsou opravené hodnoty a opravované hodnoty s jejich indexy jsou potom opět zapsány a uloženy do souboru, což v příloženém kódu již není uvedeno. Rovněž, chybí v tomto případě verze filtru pro hodinová data v příloze, protože je v podstatě stejná jako minutová verze.

Algoritmus 2: Filtr na odstranění konstantních hodnot

```
1 index = pom(:,1) == pom(:,2) & pom(:,2) == pom(:,3) &...
2     pom(:,3) == pom(:,4) & pom(:,4) == pom(:,5);
3 index(end-4:end,1) = 0;
4 x = find(index == 1);
5 VTE_oblasti.signals.values(x,i) = NaN;
6 VTE_oblasti.signals.values(x+1,i) = NaN;
7 VTE_oblasti.signals.values(x+2,i) = NaN;
8 VTE_oblasti.signals.values(x+3,i) = NaN;
9 VTE_oblasti.signals.values(x+4,i) = NaN;
10
11 indexKonst = find((isnan(VTE_oblasti.signals.values(:,i))...
12     ==1)&(isnan(old(:,i))==0));
13 opravovane_hodnoty = old(indexKonst,i);
14 y = [indexKonst'; opravovane_hodnoty'];
15 nazev = VTE_oblasti.hodnoty{i};
16 file = ['vypisy_oprav\' zpracovane_VTE_oblasti...
17     \'_\' int2str(rok) \'_Odstranění konstantních hodnot\'...
18     datestr(now,\'yyyy-mm-dd_HH-MM\') \'.txt\'];
19 fid = fopen(file, \'a\');
20 fprintf(fid,\'\n\');
21 fprintf(fid, nazev);
22 fprintf(fid,\'\n\');
23 fprintf(fid, \'%6.0f %10.4f \n\', y);
24 status = fclose(fid);
```

Algoritmus 3: Filtr na odstranění extrémně velkých hodnot

```
1 big_hodnoty = VTE_oblasti.signals.values ./ ...
2     repmat(instal_vykon, size(VTE_oblasti.signals.values,1),1)*100;
3 VTE_oblasti.signals.values(big_hodnoty > 200) = NaN;
```

Algoritmus 4: Náhrada chybějících dat hodinovými průměry

```
1 pom = [VTE_skutecnost.signals.values(1:end-2,4)...
2         VTE_skutecnost.signals.values(2:end-1,4)...
3         VTE_skutecnost.signals.values(3:end,4)];
4 prumer_online = nanmean(reshape(Struk_pom(:,4),60,[]))';
5 pom2 = [isnan(pom(:,1))==0 isnan(pom(:,2))==1 isnan(pom(:,3))==0];
6 idx = pom2(:,1)==1 & pom2(:,2)==1 & pom2(:,3)==1;
7 index_NaNu = find(idx==1)+1;
8 VTE_skutecnost.signals.values(index_NaNu,4) =...
9     prumer_online(index_NaNu,1);
10
11
12 for i=1:1:4,
13     indexH = find((isnan(VTE_skutecnost.signals.values(:,i))...
14                  ==0)&(isnan(old(:,i))==1));
15     opravovane_hodnoty = VTE_skutecnost.signals.values(indexH,i);
16     y = [indexH'; opravovane_hodnoty'];
17     nazev = VTE_skutecnost.hodnoty{i};
18     file = ['vypisy_interpolaci\' zpracovane_VTE_skutecnost '_'...
19             int2str(rok) '_1 hodina_'...
20             datestr(now,'yyyy-mm-dd_HH-MM') '.txt'];
21     fid = fopen(file, 'a');
22     fprintf(fid, '\n');
23     fprintf(fid, nazev);
24     fprintf(fid, '\n');
25     fprintf(fid, '%6.0f %10.4f \n', y);
26     status = fclose(fid);
27 end
```

4.3 Náhrada dat a lineární interpolace

Hodinová data

Vstupem algoritmu 4 jsou již opravené naměřené hodnoty činného výkonu. Parametrem je doba výskytu hodnoty NaN, což je v tomto případě jedna hodina. Algoritmus podle zadání nahrazuje pouze, takové hodnoty NaN, v jejichž okolí se vyskytuje smysluplná hodnota naměřeného výkonu. Jinými slovy prostřední ze tří čísel má hodnotu NaN. Takovou hodnotu potom nahrazujeme hodinovým průměrem naměřených dat. Tento průměr odpovídá lineární interpolaci, kdy jsou použita pouze čísla nacházející se před a po dané interpolované hodnotě. Výstupem algoritmu jsou opravené hodnoty doplněné o interpolované a textový soubor se zaznamenanými změnami. Jsou v něm uvedeny indexy hodnot nahrazených hodinovými průměry a odpovídající nová hodnota.

Minutová data

Algoritmus 5 pro minutová data má vstup v podobě filtrovaných naměřených dat. Parametrem je doba výskytu hodnoty NaN. Výstup tvoří opravené a následně interpolované hodnoty činných výkonů a rovněž textový soubor s údaji o změnách. V tomto případě jsou nahrazovány hodnoty NaN trvajících 1,2,3,4 nebo 5 minut za použití lineární interpolace. Zde je uveden pouze algoritmus pro trvalou hodnotu

Algoritmus 5: Náhrada chybějících dat lineární interpolací

```

1 pom2 = [isnan(pom(:,1))==0 isnan(pom(:,2))==1 ...
2         isnan(pom(:,3))==1 isnan(pom(:,4))==1 ...
3         isnan(pom(:,5))==1 isnan(pom(:,6))==1 isnan(pom(:,7))==0];
4 idx = pom2(:,1)==1 & pom2(:,2)==1 & pom2(:,3)==1 & ...
5       pom2(:,4)==1 & pom2(:,5)==1 & pom2(:,6)==1 & pom2(:,7)==1;
6 index_NaNu = find(idx==1)+1;
7 pocet = length(index_NaNu);
8 for j=1:1:pocet,
9     X = [index_NaNu(j)-1 index_NaNu(j)+5];
10    Y =VTE_oblasti.signals.values([index_NaNu(j)-1"...
11        index_NaNu(j)+5],i);
12    xi = index_NaNu(j)-1:1:index_NaNu(j)+5;
13    yi = interp1(X,Y,xi);
14    VTE_oblasti.signals.values(index_NaNu(j),i) = yi(2);
15    VTE_oblasti.signals.values(index_NaNu(j)+1,i) = yi(3);
16    VTE_oblasti.signals.values(index_NaNu(j)+2,i) = yi(4);
17    VTE_oblasti.signals.values(index_NaNu(j)+3,i) = yi(5);
18    VTE_oblasti.signals.values(index_NaNu(j)+4,i) = yi(6);
19 end
20 index = find((isnan(VTE_oblasti.signals.values(:,i))==0)&...
21             (isnan(old(:,i))==1));
22 opravene_hodnoty = VTE_oblasti.signals.values(index,i);
23 y = [index'; opravene_hodnoty'];
24 nazev = VTE_oblasti.hodnoty{i};
25 file = ['vypisy_interpolaci\' zpracovane_VTE_oblasti '_'...
26         int2str(rok) '_5 minut_'...
27         datestr(now,'yyyy-mm-dd_HH-MM') '.txt'];
28 fid = fopen(file, 'a');
29 fprintf(fid,'\n');
30 fprintf(fid, nazev);
31 fprintf(fid,'\n');
32 fprintf(fid, '%6.0f %10.4f \n', y);
33 status = fclose(fid);

```

NaN v délce 5 minut.

4.4 Výpisy a čišťení dat

Jak již bylo zmíněno datové změny provedené prostřednictvím filtru jsou zaznamenány a uloženy do textové souborů, v jehož názvu je uveden zdroj dat, typ opravy a datum vytvoření. V souboru je zapsán název konkrétní opravované oblasti, pod kterým se nacházejí indexy a jim příslušející naměřené hodnoty, které byly pomocí filtru opraveny.

