

**ZÁPADOČESKÁ UNIVERZITA V PLZNI
FAKULTA ELEKTROTECHNICKÁ**

Katedra technologií a měření

DIPLOMOVÁ PRÁCE

**Porovnání možností a výsledků analýzy časových řad ve
statistických SW STATISTICA a NCSS**

Tomáš Linhart

Plzeň 2015

ZÁPADOČESKÁ UNIVERZITA V PLZNI

Fakulta elektrotechnická

Akademický rok: 2014/2015

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: Bc. Tomáš LINHART
Osobní číslo: E12N0019P
Studijní program: N2612 Elektrotechnika a informatika
Studijní obor: Komerční elektrotechnika
Název tématu: Porovnání možností a výsledků analýzy časových řad ve statistických SW STATISTICA a NCSS
Zadávací katedra: Katedra technologií a měření

Zásady pro vypracování:

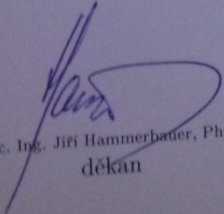
1. Vypracujte přehled metod, které se používají pro statistické vyhodnocování časových řad.
2. Definujte specifické problémy, které se objevují při analyzování časových řad.
3. Analyzujte možnosti statistických SW STATISTICA a NCSS v oblasti analýzy časových řad.
4. Proveďte analýzu časových řad z dodaných výsledků materiálových zkoušek. Analýzu proveďte oběma SW.
5. Porovnejte možnosti a výsledky provedených analýz a vyvoďte závěry.

Rozsah grafických prací: podle doporučení vedoucího
Rozsah pracovní zprávy: 30 - 40 stran
Forma zpracování diplomové práce: tištěná/elektronická
Seznam odborné literatury:

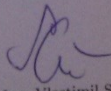
1. Kožíšek, J.: Statistická analýza, nakladatelství ČVUT, 1997
2. Parr Rud, O.: Data mining, Computer Press, 2008.
3. manuály k SW

Vedoucí diplomové práce: Ing. Lukáš Kupka, Ph.D.
Katedra technologií a měření
Konzultant diplomové práce: Ing. Martin Motyčka
Katedra technologií a měření

Datum zadání diplomové práce: 15. října 2014
Termín odevzdání diplomové práce: 11. května 2015


Doc. Ing. Jiří Hammerbauer, Ph.D.
děkan




Doc. Ing. Vlastimil Skočil, CSc.
vedoucí katedry

V Plzni dne 15. října 2014

Abstrakt

Předkládaná diplomová práce se zabývá porovnáním možností a výsledků analýzy časových řad ve statistických SW STATISTICA a NCSS. První kapitola této diplomové práce popisuje základní vlastnosti časových řad a specifické problémy, které se objevují při jejich analýze. Druhá kapitola popisuje vybrané metody, které se používají pro statistické vyhodnocování časových řad. Jsou to metody: dekompozice časových řad, spektrální analýza a Box-Jenkinsova metodologie. Třetí kapitola se zabývá porovnáním statistických SW STATISTICA a NCSS v oblasti analýzy časových řad. Ve čtvrté kapitole je provedena analýza časových řad z výsledků materiálových zkoušek tří různých senzorů. Tato analýza je provedena oběma SW.

Klíčová slova

Časová řada, SW STATISTICA, SW NCSS, Box-Jenkinsova metodologie, spektrální analýza, dekompozice časových řad.

Abstract

Presented diploma thesis is dealing with comparison of possibilities and results of time series analysis in statistical SW STATISTICA and NCSS. First chapter of this diploma thesis describes basic properties of time series and specific issues which appear during time series analysis. Second chapter describes selected methods which are used for statistical evaluation of time series. These methods are: decomposition of time series, spectral analysis and Box-Jenkins methodology. Third chapter is dealing with comparison of statistical SW STATISTICA and NCSS in field of time series analysis. In fourth chapter is performed analysis of time series from results of material tests of three various sensors. This analysis is performed by both SW.

Key words

Time serie, SW STATISTICA, SW NCSS, Box-Jenkins methodology, spectral analysis, decomposition of time series

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně s použitím odborné literatury a pramenů uvedených v seznamu, který je součástí této diplomové práce.

Dále prohlašuji, že veškerý software, použitý při řešení této diplomové práce, je legální.

Tomáš Linhart

V Plzni dne 11. 5. 2015

podpis:

Poděkování

Tímto bych rád poděkoval vedoucímu diplomové práce Ing. Lukáši Kupkovi, Ph.D. za profesionální rady, připomínky a metodické vedení práce.

Obsah

OBSAH	8
SEZNAM OBRÁZKŮ	10
SEZNAM TABULEK	11
SEZNAM SYMBOLŮ A ZKRATEK	12
ÚVOD	13
1 ČASOVÉ ŘADY	14
1.1 DĚLENÍ ČASOVÝCH ŘAD:	14
1.2 SLOŽKY ČASOVÉ ŘADY	14
1.3 CÍLE ANALÝZY ČASOVÝCH ŘAD	15
1.4 SPECIFICKÉ PROBLÉMY PŘI ANALÝZE ČASOVÝCH ŘAD	15
1.4.1 Problémy s volbou časových bodů	15
1.4.2 Problémy s kalendářem	15
1.4.3 Problémy s délkou časových řad	16
1.4.4 Problémy s nesrovnalostí jednotlivých měření	16
2 METODY PRO STATISTICKÉ VYHODNOCOVÁNÍ ČASOVÝCH ŘAD	17
2.1 DEKOMPOZICE ČASOVÝCH ŘAD	17
2.1.1 Typy modelů časových řad	17
2.1.2 Trend	18
2.1.2.1 Klouzavé průměry	18
2.1.2.2 Klouzavé mediány	18
2.1.2.3 4253H filtr	18
2.1.2.4 Jednoduché exponenciální vyrovnávání	19
2.1.2.5 Určení trendu	21
2.2 METODA CENSUS I	22
2.3 METODA X-11 CENSUS II	23
2.4 SPEKTRÁLNÍ ANALÝZA ČASOVÝCH ŘAD	24
2.4.1 Základní zápis a principy spektrální analýzy	25
2.4.2 Příprava dat pro spektrální analýzu	29
2.4.3 Testování náhodné posloupnosti	29
2.5 BOX – JENKINSOVA METODOLOGIE	29
2.5.1 Základní pojmy Box-Jenkinsovy metodologie	30
2.5.1.1 Stacionarita	30
2.5.1.2 Autokorelační funkce	31
2.5.1.3 Parciální autokorelační funkce	33
2.5.1.4 Lineární proces	34
2.5.2 Autoregresní proces	35
2.5.3 Proces klouzavých součtů	35
2.5.4 ARIMA metodologie	36
2.5.4.1 Identifikace modelu	36
2.5.4.2 Odhad parametrů modelu	37
2.5.4.3 Ověřování modelu	38
2.5.5 Sezónní modely	38

3	POROVNÁNÍ STATISTICKÝCH SW STATISTICA A NCSS V OBLASTI ANALÝZY ČASOVÝCH ŘAD	39
3.1	SW STATISTICA	39
3.1.1	ARIMA metodologie	39
3.1.2	Dekompozice časových řad	40
3.1.3	Sezónní a nesezónní exponenciální vyhlazování	40
3.1.4	Spektrální analýza	40
3.1.5	Zobrazení výsledků	41
3.2	SW NCSS	41
3.2.1	ARIMA metodologie	41
3.2.2	Dekompozice a předpověď časových řad	42
3.2.3	Exponenciální vyhlazování	42
3.2.4	Spektrální analýza	42
3.2.5	Zobrazení výsledků	42
3.2.6	Porovnání SW STATISTICA a NCSS	43
4	ANALÝZA ČASOVÝCH ŘAD Z VÝSLEDKŮ MATERIÁLOVÝCH ZKOUŠEK	44
4.1	TEXTILNÍ SENZOR	45
4.1.1	SW STATISTICA	46
4.1.2	SW NCSS	51
4.2	SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS	55
4.2.1	SW STATISTICA	56
4.2.2	SW NCSS	62
4.3	SENZOR - BI2 + SÍTOTISK	66
4.3.1	SW Statistica	67
4.3.2	SW NCSS	73
4.4	VÝSLEDKY A POROVNÁNÍ PROVEDENÝCH ANALÝZ	76
	ZÁVĚR	78
	SEZNAM POUŽITÉ LITERATURY	79

