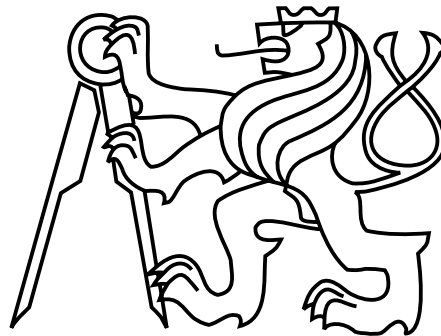


České vysoké učení technické v Praze

Fakulta strojní

Ústav přístrojové a řídicí techniky



Diplomová práce

**Detekce oscilací v signálech získaných ze  
systémů techniky prostředí budov**

*Bc. Michal Kuchař*

Vedoucí práce: Ing. Cyril Oswald, Ph.D.

Studijní program: Strojní inženýrství

Obor: Přístrojová a řídicí technika

11. června 2019

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Kuchař** Jméno: **Michal** Osobní číslo: **419774**  
Fakulta/ústav: **Fakulta strojní**  
Zadávající katedra/ústav: **Ústav přístrojové a řídicí techniky**  
Studijní program: **Strojní inženýrství**  
Studijní obor: **Přístrojová a řídicí technika**

## II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

**Detekce oscilací v signálech získaných ze systémů techniky prostředí budov**

Název diplomové práce anglicky:

**Oscillations detection in data obtained from HVAC systems**

Pokyny pro vypracování:

- 1) Proveďte literární rešerši s cílem identifikovat možné nežádoucí jevy v systémech řízení techniky prostředí, které se projevují typickými oscilacemi v měřených či řídicích signálech.
- 2) Proveďte literární rešerši s cílem vytypovat vhodné matematické nástroje pro detekci oscilací binárních i spojitých časových řad.
- 3) Implementujte vytypované matematické metody v podobě algoritmů.
- 4) Ověřte vyvinuté algoritmy na dodaných datech.

Seznam doporučené literatury:

- Tomáš Hobza. \u{Matematická statistika}. FJFI ČVUT Praha, 2011.
- Gejza Dohnal. \u{Základy stochastiky}. FS ČVUT Praha, 2001.
- Alois Klíč, Karel Volka, Miroslava Dubcová. \u{Fourierova transformace}. VŠCHT Praha, 2002
- Turner, W. J. N., Staino, A., Basu, B. \u{Residential HVAC fault detection using a system identification approach}. Energy and Buildings, 151, 1-17. 2017

Jméno a pracoviště vedoucí(ho) diplomové práce:

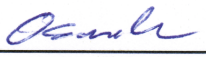
**Ing. Cyril Oswald, Ph.D., U12110.3**

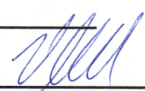
Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

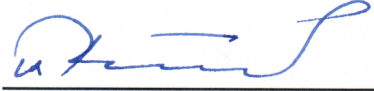
Datum zadání diplomové práce: **26.04.2019**

Termín odevzdání diplomové práce: **12.06.2019**

Platnost zadání diplomové práce: \_\_\_\_\_

  
Ing. Cyril Oswald, Ph.D.  
podpis vedoucí(ho) práce

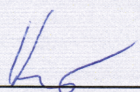
  
podpis vedoucí(ho) ústavu/katedry

  
prof. Ing. Michael Valášek, DrSc.  
podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

26.4 2019  
Datum převzetí zadání

  
Podpis studenta

# Poděkování

Děkuji vedoucímu práce za připomínky během tvorby. Dále děkuji rodině a přátelům za podporu při studiu.

# Prohlášení

Prohlašuji, že jsem práci vypracoval samostatně a použil jsem pouze podklady uvedené v příloženém seznamu.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne:

.....



## **Abstract**

This thesis deals with data analysis and fault detection of heating, ventilating, air-conditioning (HVAC) system. Key phenomenon for evaluating a fault are oscillations in data. This work analyzes binary data oscillation as well as discretely measured data of continuous quantity such as temperature or power. Main step is a data transformation from time domain into a frequency domain using fast Fourier transformation algorithm. These transformations are analyzed with tools like cluster analysis and Kohonen self-organizing map.

## **Keywords**

HVAC, fault detection, data oscillations, Fourier transform, cluster analysis, self-organizing map

## **Abstrakt**

Diplomová práce se zabývá analýzou dat a detekcí poruch ze systému techniky prostředí budov. Klíčovým jevem pro vyhodnocení poruchy jsou oscilace v datech. Práce analyzuje oscilace binárních dat i diskrétně naměřených dat spojitých veličin, jako je teplota, nebo výkon. Hlavním krokem je převedení dat z časové oblasti do oblasti frekvenční za pomoci algoritmu rychlé Fourierovy transformace. Tyto transformace jsou nadále analyzovány nástroji jako je shluková analýza a Kohonenova samoorganizační mapa.

## **Klíčová slova**

Technika prostředí, detekce poruch, oscilace v datech, Fourierova transformace, shluková analýza, samoorganizační mapa

# Obsah

<b>1</b>	<b>Úvod</b>	<b>12</b>
<b>2</b>	<b>Technika prostředí budov</b>	<b>13</b>
2.1	Poruchy v systémech techniky prostředí budov . . . . .	13
<b>3</b>	<b>Nástroje analýzy dat a detekce poruch</b>	<b>14</b>
3.1	Amplitudy v datech . . . . .	17
3.2	Změny v datech . . . . .	17
3.3	Fourierova transformace . . . . .	18
3.3.1	Rychlá Fourierova transformace . . . . .	19
3.4	Korelace . . . . .	20
3.4.1	Korelace histogramů . . . . .	21
3.5	Shluková analýza pomocí k-means . . . . .	22
3.6	Metoda hlavních komponent . . . . .	24
3.7	Kohonenova samoorganizační mapa . . . . .	27
<b>4</b>	<b>Dodaná data a jejich předzpracování</b>	<b>30</b>
4.1	Předzpracování boolovských dat . . . . .	31
4.2	Předzpracování dat typu double . . . . .	32
<b>5</b>	<b>Aplikace vybraných metod</b>	<b>33</b>
5.1	Aplikace FFT . . . . .	33
5.1.1	Filtr frekvencí . . . . .	34
5.2	Filtr amplitud . . . . .	35
5.3	Filtr rychlých změn . . . . .	36
5.4	Aplikace PCA . . . . .	36
5.5	Aplikace k-means . . . . .	39
5.6	Aplikace SOM . . . . .	40
5.7	Aplikace korelace . . . . .	43
<b>6</b>	<b>Návrh algoritmu</b>	<b>44</b>

6.1	Výpočet vlastností signálu . . . . .	44
6.2	Detekce oscilací v boolovských datech . . . . .	45
6.3	Detekce oscilací v datech typu double . . . . .	46
<b>7</b>	<b>Výsledky navrženého algoritmu na dodaných datech</b>	<b>48</b>
7.1	Výsledky boolovských dat . . . . .	49
7.2	Výsledky dat typu double . . . . .	51
<b>8</b>	<b>Závěr</b>	<b>57</b>
	<b>Literatura</b>	<b>59</b>
<b>A</b>	<b>Příloha zbylých výsledků</b>	<b>60</b>

## Seznam obrázků

1	Schéma detekce poruch na základě známého modelu [6]. . . . .	14
2	Schéma detekce poruch na základě vlastností signálu bez znalosti modelu [6].	15
3	Schéma obecné metodiky detekce poruch [6]. . . . .	17
4	Příklad aplikace Fourierovy transformace. V horní části se nachází originální signál v časové oblasti, v dolní části jeho transformace do frekvenční oblasti.	19
5	Znázornění významu korelace [10]. . . . .	21
6	Průnik histogramů [10]. . . . .	22
7	Euklidovská vzdálenost je zobrazena jako kružnice, Čebyševova vzdálenost jako největší čtverec a manhattanská vzdálenost jako vnitřní natočený čtverec. [11] . . . . .	23
8	Ukázka použití k-means na dvou-dimenzionálním příkladu - rozdělení na 4 shluky. [11] . . . . .	24
9	Modelová data pro aplikaci metody hlavních komponent. . . . .	26
10	Vizualizace hlavních komponent na modelových datech. . . . .	26
11	Použití metody hlavních komponent pro snížení rozměru dat. . . . .	27
12	Čtvercová topologie SOM. . . . .	28
13	Šestiúhelníková topologie SOM. . . . .	29
14	Limit frekvencí ve frekvenční charakteristice. . . . .	34
15	Limit prahových hodnot signálu. Práh určen na základě zákona o pracovním prostředí. [13] . . . . .	35
16	Limit rychlostí . . . . .	36
17	Aplikace metody hlavních komponent . . . . .	37
18	Rozptyly jednotlivých hlavních komponent. . . . .	38
19	Aplikace k-means a metody hlavních komponent. . . . .	40
20	Aplikace SOM - četnost vítězství. . . . .	42
21	Aplikace SOM - vzdálenosti neuronů. Černá značí největší vzdálenost, žlutá nejmenší. . . . .	43
22	Ukázka dvou histogramů poruchy s podobnou charakteristikou (první dva od shora) a histogramu normální charakteristiky (dole). . . . .	44

23	Schéma algoritmu detekce a klasifikace poruch. . . . .	47
24	Algoritmus vyhodnocení podezřelých dat. . . . .	48
25	Znázornění průměrného počtu změn jednotlivých čerpadel. . . . .	49
26	Znázornění počtů změn jednotlivých čerpadel v říjnu. Červeně - průměr, zeleně - směrodatná odchylka. . . . .	50
27	Ukázka výsledků SOM z prosince - vzdálenosti mezi neurony. . . . .	55
28	Ukázka výsledků SOM z prosince - vítězství neuronů. . . . .	55
29	Ukázka výsledků shlukové analýzy vizualizované pomocí metody hlavních kom- ponent z prosince. Normální data červeně, podezřelá modře. . . . .	56

## Seznam tabulek

1	Příklady poruchy v systémech techniky prostředí budov [4] . . . . .	13
2	Vybrané vlastnosti signálu [7]. . . . .	16
3	Dodaná data . . . . .	30
4	Proměnné v dodaných datech v lednu . . . . .	31
5	Vstupy a výstupy funkce <i>pca</i> . . . . .	37
6	Vstupy a výstupy funkce <i>kmeans</i> . . . . .	39
7	Vstupy a výstupy funkce <i>selfogram</i> [15] . . . . .	41
8	Parametry normálního chování . . . . .	45
9	Výsledky kontroly funkce čerpadel . . . . .	51
10	Parametry signálu . . . . .	52
11	Výsledky kontroly vnitřních teplot . . . . .	53
12	Výsledky klasifikace . . . . .	54
13	Průměrný počet poruch u ostatních veličin. . . . .	56

## Seznam symbolů

$\alpha(k)$  Součinitel velikosti skoků při učení neuronové sítě závislý na iteraci  $k$

$\omega$  Vlastní vektor kovarianční matice

$\Sigma$  Matice kovariancí

**FFT** Matice Fourierových transformací zkoumaných dat

**K** Vektor hlavní komponenty

**u** Vektor příslušnosti k centroidům

$w_c$  Vektor počátečních vah

$w_s$  Vektor vah sousedního neuronu

$w_v$  Vektor vah vítězného neuronu

**X** Vektor vstupních dat obecně

**Z** Vektor počtu změn v boolovských datech

$\eta(v, k, s)$  Funkce sousedství

$\Omega(k)$  Velikost okolí vítězného neuronu v závislosti na iteraci  $k$

$\sigma_X$  Směrodatná odchylka náhodné veličiny  $X$

$C_{XY}$  Kovariance náhodných veličin  $X$  a  $Y$

$E(X)$  Operátor střední hodnoty náhodné veličiny  $X$

$f$  frekvence [Hz]

$f_s$  Vzorkovací frekvence [Hz]

$f_{KM}$  Funkce k-means



$H$	Míra průniku histogramů
$I$	Histogram k porovnání
$M$	Vzorový histogram
$N$	Počet vzorků
$O(N)$	Operátor složitosti $N$ vstupů
$r_{XY}$	Korelační koeficient náhodných veličin $X$ a $Y$
$s$	Pozice sousedního neuronu
$S(f)$	Fourierův obraz do frekvenční oblasti $f$
$s(t)$	Obecná funkce v časové oblasti $t$
$t$	Čas [s]
$v$	Pozice vítězného neuronu

# 1 Úvod

Systémy techniky prostředí tvoří až 20 % celkové energetické spotřeby v Evropě [1]. Poruchy v systémech techniky prostředí spotřebují 1-2,5 % energie [2]. Z těchto faktů vychází najevo, že odstranění poruch v systémech techniky prostředí budov má velký potenciál v úspoře energie, peněz a ve zmírnění negativních ekologických dopadů. Mimo jiné mohou mít poruchy v takových systémech vliv na zdraví lidí uvnitř těchto budov [3]. Práce se zaměřuje především na detekci oscilací v těchto systémech jako na potenciální oblast detekce poruch.

Pro detekci oscilací je tedy nutné nastudovat a aplikovat takové nástroje, jež jsou schopné oscilace popsat. Jako takový nástroj se nabízí především algoritmus rychlé Fourierovy transformace. Nicméně je dále potřeba určit, které oscilace jsou ještě normální a které jsou již poruchové. Jedna z cest je skrze znalost chování systému, a tudíž je možné určit, co už se pohybuje za hranicí systému dle našich znalostí. Druhá z cest je analýza vlastností signálů bez znalosti chování systému. Mezi takové vlastnosti patří průměrná hodnota, minima, maxima, rychlost změn atd.

Na oblast detekce poruch navazuje diagnostika poruch, která vychází především ze symptomů zjištěných při detekci poruch. Proto by bylo výhodné zjištěné poruchy nějakým způsobem klasifikovat, např. zdali se detekované poruchy neshlukují, případně zdali nemají podobné vlastnosti. V jednoduchých případech je možné problém vizualizovat přímo, ale v případě vícerozměrného problému, je nutné použít dalších nástrojů. V tomto případě se jedná o nástroje jako metoda hlavních komponent, shluková analýza a Kohonenova samoorganizační mapa.

V praktické části se zabývám zpracováním a analýzou dat dodaných od společnosti ENERGOCENTRUM PLUS, s.r.o. Jako ve velké části případů měření, je potřeba se zabývat problematikou doplnění chybějících naměřených hodnot, případně opravou některých dat. Po potřebném předzpracování dat je možné aplikovat vybrané metody skrze navržený algoritmus. Výstupem je určení podezřelých pozorování v dodaných datech a určení jejich symptomů.

## 2 Technika prostředí budov

Technika prostředí je samostatná vědní disciplína, která se zabývá především větráním, klimatizací, vytápěním, ochranou ovzduší, snižováním hluku a osvětlením . Z toho plyne, že se jedná o komplexní zařízení a rozsáhlé systémy. Regulovanými veličinami jsou zejména teplota vzduchu, vlhkost vzduchu, tlak vzduchu, proudění vzduchu, koncentrace škodlivin. [3]

Na těchto systémech lze pozorovat velké množství dat, z nichž lze usuzovat, zdali daný systém pracuje správně. Pozorovanými veličinami jsou zejména teploty vzduchu, teploty médií v těchto zařízeních, spotřeba energie, spínání čerpadel a ventilů.

### 2.1 Poruchy v systémech techniky prostředí budov

Jednotlivých poruch v takto rozsáhlých systémech je mnoho, proto uvedu jen nejzákladnější z nich (tab. 1). Velkou částí je mechanické poškození, nebo ucpání v trubkách, zaseknutí ventilů, případně nežádoucí úniky médií. Dalším typem je porucha na straně senzorů. Jednoduchým příkladem je nekalibrovaný offset teploměru, což má vliv jak na tepelný komfort, tak i na spotřebu energie. [4]

Tabulka 1: Příklady poruchy v systémech techniky prostředí budov [4]

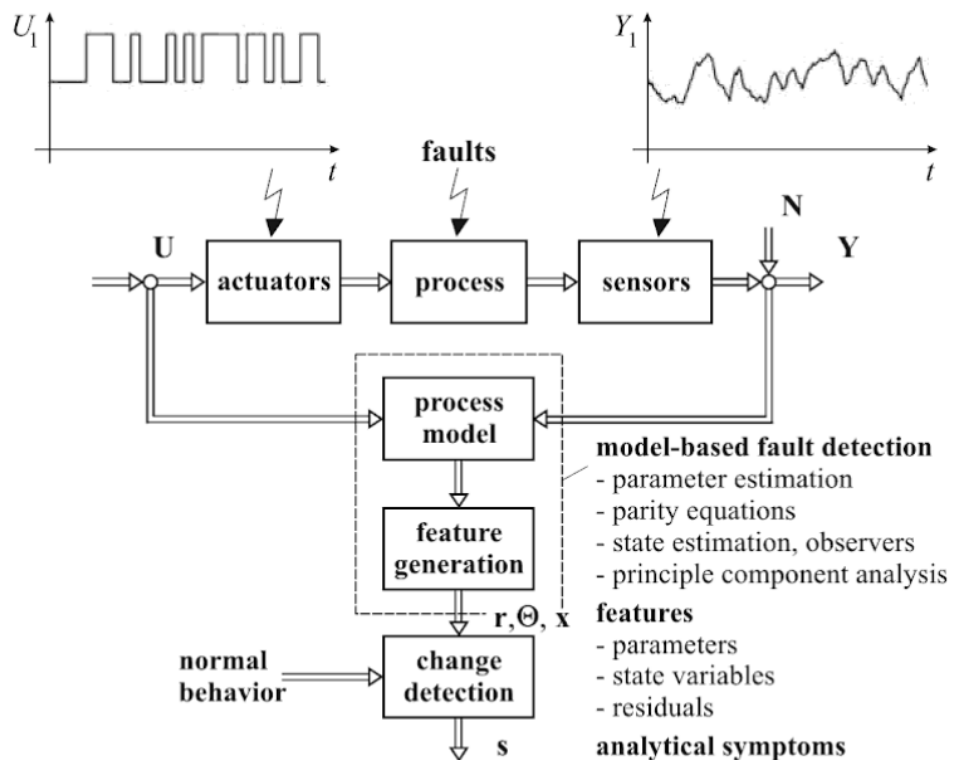
Druh poruchy:	Symptomy:
Bojler - poškozené trubky	Tepelný odpor - snížení efektivity
Zaseknutý ventil	Nadbytečný chod nebo naopak nedostatečný
Poškozené trubky ve výměníku	Topný faktor nižší o více jak 10 %
Zanesený filtr	Pokles tlaku, zvýšená spotřeba energie ventilátorů, menší průtok
Kalibrace senzorů	Tepelný diskomfort nebo např. nižší/vyšší tlak páry při nekalibrovaném senzoru v ohřívači
Poruchy řízení - regulátorů	Oscilace v systému (akčních i regulovaných veličin)

Stěžejním druhem poruch pro tuto práci jsou poruchy způsobující oscilace v systémech

techniky prostředí budov. Základními příčinami jsou nelinearita systému, špatně nastavené řízení systému (regulátorů), špatně dimenzovaná zařízení, případně neočekávané změny v systému. Špatně nastavené řízení může být důsledkem kopírování předchozích projektů do nových, čímž mohou být některé odchylky nových systémů zanedbány a nastavené řízení se snadno dostane mimo lineární oblast pracovního bodu. [5]

### 3 Nástroje analýzy dat a detekce poruch

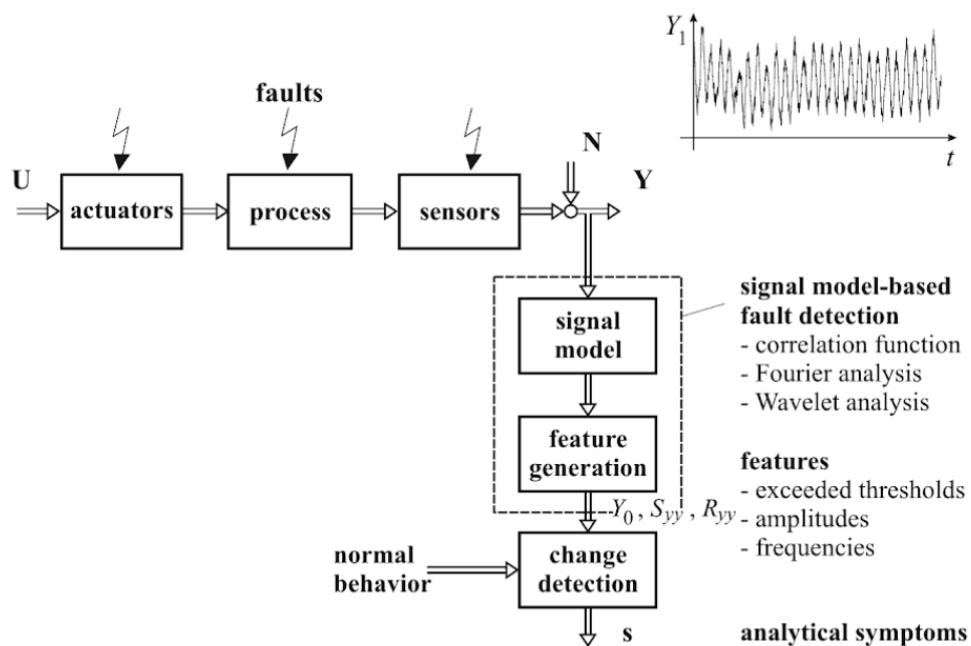
S neustálým rozvojem technologií a rozšiřováním složitých systémů přichází velký počet možných poruch. Tím pádem se rozvíjí i široká inženýrská oblast zabývající se automatickou detekcí a diagnostikou těchto poruch. Tato oblast využívá velké škály matematických nástrojů a postupů, jejichž část zde bude uvedena. Jedním z přístupů detekce poruch je skrze známý model systému (obr. 1). Porovnávají se vstupy/výstupy a parametry modelu s reálnými výstupy systému.