Dále byl vytvořen script pro výpisy (viz Algoritmus 8 v příloze), které jsou určeny pro výstupní soubory *.mat* vzniklé zpracováním dat. Při spuštění skriptu se na obrazovku vypíše průměrná a maximální délka NaN úseku a úseku s příliš malou hodnotou naměřené energie. Kód výpisu pro minutová data je uveden v příloze. Pro hodinová data je algoritmus analogický. Program vypíše vždy název měřené oblasti, následně pomocí funkce *isnan()* rozhoduje, zda naměřené data zcela chybí. V takovém případě vypíše příslušnou hlášku nebo hodnoty činného výkonu jsou pro danou oblast kompletní. Pokud nějaká data chybí nalezne začátky a konce NaN úseků, vypočte jejich délky a průměrnou délku vypíše. Obdobně tomu je v případě vypsání průměrné délky úseku s hodnotou menší než 0.01.

V této části jsou pro ukázkou uvedeny výpisy ze zpracovaného souboru, který mapuje záznamy o vyrobené energii z druhé poloviny roku 2011. Pro porovnání je uveden výpis souboru bez oprav naměřených dat a s opravami.

4.4.1 Výpis souboru dat bez filtrace

Vstupem jsou data pouze po zpracování do struktur bez jakýchkoliv náhrad naměřených hodnot.

```
-----E.ON. Východ-----
Průměrná délka useku s trvalou hodnotou NaN pro E.ON. Východ je 6 min.
Maximální délka NaN úseku: 6 min.
Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro EON Východ je
18 min.
-----SME - CEZ Morava-----
Průměrná délka useku s trvalou hodnotou NaN pro SME - CEZ Morava je 6 min.
Maximální délka NaN úseku: 6 min.
Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro SME - CEZ Morava
je 14 min.
-----SCE - CEZ Sever-----
Průměrná délka useku s trvalou hodnotou NaN pro SCE - CEZ Sever je 6 min.
Maximální délka NaN úseku: 6 min.
Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro SCE - CEZ Sever
je 32 min.
-----STE - CEZ Stred-----
Průměrná délka useku s trvalou hodnotou NaN pro STE - CEZ Stred je 6 min.
Maximální délka NaN úseku: 6 min.
Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro STE - CEZ Stred
je 155 min.
```

-----VCE - CEZ Východ-----

Průměrná délka úseku s trvalou hodnotou NaN pro VCE - CEZ Východ je 6 min.
Maximální délka NaN úseku: 6 min.

Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro VCE - CEZ Východ je 1055 min.

4.4.2 Výpis souboru dat s filtrací

Vstupem těchto výpisů jsou data po filtraci a následné interpolaci.

-----E.ON. Východ-----

Průměrná délka úseku s trvalou hodnotou NaN pro E.ON. Východ je 26 min.
Maximální délka NaN úseku: 4969 min.

Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro EON Východ je 2 min.

-----SME - CEZ Morava-----

Průměrná délka úseku s trvalou hodnotou NaN pro SME - CEZ Morava je 125 min.

Maximální délka NaN úseku: 43395 min.

Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro SME - CEZ Morava je 2 min.

-----SCE - CEZ Sever-----

Průměrná délka úseku s trvalou hodnotou NaN pro SCE - CEZ Sever je 107 min.

Maximální délka NaN úseku: 43282 min.

Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro SCE - CEZ Sever je 2 min.

-----STE - CEZ Stred-----

Průměrná délka úseku s trvalou hodnotou NaN pro STE - CEZ Stred je 171 min.

Maximální délka NaN úseku: 122288 min.

Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro STE - CEZ Stred je 3 min.

-----VCE - CEZ Východ-----

Průměrná délka úseku s trvalou hodnotou NaN pro VCE - CEZ Východ je 248 min.

Maximální délka NaN úseku: 89906 min.

Průměrná délka úseku s trvalou hodnotou menší než 0.01 pro VCE - CEZ Východ je 4 min.

4.5 Zhodnocení

Při porovnání obou výpisů podle očekávání pozorujeme nárůst úseků s hodnotou NaN a nižší počet dat s nízkou hodnotou. Jako nejkvalitnější naměřená data lze tedy určit E.ON. Východ, kde bylo odfiltrováno nejméně dat. V ostatních oblastech

sledujeme rapidní nárůst NaN hodnot. Hlavním problémem těchto dat, který řeší filtrace, jsou jejich konstantní hodnoty.

5 Datová analýza

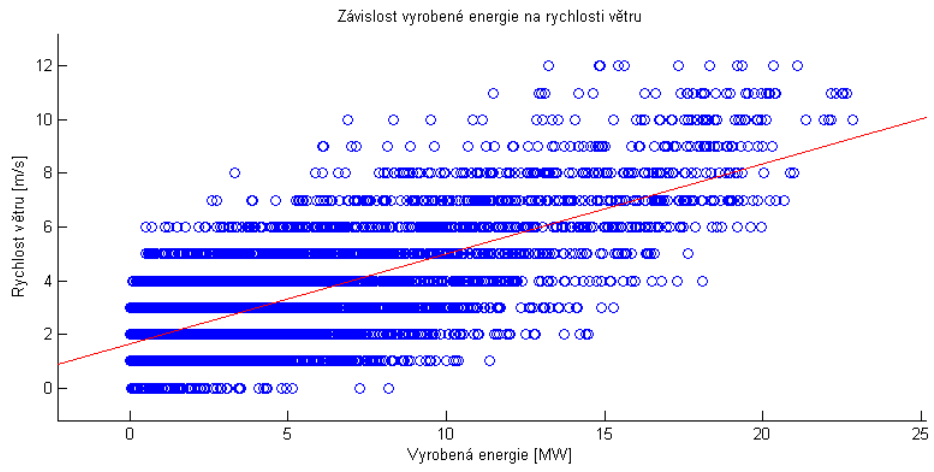
V této kapitole jsou ukázány výsledky provedených datových analýz. Analýzy byly provedeny na datech opravených pomocí filtrů z předchozí kapitoly.

5.1 Korelační analýza

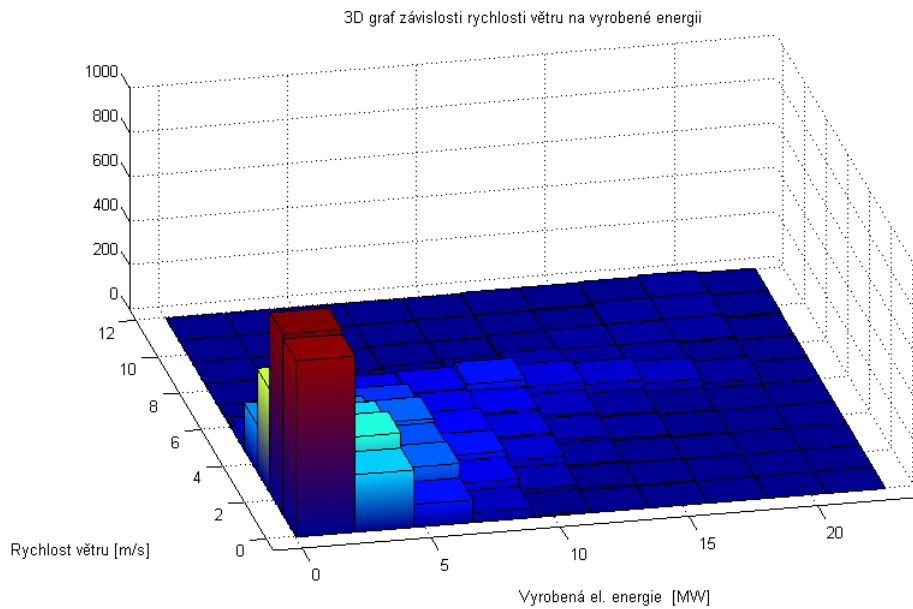
Pro určení korelací byly sestaveny tabulky s jednotlivými lokalitami nebo kraji s naměřenými prvky počasí (rychlost větru, směr větru, teplota ovzduší a oblačnost) a naměřeným činným výkonem pro danou oblast. Tabulky jsou za celé dva roky měření a dále jsou pro porovnání uvedeny také za čtyři roční období z meteorologického hlediska. Například jaro tedy začíná 1. března a končí 31. května. Tabulky jsou však příliš rozměrné a nacházejí se na přiloženém CD. Zde budou uvedeny pouze vyhodnocovací tabulky těchto korelací. Pro každou oblast bude vybrána nejlepší lokalita (resp. kraj) s nejvyšší hodnotou korelace. Není v nich uvedeno hledisko směru větru, z důvodu, že hodnoty korelací dosahovaly příliš nízkých hodnot. Lze tedy usoudit, že směr větru téměř nemá vliv na vyrobený výkon. To lze přisoudit schopnosti lopatek elektrárny natočit se proti větru. Rovněž z úsporných důvodů není uvedena oblačnost, ale bude vždy analyzováno i toto hledisko. V tabulkách je uveden **korelační koeficient**, ale při analýze je přihlédnuto i k jeho druhé mocnině, která se nazývá **koeficient determinace** a udává se v procentech.