Seznam obrázků

OBR. 2.1 UKÁZKA NESTACIONÁRNÍ ČASOVÉ ŘADY	31
OBR. 2.2 UKÁZKA STACIONÁRNÍ ČASOVÉ ŘADY (NESTACIONÁRNÍ ČASOVÁ ŘADA PŘEVEDENA NA ŘADU STACIONÁRNÍ)	31
OBR. 2.3 UKÁZKA KORELOGRAMU AUTOKORELAČNÍ FUNKCE	32
OBR. 2.4 UKÁZKA KORELOGRAMU PARCIÁLNÍ AUTOKORELAČNÍ FUNKCE	34
OBR. 4.1 GRAF ZMĚNY IMPEDANCE A RELATIVNÍ VLHKOSTI - TEXTILNÍ SENZOR	45
OBR. 4.2 ČASOVÁ ŘADA - TEXTILNÍ SENZOR – SW STATISTICA	46
OBR. 4.3 VYHLAZENÁ ČASOVÁ ŘADA - TEXTILNÍ SENZOR – SW STATISTICA	46
OBR. 4.4 ACF ČASOVÉ ŘADY - TEXTILNÍ SENZOR – SW STATISTICA	47
OBR. 4.5 JEDNOU NESEZÓNĚ DIFERENCOVANÁ ČASOVÁ ŘADA - TEXTILNÍ SENZOR – SW STATISTICA	47
OBR. 4.6 JEDNOU NESEZÓNĚ A JEDNOU SEZÓNĚ DIFERENCOVANÁ ČASOVÁ ŘADA - TEXTILNÍ SENZOR – SW STATISTICA	48
OBR. 4.7 NASTAVENÍ ARIMA MODELU - TEXTILNÍ SENZOR – SW STATISTICA	48
OBR. 4.8 PRODLOUŽENÁ ČASOVÁ ŘADA - TEXTILNÍ SENZOR – SW STATISTICA	49
OBR. 4.9 PŘEDPOVĚĎ ČASOVÉ ŘADY - TEXTILNÍ SENZOR - SW STATISTICA	49
OBR. 4.10 HISTOGRAM REZIDUÍ - TEXTILNÍ SENZOR – SW STATISTICA	50
OBR. 4.11 NORMÁLNÍ ROZDĚLENÍ REZIDUÍ - TEXTILNÍ SENZOR – SW STATISTICA	50
OBR. 4.12 ACF REZIDUÍ - TEXTILNÍ SENZOR – SW STATISTICA	51
OBR. 4.13 PACF REZIDUÍ - TEXTILNÍ SENZOR – SW STATISTICA	51
OBR. 4.14 ACF ČASOVÉ ŘADY - TEXTILNÍ SENZOR - SW NCSS	52
OBR. 4.15 PACF JEDNOU NESEZÓNĚ A JEDNOU SEZÓNĚ DIFERENCOVANÉ ČASOVÉ ŘADY - TEXTILNÍ SENZOR - SW NCSS	52
OBR. 4.16 NASTAVENÍ ARIMA MODELU - TEXTILNÍ SENZOR – SW NCSS	53
OBR. 4.17 PŘEDPOVĚĎ ČASOVÉ ŘADY - TEXTILNÍ SENZOR - SW NCSS	53
OBR. 4.18 ACF REZIDUÍ - TEXTILNÍ SENZOR - SW NCSS	54
OBR. 4.19 HISTOGRAM REZIDUÍ - TEXTILNÍ SENZOR - SW NCSS	54
OBR. 4.20 GRAF ZMĚNY IMPEDANCE A RELATIVNÍ VLHKOSTI - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS	55
OBR. 4.21 ČASOVÁ ŘADA - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	56
OBR. 4.22 LOGARITMOVANÁ ČASOVÁ ŘADA - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	56
OBR. 4.23 ACF LOGARITMOVANÉ ČASOVÉ ŘADY - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	57
OBR. 4.24 JEDNOU NESEZÓNĚ DIFERENCOVANÁ LOGARITMOVANÁ ČASOVÁ ŘADA - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	57
OBR. 4.25 PACF JEDNOU NESEZÓNĚ DIFERENCOVANÉ LOGARITMOVANÉ ČASOVÉ ŘADY - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	58
OBR. 4.26 JEDNOU NESEZÓNĚ A JEDNOU SEZÓNĚ DIFERENCOVANÁ LOGARITMOVANÁ ČASOVÁ ŘADA - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	58
OBR. 4.27 NASTAVENÍ ARIMA MODELU - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	59
OBR. 4.28 PRODLOUŽENÁ ČASOVÁ ŘADA - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	60
OBR. 4.29 PŘEDPOVĚĎ ČASOVÉ ŘADY - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	61
OBR. 4.30 GRAF REZIDUÍ - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	61
OBR. 4.31 HISTOGRAM REZIDUÍ - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	62
OBR. 4.32 NORMÁLNÍ ROZDĚLENÍ REZIDUÍ - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW STATISTICA	62
OBR. 4.33 ACF LOGARITMOVANÉ ČASOVÉ ŘADY - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW NCSS	63
OBR. 4.34 PACF LOGARITMOVANÉ ČASOVÉ ŘADY - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW NCSS	63
OBR. 4.35 NASTAVENÍ ARIMA MODELU - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW NCSS	64
OBR. 4.36 PŘEDPOVĚĎ ČASOVÉ ŘADY - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW NCSS	65
OBR. 4.37 ACF REZIDUÍ - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW NCSS	65
OBR. 4.38 HISTOGRAM REZIDUÍ - SENZOR - FLEXIBILNÍ SUBSTRÁT + PEDOT:PSS – SW NCSS	65
OBR. 4.39 GRAF ZMĚNY IMPEDANCE A RELATIVNÍ VLHKOSTI - SENZOR - BI2 + SÍTOTISK	66
OBR. 4.40 ČASOVÁ ŘADA - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	67
OBR. 4.41 LOGARITMOVANÁ ČASOVÁ ŘADA - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	67
OBR. 4.42 ACF LOGARITMOVANÉ ČASOVÉ ŘADY - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	68
OBR. 4.43 JEDNOU NESEZÓNĚ DIFERENCOVANÁ ČASOVÁ ŘADA - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	68