Obrázek 1: Schéma detekce poruch na základě známého modelu [6].

Dalším z přístupů je analýza vlastností samotného signálu bez znalosti modelu řízeného

systému (obr. 2). Vzhledem k tomu že tato práce analyzuje signály bez znalosti modelu, je kladen větší důraz na hledání poruch na základě vlastností signálu. Může se tedy jednat o libovolný systém jehož proměnné jsou analyzovány nástroji jako je například korelační funkce, Fourierova transformace, vlnková transformace atd. Sledovanými vlastnostmi jsou například průměry, hodnoty amplitud, spektrum obsažených frekvencí, trendy, rychlosti změn v datech (tab. 2) [6]. Vybrané vlastnosti jsou porovnávány s normálním, případně požadovaným chováním těchto veličin. Při odchylkách od normálního chování je možno detekovat symptomy v podobě příliš nízké nebo vysoké frekvence, příliš nízkých nebo vysokých hodnot pozorovaných veličin atd. Získané symptomy následně slouží k diagnostice poruch. Odchylna od normálního chování je nejčastěji posuzována dle vzdálenosti od aritmetického průměru, ale je možno zvolit i modus, medián, či jiný parametr.



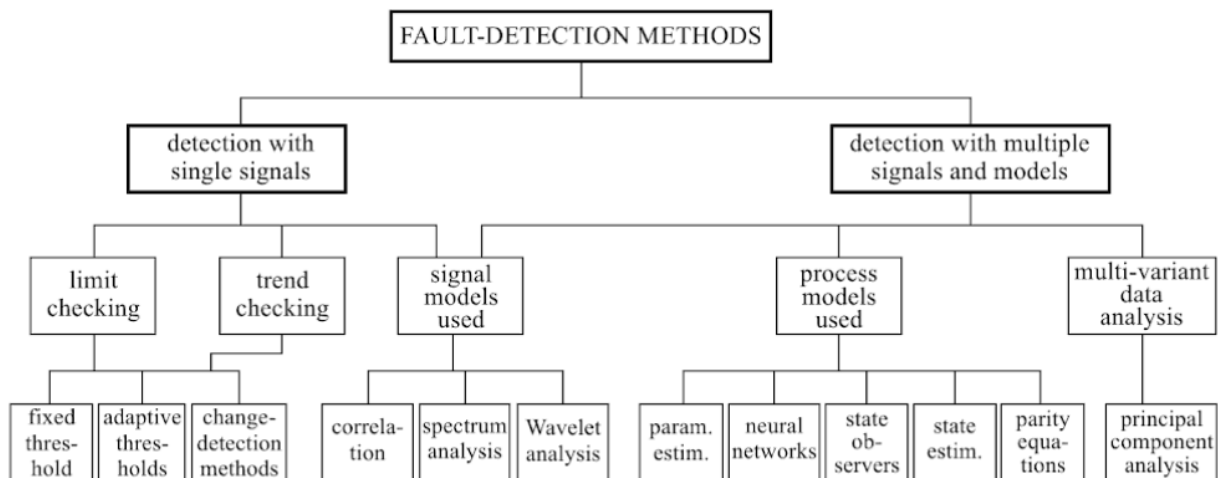
Obrázek 2: Schéma detekce poruch na základě vlastností signálu bez znalosti modelu [6].

Tabulka 2: Vybrané vlastnosti signálu [7].

Feature name	Feature definition	Value for the harmonic signal
Mean value	$\bar{x} = \frac{1}{T} \int_t^{t+T} x(u) du$	$\bar{x} = 0$
Absolute mean value	$x_{\text{AVE}} = \frac{1}{T} \int_t^{t+T}  x(u)  du$	$x_{\text{AVE}} = \frac{2}{\pi} X \cong 0,603X$
Root mean square value	$x_{\text{RMS}} = \sqrt{\frac{1}{T} \int_t^{t+T} x^2(u) du}$	$x_{\text{RMS}} = \frac{X}{\sqrt{2}} \cong 0,707X$
Absolute peak value	$x_{\text{PEAK}} = \max_{t < u \leq t+T}  x(u) $	$x_{\text{PEAK}} = X$
Positive peak value	$x_{\text{PEAK}+} = \max_{t < u \leq t+T} x(u)$	$x_{\text{PEAK}+} = X$
Negative peak value	$x_{\text{PEAK}-} = \min_{t < u \leq t+T} x(u)$	$x_{\text{PEAK}-} = -X$
Peak to peak value	$x_{P-P} = x_{\text{PEAK}+} - x_{\text{PEAK}-}$	$x_{P-P} = 2X$
Form factor	$K = \frac{x_{\text{RMS}}}{x_{\text{AVE}}}$	$K = \frac{\pi}{2\sqrt{2}} \cong 1,111$
Crest factor	$C = \frac{x_{\text{PEAK}}}{x_{\text{RMS}}}$	$C = \sqrt{2} \cong 1,414$

Hlavním rozdílem obou metod je právě znalost modelu, která v některých případech může být velice složitá a mnohdy model nemusí být přesný. Toto se může obejít přístupem identifikace daného systému, kdy skrze naměřené vstupy a výstupy jsou identifikovány parametry modelu. Pro tento přístup je sice nutná znalost správného chování systému, ale může odstranit nepřesnosti analytického modelu. Na obr. 3 je znázorněno základní schéma rozdělení metod detekce poruch. Této práci se týkají především krajní části schématu, jelikož zahrnuje práci se samotným signálem (levá část - limity hodnot, limity rychlostí změn, spektrální analýza), tak v jistém smyslu i analýzu více proměnných (pravá část - metoda hlavních komponent).





Obrázek 3: Schéma obecné metodiky detekce poruch [6].

### 3.1 Amplitudy v datech

Analýza amplitud signálu je základní kontrolou absolutní hodnoty veličiny. Překročení limitních teplot může mít i fatální následky nejen materiálně, ale i zdravotně. Primárně se nemusí zdát přímá souvislost s detekcí oscilací, ale je tím možné například sledovat rozdíl mezi minimem a maximem signálu, což už je u periodického signálu důležitý parametr určující jeho výkon. Nehledě na to, že velké výkyvy teplot ve vnitřních prostorech obývaných lidmi je nežádoucí jev, tak mj. cyklické namáhání materiálu s velkými výkyvy teplot snižuje odolnost.

### 3.2 Změny v datech

Jednou ze základních operací sledování signálu je i analýza rychlostí jeho změn, jelikož mohou být veličiny, u kterých je žádoucí stabilní průběh bez výrazných skoků. Vzhledem k tomu, že se jedná o diskretní signál, tak základním vzorcem pro výpočet rychlostí změn je výpočet diferencí (1).

$$\dot{x}_k = \frac{x_k - x_{k-1}}{\Delta t} \quad (1)$$

### 3.3 Fourierova transformace

Pro detekci oscilací se nabízí využití široce používaného nástroje - Fourierovy transformace. Fourierova transformace je základním matematickým nástrojem analýzy signálů [8]. Fourierova transformace slouží pro převod signálu z časové oblasti do oblasti frekvenční. Výsledkem Fourierovy transformace je komplexní číslo (funkce komplexní proměnné), z něhož lze výpočtem jeho absolutní hodnoty určit spektrum signálu, tj. informace o tom, jaké frekvence jsou v signálu obsaženy. Základní vztah vyjadřuje Fourierův obraz  $S(f)$  spojité funkce  $s(t)$ , kde  $f$  je frekvence a  $t$  je čas (2).

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-2\pi ift} dt \quad (2)$$

Je nutné podotknout, že v dnešní době počítačů a digitální techniky je důležitá diskrétní forma Fourierovy transformace (3), kde je integrál nahrazen sumou a spojitý čas a frekvence jsou nahrazeny diskrétními veličinami dle rovnic (4), (5). Po dosazení diskrétních veličin má Fourierova transformace podobu dle (8).

$$S_d(f_n) = \Delta t \sum_{k=0}^{N-1} s(t_k)e^{-2\pi if_n t_k} \quad (3)$$

$$t_k = k\Delta t, \quad k = 0, 1, \dots, N-1 \quad (4)$$

$$f_n = n\Delta f, \quad n = 0, 1, \dots, N-1 \quad (5)$$

$$\Delta t = \frac{T}{N} \quad (6)$$

$$\Delta f = \frac{1}{T} \quad (7)$$

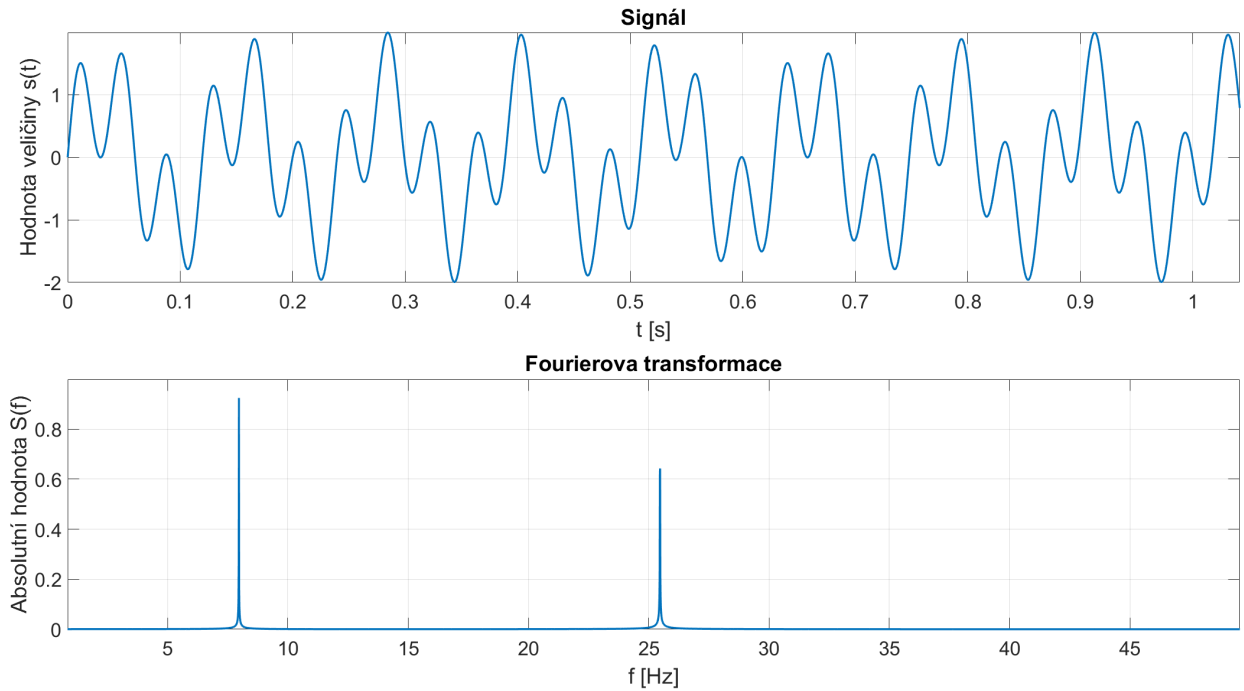
$$S_d(f_n) = \Delta t \sum_{k=0}^{N-1} s(t_k)e^{\frac{-2\pi ink}{N}}, \quad n = 0, 1, \dots, N-1 \quad (8)$$

V diskrétních systémech hrozí takzvaný aliasing a měla by být splněna Nyquistova (Shannon-Kotělnikovova) podmínka (9).

$$f \geq 2f_{max} \quad (9)$$

Kde  $f$  je vzorkovací frekvence a  $f_{max}$  je maximální frekvence obsažena v signálu. V případě nesplnění této podmínky dochází k výskytu falešných frekvencí v signálu.

Na obr. 4 je znázorněn příklad signálu, který je složen ze dvou sinových funkcí s různou frekvencí. Pod časovým průběhem signálu je znázorněn signál po Fourierově transformaci ve frekvenční oblasti, z čehož jsou zřetelné dvě různé frekvence.



Obrázek 4: Příklad aplikace Fourierovy transformace. V horní části se nachází originální signál v časové oblasti, v dolní části jeho transformace do frekvenční oblasti.

### 3.3.1 Rychlá Fourierova transformace

Vzhledem k výpočetní náročnosti diskrétní Fourierovy transformace, je vhodné využít algoritmus rychlé Fourierovy transformace, jenž složitost výpočtu změní z  $O(N^2)$  na  $O(N \log_2 N)$ . Vyjděme tedy z předpisu pro diskrétní Fourierovu transformaci (3), kde nahradíme exponenciální člen dle (10), z čehož vyplývá (11) a (12) [8].

$$e^{-2\pi i \nu t_k} = e^{\frac{-2\pi i n k}{N}} = w^{nk} \quad (10)$$

$$w^2 = e^{2(\frac{-2\pi i}{N})} = e^{\frac{-2\pi i}{N/2}} = w_{N/2} \quad (11)$$

$$S_d(f_n) = \sum_{k=0}^{N-1} s(k)w^{nk} \quad (12)$$

Princip spočívá v rozdělování na transformace dle sudých a lichých diskretních indexů  $k$  (13), (14).

$$k = \begin{cases} 2j & \text{pro sudé indexy;} \\ 2j + 1 & \text{pro liché indexy.} \end{cases}, \quad j = 0, 1, \dots, \frac{N}{2} - 1 \quad (13)$$

$$S_d(f_n) = \sum_{j=0}^{\frac{N}{2}-1} s(2j)w^{n2j} + \sum_{j=0}^{\frac{N}{2}-1} s(2j+1)w^{n(2j+1)} \quad (14)$$

Úpravou předchozích rovnic dospějeme k rovnici (15), což je součet dvou polovičních Fourierových transformací. V tomto rozdělování výpočtu na větší počet méně složitých výpočtů lze pokračovat dle (16), jelikož se vždy jedná o rozdělení na dva poloviční problémy, kde počet rozdělení je  $m$ .

$$S_d(f_n) = \sum_{j=0}^{\frac{N}{2}-1} s(2j)w^{\frac{nj}{2}} + w^n \sum_{j=0}^{\frac{N}{2}-1} s(2j+1)w^{\frac{nj}{2}} \quad (15)$$

$$m = \log_2 N \quad (16)$$

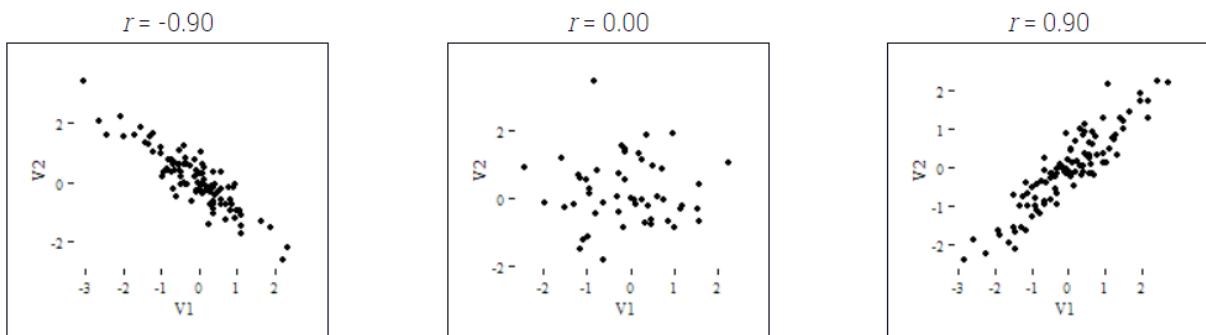
### 3.4 Korelace

Korelace je známý statistický nástroj, který je využíván k vyhodnocení míry lineární závislosti mezi proměnnými. Nejčastějším vztahem pro korelaci je výpočet Pearsonova korelačního koeficientu (17), kde  $X$  a  $Y$  jsou náhodné veličiny, jejichž míru korelovanosti chceme vypočítat.

$$r_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} \quad (17)$$

$C_{XY}$  je kovariance náhodných veličin a  $\sigma_X$ ,  $\sigma_Y$  jsou směrodatné odchylky náhodných veličin. Po dosazení za kovarianci a směrodatné odchylky, dostaneme vzorec (18), kde  $\bar{x}$  a  $\bar{y}$  jsou střední hodnoty náhodných veličin  $X$  a  $Y$ . [9]

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (18)$$



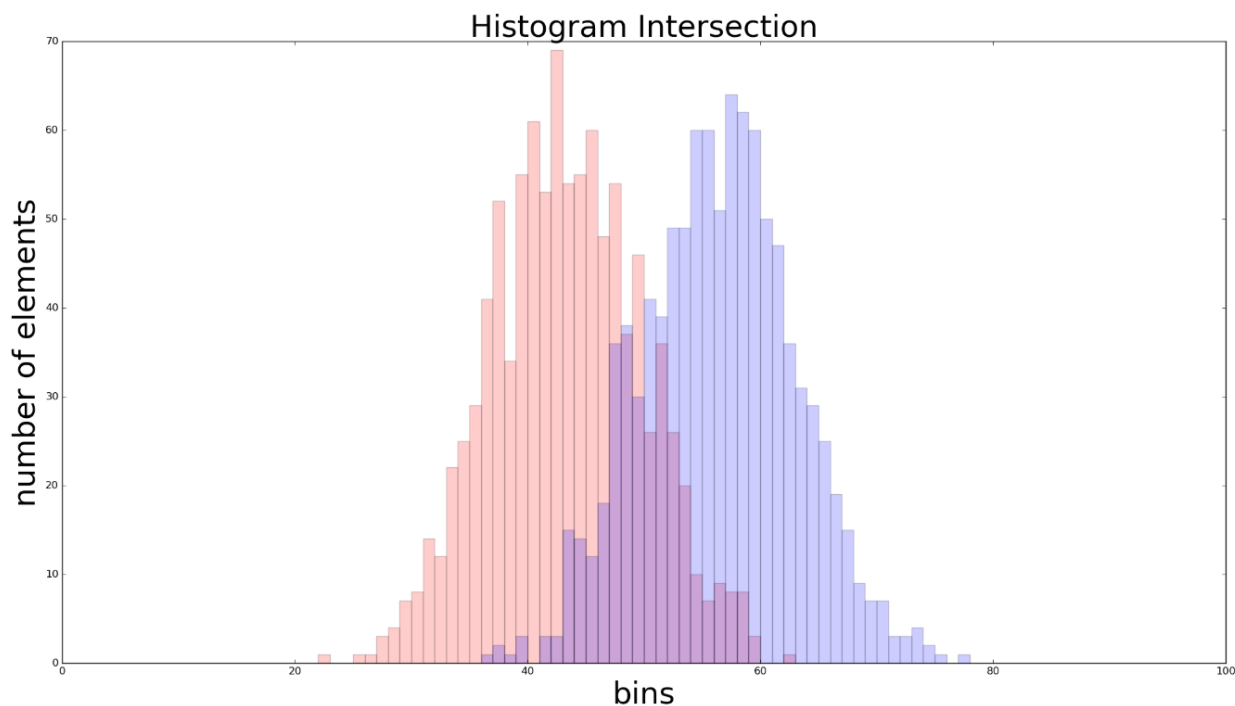
Obrázek 5: Znázornění významu korelace [10].