Následující grafy zobrazují závislost vyrobeného činného výkonu na rychlosti větru a teplotě ovzduší pro určení přímé či nepřímé úměrnosti. Hodnoty výkonu jsou použity z oblasti E.ON. Východ a prvky počasí byly naměřeny pro lokalitu Náměšť nad Oslavou.

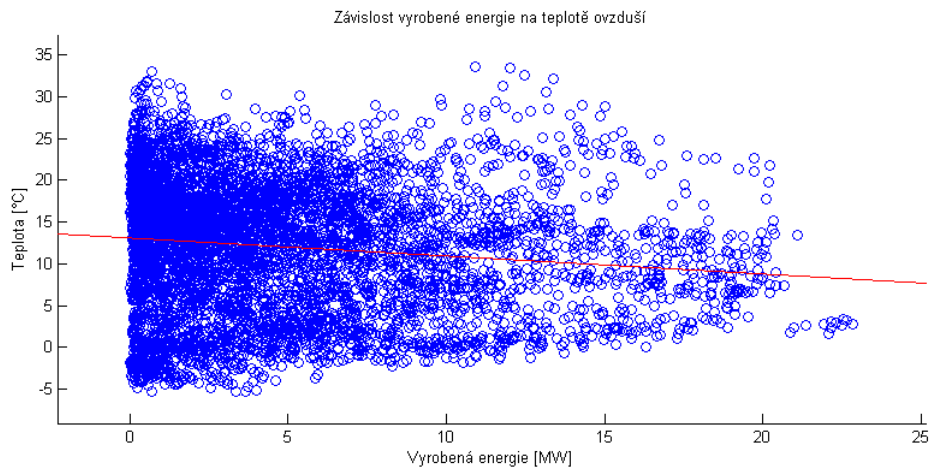
Z grafu (Obr. 1) je zřejmé, že vyrobená energie je podle očekávání přímo úměrná rychlosti větru. Jednotlivé rychlosti větru byly naměřeny v celých číslech. Nachází se zde rovněž 3D graf této závislosti (Obr. 2), kde vidíme četnost jednotlivých hodnot. Na obrázku 3 pro stejnou oblast a lokalitu již pozorujeme nepřímou úměrnost, z čehož vyplývá, že čím nižší teplota, tím vyšší je vyráběný výkon. Tato skutečnost odpovídá realitě, protože při nižších teplotách více fouká vítr. 3D graf (Obr. 4) znázorňuje, že nejčastěji se teplota pohybuje okolo 15 st. C.



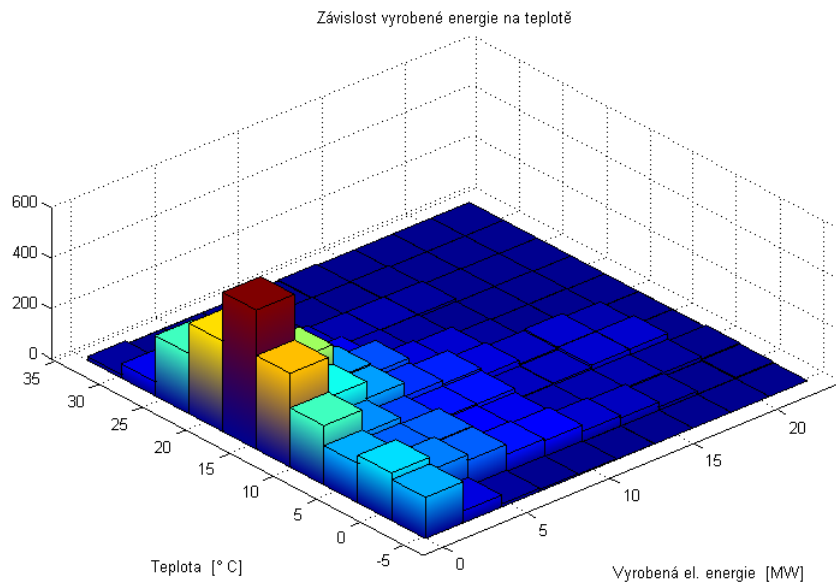
Obrázek 1: Závislost výkonu na rychlosti větru



Obrázek 2: Závislost výkonu na rychlosti větru



Obrázek 3: Závislost výkonu na teplotě ovzduší



Obrázek 4: Závislost výkonu na teplotě ovzduší

5.1.1 Korelace činného výkonu s naměřeným počasím

Nejdříve je uvedena celková tabulka korelací rychlosti větru a teploty ovzduší za roky 2011 a 2012.

2011-2012	m/s		Teplota	
	Oblast měření	Lokalita	Korelace	Lokalita
E.ON. Východ	Náměšť	0,7591	Svratouch	-0,2259
SME - ČEZ Morava	Červená	0,6285	Šerák	-0,3655
SCE - ČEZ Sever	Milešovka	0,6019	Milešovka	-0,2958
STE - ČEZ Střed	Praha - Ruzyně	0,5160	Praha - Kbely	-0,3185
VCE - ČEZ Východ	Svratouch	0,6739	Svratouch	-0,2419

Z tabulky 5.1.1 vidíme vždy nejvhodnější lokality pro měření z daného hlediska. Nejvyšší podíl na vyrobené energii má pochopitelně rychlost větru, čemuž odpovídá vysoká hodnota korelace. Vhodné lokality rovněž vyhovují geografickému rozložení oblastí, ve kterých bylo měřeno.

Z hlediska teploty vyšla korelace záporná. Panuje zde nepřímá úměrnost, z čehož vyplývá, že čím nižší teplota, tím vyšší je vyráběný výkon. Tato skutečnost odpovídá realitě, protože při nižších teplotách více fouká vítr. Zajímavé v tomto případě je, že pro měřící stanici Milešovka, která se nachází na severu republiky, vyšly vysoké hodnoty korelací i pro vzdálené oblasti. Vhodné lokality uvedené v tabulce jsou tedy zohledněny podle oblastí měření. Výjimku tvoří lokalita Svratouch, která se nachází zhruba na hranicích mezi oblastmi E.ON. Východ a VCE - ČEZ Východ, lze ji tedy zařadit do obou oblastí měření.

Co se týče oblačnosti, tak korelace dosahuje sice vyšších hodnot než u směru větru, ovšem i v zde lze tvrdit, že tento aspekt nemá vysoký vliv na vyrobenou energii. Lze říci, že vyšší oblačnost znamená vyšší rychlost větru.

Tabulka 2: Korelace za roční období jaro

Jaro	m/s		Teplota	
	Oblast měření	Lokalita	Korelace	Lokalita
E.ON. Východ	Náměšť	0,8086	Svratouch	-0,0845
SME - ČEZ Morava	Přerov	0,5188	Šerák	-0,1855
SCE - ČEZ Sever	Tušimice	0,3540	Tušimice	0,1314
STE - ČEZ Střed	Praha - Ruzyně	0,7081	Praha - Kbely	0,1137
VCE - ČEZ Východ	Svratouch	0,7253	Ústí nad Labem	-0,0271

V tabulce 2 která mapuje roční období jaro, pozorujeme změny nejvhodnějších lokalit, Pro oblast SCE - ČEZ Sever je v tomto období z obou hledisek nejvhodnější stanice Tušimice. Dále sledujeme nižší význam teploty ovzduší na vyrobenou energii, jinak oblačnost a rychlost větru zůstává na stejných číslech.