OBR. 4.44 PACF JEDNOU NESEZÓNNE DIFERENCOVANÉ ČASOVÉ ŘADY - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	69
OBR. 4.45 NASTAVENÍ ARIMA MODELU - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	69
OBR. 4.46 PRODLOUŽENÁ ČASOVÁ ŘADA - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	70
OBR. 4.47 PŘEDPOVĚĎ ČASOVÉ ŘADY - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	71
OBR. 4.48 HISTOGRAM REZIDUÍ - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	71
OBR. 4.49 NORMÁLNÍ ROZDĚLENÍ REZIDUÍ - SENZOR - BI2 + SÍTOTISK – SW STATISTICA.....	72
OBR. 4.50 ACF REZIDUÍ - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	72
OBR. 4.51 PACF REZIDUÍ - SENZOR - BI2 + SÍTOTISK – SW STATISTICA	72
OBR. 4.52 ACF LOGARITMOVANÉ ČASOVÉ ŘADY - SENZOR - BI2 + SÍTOTISK – SW NCSS.....	73
OBR. 4.53 PACF JEDNOU NESEZÓNNE DIFERENCOVANÉ LOGARITMOVANÉ ČASOVÉ ŘADY - SENZOR - BI2 + SÍTOTISK – SW NCSS.....	73
OBR. 4.54 NASTAVENÍ ARIMA MODELU - SENZOR - BI2 + SÍTOTISK – SW NCSS	74
OBR. 4.55 PŘEDPOVĚĎ ČASOVÉ ŘADY - SENZOR - BI2 + SÍTOTISK – SW NCSS.....	75
OBR. 4.56 ACF REZIDUÍ - SENZOR - BI2 + SÍTOTISK – SW NCSS	75
OBR. 4.57 HISTOGRAM REZIDUÍ - SENZOR - BI2 + SÍTOTISK – SW NCSS	75

Seznam tabulek

TAB. 2.1 ZÁKLADNÍ MODELY A CHARAKTERISTIKY ACF A PACF	37
---	----

Seznam symbolů a zkratek

Y_t [-].....	pozorovaná hodnota časové řady v čase t
T_t [-].....	trendová složka časové řady v čase t
S_t [-].....	sezónní složka časové řady v čase t
C_t [-].....	cyklická složka časové řady v čase t
ε_t [-].....	náhodná, chybová nebo nepravidelná složka časové řady v čase t
α [-].....	parametr exponenciálního vyrovnávání
a_k [-].....	kosinusový parametr
b_k [-].....	sinusový parametr
P_k [-].....	hodnota periodogramu pro danou frekvenci
N [-].....	počet pozorování
ρ_k [-].....	hodnota autokorelační funkce
ρ_{kk} [-].....	hodnota parciální autokorelační funkce
φ_x [-].....	autoregresní parametry
θ_x [-].....	parametry klouzavých součtů
ACF.....	autokorelační funkce
PACF.....	parciální autokorelační funkce
SW.....	software

Úvod

S pojmem časová řada se můžeme setkat například ve fyzice, medicíně, ekonomii, metrologii a mnoha dalších oborech. Je to posloupnost hodnot určitého ukazatele, která je chronologicky uspořádána. Mnoho informací má právě tento charakter, a proto je nutné na ně nahlížet právě jako na časové řady.

S rostoucím vývojem výpočetní techniky se k analyzování časových řad často využívá různých statistických SW.

Cílem analýzy časových řad je nalezení modelu (mechanismu), podle kterého jsou generována data časové řady. Na základě znalosti tohoto modelu je možné předpovídat budoucí vývoj daného systému a částečně ho řídit a optimalizovat.

Cílem této diplomové práce je vypracovat přehled metod, které se používají pro statistické vyhodnocování časových řad a analyzovat možnosti statistických SW STATISTICA a NCSS v oblasti analýzy časových řad.

Stěžejní částí této práce je analýza časových řad z výsledků materiálových zkoušek tří různých senzorů. Tyto senzory jsou vyvíjeny na Západočeské univerzitě v Plzni, Fakultě elektrotechnické, Katedře technologií a měření. Analýza je provedena oběma statistickými SW pomocí Box-Jenkinsovy metodologie (ARIMA modely). Cílem je nalézt a ověřit model, který konkrétní časovou řadu nejlépe popisuje. Dále za pomoci tohoto modelu provést předpověď vývoje časové řady. Závěrem shrnout a porovnat výsledky provedených analýz pomocí obou SW.

1 Časové řady

Časová řada je posloupnost hodnot určitého ukazatele, která je chronologicky uspořádána. Někdy se také uvažuje uspořádání prostorové.

1.1 Dělení časových řad:

Časové řady lze dělit z několika hledisek.

- a) podle náhodné složky
 - deterministické – neobsahují náhodnou složku
 - stochastické – obsahují náhodnou složku
- b) podle časových intervalů
 - ekvidistantní – mají stejné časové intervaly
 - neekvidistantní – nemají stejné časové intervaly
- c) podle doby trvání
 - dlouhodobé
 - krátkodobé
- d) podle doby zachycení sledované veličiny
 - intervalové – hodnota veličiny je kumulovaná za celé období
 - okamžikové – hodnota veličiny je aktuální (okamžiková)[1]

1.2 Složky časové řady

Obecně lze časovou řadu rozložit na čtyři složky.

- a) trendové složky – T_t
- b) sezónní složky – S_t
- c) cyklické složky – C_t
- d) náhodné, chybové nebo nepravidelné složky – ε_t

Trendová složka představuje systematickou lineární nebo nelineární (polynomiální, exponenciální, ...) změnu během sledovaného časového úseku, která se neopakuje. Sezónní složka je odchylka časové řady od její trendové složky a periodicky se opakuje. Rozdíl mezi sezónní a cyklickou složkou je ten, že sezónní složka se vyskytuje v pravidelných intervalech (délka intervalu do jednoho roku), zatímco cyklická složka má delší dobu trvání (několik let) a její délka se mění většinou s každým cyklem. Náhodná složka obsahuje nesystematické odchylky a výkyvy. [2]

1.3 Cíle analýzy časových řad

Cílem analýzy časových řad je nalezení modelu (mechanismu), podle kterého jsou generována data časové řady. Na základě znalosti tohoto modelu je možné předpovídat budoucí vývoj daného systému a částečně ho řídit a optimalizovat pomocí vhodně zvolených hodnot vstupních parametrů. [3]

1.4 Specifické problémy při analýze časových řad

Při analýze časových řad se mohou objevit některé specifické problémy. Ty lze shrnout do následujících 4 bodů.

- a) problémy s volbou časových bodů
- b) problémy s kalendářem
- c) problémy s délkou časových řad
- d) problémy s nesrovnalostí jednotlivých měření

1.4.1 Problémy s volbou časových bodů

Časové řady mohou být tvořeny pozorováními v určitých nespojitých časových bodech. Jsou to tzv. diskrétní časové řady. Ty mohou vznikat třemi způsoby.

- a) Jsou diskrétní již svou povahou.
- b) Vznikají diskretizací spojité časové řady.
- c) Vznikají akumulací nebo průměrováním hodnot za dané časové období.

Někdy není možnost volby časových bodů, ale pokud tato možnost existuje, tak je nutné jí věnovat pozornost. Pokud by časové body nebyly zvoleny správně, tak by daná časová řada nemusela zachytit skutečný průběh sledovaného děje. [4]

1.4.2 Problémy s kalendářem

Za většinu těchto problémů je zodpovědný sám člověk. Problémy vznikají tím, že je různá délka kalendářních měsíců, různý počet pracovních dnů v měsíci, čtyři nebo pět víkendů v měsíci, pohyblivé svátky a podobně.

Dalším problémem, za který člověk nemůže, je délka slunečního roku, která není celočíselná.

Tyto anomálie mohou způsobovat při analýze časových řad nezanedbatelné následky a je tedy někdy nutné provést korekci dat. Těchto korekcí je celá řada. Možností je například použití měsíce o délce 30 dnů, vyrovnání různého počtu pracovních dnů nebo některé nepravidelnosti mohou být odstraněny pomocí akumulace dat. [4]

1.4.3 Problémy s délkou časových řad

Tento problém je dvojího typu. Může se stát, že daná časová řada je příliš krátká nebo naopak dlouhá.

Problém krátké časové řady je například u některých metod (Box-Jenkinsova metodologie), které vyžadují určitý minimální počet pozorování.