### 3.4.1 Korelace histogramů

Korelace může být také určena například porovnáním histogramů signálu. V našem případě uvažujeme jako histogram frekvencí obraz Fourierovy transformace naměřených dat. V případě že je rozložení frekvencí podobné, bude se korelace blížit k jedné. Toho se dá využít při posouzení neobvyklých signálů, jelikož předpokládáme, že rozložení frekvencí bude pro stejné veličiny podobné. Může to být použito jako filtr, pokud známe požadované rozložení frekvencí.

Mimo vypočtení korelačního koeficientu z předchozí kapitoly, je možné využít i průniku histogramů dle (19), kde  $M_j$  značí  $j$ -tý prvek modelového histogramu a  $I_j$  značí histogram, s kterým porovnáváme histogram modelový, což je graficky znázorněno na obr. 6.

$$H = \frac{\sum_{j=0}^n \min(I_j, M_j)}{\sum_{j=0}^n M_j} \quad (19)$$



Obrázek 6: Průnik histogramů [10].

### 3.5 Shluková analýza pomocí k-means

Shluková analýza je dalším statistickým nástrojem, jehož cílem je klasifikace objektů tj. rozřazení objektů do skupin, v níž jsou si k sobě objekty blíže než k objektům mimo skupinu, z čehož se může usuzovat jejich společná charakteristika.

Metoda k-means uvažuje předem daný počet shluků ( $k$  značí počet uvažovaných shluků), ke kterým jsou iterativním postupem přiřazovány objekty. V prvním kroku jsou středy shluků (centroidy) určeny náhodně (existují ale i algoritmy, které určují první polohu centroidů nenáhodně). Následně jsou spočítány vzdálenosti (může být dle různých metrik) každého objektu od centroidu. Objekt je přiřazen k centroidu, ke kterému je nejbliže. Po přiřazení následuje přepočítání polohy centroidů. Nová poloha centroidu je spočítána jako průměrná hodnota poloh všech bodů k němu přiřazených. Algoritmus se opakuje dokud se příslušnost objektů k jednotlivým shlukům nepřestane měnit (nebo počet iterací nepřesáhne stanovenou mez). Jedná se tedy o minimalizaci funkce (20), kde  $u_{ih} \in \{0, 1\}$  značí příslušnost  $i$ -tého objektu k  $h$ -tému shluku,  $\mathbf{x}_i$  je  $i$ -tý objekt (pozorování) a  $\bar{\mathbf{x}}_h$  je vektor průměrných hodnot  $h$ -tého shluku (pozice centroidu). [11]



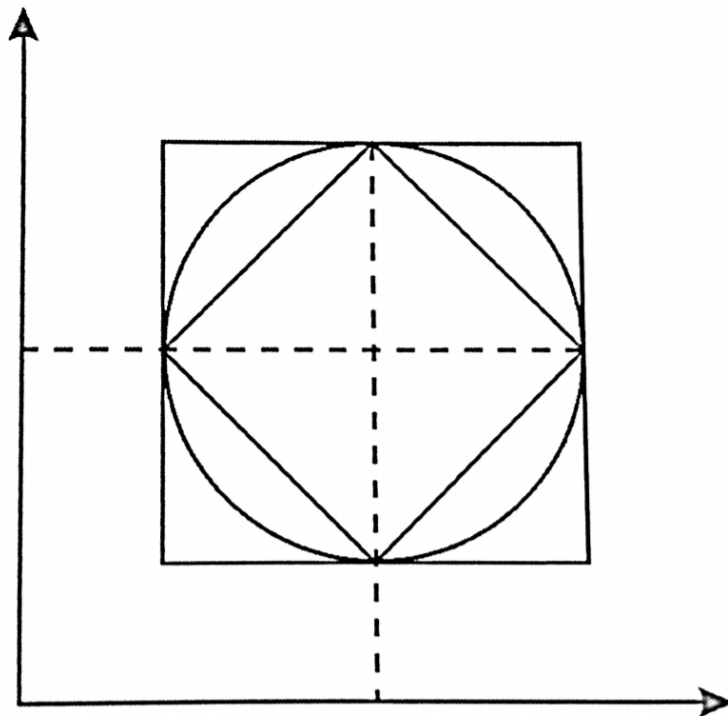
$$f_{KM} = \sum_{h=1}^k \sum_{i=1}^n u_{ih} \|\mathbf{x}_i - \bar{\mathbf{x}}_h\|^2 \quad (20)$$

Příklady jednotlivých metrik jsou ukázány dle rovnic (21) - tzv. manhattanská vzdálenost, (22) - euklidovská vzdálenost, (23) - Čebyševova vzdálenost. Graficky jsou tyto metriky znázorněny na obr. 7. V této práci je využita euklidovská norma.

$$\|\mathbf{x}_1 - \mathbf{x}_2\|_1 = \sum_{j=1}^p |x_{1j} - x_{2j}| \quad (21)$$

$$\|\mathbf{x}_1 - \mathbf{x}_2\|_2 = \sqrt{\sum_{j=1}^p (x_{1j} - x_{2j})^2} \quad (22)$$

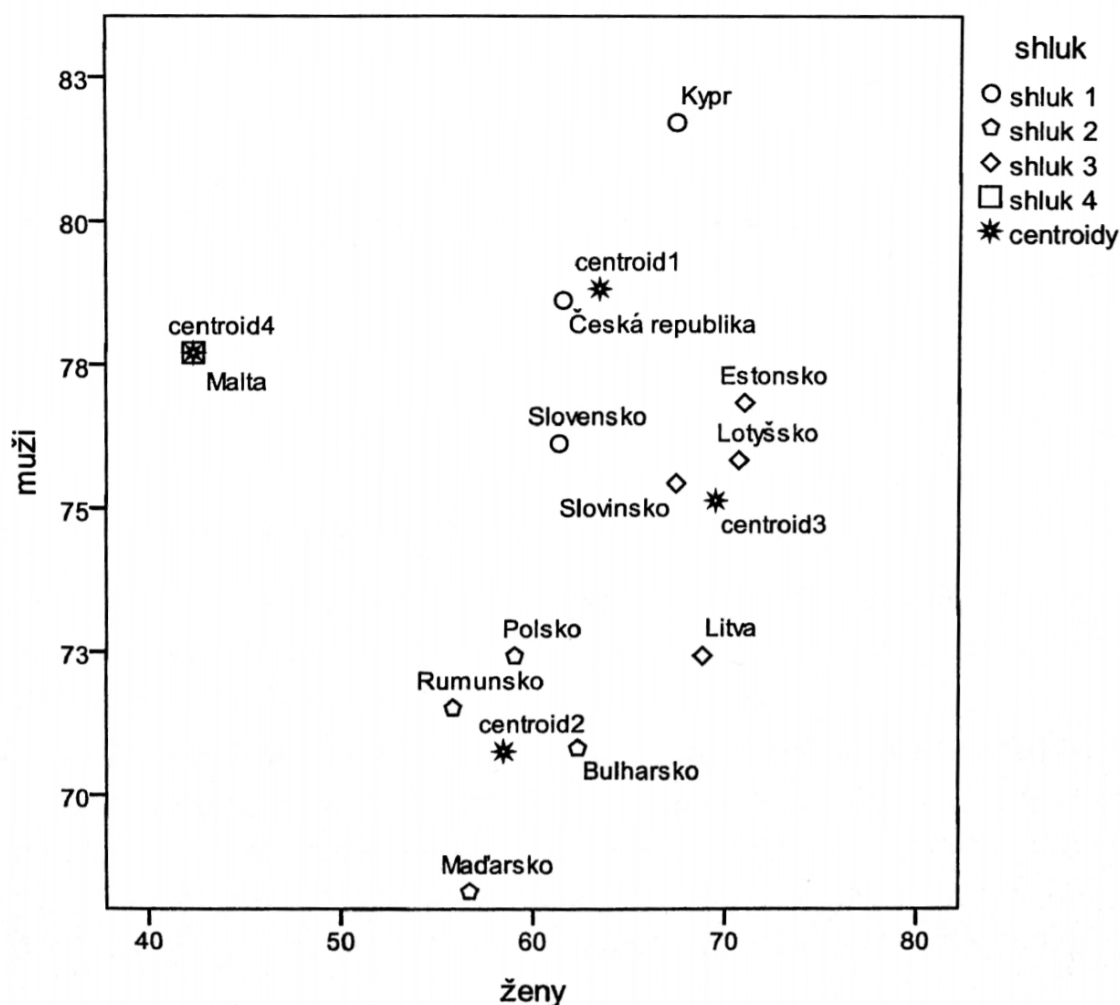
$$\|\mathbf{x}_1 - \mathbf{x}_2\|_\infty = \max_j |x_{1j} - x_{2j}| \quad (23)$$



Obrázek 7: Euklidovská vzdálenost je zobrazena jako kružnice, Čebyševova vzdálenost jako největší čtverec a manhattanská vzdálenost jako vnitřní natočený čtverec. [11]

Na obr. 8 je znázorněn jednoduchý dvourozměrný příklad shlukové analýzy. Z grafického zobrazení závislosti dvou proměnných jsou zřetelné vzdálenosti mezi pozorováními. Na

příkladu jsou zvoleny 4 shluky, jejichž příslušníci mají k sobě nejbližší.



Obrázek 8: Ukázka použití k-means na dvou-dimenzionálním příkladu - rozdělení na 4 shluky.

[11]

### 3.6 Metoda hlavních komponent

Metoda hlavních komponent je užitečný nástroj především při průzkumové analýze dat. Má uplatnění v analýze vícerozměrných úloh, kdy chceme prvotně analyzovat a vizualizovat vícerozměrná data (tj. především více než třírozměrná). Může odpovědět na otázky typu, zda jsou vícerozměrná data opravdu nezávislá, či identifikovat odlehlá pozorování. Tato metoda je ovšem i velice užitečná jako podklad pro shlukovou analýzu, což je i případ této práce. Metoda hlavních komponent dokáže vizualizovat vícerozměrná data v námi pozorovatelných

dimenzích (obvykle v rovině). Podle průzkumové vizualizace jsme pak schopni odhadnout počet shluků pro shlukovou analýzu, případně zdali má shluková analýza vůbec smysl.

Metoda hlavních komponent se snaží nalézt nejlépe vzájemně nezávislé lineární kombinace původních proměnných. Nově nalezené takzvané hlavní komponenty mohou vyjádřit původní informaci s minimální ztrátou v rámci menšího počtu proměnných než bylo na počátku tzn. sníží se rozměr dat. Samozřejmě v nepříznivém případě může být ztráta informací větší, ale alespoň může dojít k prvotní průzkumové analýze. [11]

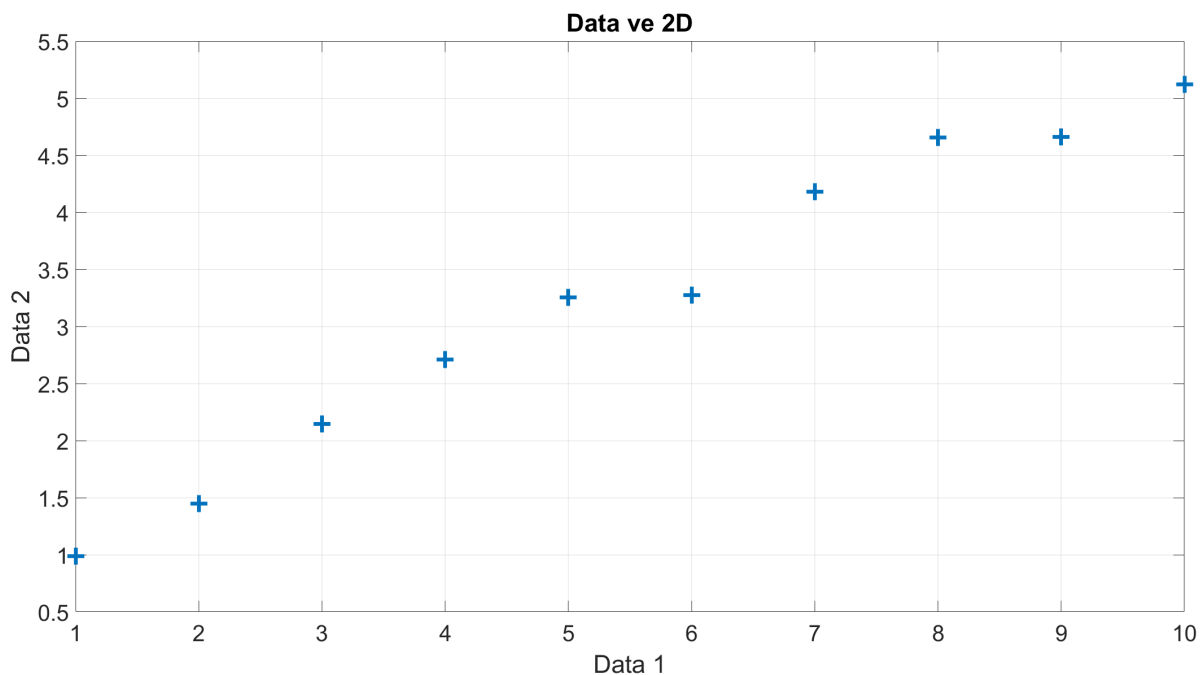
Pro metodu hlavních komponent je určující velikost rozptylu ve zkoumaných datech. První hlavní komponenta je ta komponenta s největším rozptylem. Pro vysvětlení je ukázáno promítnutí rovinné závislosti do přímkové. Na obr. 9 je zobrazena závislost dvou proměnných. K vyjádření této závislosti je tedy potřeba i dvou dimenzí.

Nalezení první hlavní komponenty  $K_1$  je popsáno dle (25), kde jsou vycentrovaná data násobena vektorem  $\omega_1^T$ . Tento vektor byl nalezen jako vlastní vektor kovarianční matice (24) zkoumaných dat, tak aby rozptyl hlavní komponenty byl maximalizován.

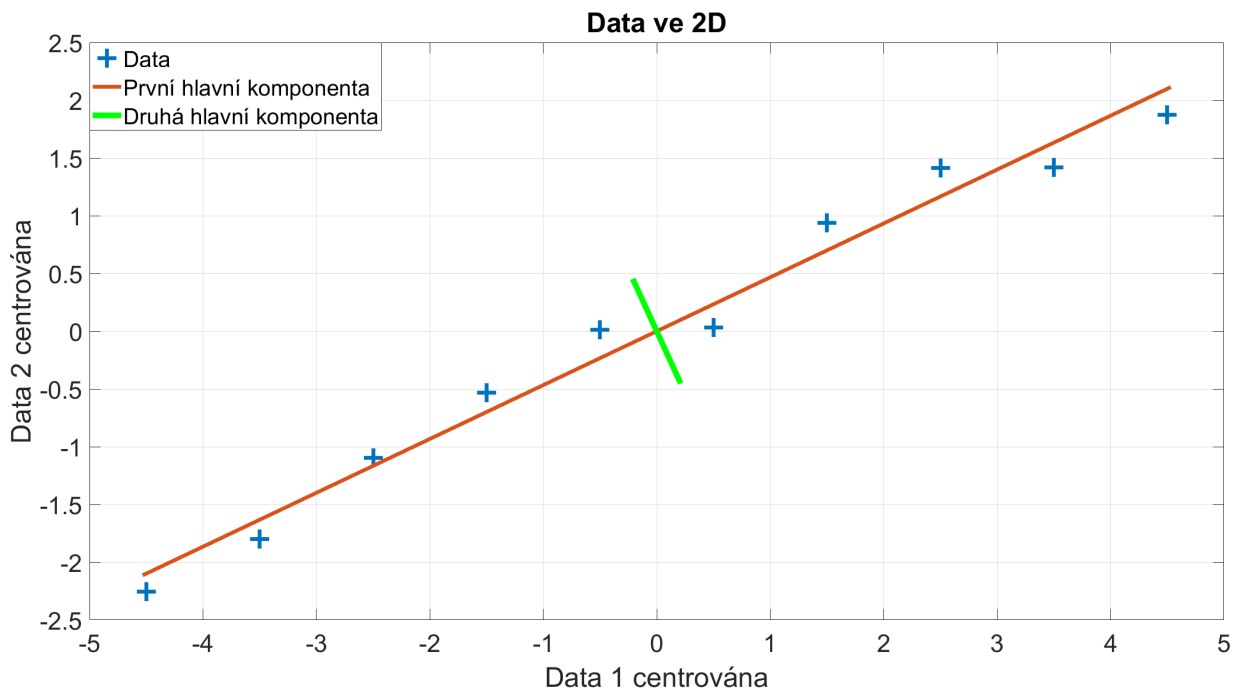
$$\Sigma = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T] \quad (24)$$

$$K_1 = \omega_1^T(\mathbf{x} - \bar{\mathbf{x}}) \quad (25)$$

Grafický význam této operace je zjevný na obr. 10. Na tomto obrázku je vidět vycentrování dat a také, že první hlavní komponenta je směr pokrývající největší rozptyl. Přesněji řečeno je součet vzdáleností mezi body promítnutých do první hlavní komponenty a počátkem maximální. Druhá hlavní komponenta je kolmá na první a jedná se o vektor, jenž by pokrýval druhý největší rozptyl. Analogicky lze tuto úlohu rozšířit do libovolného konečného počtu dimenzí.