Tabulka 3: Korelace pro roční období léto

Léto	m/s		Teplota	
	Oblast měření	Lokalita	Korelace	Lokalita
E.ON. Východ	Náměšť	0,7435	Svratouch	-0,2286
SME - ČEZ Morava	Červená	0,6392	Šerák	-0,0880
SCE - ČEZ Sever	Milešovka	0,6708	Milešovka	-0,1577
STE - ČEZ Střed	Praha - Kbely	0,4316	Čáslav	-0,1515
VCE - ČEZ Východ	Svratouch	0,6305	Svratouch	-0,1195

V letním období vychází často jako vhodná lokalita Milešovka, přičemž pro oblast SCE - ČEZ Sever má největší hodnotu korelačního koeficientu stanice Přimda. V tabulce 3 je tedy opět výrazněji přihlédnuto ke geografickému rozložení lokalit. Zajímavostí je poměrně vysoká korelace oblačnosti v Liberci s měřením výkonu v SCE - ČEZ Sever.

Tabulka 4: Korelace pro roční období podzim

Podzim	m/s		Teplota	
	Oblast měření	Lokalita	Korelace	Lokalita
E.ON. Východ	Svratouch	0,7334	Holešov	0,0870
SME - ČEZ Morava	Šerák	0,6382	Šerák	-0,1439
SCE - ČEZ Sever	Milešovka	0,6625	Milešovka	-0,1693
STE - ČEZ Střed	Praha - Ruzyně	-0,4228	Praha - Kbely	-0,2544
VCE - ČEZ Východ	Svratouch	0,4027	Pardubice	-0,2865

Na podzimní tabulce 4 je zarážející záporná hodnota korelace v oblasti STE - ČEZ Střed s rychlostí větru. Záporná hodnota neplatí pouze pro lokalitu Praha - Ruzyně, kde je největší, ale také pro všech zbývajících 38 lokalit. Lze tedy tvrdit, že došlo k chybnému měření. Dále pozorujeme poprvé změnu pro oblast E.ON. Východ a SME - ČEZ Morava z hlediska rychlosti větru. Vyskytuje se zde i záporné znaménko v případě oblačnosti, nemusí však nutně znamenat chybu v měření, protože s klesající oblačností může růst síla větru.

Tabulka 5: Korelace pro roční období zima

Zima	m/s		Teplota	
	Oblast měření	Lokalita	Korelace	Lokalita
E.ON. Východ	Náměšť	0,7748	Svratouch	0,3115
SME - ČEZ Morava	Červená	0,7393	Šerák	-0,2471
SCE - ČEZ Sever	Milešovka	0,6539	Milešovka	0,1833
STE - ČEZ Střed	Praha - Kbely	0,5291	Praha - Ruzyně	0,3156
VCE - ČEZ Východ	Pardubice	0,2555	Svratouch	0,2070

Poslední tabulka 5 pro zimu se liší od ostatních kladnými hodnotami korelací ve většině oblastí v ohledu na teplotu, což je patrně způsobeno vyšším počtem teplot

ovzduší pod bodem mrazu. Zimní měsíce mají velký vliv na výrobu větrné energie, což bude potvrzeno v kapitole 5.2.

Zhodnocení

Zde je uvedena tabulka 6, která určuje vždy nejvhodnější lokalitu pro danou oblast měření. Jako druhá je uvedena lokalita, jejíž vhodnost byla určena zpravidla z hlediska teploty ovzduší, kde občas musela být vybrána lokalita, která geograficky odpovídá dané oblasti měření, ačkoliv její hodnota korelace nebyla největší. Nejčastěji tato skutečnost nastávala pro lokalitu Milešovka, které se jevila jako vhodná i pro vzdálené oblasti.

Tabulka 6: Nejvhodnější lokality měření pro danou oblast

Oblast měření	Lokalita
E.ON. Východ	Náměšť nad Oslavou, Svratouch
SME - ČEZ Morava	Červená u Libavé, Šerák
SCE - ČEZ Sever	Milešovka, Tušimice
STE - ČEZ Střed	Praha - Ruzyně, Praha - Kbely
VCE - ČEZ Východ	Svratouch, Pardubice

5.1.2 Korelace činného výkonu s předpověďmi počasí

Předmětem této korelační analýzy bude opět pozorování závislosti na vyráběné energii z hledisek rychlosti větru a teploty ovzduší, ovšem tentokrát bude použita předpověď počasí pro jednotlivé kraje České republiky. Budeme tedy hledat kraj s nejvhodnější předpovědí pro danou oblast.

Tabulka 7: Korelace pro celé roky 2011 a 2012

2011-2012	m/s		Teplota	
	Oblast měření	Kraj	Kraj	Korelace
E.ON. Východ	Vysočina	0,8344	Jihomoravský	0,0780
SME - ČEZ Morava	Jihomoravský	0,6600	Olomoucký	-0,3908
SCE - ČEZ Sever	Plzeňský	0,6708	Ústecký	-0,3593
STE - ČEZ Střed	Středočeský	0,4153	Středočeský	-0,3600
VCE - ČEZ Východ	Vysočina	0,5722	Pardubický	-0,3168

Při pohledu na tabulku 7 pro celé dva roky, pozorujeme velmi vysokou korelaci pro kraj Vysočina z hlediska rychlosti větru. Dále pak sledujeme celkově větší závislost vyrobené energie na teplotě ovzduší, než tomu bylo z pohledu lokalit. Zajímavé je, že vysoké korelace zde dosahuje Karlovarský a Plzeňský kraj, přestože geograficky nepatří do dané oblasti. Oba kraje spadají do oblasti ČEZ západ, pro kterou ovšem nebyla naměřená data k dispozici. Jsou zde tedy uvedeny takové kraje přibližně odpovídají svojí oblasti, ve které bylo měřeno. Plzeňský kraj příliš nezapadá do oblasti SCE - ČEZ Sever, ale hodnota korelačního koeficientu byla pro tento kraj největší.

Tabulka 8: Korelace pro roční období jaro

Jaro	m/s		Teplota	
	Oblast měření	Kraj	Kraj	Korelace
E.ON. Východ	Vysočina	0,8320	Vysočina	-0,0744
SME - ČEZ Morava	Jihomoravský	0,5832	Olomoucký	-0,2068
SCE - ČEZ Sever	Ústecký	-0,1570	Plzeňský	0,6982
STE - ČEZ Střed	Středočeský	0,8155	Středočeský	-0,2295
VCE - ČEZ Východ	Vysočina	0,7704	Vysočina	-0,0259

Pro roční období jaro se v aspektu rychlosti v tabulce 8 vyskytují totožné kraje, jako v předchozí tabulce, pouze Plzeňský kraj na Ústecký, který správně patří do této oblasti. Ovšem nastává u něj, stejně u ostatních krajů záporná hodnota korelace. Rovněž z pohledu teploty nastává pro oblast SCE - ČEZ Sever neobvyklá míra závislost, která je nečekaně vysoká a navíc kladná. Lze tedy tvrdit, že měření v této oblasti bylo na jaře chybné.

V letní tabulce 9 můžeme vidět rekordní hodnotu korelace. Nastává pochopitelně v závislosti na rychlosti větru pro kraj Vysočina. Z hlediska oblačnosti je vhodné

Tabulka 9: Korelace pro roční období léto

Léto		m/s		Teplota	
Oblast měření	Kraj	Korelace	Kraj	Korelace	
E.ON. Východ	Vysočina	0,8866	Vysočina	-0,2436	
SME - ČEZ Morava	Zlínský	0,6783	Olomoucký	0,1194	
SCE - ČEZ Sever	Ústecký	0,7078	Ústecký	-0,4529	
STE - ČEZ Střed	Středočeský	0,3076	Středočeský	-0,1552	
VCE - ČEZ Východ	Vysočina	0,5749	Pardubický	0,2598	

upozornit na Královehradecký kraj, ovšem ve všech ostatních krajích byly pro SCE - ČEZ Sever korelace abnormálně vysoké.

Tabulka 10: Korelace pro roční období podzim

Podzim		m/s		Teplota	
Oblast měření	Kraj	Korelace	Kraj	Korelace	
E.ON. Východ	Vysočina	0,8292	Jihomoravský	0,0677	
SME - ČEZ Morava	Jihomoravský	0,5949	Olomoucký	0,0280	
SCE - ČEZ Sever	Ústecký	0,6668	Liberecký	-0,1250	
STE - ČEZ Střed	Praha	-0,3447	Praha	-0,3208	
VCE - ČEZ Východ	Vysočina	0,4541	Královehradecký	-0,4447	

V následující tabulce 10 pro podzim pozorujeme záporné hodnoty v oblasti STE - ČEZ Střed z hlediska rychlosti větru, z nichž největší platí pro kraj Praha. Tuto skutečnost lze přisoudit opět k chybě v měření. K zajímavosti dochází také pro oblast SCE - ČEZ Sever, kde velmi vysoké korelaci dochází pro Jihočeský kraj (0.8208), ale z geografických důvodů v tabulce není uveden.