U velmi dlouhých časových řad se může vyskytnout riziko, že s průběhem časové řady se mění charakteristika modelu, který tuto řadu popisuje. Je tedy někdy nutné připustit změny parametrů modelu v průběhu časové řady. [4]

1.4.4 Problémy s nesrovnalostí jednotlivých měření

S rostoucím technickým rozvojem se také zvyšuje technická vybavenost průmyslových výrobků. Není tedy například vhodné srovnávat hodnoty průmyslové produkce ze současnosti a z let dřívějších (více let zpátky). Jedná se o tzv. věcnou nesrovnalost údajů.

Někdy může tento problém vzniknout také různým počtem získaných údajů za jednotlivá časová období. [4]

2 Metody pro statistické vyhodnocování časových řad

Existuje několik metod pro analýzu časových řad. Každá tato metoda vychází z různých předpokladů, je různě složitá a vhodná pro různé typy časových řad. Příklady těchto metod jsou dekompozice časové řady, metoda Census I, metoda Census II, Box – Jenkinsova metodologie a spektrální analýza časových řad.

2.1 Dekompozice časových řad

Při analýze reálných časových řad se předpokládá, že data obsahují tzv. vzor (systematické složky: T_t , S_t , C_t) a šum (náhodné složky: ε_t). Právě na tomto předpokladu je založena tato metoda. Jejím cílem je identifikace jednotlivých složek, přičemž konkrétní časové řady mohou nebo nemusí obsahovat všechny tyto složky.

Náhodná složka způsobuje obtížnou identifikaci vzoru. Z tohoto důvodu jsou také využívány metody pro odfiltrování šumu, aby se vzor stal lépe identifikovatelný. [2]

2.1.1 Typy modelů časových řad

Jednotlivé složky časových řad mohou mít mezi sebou různé funkční vztahy. Je možné je kombinovat aditivně, multiplikativně nebo jejich jinou kombinací. Tyto vztahy jsou popsány pomocí tří základních modelů. [2] [3]

a) aditivní model

Výsledná časová řada je tvořena součtem jednotlivých složek a popsána vztahem 2.1.

$$Y_t = T_t + S_t + C_t + \varepsilon_t \quad (2.1)$$

b) multiplikativní model

Výsledná časová řada je tvořena součinem jednotlivých složek a popsána vztahem 2.2.

$$Y_t = T_t \times S_t \times C_t \times \varepsilon_t \quad (2.2)$$

c) smíšený model

Časová řada je tvořena kombinací součtu a součinů jednotlivých složek (vztah 2.3.)

$$Y_t = T_t \times S_t \times C_t + \varepsilon_t \quad (2.3)$$

2.1.2 Trend

Identifikace trendu v reálných datech není jednoduchá. Pokud data obsahují významnou náhodnou složku, tak je nutné pro nalezení trendu a jeho předpovědi použít některý ze způsobů vyhlazování (klouzavé průměry, klouzavé mediány, 4253H filtr nebo jednoduché exponenciální vyrovnávání).

2.1.2.1 Klouzavé průměry

Metoda klouzavých průměrů je založena na tom, že vyrovnávání krátkých úseků časové řady pomocí polynomických funkcí. To znamená, že jednotlivé prvky časové řady jsou nahrazeny průměrem nebo váženým průměrem právě vyrovnávané hodnoty a m sousedních prvků (zleva a zprava). Tato metoda má dva parametry.

- a) délka klouzavých průměrů
- b) řád klouzavých průměrů

Délka klouzavých průměrů značí počet členů, ze kterých je průměr počítán. Tato délka by měla být liché číslo ($2m + 1$, kde $m = 1, 2, 3, \dots$). Řád klouzavých průměrů označuje stupeň vyrovnávacího polynomu. Nevýhodou této metody je, že nová křivka klouzavých průměrů je kratší než původní řada. Při zvolené délce klouzavých průměrů ($2m + 1$) dochází ke ztrátě m nevyrovnaných hodnot na začátku a na konci časové řady. Ve výsledku je nová řada kratší o $2m$ prvků oproti řadě původní. [2] [5]

2.1.2.2 Klouzavé mediány

Tato metoda je založena na principu, že krátké úseky časové řady jsou vyrovnány pomocí mediánů. Princip je podobný jako u klouzavých průměrů. Jednotlivé prvky časové řady jsou nahrazeny mediánem vyrovnávané hodnoty a m sousedních prvků (zleva a zprava). Metoda je vhodná pro data, ve kterých se vyskytují odlehle hodnoty (například chybná měření). Použití této metody na taková data poskytuje věrohodnější křivku oproti metodě klouzavých průměrů, jelikož není tolik ovlivněna extrémními hodnotami. Má pouze jeden parametr, a to délku klouzavých mediánů. Stejně jako u metody klouzavých průměrů je doporučeno, aby délka klouzavých mediánů byla liché číslo ($2m + 1$, kde $m = 1, 2, 3, \dots$). Nevýhodou je, že nelze použít vážení jednotlivých hodnot. [2] [5]

2.1.2.3 4253H filtr

Jedná se o vyhlazovací metodu, která se skládá z několika kroků, přičemž danou řadu vyhledá a zároveň zachová klíčové charakteristiky řady původní. Název je odvozen z jednotlivých kroků aplikace této metody.

- a) 4 bodový klouzavý medián
- b) 5 bodový klouzavý medián
- c) 3 bodový klouzavý medián
- d) 3 bodový vážený klouzavý průměr s Hammingovými váhami 0,25 – 0,5 – 0,25
- e) spočtení reziduí jako rozdíl hodnot původní řady a hodnot nové transformované řady
- f) uplatnění kroků 1 až 4 na spočtená rezidua
- g) přičtení transformovaných reziduí k transformované řadě [2] [6]

2.1.2.4 Jednoduché exponenciální vyrovnávání

Jedná se o často využívanou vyhlazovací metodu pro širokou škálu časových řad, která dává v praxi poměrně věrohodné předpovědi. Jednoduché exponenciální vyrovnávání se používá v případech, kdy časová řada má konstantní trend.

Jednoduchý model může být popsán jako časová řada složená z konstanty a náhodné složky (vztah 2.4).

$$Y_t = b + \varepsilon_t \quad (2.4)$$

Konstanta b se může zvolna měnit v čase, ale v každém úseku časové řady musí být relativně stabilní. Z toho vyplývá, že časová řada je v krátkých úsecích konstantní. Cílem je najít odhad parametru b , a tím i předpověď časové řady do budoucna.

Jednou z možností jak určit b je spočítat speciální druh klouzavého průměru (vztah 2.5). Počítá se vážený průměr ze všech předešlých pozorování (předcházejících tomu, které je právě vyrovnáváno) s exponenciálně se snižujícími váhami do minulosti.

$$S_t = \alpha Y_t + (1 - \alpha) S_{t-1} \quad (2.5)$$

S_t je vyrovnávaná hodnota v čase t , α je parametr exponenciálního vyrovnávání, kde $\alpha \in (0; 1)$. Každá vyrovnaná hodnota je spočtena jako vážený průměr současné hodnoty pozorování a vyrovnané hodnoty předchozího pozorování. Pokud $\alpha = 1$, pak jsou minulé pozorování úplně ignorována. Pokud $\alpha = 0$, pak je ignorováno současné pozorování a všechny vyrovnané hodnoty jsou rovny počáteční vyrovnané hodnotě S_0 .