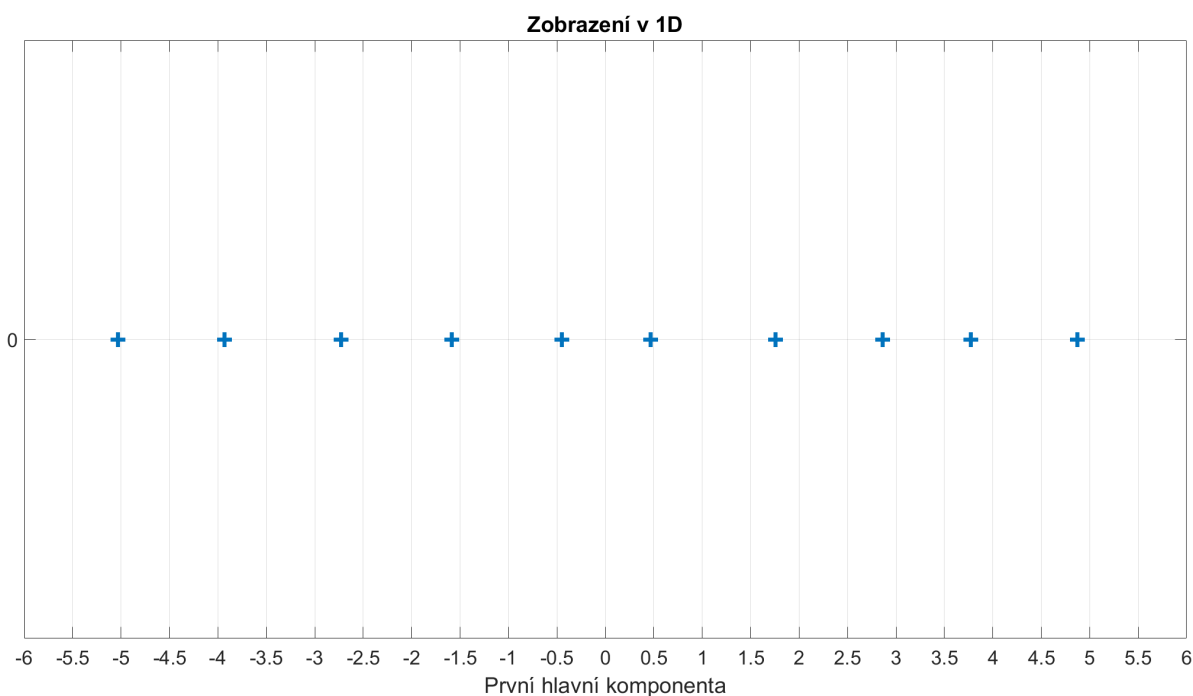
Obrázek 9: Modelová data pro aplikaci metody hlavních komponent.



Obrázek 10: Vizualizace hlavních komponent na modelových datech.

Vypovídajícím parametrem pro použitelnost metody hlavních komponent je rozptyl jednotlivých hlavních komponent. V uvedeném příkladu je rozptyl první hlavní komponenty

11, 15 a druhé hlavní komponenty 0,04, z čehož plyne, že první hlavní komponenta jasně převyšuje rozptyl druhé, a tudíž má smysl snížit dimenzi promítnutím dat pouze do první hlavní komponenty (obr. 11). Pokud by tento rozdíl v rozptylech nebyl významný, pak dochází k velké ztrátě informací při snižování dimenzí. Obecně by se mělo promítat do komponent s nejvýraznějšími rozptily a je v podstatě štěstí, když počet nejvýraznějších komponent je menší nebo roven třem.



Obrázek 11: Použití metody hlavních komponent pro snížení rozměru dat.

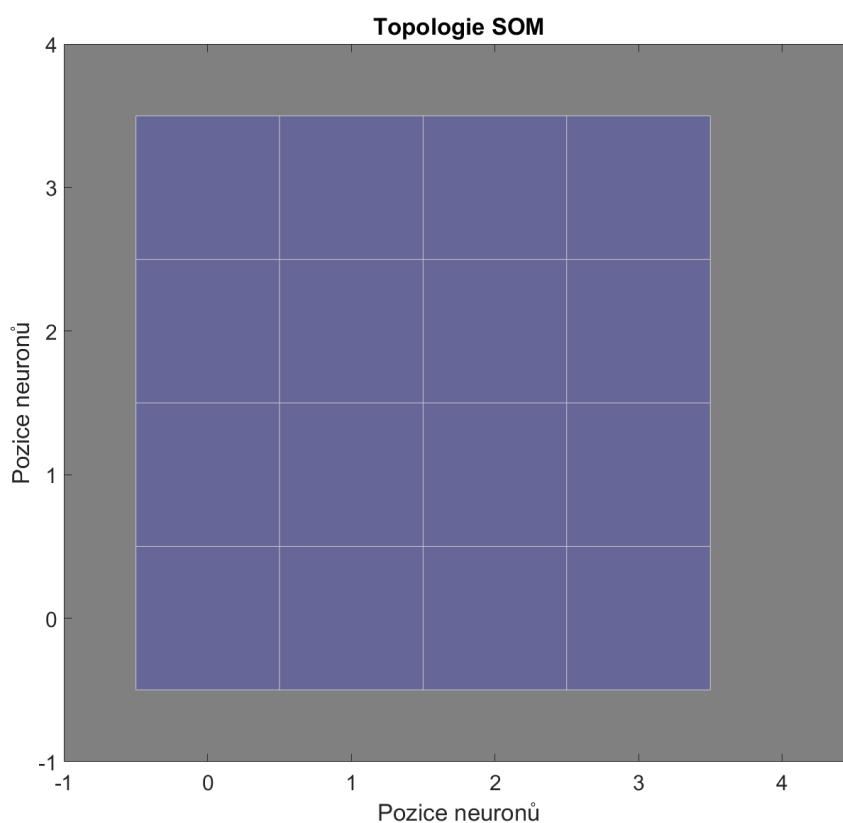
Kdyby data způsobující rozptyl druhé hlavní komponenty byl šum, je možné zpětně transformovat promítnutá data z první komponenty. V našem případě bychom dostali data na vyhlazené přímce.

### 3.7 Kohonenova samoorganizační mapa

Kohonenova samoorganizační mapa je opět nástrojem vícerozměrné analýzy dat. Kohonenovy mapy patří mezi neuronové sítě bez učitele. Podobně jako u metody hlavních komponent je tedy cílem promítnutí vícerozměrných dat obvykle do roviny. Body v Kohonenově mapě, které leží u sebe, leží u sebe i v původním prostoru dat (což se obecně u metody hlavních

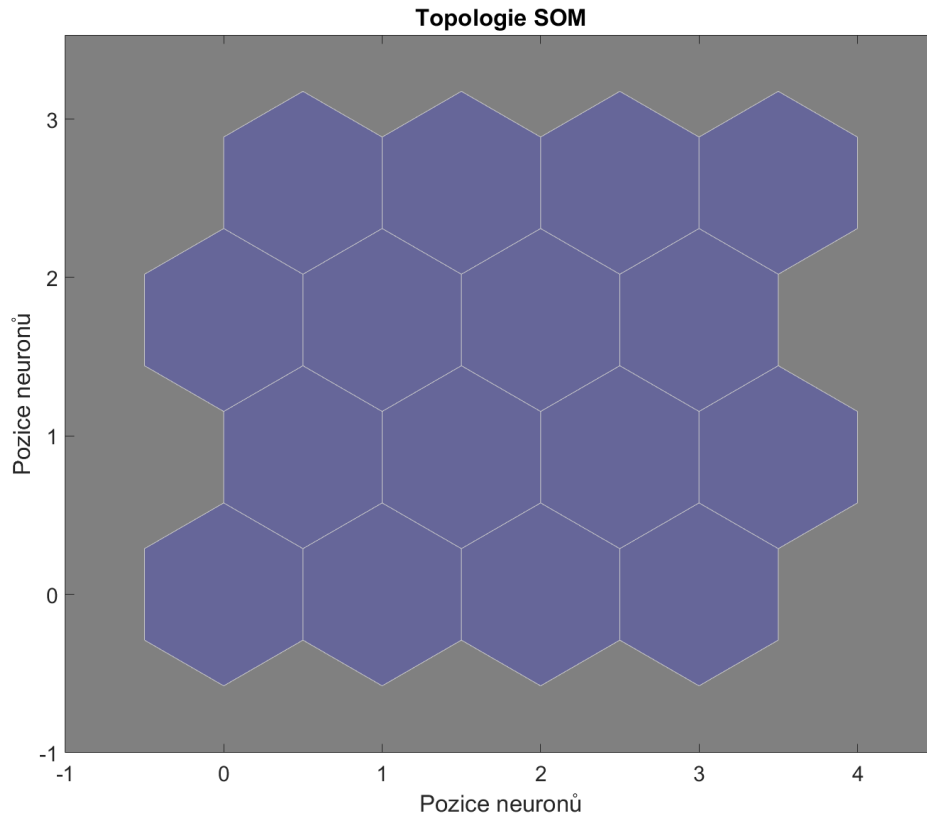
komponent nemusí vždy povést). Další výhodou je i adaptivnost tj. pokud chceme mapu obohatit o další pozorování, je doučení mapy mnohem méně náročné, než například přepočítat celou metodu hlavních komponent (to je důležité především u opravdu nadměrně velkém počtu dat). [11]

Základem Kohonenových map je (nejčastěji) dvourozměrná síť neuronů. Tato síť může mít například čtvercovou (obr. 12), nebo šestiúhelníkovou topologii (obr. 13)



Obrázek 12: Čtvercová topologie SOM.





Obrázek 13: Šestiúhelníková topologie SOM.

Je tedy definována síť a její topologie. V algoritmu učení této sítě se v prvním kroku přiřadí váhy neuronů náhodně (případně se pro zrychlení mohou využít data z jiných analýz). Dále se vybere jedno pozorování ze souboru dat a hledá se pozice neuronu, jenž je svými vahami  $\mathbf{w}_c$  nejbližší vybranému vektoru pozorování  $\mathbf{x}(k)$  (26). Výpočet vzdálenosti mezi těmito vektory je možno zvolit (jak již bylo řečeno v předchozích kapitolách), ale já vybírám tradiční euklidovskou vzdálenost.

$$v = \arg \min \|\mathbf{x}(k) - \mathbf{w}_c\| \quad (26)$$

Po nalezení vítězného neuronu se aktualizují jeho váhy  $\mathbf{w}_v$  tak, aby se vektor vah přiblížil příslušnému pozorování  $\mathbf{x}(k)$  (27), kde  $\alpha(k)$  je parametr určující rychlost učení. Tento parametr se s narůstajícím počtem učících iterací zmenšuje tzn. na začátku se síť učí velkými skoky a postupně se aktualizace vah zjemňuje.

$$\mathbf{w}_v(k+1) = \mathbf{w}_v(k) + \alpha(k)[\mathbf{x}(k) - \mathbf{w}_v(k)] \quad (27)$$

Kromě učení samotného vítězného neuronu se aktualizují váhy i jeho sousedních neuronů, a to z důvodu zvýšení pravděpodobnosti vítězství blízkých neuronů v případě podobného pozorování. Tento jev již naznačuje shlukování sobě blízkých objektů. Aktualizační rovnice vah sousedních neuronů (28) je podobná aktualizační rovnici vítězného neuronu. Je však doplněna o parametr  $\eta$ , který určuje míru ovlivnění sousedních neuronů. Z rovnice (29) je patrné, že se parametr  $\eta$  zmenšuje se zvyšující se vzdáleností sousedních neuronů  $s$  od vítězného neuronu  $v$ . Parametr  $\Omega$  definuje velikost okolí, ve kterém má vítězný neuron vliv. Tento parametr se stejně jako parametr  $\alpha$  zmenšuje s počtem učících iterací. [11]

$$\mathbf{w}_s(k+1) = \mathbf{w}_s(k) + \alpha(k)\eta(v, k, s)[\mathbf{x}(k) - \mathbf{w}_v(k)] \quad (28)$$

$$\eta = \exp\left(-\frac{\|s - v\|^2}{2\Omega(k)^2}\right) \quad (29)$$

## 4 Dodaná data a jejich předzpracování

Data dodaná společností ENERGOCENTRUM PLUS, s.r.o. byla poskytnuta ve formátu *xlsx*. Bylo zasláno 12 souborů, z nichž každý jeden soubor obsahoval data za jeden kalendářní měsíc. Vzorkovací perioda byla 5 minut, celkem bylo naměřeno 101 676 vzorků a zaznamenáno chování až 2860 proměnných (v každém měsíci se počet proměnných měnil) (tab. 3).

Tabulka 3: Dodaná data

<b>Typ souboru:</b>	xlsx
<b>Vzorkovací perioda:</b>	5 min
<b>Počet vzorků (jeden rok):</b>	101676
<b>Počet proměnných:</b>	2173–2860

Jednotlivé proměnné jsou řazeny do sloupců a jsou složeny z 9 typů veličin (tab. 4), první

sloupec obsahuje časové značky v excelovském formátu (číslo 1 znamená 1. 1. 1900 - 0:00, trvání jedno dne je vyjádřeno velikostí jedna, tzn. jedna hodina je vyjádřena jako  $\frac{1}{24}$ )

Tabulka 4: Proměnné v dodaných datech v lednu

Název parametru:	Datový typ:	Jednotka:	Počet výskytů:
Indoor air temperature	Double	°C	729
Outdoor air temperature	Double	°C	169
Supply water temperature	Double	°C	309
Return water temperature	Double	°C	74
Pump operation	Bool	-	261
Heating or cooling regime indication	Bool	-	15
Warm service	Double	°C	56
Energy meter actual	Double	kW	135
Energy meter cumulative	Double	kWh	425

První úpravou, kterou jsem udělal, bylo převedení dat z formátu *xlsx* na matlabovský formát *mat*. Důvodem bylo mnohonásobně delší nahrávání excelovských dat, k čemuž byla použita funkce *readtable* z prostředí aplikace Matlab. Dále byla data roztříděna do kategorií podle svého významu (viz. tab. 4). K tomu jsem naprogramoval jednoduchý algoritmus, který hledal klíčová slova v názvu sloupce.

#### 4.1 Předzpracování boolovských dat

Proměnné, jejichž datový typ je bool (tj. *Pump operation* a *Heating or cooling regime indication*) byly předzpracovány zvlášť. Primární problém dat jsou chybějící naměřené hodnoty. Nabízí se tyto hodnoty buď vyškrtnout, nebo interpolovat (extrapolovat). Vzhledem k případné časové nekonzistentnosti naměřených dat jsem se rozhodl k interpolaci. Jak už z názvu práce vyplývá, jedná se o detekci oscilací, a tudíž jsem zvolil metodu interpolace takovou, která nezanese přidání oscilací do naměřených dat. Data jsou tedy doplněna o před-

chozí naměřenou hodnotu. V případě že chybí série dat na začátku souboru (což se reálně děje), jsou data uvažána od první naměřené hodnoty, tudíž nedochází k extrapolaci. Občas bohužel data obsahovala čísla, která nebyla pouze nula nebo jedna. Tato čísla byla např. 193477232 a 193473668. Dle mého úsudku tato čísla měla reprezentovat binární informaci, akorát došlo pravděpodobně k chybě při zápisu dat. Rozhodl jsem se tato čísla nahradit nulami a jedničkami tak, že větší z čísel reprezentuje jedničku a menší nulu.

## 4.2 Předzpracování dat typu double

Data typu double, tedy data, která nesou diskrétní informaci o spojitých fyzikálních veličinách, obsahují stejný problém jako data typu bool, tedy chybějící naměřené hodnoty. Opět jsem se rozhodl k interpolaci a to tentokrát k interpolaci lineární. Lineární interpolaci jsem zvolil z důvodu alespoň částečného zachování charakteru průběhu veličiny. Metoda, která byla zvolena u dat typu bool, by mohla do dat zanášet významnější skoky, což je nežádoucí i vzhledem k tomu, že je sledována derivace fyzikálních veličin. Dále jsem se rozhodl sledovat, zdali uprostřed dat nechybí větší série hodnot, jelikož by to mohlo do následného vyhodnocení zanechat chybu, nicméně nebyla odhalena delší série než 3 chybějící vzorky za sebou. Stejně jako u dat typu bool nedochází k extrapolaci a případná chybějící série na začátku datového souboru je ignorována. Pro analýzu pomocí rychlé Fourierovy transformace, je potřeba data vystředit a standardizovat dle (30), kde  $\mathbf{x}$  je původní vektor naměřených dat,  $\bar{\mathbf{x}}$  je střední hodnota naměřených dat a  $\sigma(\mathbf{x})$  je směrodatná odchylka naměřených dat.

$$\mathbf{x}_c = \frac{\mathbf{x} - \bar{\mathbf{x}}}{\sigma(\mathbf{x})} \quad (30)$$

Nutno podotknout, že data Energy meter actual a Energy meter cumulative byla sloučena. Jak již z tab. 4 (i z názvu proměnných) vyplývá, jedná se o hodnoty okamžitého výkonu (Energy meter actual) případně o hodnoty spotřebované energie (Energy meter cumulative). Přepočítání je tedy pouze výpočet diferencí parametru Energy meter cumulative.

## 5 Aplikace vybraných metod

### 5.1 Aplikace FFT

Prvním a zároveň stěžejním krokem detekce oscilací na diskretním signálu spojitě veličiny je provedení rychlé Fourierovy transformace. K tomu jsem použil funkci *fft* v prostředí Matlab [12]. Vstupem do této funkce je vektor dat, které chceme převést z časové do frekvenční oblasti. Výstupem je vektor komplexních čísel o stejné délce, jako je vektor vstupních dat. Pro reálný výstup je potřeba vypočítat absolutní hodnotu těchto komplexních čísel dle rovnice (31). Pro vyhovění Nyquistovu teorému (9) jsem použil pouze první polovinu vektoru transformovaných dat. Nyní už je frekvenční charakteristika připravena k další analýze. Pro přiřazení konkrétních frekvencí k jednotlivým hodnotám transformovaného vektoru je potřeba vytvořit vektor frekvencí  $\mathbf{f}$  a provést přepočít přes vzorkovací frekvenci  $f_s$  a délku naměřených hodnot  $L$  dle (32), kde  $k$  značí pozici ve vektoru frekvencí.

$$|S(f_k)| = \sqrt{Im_k^2 + Re_k^2} \quad (31)$$

$$f_k = k \frac{f_s}{L}, \quad k = 0, 1 \dots L/2 \quad (32)$$

Tento výpočet je udělán pro všechny sloupce typu double (roztříděno podle názvu). Tudiž je vytvořena matice všech Fourierových transformací (34) např. pro vnitřní teploty dle (33), kde jsou údaje vnitřních teplot z  $n$  senzorů, naměřené  $m$  počtem vzorků.