Tabulka 11: Korelace pro roční období zima

Zima		m/s		Teplota	
Oblast měření	Kraj	Korelace	Kraj	Korelace	
E.ON. Východ	Vysočina	0,7951	Vysočina	0,2840	
SME - ČEZ Morava	Zlínský	0,6542	Zlínský	-0,2370	
SCE - ČEZ Sever	Středočeský	0,6988	Ústecký	0,3798	
STE - ČEZ Střed	Středočeský	0,6308	Praha	0,5042	
VCE - ČEZ Východ	Královehradecký	0,2868	Pardubický	0,2672	

V tabulce 11 pro zimu pozorujeme, podobně jako v případě lokalit, i zde kladné a poměrně vysoké hodnoty korelací v aspektu teplota ovzduší. Lze to opět přisoudit vyššímu počtu hodnot teploty pod bodem mrazu, tedy se záporným znaménkem.

Zhodnocení

Následující tabulka 12 určuje vždy nejvhodnější kraj pro danou oblast měření. Kraj, jehož předpověď vykazovala nejvyšší korelaci s vyrobenou energií byl kraj Vysočina, který byl nejvhodnější i pro oblast VCE - ČEZ Východ. Při analýze ovšem bylo opět nutné zohledňovat geografické rozložení krajů.

Tabulka 12: Nejvhodnější kraje měření pro danou oblast

Oblast měření	Kraj
E.ON. Východ	Vysočina
SME - ČEZ Morava	Jihomoravský, Zlínský
SCE - ČEZ Sever	Ústecký
STE - ČEZ Střed	Praha, Středočeský
VCE - ČEZ Východ	Vysočina, Královehradecký

5.2 Statistická analýza

V této kapitole se nacházejí výsledky statistických analýz. Pro porovnání jsou rozděleny do třech kategorií a to pro jednotlivé kalendářní roky roční období a měsíce.

5.2.1 Vyrobená energie v jednotlivých rocích

Následující statistiky udávají střední hodnoty a směrodatné odchylky vyrobeného činného výkonu v MW odděleně pro roky 2011 a 2012.

Rok 2011		
Oblast měření	Střední hodnota [MW]	Směrodatná odchylka [MW]
E.ON. Východ	5,1367	4,5412
SME - ČEZ Morava	6,6902	6,2570
SCE - ČEZ Sever	12,9163	12,3230
STE - ČEZ Střed	1,7192	1,1362
VCE - ČEZ Východ	1,2060	0,4897
ČR	23,3136	19,2598

Tabulka 13: Statistiky 2011

Rok 2012		
Oblast měření	Střední hodnota [MW]	Směrodatná odchylka [MW]
E.ON. Východ	5,7269	4,9137
SME - ČEZ Morava	8,5147	7,3869
SCE - ČEZ Sever	8,8075	14,8900
STE - ČEZ Střed	0,9534	1,3886
VCE - ČEZ Východ	0,5737	0,5633
ČR	14,3213	16,6332

Tabulka 14: Statistiky 2012

Při porovnání tabulek 13 a 14 sledujeme, že pro oblast E.ON. Východ hodnoty naměřených výkonů za rok 2011 nabývají řádově stejných hodnot jako v roce následujícím. Hodnoty výkonů se pohybují nejčastěji okolo 5.5 MW.

Následuje oblast SME - ČEZ Morava, kde lze pozorovat mírný rozdíl mezi měřenými roky, který činí přibližně 2 MW. Nejvyšší hodnoty výkonů jsou zde v řádu 28 MW.

V další oblasti SCE - ČEZ Sever dosahují hodnoty hranice až 40 MW, ovšem zde pozorujeme poměrně vysokou směrodatnou odchylku, a to zejména v roce 2012. Od druhé čtvrtiny tohoto roku jsou hodnoty výkonu buď nulové nebo blízké nule. Toto bude blíže rozebráno z pohledu ročních období a měsíců.

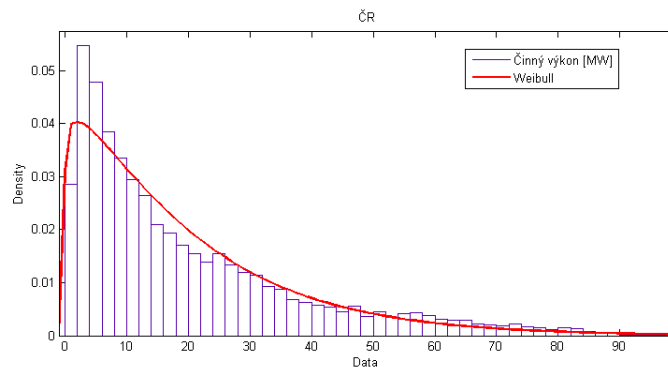
Naměřená data z oblastí STE - ČEZ Střed a VCE - ČEZ Východ již nejsou moc kvalitní. Značné množství hodnot bylo odfiltrováno (konstantní a blízké nule). Jak napovídá střední hodnota, nedochází zde k nijak vysokým naměřeným výkonům porovnatelným s ostatními oblastmi.

Co se týče hodinových dat měřených pro celou Českou Republiku, dochází zde ke značným rozdílům jejich střední hodnoty. Pokles z 23 MW na 14 MW lze vysvětlit patrně změnou instalovaného výkonu.

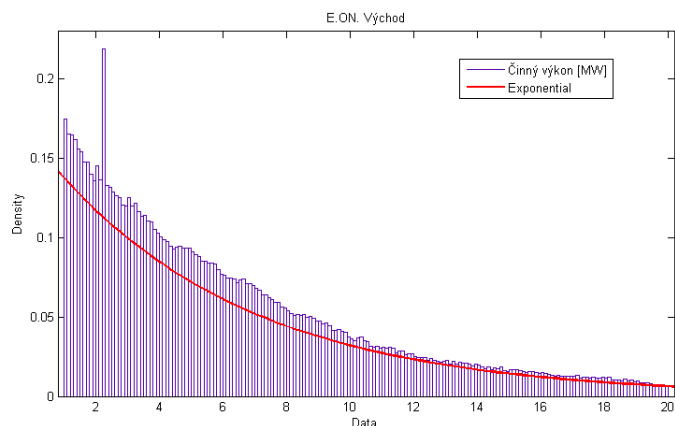
Histogramy

V této práci jsou uvedeny histogramy v porovnání s Weibullovým rozdělením, které nejlépe aproximuje naměřená data ČR a Exponenciálním rozdělením pro data E.ON. Východ. Obrázky byly vytvořeny pomocí nástroje *dfittool.m*, který pracuje s normovaným histogramem, aby šel porovnat s hustotou pravděpodobnosti.

Zde jsou uvedeny histogramy pro celou Českou republiku a oblast E.ON. Východ, které vycházejí podle očekávání. Histogramy pro jednotlivá roční období měření ČR se nacházejí v příloze na obrázku 10. Z tohoto obrázku pozorujeme výrazně vyšší četnost vysokých hodnot naměřeného výkonu u zimního histogramu než v ostatních ročních obdobích. Tato skutečnost bude potvrzena denními diagramy.



Obrázek 5: Histogram ČR pro celé dva roky



Obrázek 6: Histogram oblasti E.ON. Východ pro celé dva roky

Denní diagramy

Denní diagram je typický denní průběh sledovaného signálu. V našem případě budeme zobrazovat průběh vyrobené větrné energie v průběhu dne. K získání denních diagramů je v této práci použita střední hodnota a platí následující vzorec:

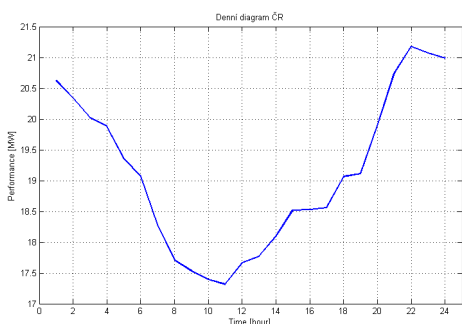
$$DD(i) = \frac{1}{PD} \sum_{j=1}^{PD} x(i + (j-1) \cdot P),$$

kde $i = 1, \dots, P$, $P = PM$, resp. $P = PH$ je počet minut, resp. počet hodin ($PM = 1440$ minut, $PH = 24$ hodin), PD je počet dní ($PD = 365$ dní).