Důležitou částí je vhodná volba parametru α . Nejlepším řešením je odhadnout hodnotu α z dat. [2]

Vizuální technikou může být například graf pozorovaných hodnot a jejich předpovědí. Tento graf může být doplněn o rezidua, s jejichž pomocí lze identifikovat oblasti, kde model odpovídá datům lépe a kde naopak hůře. Pro optimální volbu hodnoty α se mohou využít také různé míry chyby.

a) průměr čtverců odchylek

Je to nejpoužívanější míra vhodnosti a je popsána vztahem 2.6.

$$\frac{1}{N} \sum_{i=1}^N \{Y_t - [\alpha Y_t + (1 - \alpha)S_{t-1}]\}^2 \quad (2.6)$$

b) jednoduchý průměr odchylek

Nevýhodou této míry (vztah 2.7) je, že kladné a záporné chyby se navzájem vruší.

$$\frac{1}{N} \sum_{i=1}^N \{Y_t - [\alpha Y_t + (1 - \alpha)S_{t-1}]\} \quad (2.7)$$

c) průměr absolutních odchylek

Oproti průměru čtverců odchylek je snížen důraz na odlehlá pozorování (potlačení vlivu extrémních a ojedinělých chyb). Tato míra vhodnosti je popsána vztahem 2.8.

$$\frac{1}{N} \sum_{i=1}^N |Y_t - [\alpha Y_t + (1 - \alpha)S_{t-1}]| \quad (2.7)$$

V praxi se α volí pomocí síťového vyhledávání nebo pomocí automatického vyhledávání.

Při síťovém vyhledávání se začne s hodnotou parametru $\alpha = 0,1$ a velikostí kroku 0,1. Takto se pokračuje až do hodnoty 0,9. Jako optimální α se zvolí to, které minimalizuje průměr čtverců odchylek (vztah 2.8). [2]

$$\alpha = \min_{\alpha \in (0,1)} \frac{1}{N} \sum_{i=1}^N \{Y_t - [\alpha Y_t + (1 - \alpha)S_{t-1}]\}^2 \quad (2.8)$$

Automatické vyhledávání optimálního parametru α je založeno na minimalizaci průměru čtverců odchylek (případně průměru absolutních odchylek) pomocí kvazi – Newtonovy procedury pro minimalizaci funkce.

Poslední částí je volba počáteční hodnoty S_0 (předpověď pro první pozorování). Díky exponenciálně klesajícím váhám volba hodnoty S_0 současné předpovědi ovlivní minimálně. Z toho vyplývá, že pro věrohodnou předpověď je nejdůležitější volba parametru α . [2]

2.1.2.5 Určení trendu

Nejčastěji používanými funkcemi pro modelování trendu jsou následující čtyři.

a) lineární trend (vztah 2.9)

Parametr β_1 určuje, o kolik se změní hodnota časové řady, pokud čas se změní o jednu jednotku. Parametr β_1 udává posuv po ose y .

$$Y_t = \beta_0 + \beta_1 t \quad (2.9)$$

b) polynomiální trend (vztah 2.10)

Další možností je modelovat trend pomocí paraboly nebo jiné polynomiální křivky.

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_k t^k \quad (2.10)$$

c) exponenciální trend (vztah 2.11)

Při exponenciálním trendu se časová řada chová jako geometrická posloupnost. Logaritmováním rovnice je problém převeden na odhad parametrů pomocí lineární regrese.

$$Y_t = \beta_0 \beta_1^t \quad (2.11)$$

d) modifikovaný exponenciální trend (vztah 2.12)

Tato funkce má vodorovnou asymptotu a lze pomocí ní lépe modelovat jevy, u nichž existuje mez nasycení. [2] [5]

$$Y_t = \beta_0 \beta_1^t + C \quad (2.11)$$

2.2 Metoda Census I

Tato metoda vychází z klasického modelu, který byl popsán v kapitole 2.1. Tento model popisuje, že časová řada se skládá ze čtyř základních složek.

- a) trendová složka – T_t
- b) sezónní složka – S_t
- c) cyklická složka – C_t
- d) náhodná, chybová nebo nepravidelná složka – ε_t

V metodě Census I jsou trendová a cyklická složka sloučeny do jedné složky (tzv. trendově-cyklická složka TC_t).

Funkční vztah mezi jednotlivými složkami je aditivní nebo multiplikativní. Aditivní vztah mezi jednotlivými složkami je popsán vztahem 2.12.

$$Y_t = TC_t + S_t + \varepsilon_t \quad (2.12)$$

Multiplikativní vztah mezi jednotlivými složkami je popsán vztahem 2.13.

$$Y_t = T_t \times S_t \times C_t \times \varepsilon_t \quad (2.13)$$

Ve většině případů se používá exponenciální vyhlazování, které analyzuje trendovou i cyklickou složku zároveň. Nejprve se pro časovou řadu spočtou klouzavé průměry se šířkou okna 1 sezóna. V případě, že je délka sezóny sudá, je možno použít váhy stejné nebo různé váhy, přičemž první a poslední pozorování v okně jsou zprůměrována. Klouzavé průměry v časových řadách eliminují sezónní variabilitu. Rozdíly (aditivní vztah) nebo zlomky (multiplikativní vztah) pozorovaných a vyrovnaných řad izolují sezónní a nepravidelné složky. Pokud je klouzavý průměr odečten od pozorované řady (aditivní vztah), nebo je pozorovaná řada vydělena hodnotou klouzavého průměru (multiplikativní vztah), pak je sezónní složka pro každý bod v sezóně vypočtena jako průměr (aditivní vztah) nebo jako středový průměr (multiplikativní vztah).

Očištěná řada od sezónních výkyvů se získá tak, že od původní řady je odečtena (aditivní vztah) sezónní složka nebo je původní řada vydělena sezónní složkou.

Kombinovaná trendově-cyklická složka je aproximována tak, že na sezónně očištěnou řadu je aplikován pětibodový centrováný vážený klouzavý průměr s váhami 1, 2, 3, 2, 1.

Nepravidelná složka se získá tak, že od sezónně očištěné řady je odečtena trendově-cyklická složka (aditivní vztah) nebo je sezónně očištěná řada vydělena trendově-cyklickou složkou (multiplikatívni vztah). [2] [3]

2.3 Metoda X-11 Census II

Metoda Census II je zdokonalením a vylepšením základní metody Census I a dává velmi dobré výsledky pro mnoho problémů z reálného světa. Obsahuje 4 hlavní vylepšení.

a) očištění od obchodních dnů

Jednotlivé měsíce mají rozdílný počet pracovních dnů a rozdílný počet sobot a nedělí. Například při analýze měsíčních zisků to může přispívat k značné variabilitě. Tato metoda umožňuje testování, zda je tento druh variability v časové řadě přítomný. Pokud ano, tak příslušnou časovou řadu očistí.

b) extrémní hodnoty

Většina reálných časových řad obsahuje odlehlé a extrémní hodnoty, které jsou způsobeny vzácnými událostmi. Tyto extrémy mohou ovlivnit odhady trendových nebo sezónních složek. Metoda Census II obsahuje nástroje statistické kontroly, které umožňují analýzu časové řady s ohledem právě na tyto extrémy. Hodnoty nad nebo pod určitým rozsahem (vyjádřeným v směrodatné odchylce) mohou být před finálním odhadem sezónnosti upraveny nebo vypuštěny.

c) vícenásobná vylepšení

Vylepšení, která jsou navržena pro extrémní a odlehlé hodnoty pro počet obchodních dní mohou být aplikována opakovaně. Díky tomu je možné získat vylepšené odhady sezónní, cyklické a nepravidelné složky a sezónně očištěné řady.

d) testy a souhrnné statistiky

Kromě odhadů hlavních složek časové řady je možno také spočítat její různé souhrnné statistiky. Možností je například analýza rozptylu, která umožňuje testování signifikantnosti sezónní variability a vlivu počtu obchodních dnů, nebo procentuální změnu z měsíce na měsíc v trendově-cyklické nebo náhodné složce. [2] [3]