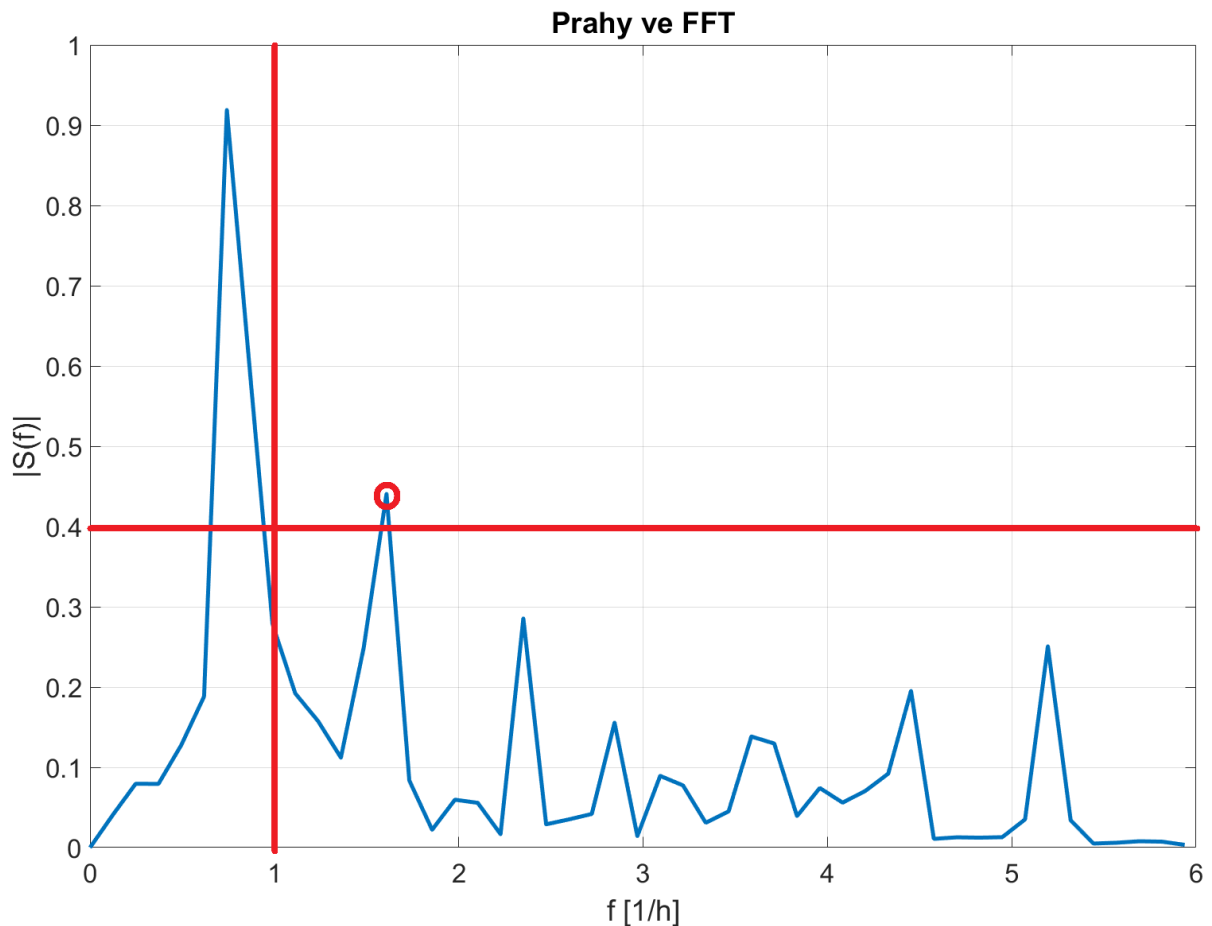
$$\mathbf{T}_{indoor} = \begin{bmatrix} T_{1,1} & T_{1,2} & \dots & T_{1,n} \\ T_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ T_{m,1} & T_{m,2} & \dots & T_{m,n} \end{bmatrix} \quad (33)$$

Toto je provedeno pro všech sedm typů proměnných typu double ( $\mathbf{T}_{indoor}$ ,  $\mathbf{T}_{outdoor}$ ,  $\mathbf{T}_{supply}$ ,  $\mathbf{T}_{return}$ ,  $\mathbf{T}_{service}$ ,  $\mathbf{E}_{actual}$ ,  $\mathbf{E}_{cumul}$ ). Stejně tak platí pro matici Fourierových transformací.

$$\mathbf{FFT}_{indoor} = \begin{bmatrix} FFT_{1,1} & FFT_{1,2} & \dots & FFT_{1,n} \\ FFT_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ FFT_{m,1} & FFT_{m,2} & \dots & FFT_{m,n} \end{bmatrix} \quad (34)$$

### 5.1.1 Filtr frekvencí

Jako prvotní filtr frekvencí lze použít limit hodnot lokálních maxim dat po Fourierově transformaci. Nejprve je potřeba lokalizovat maxima a následně zkontrolovat jejich hodnoty (obr. 14).



Obrázek 14: Limit frekvencí ve frekvenční charakteristice.

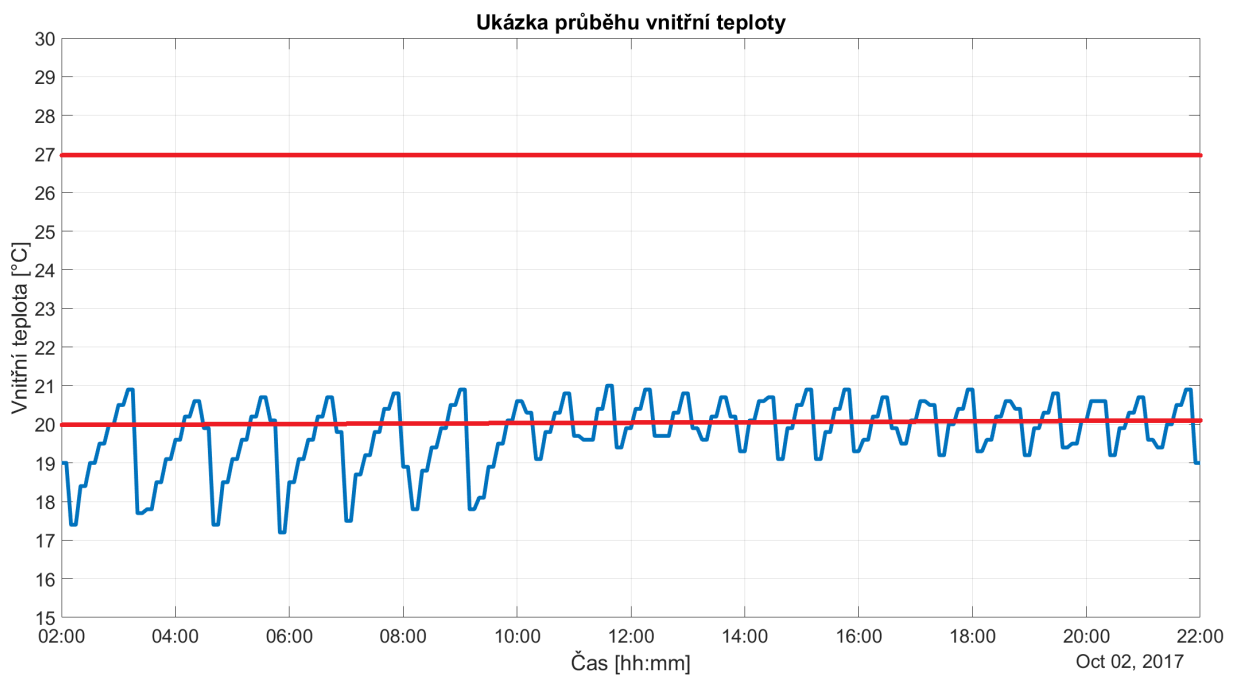
Tato metoda je základním filtrem vyšších frekvencí v signálu. Limity jsou nastaveny

aplikováním Fourierovy transformace na všechna dodaná data a vypočtením průměrné dominantní frekvence a směrodatné odchylky (filtr na vodorovné ose) a výpočtem průměrné hodnoty absolutních hodnot Fourierovy transformace (filtr na svislé ose).

## 5.2 Filtr amplitud

Filtr amplitud není nic jiného než kontrola překročení prahové hodnoty pozorované veličiny. Jak již bylo řečeno, prahová hodnota může být určena uživatelem na základě např. norem a zákonů (obr. 15), nebo přes výpočet průměru a směrodatné odchylky. Jakmile je hodnota veličiny vzdálena od průměru více než o velikost směrodatné odchylky, je hodnota vyhodnocena jako podezřelá (35).

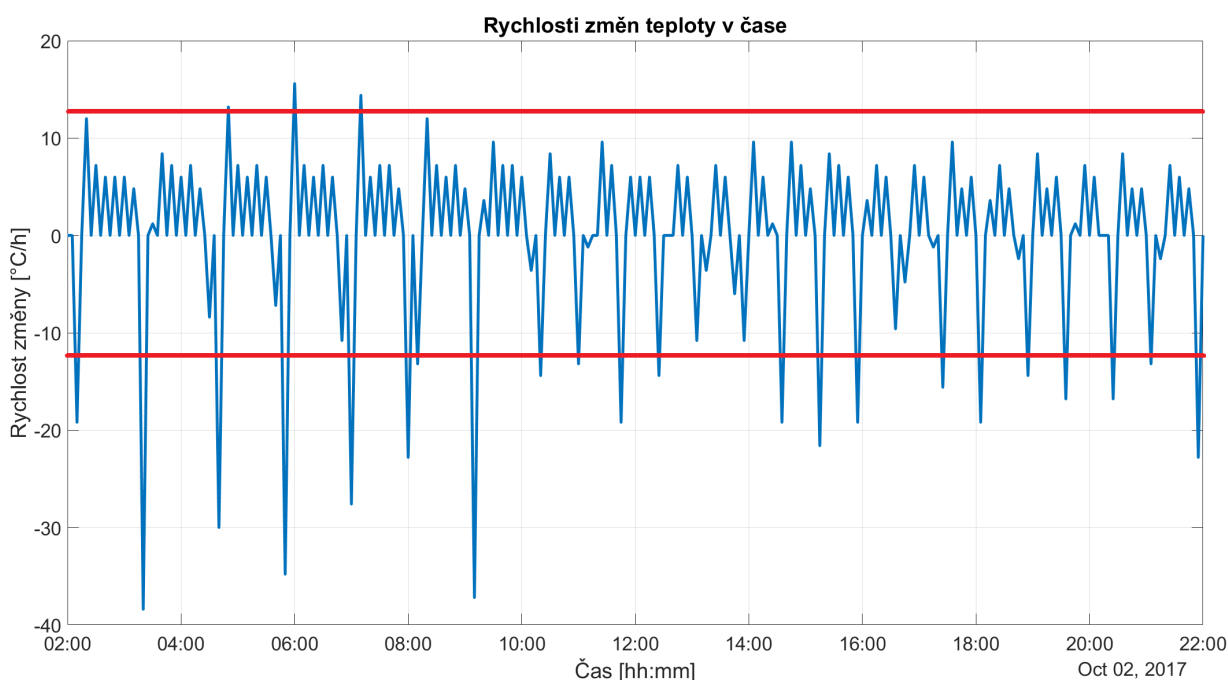
$$|x(t) - \bar{x}| \geq \bar{x} + \sigma_x \quad (35)$$



Obrázek 15: Limit prahových hodnot signálu. Práh určen na základě zákona o pracovním prostředí. [13]

### 5.3 Filtr rychlých změn

Filtr rychlých změn je aplikován úplně stejným způsobem jako filtr amplitud. Platí pro něj stejný vztah (35) s tou výjimkou, že je aplikován na difference dat (1), nikoliv na data samotná. Vypočtený průměr i směrodatná odchylka jsou určeny z lokálních maxim absolutní hodnoty průběhu diferencí (tudíž není to průměrná hodnota všech diferencí).



Obrázek 16: Limit rychlostí

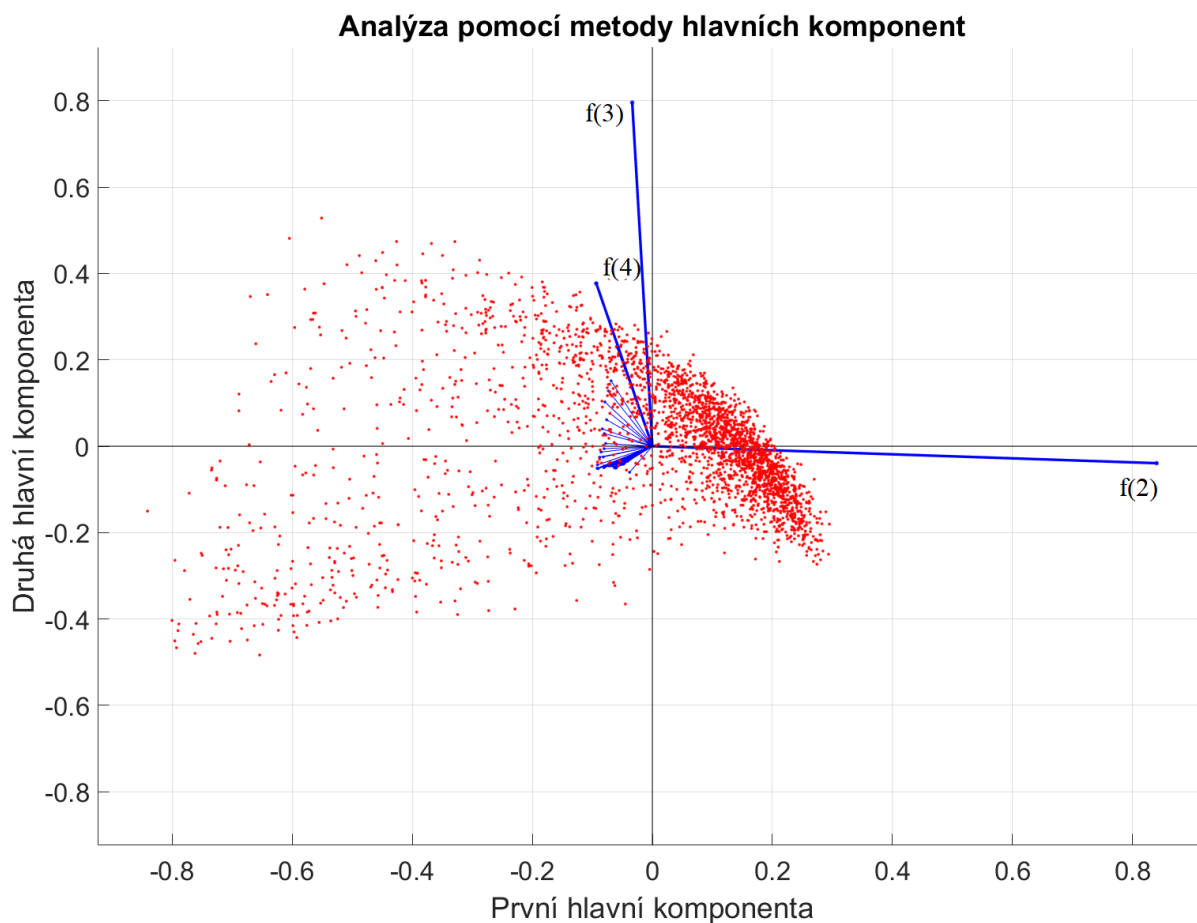
### 5.4 Aplikace PCA

Po aplikaci FFT máme k dispozici cca 2000 transformovaných vektorů o délce 96, což znamená závislost 96 proměnných. Proto jako prvotní průzkumovou analýzu použijí metodu hlavních komponent, čímž se pokusím tuto vícedimenzionální závislost vizualizovat v rovině. Opět k tomu využijí již existující funkci Matlabu s názvem *pca*. Povinným vstupem této funkce je matice našich transformovaných dat. Funkce má volitelné vstupy, čímž je možno specifikovat například maximální počet iterací, metodu algoritmu, počáteční odhady aj. Výstupem této funkce je matice koeficientů, znamenající složky vektorů jednotlivých komponent, dále matici . [14]

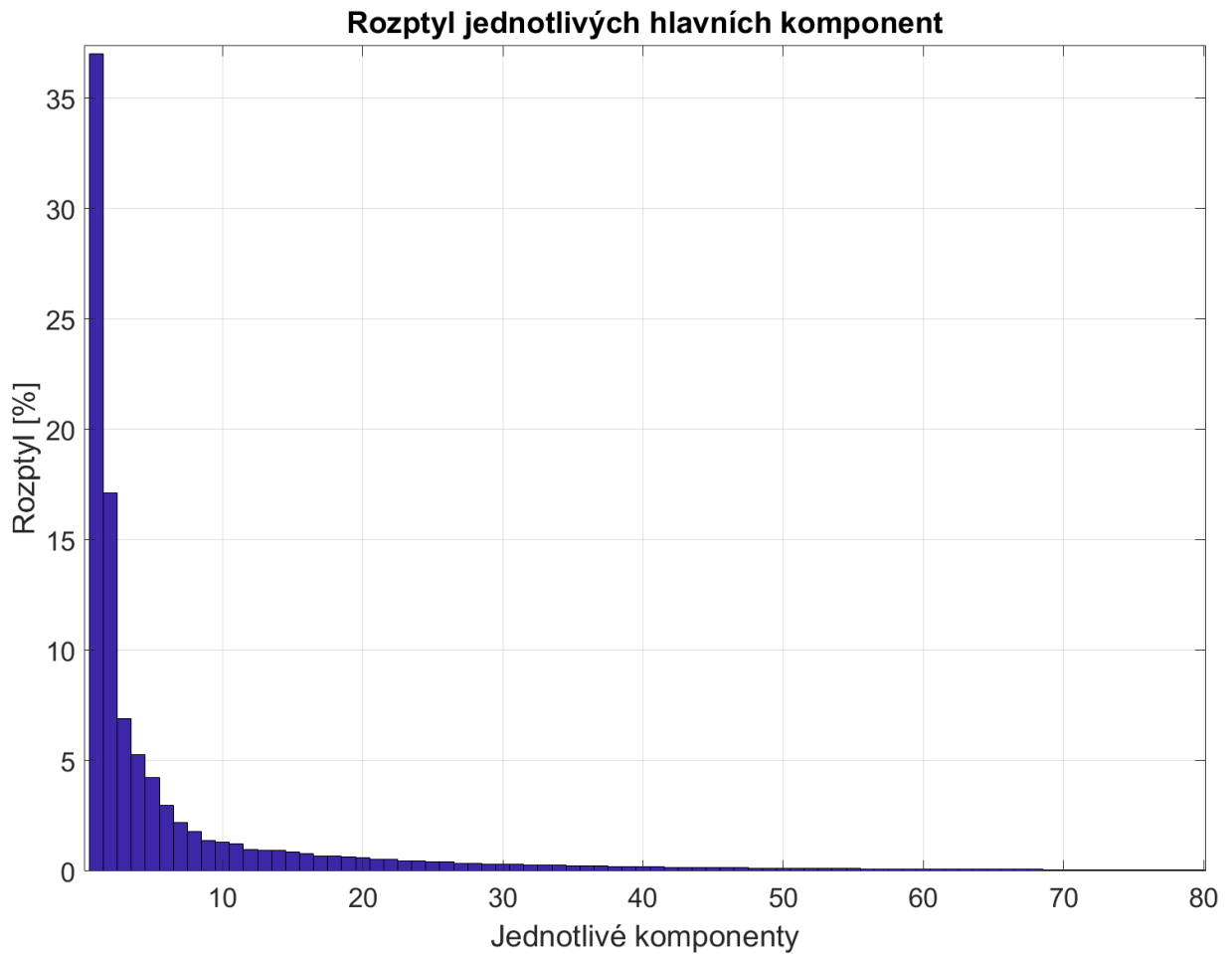


Tabulka 5: Vstupy a výstupy funkce *pca*

Vstupy:	Výstupy:
Matice dat ( $m \times n$ )	Koeficienty hlavních komponent ( $m \times m$ )
Limitní počet iterací	Skóre jednotlivých pozorování v hl. komponentách ( $m \times n$ )
Algoritmus výpočtu (SVD, EIG, ALS)	Vektor rozptylů jednotlivých komponent ( $1 \times m$ )



Obrázek 17: Aplikace metody hlavních komponent



Obrázek 18: Rozptyly jednotlivých hlavních komponent.