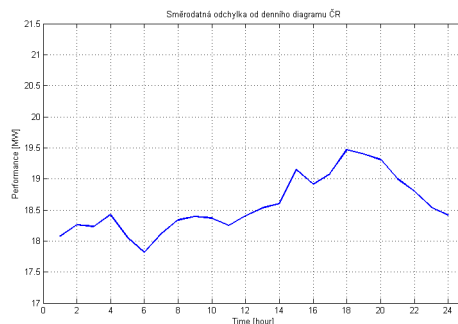
Dále jsou uváděny grafy směrodatné odchylky od denního diagramu, pro které platí vzorec:

$$\sigma_{DD}(i) = \sqrt{\frac{1}{PD-1} \sum_{j=1}^{PD} [x(i + (j-1) \cdot P) - DD(i)]^2}$$

kde $i = 1, \dots, P$, $P = PM$, resp. $P = PH$ je počet minut, resp. počet hodin ($PM = 1440$ minut, $PH = 24$ hodin), PD je počet dní ($PD = 365$ dní).



(a) Denní diagram



(b) Směrodatná odchylka od denního diagramu

Obrázek 7: Diagramy za roky 2011 - 2012 pro celou ČR

Z obrázku 7(a) můžeme vidět vývoj vyráběné energie během dne. K nejvyšší produkci dochází v nočních a večerních hodinách, kdy zpravidla více fouká. Naproti tomu v poledním čase jsou hodnoty činného výkonu nejnižší. Maximální rozdíl večerních a poledních hodnot činí přibližně 3.8 MW, což vzhledem k tomu, že instalovaný výkon je 93 MW, není velký rozdíl (cca 4%). Na obrázku 7(b) vidíme, že k největšímu kolísání hodnot dochází okolo šesté hodiny večerní, kdy směrodatná odchylka je až 19.5 MW. Rozdíl mezi minimální a maximální směrodatnou odchylkou jsou necelé 2 MW (tj. necelé 2% instalovaného výkonu).

5.2.2 Vyrobená energie v ročních obdobích

V následující části budou uvedeny statistické tabulky pro roční období sloučená za oba roky. Hlavní pozornost je zde věnována datům z oblasti E.ON. Východ, která se jeví jako nejlepší. Kompletní tabulky a jednotlivá roční období pro každý rok zvlášť jsou obsahem příloženého CD.

Jaro 2011-2012		
Oblast měření	Střední hodnota [MW]	Směrodatná odchylka [MW]
E.ON. Východ	5,5914	4,8920
ČR	17,4715	15,4785

Tabulka 15: Jaro

Léto 2011-2012		
Oblast měření	Střední hodnota [MW]	Směrodatná odchylka [MW]
E.ON. Východ	4,0911	3,7518
ČR	13,5703	14,8347

Tabulka 16: Léto

Podzim 2011-2012		
Oblast měření	Střední hodnota [MW]	Směrodatná odchylka [MW]
E.ON. Východ	4,8994	4,2599
ČR	15,6629	16,3287

Tabulka 17: Podzim

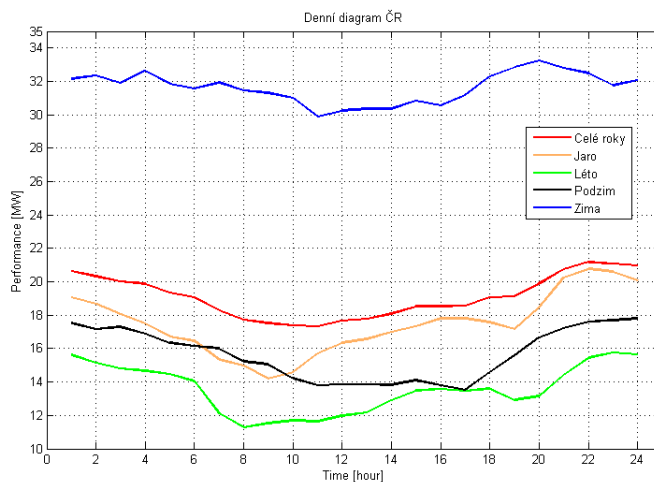
Z uvedených tabulek vidíme, že v létě (Tab. 16) v porovnání s jarem (Tab. 15) dochází k poklesu vyrobené energie. Naopak na podzim (Tab. 17) pozorujeme nárůst a v zimě (Tab. 18) nastává nejvyšší produkce větrných elektráren. Jak již bylo uvedeno, se snižující se teplotou roste rychlost a síla větru. Tomuto tvrzení odpovídá i zmíněná analýza.

Zima 2011-2012		
Oblast měření	Střední hodnota [MW]	Směrodatná odchylka [MW]
E.ON. Východ	6,9964	5,2393
ČR	31,6331	22,5976

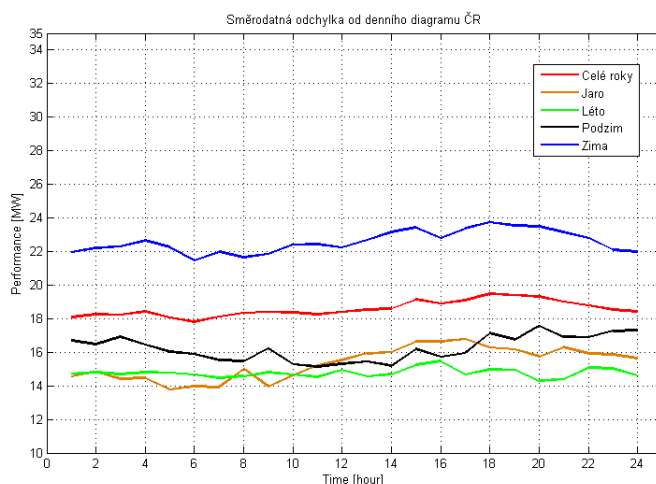
Tabulka 18: Zima

Denní diagramy

Na následujícím obrázku 8 vidíme denní diagramy čtyř ročních období měřených pro celou Českou republiku. Pro porovnání je uveden i diagram celkový. Jak již se dříve ukázalo, vyráběná energie v dopoledních hodinách klesá a během odpoledních roste. Dále můžeme dobře pozorovat, že průměr vyráběné energie za oba roky výrazně zvedají zimní měsíce, při kterých se hodnota vyráběné energie pohybuje okolo 32 MW. Ve zbytku roku se hodnoty výkonu vyskytují okolo hranice 16 MW, pouze v jarním období se ve večerních hodinách zvedají k 20 MW. Rovněž lze tvrdit, že činný výkon se příliš nemění v rámci jednoho dne, ale ke změnám dochází v závislosti na ročním období. Na obrázku 9, podle očekávání vidíme, že k největším odchylkám o denního diagramu dochází v zimě (22 MW), což činí o 4 MW více než je průměrná hodnota za celé roky. Pro úplnost se v příloze nachází diagramy online měření (Obr. 11) a pro oblast E.ON. Východ (Obr. 12), které mají očekávaný vývoj.



Obrázek 8: Denní diagram ČR



Obrázek 9: Směrodatná odchylka od denního diagramu ČR

5.2.3 Vyrobená energie v jednotlivých měsících

Zde jsou uvedeny opět pouze tabulky pro oblast EON Východ a pro Českou Republiku za sdružené roky 2011-2012. Rozměrově náročné kompletní tabulky se nachází na příloženém CD.