Výpočty touto metodou jsou poměrně složité a proces očištění je rozdělen do sedmi kroků.

a) měsíční sezónní očištění

Používá se před jakýmkoliv jiným očištěním měsíční časové řady. Je možné specifikovat druhou řadu, která obsahuje přednostní očišťující faktory. Poté se hodnoty této řady odečtou od hodnot řady původní (aditivní vztah) nebo se hodnoty původní časové řady vydělí hodnotami druhé časové řady (multiplikatívni vztah).

b) prvotní očištění pro obchodní dny

V tomto kroku je spočteno prvotní očištění od obchodních dnů a jsou také spočteny váhy pro redukci extrémních a odlehlých hodnot.

c) závěrečný odhad rozptylu způsobeného obchodními dny a odhad nepravidelných vah

Váhy a očištění spočtené v předchozím bodě jsou použity pro odvození vylepšených sezónních a trendově-cyklických odhadů. Tyto vylepšené odhady jsou použity pro výpočet konečného faktoru obchodních dnů a vah.

d) závěrečný odhad sezónních faktorů, nepravidelné složky, trendově-cyklické složky a sezónně očištěné řady

Odhady v předchozím kroku jsou použity pro konečný odhad jednotlivých složek.

e) modifikace pro extrémní hodnoty

Původní časová řada a sezónně očištěná řada a nepravidelná složka časové řady jsou modifikovány pro extrémní hodnoty. Modifikovaná řada umožňuje zkoumání stability sezónního očištění.

f) souhrnné míry

V tomto kroku jsou spočteny různé souhrnné míry. Pomocí nich lze zkoumat důležitost konkrétních složek, průměrný počet po sobě jdoucích změn v jednom směru, průměrnou fluktuaci z měsíce na měsíc.

g) tabulky a grafy

Jsou použity pro vizuální vyhodnocení výsledků a jejich lepší prezentaci. [2] [3]

2.4 Spektrální analýza časových řad

Pomocí spektrální analýzy lze zkoumat cyklické vzory v časových řadách. Cílem analýzy je rozložit časovou řadu obsahující cyklickou složku do několika základních sinusoidních funkcí o různých vlnových délkách (frekvencích).

Oproti ARIMA modelům nebo exponenciálnímu vyhlazování je cílem spektrální analýzy identifikace sezónních fluktuací o různých délkách. U ARIMA modelů a exponenciálního vyhlazování je většinou délka sezónní složky známa předem a je zahrnuta do teoretického modelu. Naopak spektrální analýzy se využívá pro nalezení délky sezónní složky.

Spektrální analýza může být rozšířena a použita také pro zkoumání vztahu mezi dvěma časovými řadami. Cílem je identifikovat korelace na různých frekvencích mezi dvěma konkrétními časovými řadami. [2] [7]

2.4.1 Základní zápis a principy spektrální analýzy

Součtem funkcí sinus a cosinus vzniká tzv. sinusoidní funkce. Počet cyklů za jednotkový čas je frekvence a je značena písmenem v . Pokud je zvolena jednotka analýzy jeden rok, pak počet cyklů je například roven 12 pro měsíční cykly ($v=12$) nebo je roven 52 pro týdenní cykly ($v=52$).

Dalším termínem je perioda a je značena písmenem T . Perioda sinusové nebo kosinusové funkce udává délku času, který je zapotřebí pro jeden celý cyklus. Tedy vztah mezi periodou a frekvencí je nepřímá úměra (vztah 2.14).

$$T = \frac{1}{v} \quad (2.14)$$

Pokud v časové řadě je obsažen měsíční cyklus, pak je v dané časové řadě obsažena perioda o délce $T = \frac{1}{12} = 0,0833$ roku.

Cílem spektrální analýzy je rozložit časovou řadu na základní sinusoidní funkce o různých frekvencích a určit frekvence, které jsou významné. Jedním ze způsobů je nahlížet na tento problém jako na problém lineární vícenásobné regrese. Závisle proměnnou je daná časová řada a nezávislými proměnnými jsou sinusoidní funkce o různých frekvencích. Tento model je popsán vztahem 2.15.

$$y_t = a_0 + [a_k \cos(\alpha_k t) + b_k \sin(\alpha_k t)]; k = 1, \dots, q \quad (2.15)$$

Ve vztahu 2.15 značí α frekvenci v radiánech za jednotkový čas. Kosinusový parametr a_k a sinusový parametr b_k jsou tzv. regresní parametry, které udávají míru korelace mezi daty časové řady a příslušné sinusové nebo kosinusové funkce. Celkem je q sinusových a kosinusových funkcí (nemůže jich být více než je dat v časové řadě). Pokud časová řada obsahuje N pozorování, pak bude $\frac{N}{2} + 1$ sinusových funkcí a $\frac{N}{2} - 1$ kosinusových funkcí. Tedy bude tolik různých sinusoidních funkcí jako je dat a bude možno úplně nahradit časovou řadu pomocí základních funkcí. Pokud je N liché, pak je většinou poslední pozorování vypuštěno. Pro identifikaci sinusoidní funkce jsou zapotřebí alespoň dva body (vrchol nahoře a dole). [2] [7]

Spektrální analýza identifikuje korelaci sinusových a kosinusových funkcí o různých frekvencích. Pokud je nalezena velká korelace (velký sinusový nebo kosinusový koeficient), pak lze usoudit, že je v datech obsažena silná periodičita na dané frekvenci.

Sinusové a kosinusové funkce jsou na sobě navzájem nezávislé. Je tedy možno sečíst druhé mocniny koeficientů pro všechny frekvence. Tento součet je označován jako periodogram a jeho hodnoty pro frekvence ν_k se spočtou podle vztahu 2.16.

$$P_k = (a_k^2 + b_k^2) \frac{N}{2} \quad (2.16)$$

Hodnoty periodogramu mohou být interpretovány jako rozptyl (součet čtverců) dat na dané periodě nebo frekvenci. Do grafu se hodnoty periodogramu zakreslují ve vztahu k frekvenci nebo periodě.

V reálných případech se poměrně často stává, že se žádná z frekvencí netrefí přímo do frekvence obsažené v časové řadě. To znamená, že skutečně obsažená frekvence uniká do sousedních frekvencí. Jsou tři základní způsoby, kterými lze přistupovat k tomuto problému. [2] [7]

- a) vyplnění časové řady
- b) zužování (zahrocení) časové řady
- c) vyhlazení periodogramu

Vyplnění časové řady

Jelikož hodnoty frekvencí jsou počítány jako $\frac{N}{t}$ (počet jednotek za čas), lze časovou řadu vyplnit konstantou (například nulou) a vložit tak menší přírůstky frekvenčních hodnot. Vyplnění časové řady umožňuje na data aplikovat podrobnější (jemnější) seznam frekvencí. Mnohdy je vyplnění vhodné také z důvodu větší efektivity výpočtů. [2]

Zahrocení časové řady

Jedná se o proces zahrocení kosinového zvonu. Tato transformace časové řady se doporučuje ještě před samotnou spektrální analýzou. Postup je takový, že se transformuje nějaký konkrétní podíl p dat na začátku a na konci časové řady tak, že tato data jsou vynásobena váhami podle vztahů 2.17 a 2.18.

$$\omega_t = \frac{1}{2} \left[1 - \cos \left\{ \pi \left(t - \frac{1}{2} \right) \frac{1}{m} \right\} \right]; t = 0, \dots, m - 1 \quad (2.17)$$

$$\omega_t = \frac{1}{2} \left[1 - \cos \left\{ \pi \left(N - t + \frac{1}{2} \right) \frac{1}{m} \right\} \right]; t = N - m, \dots, N - 1 \quad (2.18)$$