Grafický výsledek metody hlavních komponent je znázorněn na obr. 17. Data byla touto metodou vizualizována do roviny, tedy pomocí prvních dvou hlavních komponent. Jednotlivé červené body reprezentují jednotlivá pozorování promítnutá do dvou hlavních komponent. Modré vektory znázorňují míru zachyceného rozptylu jednotlivých proměnných. Dle součtu rozptylů reprezentují první dvě hlavní komponenty 54 % celkového rozptylu, což je diskutabilní, zdali je to dostatek pro hlubší analýzu. Dle mého názoru se dají udělat následující závěry:

1. Je patrný jeden dominantní shluk bodů
2. Dají se již identifikovat některá vzdálená pozorování

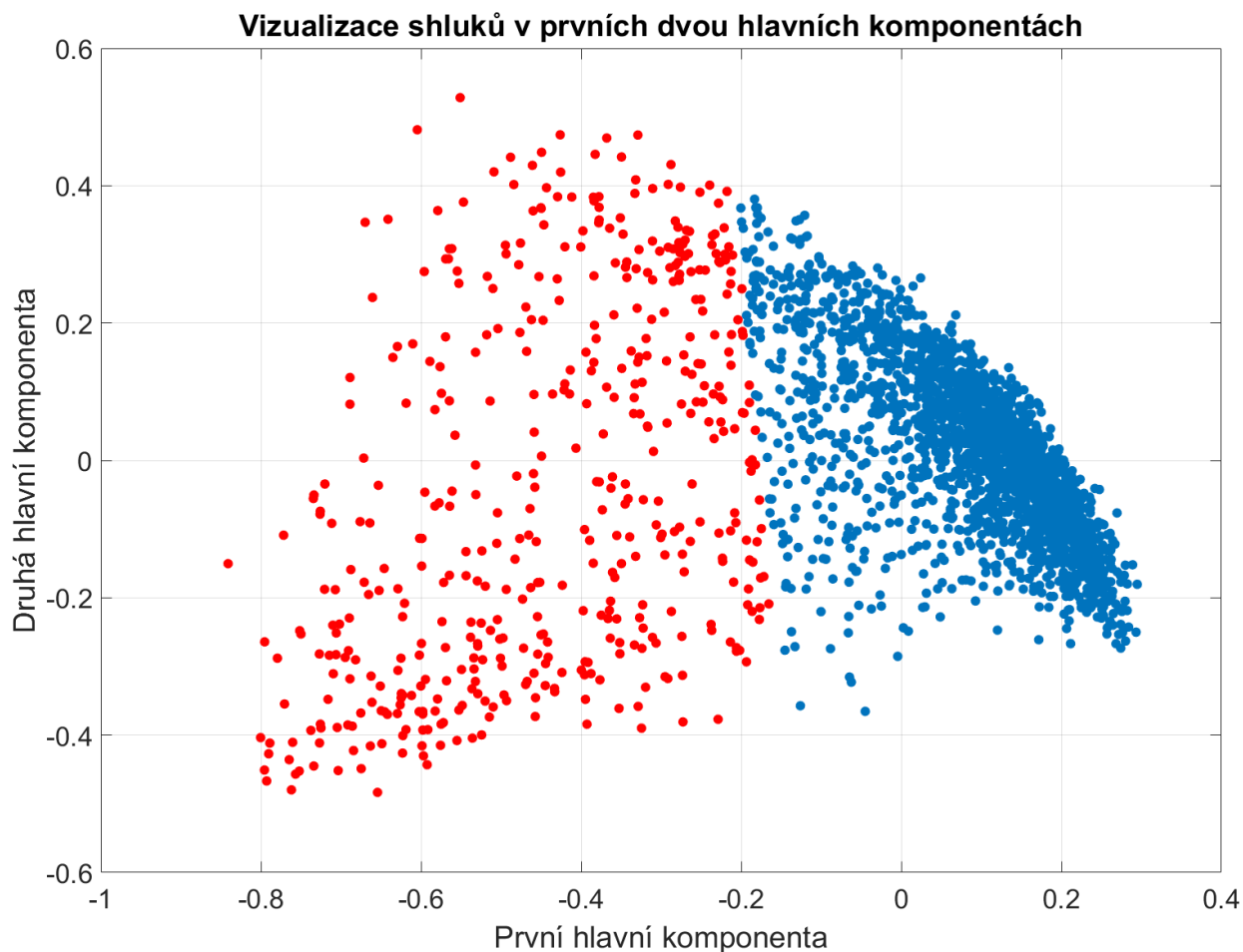
## 5.5 Aplikace k-means

Po průzkumové analýze jsem usoudil, že se ve frekvenčních charakteristikách naměřených hodnot vnitřních teplot nachází dominantní shluk reprezentující normální chování. Následně tedy tento shluk chci identifikovat pomocí metody *k-means* zavoláním matlabovské funkce *kmeans* (tab. 6).

Tabulka 6: Vstupy a výstupy funkce *kmeans*

Vstupy:	Výstupy:
Matice dat ( $m \times n$ )	Příslušnost jednotlivých pozorování k danému shluku ( $n \times 1$ )
Počet shluků $k$	Pozice centroidů jednotlivých shluků ( $k \times m$ )
Definovaná vzdálenost	Jednotlivé vzdálenosti ( $n \times k$ )
Limitní počet iterací	Součet vzdáleností od příslušného centroidu ( $k \times 1$ )
Počet opakování s různými počátečními odhady	-
Metoda určení počátečních odhadů	-

Z grafického znázornění (obr. 19) jsou tedy odděleny dva shluky. Jeden shluk s hustě uspořádanými objekty a druhý rozptýlenější. Do prvního shluku náleží většina objektů (cca 80 %), čímž ho klasifikuji jako shluk s normálními daty. Objekty příslušící druhému shluku považuji za podezřelé a vzdálené normálním.



Obrázek 19: Aplikace k-means a metody hlavních komponent.

## 5.6 Aplikace SOM

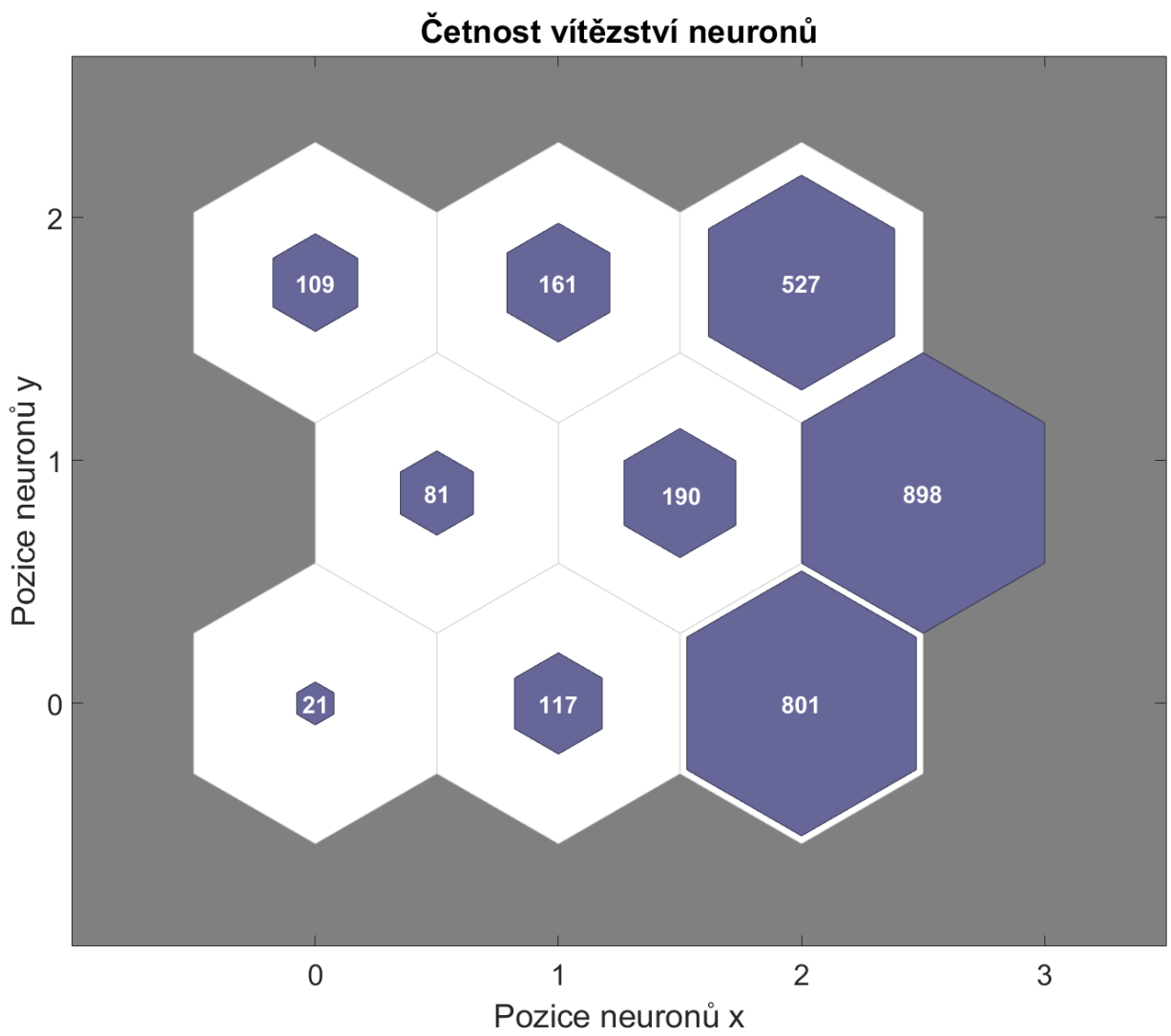
Aplikace Kohonenovy samoorganizační mapy (dále jen SOM) je realizováno opět skrze knihovnu Matlabu *selforgmap* (tab. 7). Vstupem je matice našich pozorování, dále je třeba konfigurovat rozměr mapy, který může být až třírozměrný. Já volím rovinu a velikost  $3 \times 3$ . Tento rozměr mapy jsem volil i na základě metody hlavních komponent a shlukové analýzy, z nichž jsem dospěl k závěru, že dominantní shluk bodů je jen jeden a podezřelé body jsou spíše rozptýlené. Z tohoto důvodu není nutné volit větší mapu s větším rozlišením, jelikož stačí klasifikovat pouze 9 neuronů. Dále jsem volil počet iterací (na 10 000 iterací) a zbytek nechal na základních hodnotách (tedy šestiúhelníková topologie, euklidovská vzdálenost a funkce sousedství rovna třem).

Tabulka 7: Vstupy a výstupy funkce *selfogmap* [15]

Vstupy:	Výstupy:
Matice dat ( $m \times n$ )	Příslušnost jednotlivých pozorování k daným neuronům ( $p_n \times n$ )
Rozměr mapy $l \times l = p_n$	Matice vah ( $p_n \times m$ )
Topologie mapy	-
Definovaná vzdálenost	-
Limitní počet iterací	-
Počáteční hodnota okolí f. sousedství	-

Výstupem je tedy matice vah neuronů a matice příslušnosti pozorování k daným vítězným neuronům. Na obr. 20 jsou vizualizovaná vítězství jednotlivých neuronů, z čehož jsou patrné i dominantní neurony, které zvítězili s největší četností. Tento obrázek je vhodné porovnat i s obr. 21, kde jsou znázorněny jednotlivé vzdálenosti mezi neurony. Jak vidno, je zřejmé, že dominantní neurony k sobě mají nejblíže a zbytek je víceméně rozprostřen s jedním výrazným vzdáleným pozorováním (neuron s pozicí [0 0]).

Aplikace SOM je velice podobná kombinaci aplikace metody hlavních komponent a shlukové analýzy. SOM vizualizovala do roviny vícerozměrný problém a určila vzdálenosti jednotlivých pozorování. Dá se tedy říci, že SOM je jakési zobecnění přístupu shlukové analýzy.



Obrázek 20: Aplikace SOM - četnost vítězství.

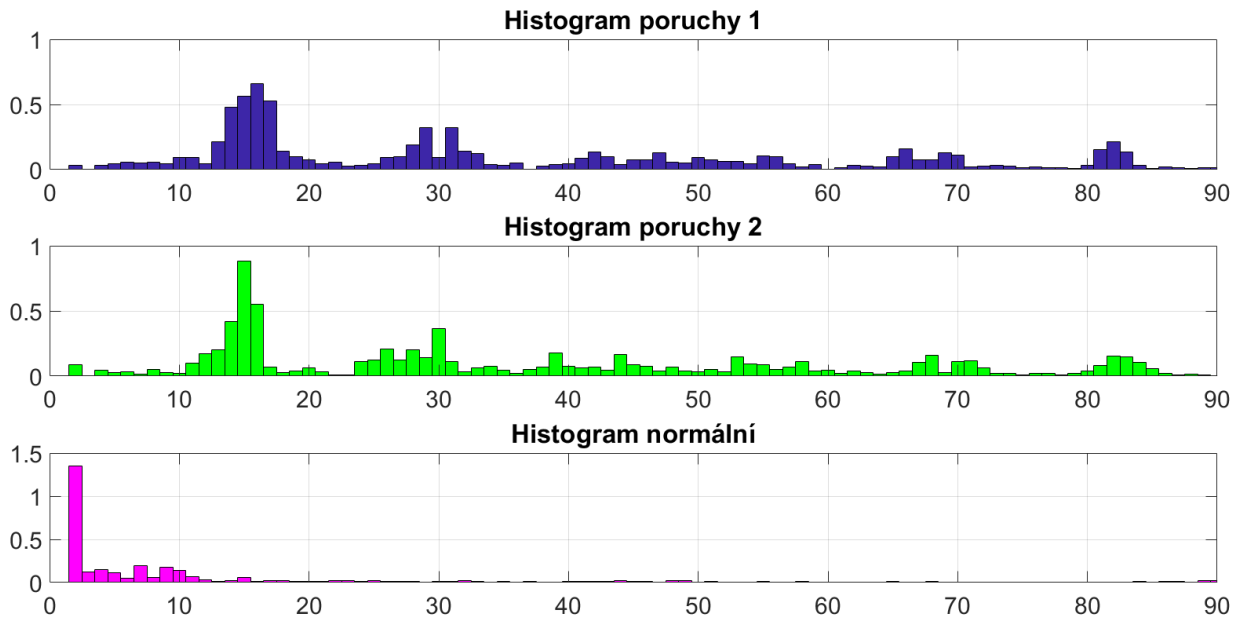


Obrázek 21: Aplikace SOM - vzdálenosti neuronů. Černá značí největší vzdálenost, žlutá nejmenší.

## 5.7 Aplikace korelace

Korelaci jsem aplikoval pouze pro porovnání výsledků. Korelaci Fourierových transformací lze aplikovat skrze srovnání modelové transformace s transformacemi zkoumaných dat. Modelová transformace může reprezentovat požadovanou charakteristiku veličiny, nebo například specifický druh poruchy. Tím lze realizovat filtr na základě frekvenční charakteristiky. Na obr. 22 jsou vidět příklady tří histogramů. Dvou poruchových a jednoho, který byl vyhodnocen jako normální. Koeficient průniku histogramů byl v případě srovnání prvních dvou roven

0,68 a v případě první poruchy a normálního histogramu 0,18.



Obrázek 22: Ukázka dvou histogramů poruchy s podobnou charakteristikou (první dva od shora) a histogramu normální charakteristiky (dole).

## 6 Návrh algoritmu

Pro celkovou analýzu dat skrze aplikaci vybraných metod, je potřeba tyto metody implementovat do algoritmu. Pro analýzu jsem vybral pouze část dodaných dat, a to všechny pondělí v roce 2017. Jeden pracovní den jsem vybral z důvodu odstranění dlouhodobých oscilací spojených s ročním obdobím a také oscilací vzniklých kvůli víkendům. Dále jsem se na základě doporučení odborníka z firmy ENERGOCENTRUM PLUS, s.r.o. zaměřil jen na parametry: Pump operation (činnost čerpadel), Indoor temperature (vnitřní teploty), Supply water temperature (teplota topné vody) a Energy meter (cumulative i actual) (výkon).

### 6.1 Výpočet vlastností signálu

Dle schématu na obr. 2 v kapitole č. 3 je porucha detekována jako odchylka od normálního chování. Normální chování určíme z vlastností signálu (tab. 2 v kapitole č. 3). Typický parametr pro porovnání je průměrná hodnota a dále například vzdálenost mezi extrémami signálu.



Tyto parametry se postupem času s nově nasbíranými daty také aktualizují. Tato práce je ale tvořena na základě offline dat, tudíž vypočte parametr jednou a v offline datech hledá poruchy. Je samozřejmě možné v budoucnu navrhnout i online detekci, kde by se tyto parametry normálního chování aktualizovaly. V tab. 8 jsou uvedeny parametry, které jsou určující pro detekci poruch a ke klasifikaci symptomů poruchy.

Tabulka 8: Parametry normálního chování

Název parametru:	Datový typ:
Průměrná hodnota veličiny	Double
Průměrný rozdíl extrémů	Double
Průměrná rychlost změny	Double
Průměrná dominantní frekvence	Double
Průměrný počet změn za jednotku času	Bool

Tyto parametry jsou spočteny pro každou pozorovanou veličinu v každém měsíci (samozřejmě pro příslušný datový typ).

## 6.2 Detekce oscilací v boolovských datech

Vzhledem k jednoduchosti binárních dat, stačí pro detekci oscilací sledovat rozdíly mezi sousedními prvky vektoru naměřených dat. Když změna nastane, rozdíl nabývá hodnot  $\pm 1$  (znaménko určí náběžnou resp. sestupnou hranu), jinak je hodnota rozdílu nulová. Celková hodnota počtu změn  $Z$  je dána jednoduchým vztahem (36), kde  $n$  je délka vektoru. Jedná se zkrátka o součet absolutních hodnot diferencí mezi jednotlivými prvky vektoru naměřených hodnot.

$$Z = \sum_{k=1}^{n-1} |x_k - x_{k-1}| \quad (36)$$

První kontrola poruch nastává zkontrolováním celkového počtu změn u každého čerpadla. Limit může nastavit odborník, případně lze sledovat čerpadla, která jsou velmi vzdálená od průměru. Nicméně velké číslo celkového počtu změn sice může značit poruchu, ale nemusí

tomu být vždy a je potřeba ještě zkontrolovat frekvenci oscilací čerpadla v kratším časovém období.