Střední hodnota 2011 - 2012		
	E.ON. Východ [MW]	ČR [MW]
Leden	7,3732	35,8443
Únor	6,6713	21,9652
Březen	5,8481	18,8366
Duben	6,1469	19,3865
Květen	4,7948	14,2379
Červen	4,5146	14,6754
Červenec	4,2744	16,5594
Srpen	3,4705	9,4238
Září	3,9182	11,5683
Říjen	5,7024	16,9613
Listopad	5,1889	19,8934
Prosinec	6,8571	40,6466

Tabulka 19: Střední hodnota

V následující tabulce 19 se pouze potvrzuje, skutečnost zjištěná z analýzy ročních období, tedy v chladných měsících je výroba energie nejvyšší a postupem roku se snižuje. Výjimku tvoří pouze Červenec pro ČR, který svými hodnotami spadá spíše do podzimních měsíců. V případě oblasti E.ON. východ tomu ovšem tak není. Tuto nesrovnalost lze přisoudit buď tomu, že někde hodně nárazově zafoukalo a došlo k vysoké výrobě energie, nebo nastala chyba měření. V příloze na obrázku 13 a 14 můžeme vidět denní diagramy pro jednotlivé měsíce. Podle očekávání se nejvíce

energie vyrábí v zimních měsících, konkrétně v prosinci a lednu.

5.2.4 Zhodnocení

Z výše uvedených statistik a denních diagramů lze učinit závěr, že z celoročního pohledu mají největší podíl na vyráběné energii zimní měsíce a to zejména leden a prosinec, kdy nejvíce fouká. K největší proměnlivosti naměřených hodnot dochází z hlediska ročních období. V rámci dní se vyráběný výkon příliš nemění, jak dokazují vytvořené denní diagramy.

6 Závěr

Cílem této práce bylo vytvoření filtrů pro čištění dat, která obsahovala naměřený činný výkon z větrných farem a následně provést datovou analýzu. Kapitola 4 popisuje vytvoření filtrů pro odstranění příliš malých hodnot, konstantních hodnot s danou dobou trvání a nesmyslně vysokých hodnot. Tyto nežádoucí hodnoty byly z důvodu zachování ekvidistance dat nahrazovány podle potřeby nulou nebo hodnotou NaN. Další fází byla náhrada NaN hodnot trvajících maximální požadovanou dobu pomocí lineární interpolace. Změny dat způsobené filtry a následně interpolacemi byly vždy zaznamenány do textových souborů. Jako nejkvalitnější naměřená data bylo možno určit E.ON. Východ, kde bylo odfiltrováno nejméně dat. V ostatních oblastech byl pozorován rapidní nárůst NaN hodnot. Hlavním problémem těchto dat, který řeší filtrace jsou jejich konstantní hodnoty.

V kapitole 5 byla řešena korelační analýza, kde nejprve byla vybírána nejlepší lokalita pro danou oblast měření. Rozhodnutí bylo učiněno na základě nejvyšší hodnoty korelace naměřeného výkonu v oblasti s rychlostí větru a teplotou ovzduší pro tu kterou lokalitu. Výsledky této analýzy jsou uvedeny v tabulce 6, kde jako nejlepší vycházely lokality Náměšť nad Oslavou, Červená u Libavé, Svratouch a Milešovka. Dále byla podobným způsobem posuzována závislost vyrobené energie na předpovědích počasí pro kraje České republiky, kde vycházel jako nejlepší pro oblast E.ON. Východ kraj Vysočina. V tomto kraji se rovněž nachází nejvhodnější lokalita pro E.ON. Východ Náměšť nad Oslavou.

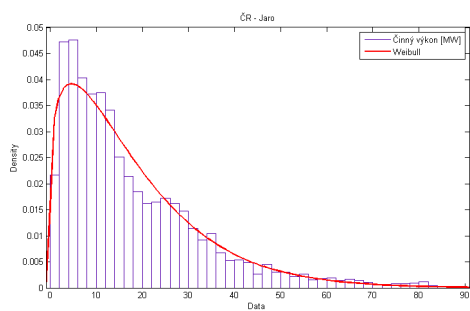
Poslední část práce se týkala statistické analýzy naměřených a zpracovaných dat. Data byla analyzována pomocí středních hodnot, směrodatných odchylek, histogramů a denních diagramů naměřených dat a bylo dosaženo závěru, že z celoročního pohledu mají největší podíl na vyráběné energii zimní měsíce a to zejména leden a prosinec, kdy nejvíce fouká. K největší proměnlivosti naměřených hodnot dochází z hlediska ročních období. V rámci dne se vyráběný výkon příliš nemění, jak dokazují vytvořené denní diagramy.

7 Literatura

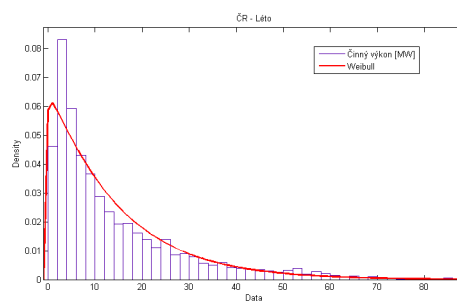
Reference

- [1] J. Reif, Z. Kobeda, *Úvod do pravděpodobnosti a spolehlivosti*, FAV ZČU, Plzeň, 2006
- [2] L. Cyhelský, E. Souček, *Základy statistiky*, VŠFS, Plzeň, 2009
- [3] J. Anděl, *Základy matematické statistiky*, MATFYZPRESS UK, Praha, 2005
- [4] E. Meijering, *A Chronology of Interpolation. From Ancient Astronomy to Modern Signal and Image Processing.*, 2002
- [5] I. Hochmannová, *Matematický model vyrobené větrné energie*, Diplomová práce, FAV ZČU, Plzeň, 2010
- [6] Internetové stránky společnosti ČEZ, a.s., 2013 Dostupných z WWW:
<http://www.cez.cz/cs/vyroba-elektriny/obnovitelne-zdroje/vitr/flash-model-jak-funguje-vetrna-elektrarna.html>
- [7] Internetové stránky společnosti ČEPS, a.s., Dostupných z WWW:
<http://www.ceps.cz/CZE/DATA/VSECHNA-DATA/Stranky/Vyroba.aspx>