V těchto vztazích je m voleno tak, aby platilo $\frac{2m}{N} = p$. [2]

Vyhlazení periodogramu

V praxi se často vyskytuje problém, kdy periodogram obsahuje mnoho chaotických výstupků. Cílem je najít pouze frekvence s největšími spektrálními hustotami. To jsou oblasti frekvencí skládající se z mnoha sousedních frekvencí, které přispívají k celkovému periodickému chování časové řady největší mírou. Jednou z možností jak tyto oblasti identifikovat je vyhladit hodnoty periodogramu pomocí transformace váženými klouzavými průměry. Za předpokladu, že okno klouzavých průměrů má šířku m (m musí být liché číslo, $p = \frac{m-1}{2}$), pak se nejčastěji používá těchto pět oken pro vyhlazování.

a) Daniellovo okno

Hodnoty periodogramu jsou transformovány pomocí klouzavých průměrů se stejnými váhami. Odhady spektrální hustoty jsou počítány jako průměr $\frac{m}{2}$ předcházejících a následujících hodnot periodogramu.

b) Tukeyovo okno

Pro každou frekvenci jsou spočteny váhy pro klouzavé průměry pomocí vztahů 2.19 a 2.20.

$$\omega_j = 0,5 + 0,5 \cos \left(\pi \frac{j}{p} \right); j = 0, \dots, p \quad (2.19)$$

$$\omega_j = -\omega_j; j \neq 0 \quad (2.20)$$

c) Hammingovo okno

Pro každou frekvenci jsou spočteny váhy pro klouzavé průměry pomocí vztahů 2.21 a 2.22.

$$\omega_j = 0,54 + 0,46 \cos\left(\pi \frac{j}{p}\right); j = 0, \dots, p \quad (2.21)$$

$$\omega_j = -\omega_j; j \neq 0 \quad (2.22)$$

d) Parzenovo okno

Pro každou frekvenci jsou spočteny váhy pro klouzavé průměry pomocí vztahů 2.23, 2.24 a 2.25.

$$\omega_j = 1 - 6\left(\frac{j}{p}\right)^2 + 6\left(\frac{j}{p}\right)^3; j = 0, \dots, \frac{p}{2} \quad (2.23)$$

$$\omega_j = 2\left(1 - \frac{j}{p}\right)^3; j = \frac{p}{2} + 1, \dots, p \quad (2.24)$$

$$\omega_j = -\omega_j; j \neq 0 \quad (2.25)$$

e) Bartlettovo okno

Pro každou frekvenci jsou spočteny váhy pro klouzavé průměry pomocí vztahů 2.26 a 2.27.

$$\omega_j = 1 - \left(\frac{j}{p}\right); j = 0, \dots, p \quad (2.26)$$

$$\omega_j = -\omega_j; j \neq 0 \quad (2.27)$$

Všechna vyhlazovací okna, s výjimkou Daniellova vyhlazovacího okna, přiřazují vyšší váhy pozorováním ve středu okna a nižší váhy pozorováním dále od středu. Ve spoustě případů dosahují všechna tato okna podobných výsledků. [2] [8]

2.4.2 Příprava dat pro spektrální analýzu

Před samotnou spektrální analýzou je zapotřebí od časové řady odečíst její průměr a zároveň také z řady odstranit trend tak, aby řada byla stacionární (neměnila se vlastnosti řady v čase). Pokud by tato úprava nebyla provedena, tak by ve většině případech byly hodnoty periodogramu silně ovlivněny vysokým prvním kosinovým koeficientem (koeficient pro frekvenci nula). V časové řadě je snaha najít periodicity, a tedy ani trend není pro tuhle analýzu žádoucí. Oba tyto efekty mohou naopak zamaskovat periodicitu v datech. V některých případech je také vhodné data před analýzou vyhladit, tak aby byl odstraněn velký vliv šumu, který může také omezit vyhledávání periodických cyklů v periodogramu. [2]

2.4.3 Testování náhodné posloupnosti

Může se stát, že v datech nejsou objeveny žádné opakující se cykly (všechna pozorování jsou nezávislá oproti ostatním). Pokud mají pozorování normální rozdělení, tak je tato řada označována jako bílý šum, pro který je charakteristické, že hodnoty periodogramu mají exponenciální rozdělení. [2]

2.5 Box – Jenkinsova metodologie

Jedná se o velmi populární, silnou a flexibilní metodologii, ale díky její komplexnosti je obtížné její používání. Výsledky do velké míry závisí na zkušenostech a rozhodnutích osoby provádějící analýzu.

Box-Jenkinsova metodologie považuje za základní prvek konstrukce modelů časových řad reziduální složku, která může být tvořena korelovanými (závislými) hodnotami. Umožňuje tedy práci s časovými řadami, které obsahují navzájem závislá pozorování. Jedním z hlavních pilířů této metodologie je využití korelační analýzy.

Postup analýzy časové řady je rozdělen do tří základních po sobě jdoucích kroků

- a) identifikace modelu
- b) odhad parametrů modelu
- c) ověřování modelu [5] [9]

2.5.1 Základní pojmy Box-Jenkinsovy metodologie

Důležitými pojmy v Box-Jenkinsově metodologii jsou stacionarita, autokorelační funkce (ACF) a parciální autokorelační funkce (PACF). Tyto pojmy jsou vysvětleny v následujících podkapitolách.

2.5.1.1 Stacionarita

Stacionarita je jednou z nejdůležitějších vlastností časové řady. To znamená, že chování časové řady je v jistém smyslu stochasticky ustálené. Pomocí Box-Jenkinsovy metodologie lze analyzovat pouze stacionární časové řady. Proto je nutné nestacionární řady převést na řady stacionární. K tomu se využívají různé transformace (nejčastěji diferencování časové řady). Ukázka nestacionární časové řady je na obr. 2.1 a ukázka stacionární časové řady na obr 2.2.

Rozlišují se dva základní druhy stacionarity.

a) silná (striktní) stacionarita

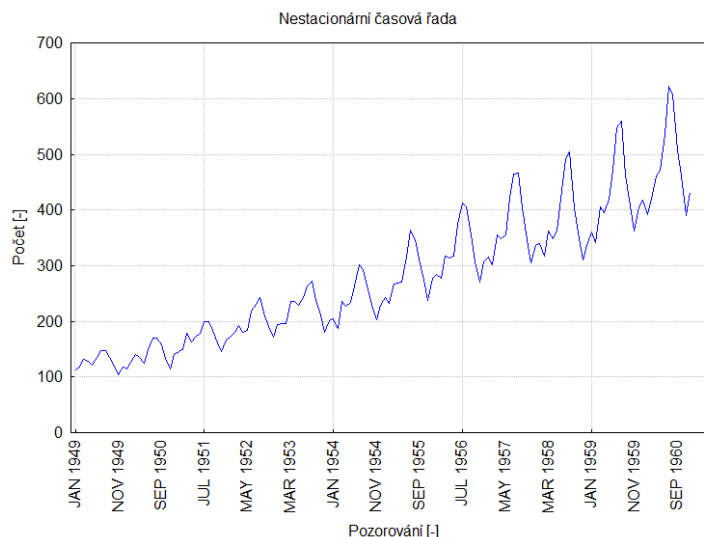
Pravděpodobnostní chování daného stochastického procesu je invariantní vůči posunům v čase, tzn. pravděpodobnostní rozdělení náhodného vektoru $(y_{t1}, y_{t2}, \dots, y_{tk})$ je stejné jako rozdělení vektoru $(y_{t1+h}, y_{t2+h}, \dots, y_{tk+h})$ pro libovolné h .