Pro kontrolu samotné limitní frekvence v kratším časovém období jsem navrhl kontrolu plovoucím oknem. Toto plovoucí okno prochází data a kontroluje počet změn v tomto okně. (alg. 1)

---

**Algoritmus 1** Plovoucí okno

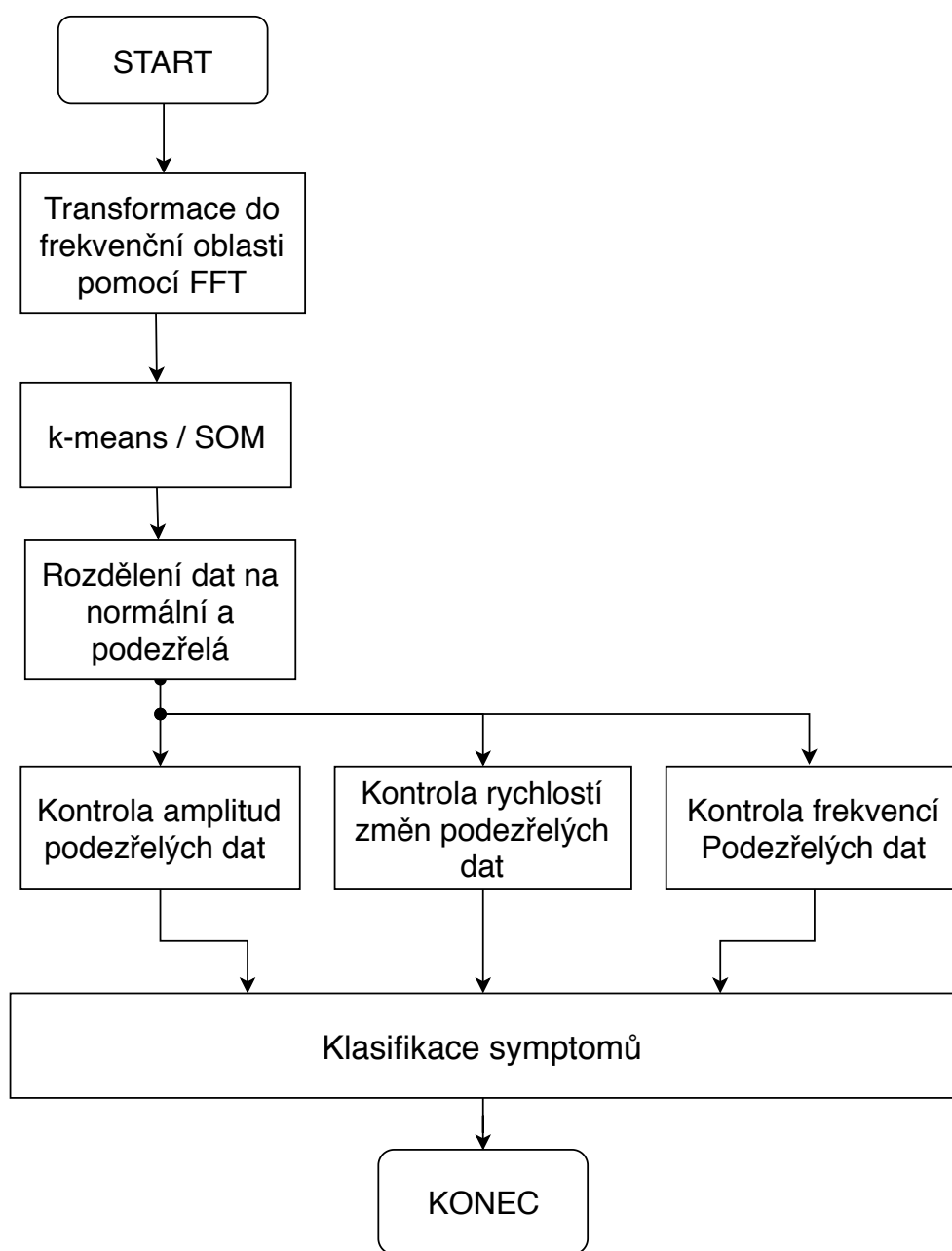
---

```
1: okno ← 12                                ▷ Nastavení velikosti okna
2: for  $n \leftarrow 1, \dots, L - \textit{okno}$  do           ▷ L značí délku dat
3:    $m \leftarrow n + \textit{okno}$ 
4:    $\textit{zmeny}(n) \leftarrow \textit{sum}(\textit{abs}(\textit{dif}f(\textit{data}(n : m))))$   ▷ Součet absolutních hodnot diferencí
```

---

### 6.3 Detekce oscilací v datech typu double

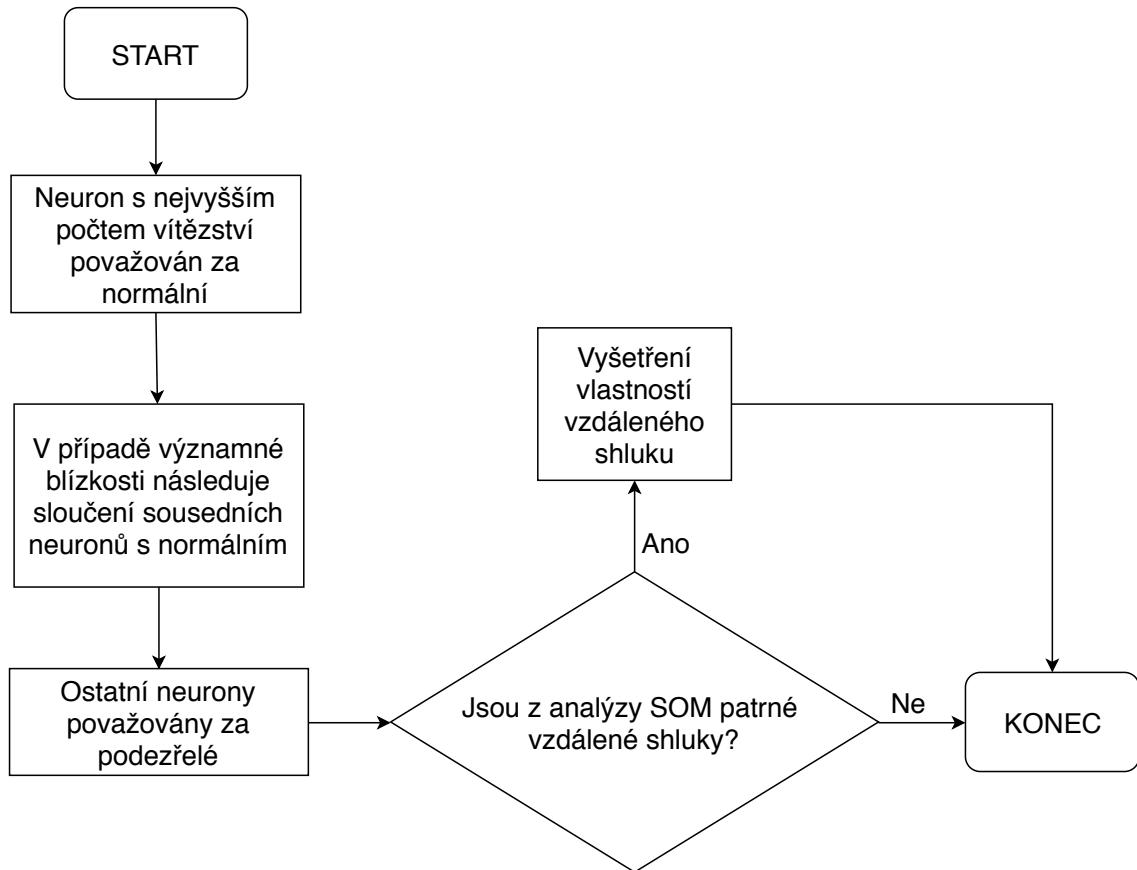
Všechna data naměřených teplot a výkonů jsou nejprve převedena do frekvenční oblasti algoritmem rychlé Fourierovy transformace. Tyto obrazy jsou nadále analyzovány třemi způsoby, které se nakonec porovnají. Prvním způsobem je shluková analýza pomocí *k - means*. Na základě průzkumové analýzy pomocí metody hlavních komponent jsem se rozhodl hledat pouze dva shluky, jeden dominantní, který považuji za normální a jeden vzdálený, který reprezentuje podezřelá data. Tato data, jichž je několiknásobně méně, jsou dále analyzována a roztříděna do třech kategorií. Jedna kategorie je prahová kontrola veličiny, kdy se vyšetřuje, zdali veličina nepřesáhla prahové hodnoty. Další kontrolou je prahová kontrola rychlosti změn veličiny a poslední kontrolou je vyšetření dominantních frekvencí dané veličiny. Dle těchto kontrol jsou klasifikovány tři základní symptomy poruch (přesažený práh, příliš rychlé změny, příliš vysoké frekvence) (obr. 23).



Obrázek 23: Schéma algoritmu detekce a klasifikace poruch.

Další metodou je nahrazení shlukové analýzy *k – means* Kohonenovou samoorganizační mapou (SOM). I v tomto případě se vybere skupina s největším zastoupením, v tomto případě tedy neuron s nejvyšším počtem vítězství. V případě, že by sousední neurony byly velice blízko vítěznému, je možné tyto neurony sloučit a považovat jejich pozorování za normální skupinu. Vzhledem k tomu, že SOM je velice dobrý vizualizační nástroj vícerozměrného problému, je možné pomocí SOM provádět i průzkumovou analýzu podobně jako je tomu

u metody hlavních komponent. Jeli tedy zratelná nějaká tendence shlukování vzdálených dat, je možné je identifikovat pomocí SOM (obr. 24). Následující postup je stejný jako u předchozí metody, tudíž podezřelá data jsou vyšetřena dle tří základních druhů symptomů.



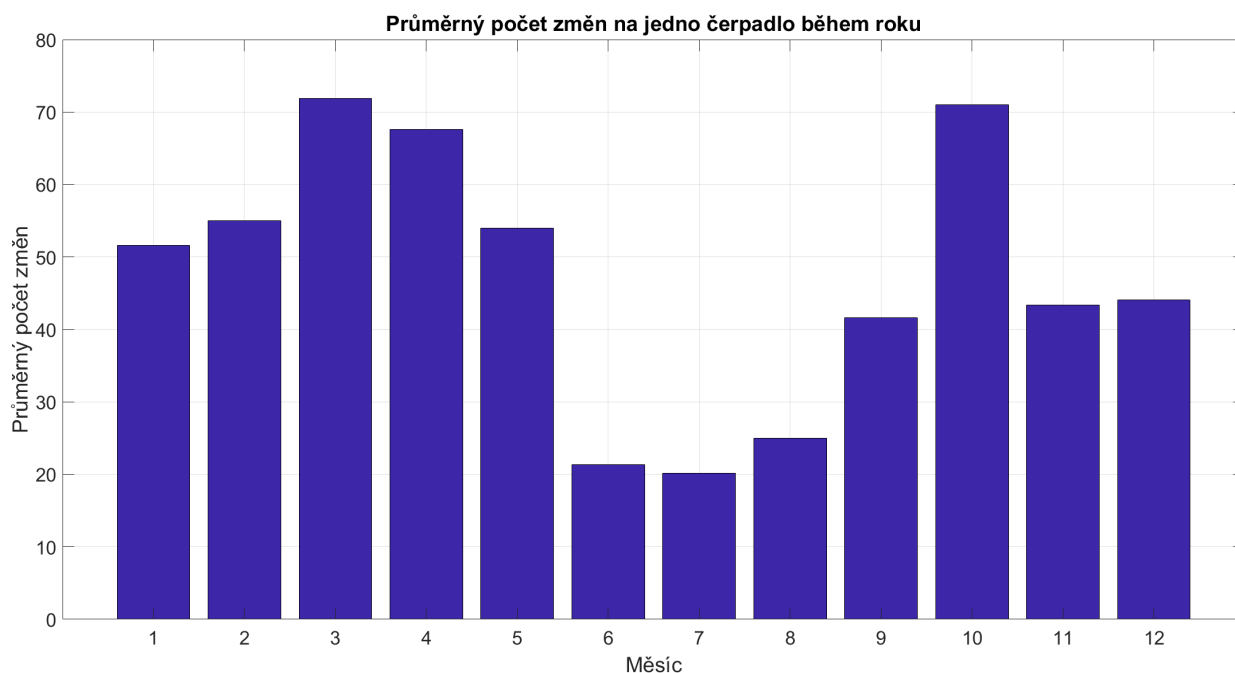
Obrázek 24: Algoritmus vyhodnocení podezřelých dat.

## 7 Výsledky navrženého algoritmu na dodaných datech

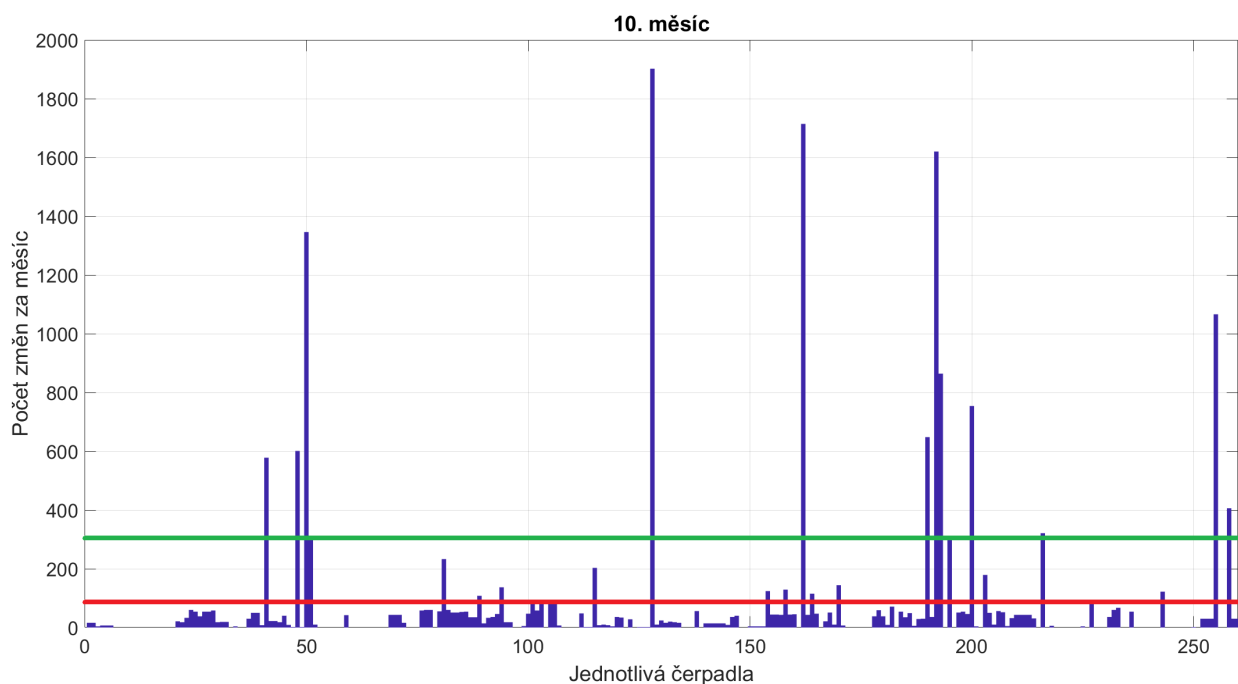
Algoritmus byl aplikován v prostředí Matlab. Výsledky jsou interpretovány graficky, či tabulkou. Na příloženém CD je možno nalézt soubor s indexy jednotlivých podezřelých pozorování. Číslo indexu značí pořadí sloupce v originálním excelovském souboru, index začíná od nuly, kde nultý sloupec je datum a čas. Vzhledem k tomu, že se data musí interpretovat po měsících, je vždy vybrán jen jeden příklad a zbytek výsledků se nachází v příloze. Výsledky v podobě grafů, jsou z důvodu velkého počtu obrázků přiloženy pouze na CD.

## 7.1 Výsledky boolovských dat

Na obr. 25 jsou znázorněny průměry počtu změn činnosti čerpadel. Největší četnost oscilací pochází z měsíce března a nejmenší z července. Na obr. 26 jsou znázorněny oscilace jednotlivých čerpadel v měsíci říjnu. Čerpadla nad zelenou čarou jsou označena za podezřelá. Z obrázku je vidět dvanáct výrazně vyčnívajících čerpadel, jejichž počet oscilací mnohonásobně převyšuje ostatní čerpadla.



Obrázek 25: Znázornění průměrného počtu změn jednotlivých čerpadel.



Obrázek 26: Znázornění počtů změn jednotlivých čerpadel v říjnu. Červeně - průměr, zeleně - směrodatná odchylka.

V tabulce 9 jsou uvedeny počty podezřelých čerpadel v každém měsíci. Krom celkového počtu v měsíci jsou uvedeny i výsledky algoritmu plovoucího okna, které vyhodnotilo v každém měsíci vždy více podezřelých čerpadel. Procentuálně se počet podezřelých pohybuje v řádu jednotek procent při vyhodnocení z celkového počtu změn v měsíci a kolem deseti procent při vyhodnocení plovoucím oknem.

Tabulka 9: Výsledky kontroly funkce čerpadel

Měsíc	Počet podezřelých	V procentech	Počet podezřelých - okno	V procentech
Leden	9	3,9 %	24	10,3 %
Únor	9	3,8 %	36	15,2 %
Březen	13	5,5 %	26	11 %
Duben	11	4,6 %	30	12,4 %
Květen	6	2,5 %	25	10,3 %
Červen	6	2,5 %	10	4 %
Červenec	6	2,4 %	10	3,9 %
Srpen	4	1,5 %	7	2,7 %
Září	9	3,5 %	26	10 %
Říjen	12	4,6 %	29	11,1 %
Listopad	12	4,7 %	27	10,6 %
Prosinec	12	4,8 %	30	12 %

## 7.2 Výsledky dat typu double

Největším zástupcem této skupiny jsou data vnitřních teplot. V tabulce 10 jsou pro představu uvedeny normální parametry tohoto signálu za celý rok. Pro vyhodnocení jsou ale důležité tyto parametry v každém měsíci (v příloze). Z tabulky lze vidět, že vnitřní teplota je v rámci roku poměrně stabilní (směrodatná odchylka 2,6 °C). Zbylé veličiny mají směrodatnou odchylku podstatně větší, z čehož lze usuzovat větší míra oscilace závisící na ročním období, což je, dle mého názoru, z podstaty funkce tohoto systému logické.

Tabulka 10: Parametry signálu

<b>Veličina</b>	<b>Průměrná hodnota</b>	<b>Směrodatná odchylka</b>	<b>Průměrná dominantní frekvence</b>	<b>Průměrná rychlost změn</b>
Vnitřní teplota	23,1 °C	2,6 °C	4,5 1/den	0,2 °C/min
Topná voda	34,2 °C	10,8 °C	7,5 1/den	2,4 °C/min
Výkon	6,3 kW	10 kW	2,6 1/h	0,8 kW/min

Výsledky v každém měsíci pro vnitřní teploty jsou uvedeny v tabulce 11. V tabulce jsou uvedeny výsledky jak shlukové analýzy, tak analýzy pomocí SOM. Na rozdíl od boolovských dat, zde je již procento podezření o mnoho vyšší - v řádech desítek procent. Metoda pomocí SOM je mírně přísnější a vyhodnocuje o jednotky procent vyšší počet podezřelých pozorování. Nejvíce podezření je v červenci, nejméně v prosinci dle k-means a v lednu dle SOM.