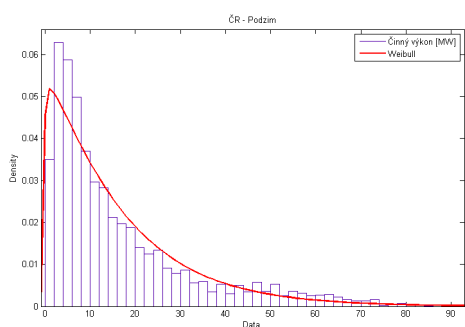
8 Příloha - Histogramy



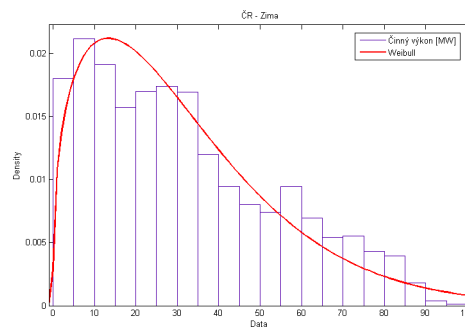
(a) Jaro



(b) Léto



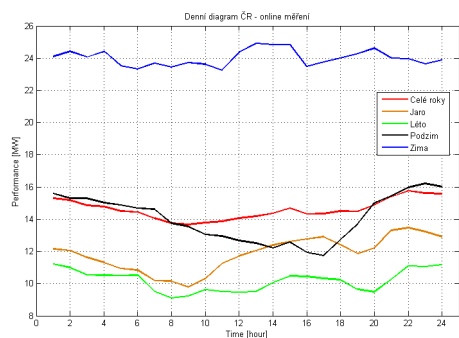
(c) Podzim



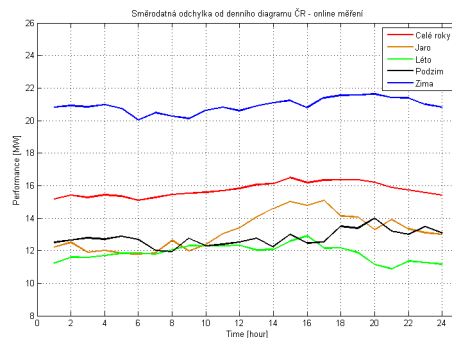
(d) Zima

Obrázek 10: Histogramy ročních období pro celou ČR

9 Příloha - Denní diagramy

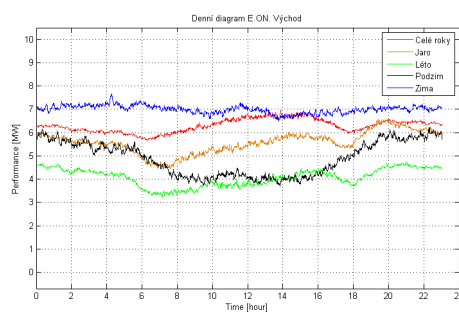


(a) Denní diagram

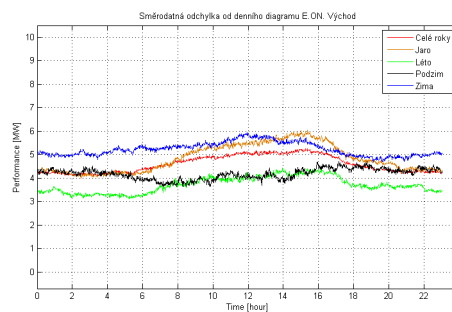


(b) Směrodatná odchylka od denního diagramu

Obrázek 11: Diagramy ročních období pro celou ČR

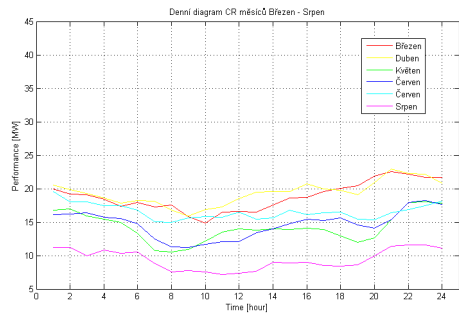


(a) Denní diagram

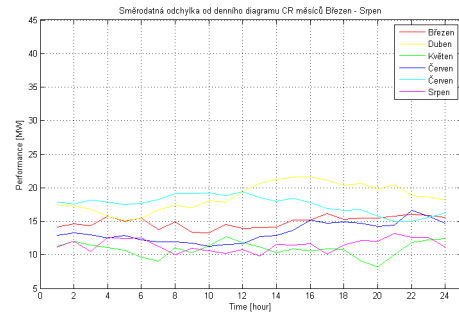


(b) Směrodatná odchylka od denního diagramu

Obrázek 12: Diagramy ročních období pro oblast E.ON. Východ

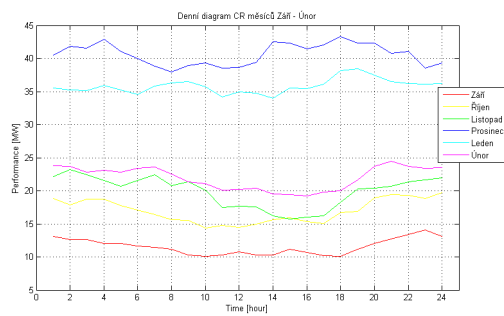


(a) Denní diagram

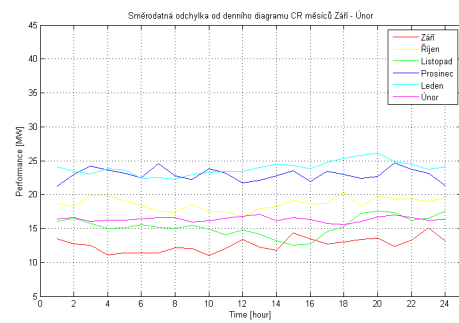


(b) Směrodatná odchylka od denního diagramu

Obrázek 13: Březen - Srpen



(a) Denní diagram



(b) Směrodatná odchylka od denního diagramu

Obrázek 14: Září - Únor

10 Příloha - Algoritmy

Algoritmus 6: Filtr na odstranění malých hodnot

```

1 if instal_vykon(i)>0,
2     pom = pom/instal_vykon(i)*100;
3     index0 = find(pom < 1);
4     opravovane_hodnoty = ...
5         VTE_skutecnost.signals.values(index0,i);
6     VTE_skutecnost.signals.values(index0,i) = 0;
7
8     y = [index0'; opravovane_hodnoty'];
9     nazev = VTE_skutecnost.hodnoty{i};
10    file = ['vypisy_oprav\' zpracovane_VTE_skutecnost ...
11            \'_\' int2str(rok) \'_Odstranění malých hodnot_\'...
12            datestr(now,\'yyyy-mm-dd_HH-MM\' ) \'.txt\'];
13    fid = fopen(file, \'a\');
14    fprintf(fid,\'\\n\');
15    fprintf(fid, nazev);
16    fprintf(fid,\'\\n\');
17    fprintf(fid, \'%6.0f %10.4f \\n\', y);
18    status = fclose(fid);
19 end

```

Algoritmus 7: Filtr na odstranění konstantních hodnot

```

1 index = pom(:,1) == pom(:,2) & pom(:,2) == pom(:,3);
2     VTE_skutecnost.signals.values(index,i) = NaN;
3     index(end-2:end,1) = 0;
4     x = find(index == 1);
5     VTE_skutecnost.signals.values(x+1,i) = NaN;
6     VTE_skutecnost.signals.values(x+2,i) = NaN;
7
8     indexKonst = find((isnan(VTE_skutecnost.signals.values(:,i))...
9                        ==1)&(isnan(old(:,i))==0));
10    opravovane_hodnoty = old(indexKonst,i);
11    y = [indexKonst'; opravovane_hodnoty'];
12    nazev = VTE_skutecnost.hodnoty{i};
13    file = ['vypisy_oprav\' zpracovane_VTE_skutecnost \'_\'...
14            int2str(rok) \'_Odstranění konstantních hodnot_\'...
15            datestr(now,\'yyyy-mm-dd_HH-MM\' ) \'.txt\'];
16    fid = fopen(file, \'a\');
17    fprintf(fid,\'\\n\');
18    fprintf(fid, nazev);
19    fprintf(fid,\'\\n\');
20    fprintf(fid, \'%6.0f %10.4f \\n\', y);
21    status = fclose(fid);

```

Algoritmus 8: Výpisy

```

1
2 nazev = data.hodnoty{i};
3 fprintf('%s', '-----', nazev, '-----');
4 fprintf('\n');
5
6 if isnan(data.signals.values(:,i))==1;
7     fprintf('%s', 'Naměřená data ', nazev, ' chybí. ');
8     fprintf('\n');
9 else
10    if isnan(data.signals.values(:,i))==0;
11        fprintf('%s', 'V měření ', nazev, ...
12            ' žádná data nechybí. ');
13        fprintf('\n');
14    else
15        je_nan = find(isnan(data.signals.values(:,i)) == 1);
16        zacatek = [je_nan(1); je_nan(find ...
17            ((je_nan(1:end-1)+1)~=je_nan(2:end))+1)];
18        konec = [je_nan(find((je_nan(1:end-1)+1)...
19            ~=je_nan(2:end))) je_nan(end)];
20        delky = konec - zacatek + 1;
21        maximal = int2str(max(delky));
22        [pocet,n] = size(delky);
23        prumer_delka = int2str(round((sum(delky)/pocet)));
24        fprintf('%s', ...
25            'Průměrná délka úseku s trvalou hodnotou NaN pro ', ...
26            , nazev, ' je ', prumer_delka, ' min. ');
27        fprintf('\n');
28        fprintf('%s', 'Maximální NaN úseku: ', maximal, ' min. ');
29        fprintf('\n');
30    end
31
32    je_zero = find(data.signals.values(:,i) <= 0.01);
33    if isempty(je_zero) == 0;
34        zacatek0 = [je_zero(1); je_zero(find...
35            ((je_zero(1:end-1)+1)~= je_zero(2:end))+1)];
36        konec0 = [je_zero(find((je_zero(1:end-1)+1)~=...
37            je_zero(2:end))); je_zero(end)];
38        delky0 = konec0 - zacatek0 + 1;
39        [pocet0,n0] = size(delky0);
40        prumer_delka0 = int2str(round((sum(delky0)/pocet0)));
41        fprintf('%s', ...
42            'Průměrná délka úseku s trvalou hodnotou menší než', ...
43            ' 0.01 pro ', nazev, ' je ', prumer_delka0, ' min. ');
44        fprintf('\n');
45    else
46        fprintf('%s', 'Data ', nazev...
47            , ' neobsahují žádnou hodnotu menší než 0.01. ');
48        fprintf('\n');
49    end
50
51 end

```