Tabulka 11: Výsledky kontroly vnitřních teplot

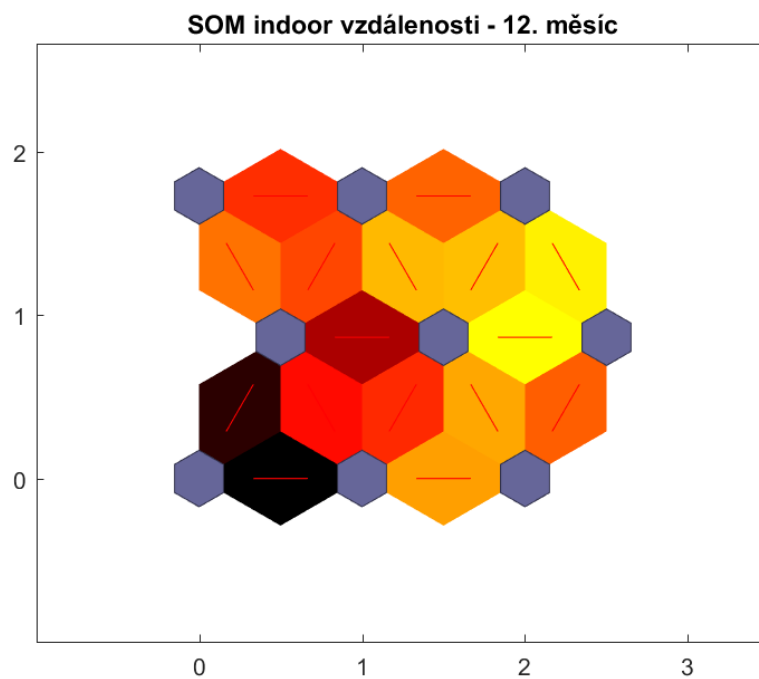
Měsíc	Počet podezřelých k-means	V procentech	Počet podezřelých SOM	V procentech
Leden	101	14,8 %	124	18,2 %
Únor	128	18,5 %	154	22,3 %
Březen	120	17,3 %	244	35,2 %
Duben	141	20,3 %	174	24,9 %
Květen	119	17 %	190	27,1 %
Červen	144	20,5 %	175	25 %
Červenec	207	30,4 %	242	35,7 %
Srpen	188	26,5 %	148	21 %
Září	136	19,1 %	149	20,9 %
Říjen	127	17,5 %	172	23,7 %
Listopad	129	15 %	223	26 %
Prosinec	128	10,7 %	404	34 %

Po vyhodnocení podezřelých pozorování následuje klasifikace dle tří základních symptomů (limit hodnot, limit rychlostí, limit frekvencí) (tab. 12). Hodnoty jsou zhruba rovnoměrně rozloženy a je nutné podotknout, že průměrně ve 29 % bylo klasifikováno jedno pozorování do všech tří kategorií. Za povšimnutí stojí prosinec s nejmenším zařazením do kategorií.

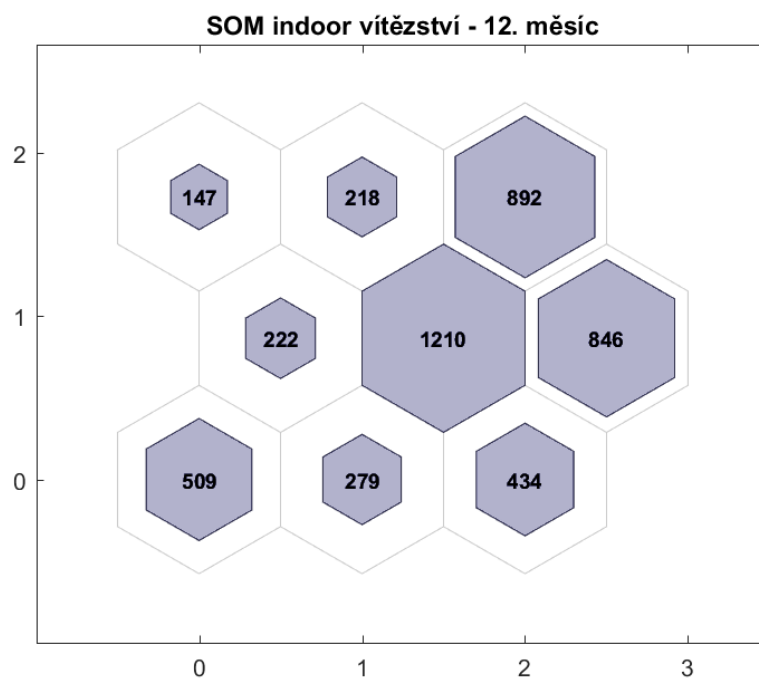
Tabulka 12: Výsledky klasifikace

Měsíc	Filtr teplot	Filtr rychlostí	Filtr frekvencí
Leden	26,3 %	30,5 %	33,8 %
Únor	27,4 %	30,1 %	31,7%
Březen	33,6 %	32,4 %	26,9 %
Duben	37,5 %	42,1 %	30,6 %
Květen	38,7 %	25 %	47,3 %
Červen	28,1 %	33,1 %	28,6 %
Červenec	40,4 %	32,9 %	25,1 %
Srpen	36,3 %	33,2 %	27,3 %
Září	40,6 %	38,2 %	30,2 %
Říjen	29,5 %	35,6 %	29,1 %
Listopad	28,5 %	31,2 %	27,7 %
Prosinec	0,6 %	3 %	0 %

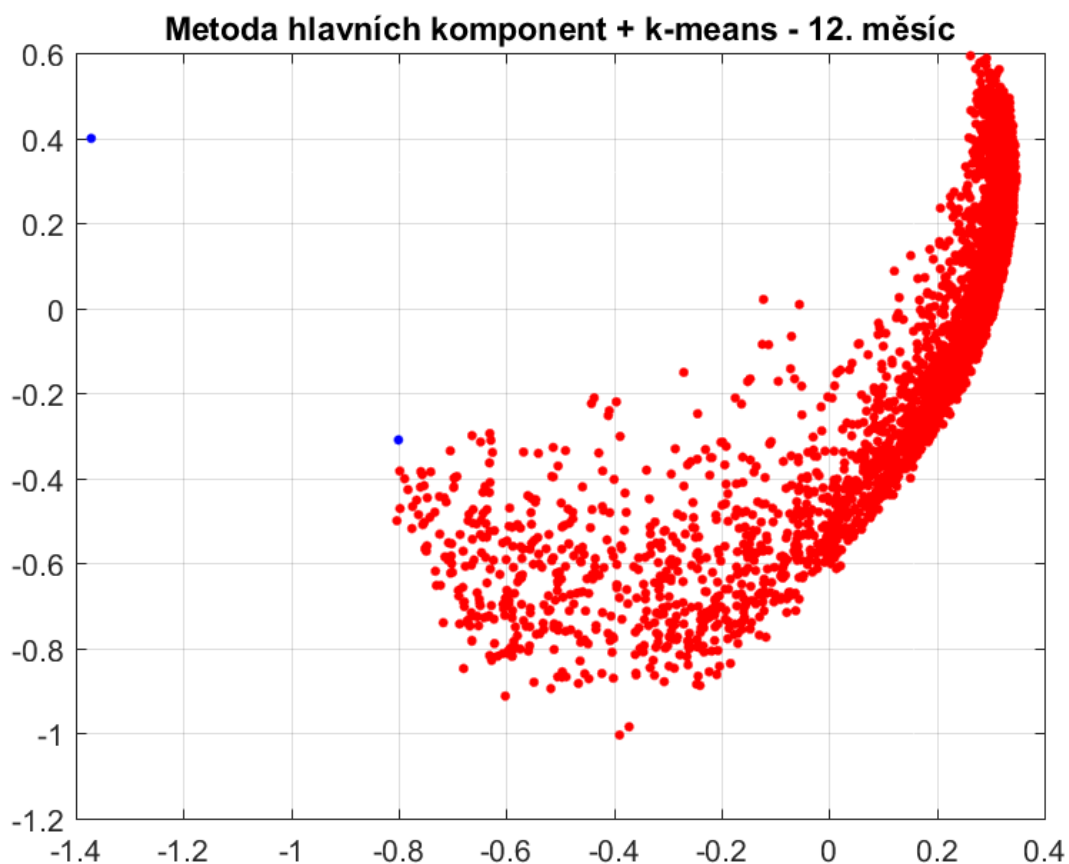
Pro ukázkou uvádím grafické výsledky SOM a shlukové analýzy z prosince (zbytek v příloze na CD). Prosinec jsem vybral jako zvláštní případ, jelikož jeho klasifikace ukázala zvláštní vlastnosti. Na obr. 27 jsou vidět zřejmé vzdálenosti mezi neurony (černá značí nejvzdálenější) a na obr. 28 četnost vítězství jednotlivých neuronů. Je zřejmý jeden velký blízký shluk a jeden malý vzdálený. Pro porovnání je uvedena i shluková analýza vizualizovaná metodou hlavních komponent do roviny (obr. 29). V ostatních případech se většina výsledků SOM a shlukové analýzy většinou shodovala. V prosinci ovšem shluková analýza označila za poruchy jen jeden menší shluk. Po vyšetření tohoto shluku vyšlo najevo, že se jedná o teploty s konstantním průběhem, tedy bez žádné změny. Tyto teploty se vyskytují ve všech měsících, jenom v prosinci získaly dominantní převahu. Z toho důvodu se klasifikace symptomů v prosinci blížila k nule.



Obrázek 27: Ukázka výsledků SOM z prosince - vzdálenosti mezi neurony.



Obrázek 28: Ukázka výsledků SOM z prosince - vítězství neuronů.



Obrázek 29: Ukázka výsledků shlukové analýzy vizualizované pomocí metody hlavních komponent z prosince. Normální data červeně, podezřelá modře.

Jak již bylo řečeno, ukázka všech výsledků by musela být interpretována po měsíci, tak zde alespoň uvedu roční průměrný počet podezřelých pozorování ve zbylých veličinách (tab. 13). Parametr Supply water temperature (teplota topné vody) má podobné výsledky jako Indoor temperature (vnitřní teplota). Parametr značící výkon (Energy meter) má podstatně nižší procento podezřelých pozorování.

Tabulka 13: Průměrný počet poruch u ostatních veličin.

Veličina	Průměrný počet poruch k-means	Průměrný počet poruch SOM
Supply water temperature	24,6 %	30,4 %
Energy meter	11,6 %	10,6 %

## 8 Závěr

Ve své diplomové práci jsem shrnul teorii o možných poruchách v technice prostředí budov a postupech obecné detekce poruch. Vycházel jsem z metody detekce poruch založené na vlastnostech samotného signálu bez znalosti modelu. Dále jsem nastudoval vybrané nástroje analýzy dat jako metoda hlavních komponent, shluková analýza pomocí k-means a Kohonenova samoorganizační mapa. Tyto nástroje sloužily pro analýzu frekvenčních charakteristik dat, které vycházely z převedení signálu z časové oblasti do oblasti frekvenční pomocí algoritmu rychlé Fourierovy transformace, což byl stěžejní nástroj pro detekci oscilací v datech.

Po teoretické části následuje aplikace shrnutých metod a průzkumová analýza dodaných dat od společnosti ENERGOCENTRUM PLUS, s.r.o. Celkem bylo dodáno dvanáct souborů, které po měsících zaznamenávaly údaje z celého roku 2017 se vzorkovací periodou 5 *min*. Data monitorovala až 2860 parametrů z devíti kategorií a dvou datových typů - bool a double. Data musela projít předzpracováním především kvůli chybějícím naměřeným hodnotám, což jsem vyřešil doplněním pomocí lineární interpolace (pro data typu double) a doplněním předchozí hodnoty v boolovských datech. Z průzkumové analýzy pomocí metody hlavních komponent a samoorganizační mapy vyšlo najevo, že se shlukují především normální data a data vzdálená se spíše rovnoměrně rozprostírají, z čehož jsem provedl návrh na rozdělení do dvou shluků, a to do shluku který obsahuje většinu pozorování a do zbytku rozprostřeného daleko od shluku prvního. Právě menšinová rozprostřená data jsou považována za podezřelá.

Na základě poznatků z průzkumové analýzy jsem navrhl finální algoritmus, který rozdělil data na normální a podezřelá pomocí dvou metod (k-means a SOM). Tato data jsou následně vyšetřena a zařazena do tří kategorií (překročení limitu veličiny, překročení rychlosti, vysoké frekvence ve Fourierově transformaci).

Aplikace mnou navrženého algoritmu odhalila v průměru 3,7 % podezřelých čerpadel pouze na základě počtu změn jejich činnosti v jednom měsíci a 9,5 % čerpadel po aplikaci plovoucího okna. Obecně bylo méně poruch čerpadel detekováno v létě a více v zimě. Aplikace na data typu double odhalila 19 % podezřelých pozorování vnitřních teplot (Indoor temperature) pomocí k-means a 26,1 % pomocí SOM, dále 24,6 % topné vody (Supply water temperature) pomocí k-means a 30,4 % pomocí SOM. Nakonec 11,6 % podezřelých dat výkonu (Energy meter) pomocí k-means a 10,6 % pomocí SOM. Podezřelá data byla

roztřívána do tří kategorií symptomů, a to: Přesah limitní hodnoty veličiny, přesah limitu rychlosti změny, rychlé frekvence ve Fourierových transformacích. Metoda SOM obecně klasifikovala více podezřelých pozorování, stejně tak metoda plovoucím oknem. Metoda SOM se obecně hodí v případě dalšího přidávání nových dat, jelikož se algoritmus nemusí projíždět od začátku, ale stačí SOM pouze doučit.

V této práci jsem detekoval poruchy v datech postavené na analýze oscilací v signálu. Výstupem jsou tedy indexované pozice v datech, které byly vyhodnoceny jako porucha. Tato data mohou být dále zkoumána z hlediska diagnostiky poruch.

## Literatura

- [1] WJN Turner, A Staino a B Basu. “Residential HVAC fault detection using a system identification approach”. In: *Energy and Buildings* 151 (2017), s. 1–17.
- [2] Matthew Wiggins a James Brodrick. “HVAC fault detection”. In: *ASHRAE Journal* 54.2 (2012), s. 78–80.
- [3] Richard Nový. *Technika prostředí*. Nakladatelství ČVUT, 2000. ISBN: 80-01-03492-5.
- [4] Mangesh Basarkar et al. *Modeling and simulation of HVAC faults in EnergyPlus*. Tech. zpr. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2011.
- [5] Sasa Kojic a Danfoss Trata d.o.o. *Reducing oscillations in a HVAC system*. Tech. zpr. 2011.
- [6] Rolf Isermann. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer, 2006. ISBN: 978-3-540-24112-6.
- [7] Józef Korbicz a kolektiv. *Fault Diagnosis: Models, Artificial Intelligence, Applications*. Springer, 2012. ISBN: 3-540-40767-7.
- [8] Alois Klíč, Karel Volka a Miroslava Dubcová. *Fourierova transformace - s příklady z infračervené spektroskopie*. VŠCHT v Praze, 2002. ISBN: 80-7080-478-5.
- [9] Viktor Beneš a Gejza Dohnal. *Pravděpodobnost a matematická statistika*. Nakladatelství ČVUT, 1993. ISBN: 80-01-00911-4.
- [10] Kent State University. *SPSS Tutorials: Pearson Correlation*. Dostupné z: <https://libguides.library.kent.edu/SPSS/PearsonCorr> (cit. 08.05.2019).
- [11] Petr Hebák a kolektiv. *Statistické myšlení a nástroje analýzy dat*. Informatorium, 2013. ISBN: 978-80-7333-105-4.
- [12] The MathWorks Inc. *Fast Fourier Transform*. Dostupné z: <https://www.mathworks.com/help/matlab/ref/fft.html> (cit. 08.05.2019).
- [13] “Zákon č. 361/2007 Sb.” In: *Sbírka zákonů 2007, částka 111* (). ISSN: 1211-1244.
- [14] The MathWorks Inc. *Principle Component Analysis*. Dostupné z: <https://www.mathworks.com/help/stats/pca.html> (cit. 08.05.2019).
- [15] The MathWorks Inc. *Self-organizing map*. Dostupné z: <https://www.mathworks.com/help/deeplearning/ref/selforgmap.html> (cit. 08.05.2019).

## A Příloha zbylých výsledků

Počet podezřelých parametru Supply water temperature

Měsíc	Počet podezřelých k-means	V procentech	Počet podezřelých SOM	V procentech
Leden	77	28,6 %	83	30,9 %
Únor	73	26,5 %	78	28,4 %
Březen	65	23,7 %	103	37,7 %
Duben	81	28,7 %	85	30,1 %
Květen	83	25,1 %	101	30,5 %
Červen	41	14,2 %	79	27,4 %
Červenec	52	17,7 %	79	26,7 %
Srpen	37	12,6 %	55	18,7 %
Září	82	27 %	78	26 %
Říjen	71	23 %	68	22,3 %
Listopad	105	34,8 %	109	36,3 %
Prosinec	99	33,4%	115	38,5 %



Počet podezřelých parametru Energy meter

Měsíc	Počet podezřelých k-means	V procentech	Počet podezřelých SOM	V procentech
Leden	40	11,7 %	40	11,8 %
Únor	38	11,1 %	59	17,2 %
Březen	32	9,4 %	63	18,6 %
Duben	36	10,4 %	33	9,8 %
Květen	36	10,8 %	29	8,9 %
Červen	35	11,8 %	12	4,1 %
Červenec	35	11,7 %	36	12,1 %
Srpen	40	13,2 %	39	12,9 %
Září	35	11,2 %	41	13 %
Říjen	36	10,3 %	34	9,6 %
Listopad	48	12 %	57	14,3 %
Prosinec	67	15 %	53	12 %

Výsledky klasifikace parametru Supply water temperature

<b>Měsíc</b>	<b>Filtr teplot</b>	<b>Filtr rychlostí</b>	<b>Filtr frekvencí</b>
Leden	10,7 %	11,1 %	10,3 %
Únor	10,6 %	10,6 %	9,8%
Březen	8,6 %	8,9 %	7,8 %
Duben	9,9 %	10,1 %	8,7 %
Květen	9,4 %	9,6 %	10,1 %
Červen	11 %	11,2 %	10,2 %
Červenec	11,2 %	11,4 %	10,8 %
Srpen	11,8 %	12 %	11 %
Září	10,5 %	10,5 %	8,5 %
Říjen	9,6 %	9,9 %	9,1 %
Listopad	11,5 %	11,7 %	11,1 %
Prosinec	12 %	12,4 %	13,2 %

Výsledky klasifikace parametru Energy meter

Měsíc	Filtr výkonu	Filtr rychlostí	Filtr frekvencí
Leden	10,1 %	5,9 %	17,6 %
Únor	9,9 %	6,5 %	15,1 %
Březen	9,6 %	5,8 %	12,1 %
Duben	13,4 %	8 %	12 %
Květen	14,8 %	10,2 %	10,1 %
Červen	3,7 %	2,8 %	4,1 %
Červenec	4,3 %	3,4 %	5,3 %
Srpen	3,7 %	2,8 %	4,1 %
Září	14,3 %	8,5 %	7,5 %
Říjen	9,2 %	6,5 %	13,1 %
Listopad	10,5 %	6,8 %	19,4 %
Prosinec	12,6 %	9,1 %	20,8 %