b) slabá stacionarita

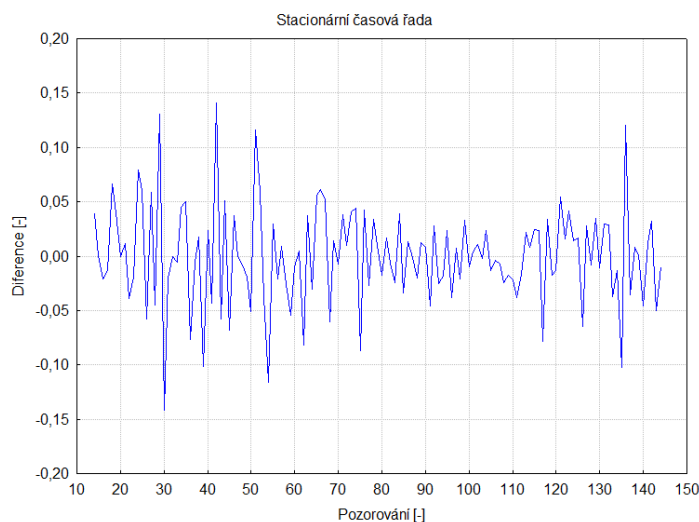
Stochastický proces má konstantní úroveň (konstantní střední hodnotu). Dále je rovnoměrně vyvážen kolem své konstantní střední hodnoty. Z toho vyplývá, že má konstantní rozptyl. Kovariační struktura druhého řádu je invariantní vůči posunům v čase (vztah 2.28).

$$cov(y_t, y_s) = cov(y_{t+h}, y_{s+h}) \text{ pro libovolné } h \quad (2.28)$$

V případě normálního procesu slabá a silná stacionarita splývají. Pokud proces není normální, uvažuje se stacionarita slabá, jelikož je méně omezující. [5] [9]



Obr. 2.1 Ukázka nestacionární časové řady



Obr. 2.2 Ukázka stacionární časové řady (nestacionární časová řada převedena na řadu stacionární)

2.5.1.2 Autokorelační funkce

Autokorelační funkce napomáhá k identifikaci modelu analyzované časové řady. Je definována pomocí autokovarianční funkce. Pro slabě stacionární časovou řadu, s hodnotami jednotlivých pozorování y_t , je hodnota její autokovarianční funkce γ_k v bodě k definována pomocí vztahu 2.29.

$$\gamma_k = cov(y_t, y_{t+h}) = E(y_t - \mu)(y_{t+h} - \mu); k = \dots, -1, 0, 1, \dots \quad (2.29)$$

Hodnota autokorelační funkce ρ_k v bodě k je definována pomocí vztahu 2.30.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_y^2}; k = \dots, -1, 0, 1, \dots \quad (2.29)$$

$\mu = Ey_t$ představuje střední hodnotu a $\sigma_y^2 = \gamma_0 = \text{var}(y_t)$ udává rozptyl dané stacionární časové řady.

Za předpokladu stacionárního stochastického procesu má autokorelační a autokovariační funkce následující tři vlastnosti.

a) $\rho_0 = 1$

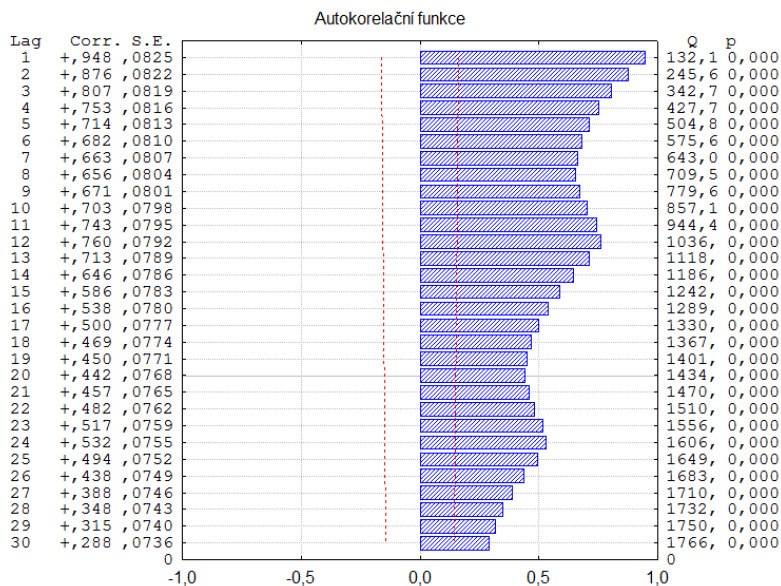
b) $|\gamma_k| \leq \gamma_0; |\rho_k| \leq 1$

c) $\gamma_k = \gamma_{-k}$ a $\rho_k = \rho_{-k}$ pro všechna k

Tato vlastnost říká, že autokorelační funkce a autokovariační funkce jsou sudé. Z toho vyplývá, že k lze omezit na $k \geq 0$.

Grafické vyjádření hodnot ρ_k pro jednotlivá $k \geq 0$ se označuje jako korelogram. Ukázka konkrétního korelogramu autokorelační funkce je na obr. 2.3.

V korelogramu autokorelační funkce je kladen důraz převážně na významné autokorelace, protože právě ty napovídají mezi kterými hodnotami časové řady je vztah. Vyjde-li významná autokorelace pro zpoždění 12, tak to znamená, že je vztah mezi hodnotami 1 a 12, 2 a 13, 3 a 14, atd.[5] [9]



Obr. 2.3 Ukázka korelogramu autokorelační funkce

2.5.1.3 Parciální autokorelační funkce

V případě autokorelační funkce může nastat problém, kdy je korelovaná hodnota první s druhou a zároveň je také korelovaná hodnota první se třetí. Řešení tohoto problému se sériovou korelací po sobě jdoucích pozorováních je právě parciální autokorelační funkce. Ukázka konkrétního korelogramu parciální autokorelační funkce je na obr. 2.4.

Pomocí parciální autokorelační funkce lze sledovat korelace mezi dvěma hodnotami časové řady a to při eliminaci vlivu mezilehlých pozorování. Za předpokladu, že zpoždění je rovno 1, pak ACF a PACF splývají.[5] [9]

Parciální autokorelační funkce ρ_{kk} je definována pomocí vztahu 2.30.

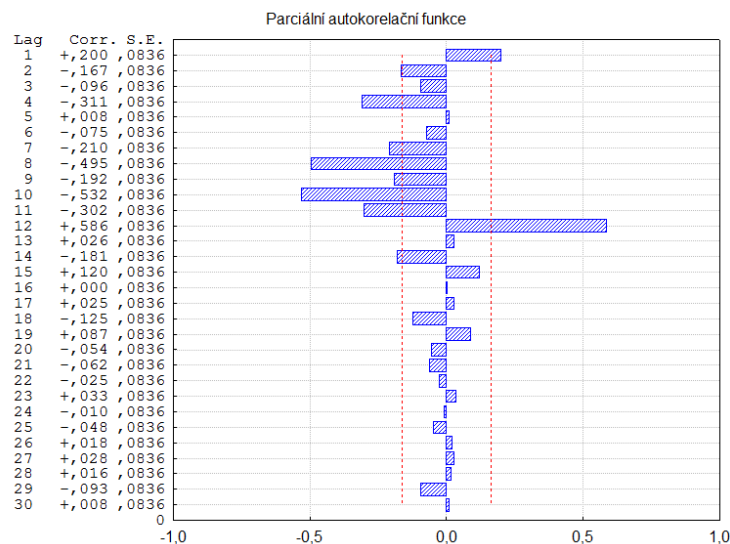
$$\rho_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}} \quad (2.30)$$

V bodě $k = 1$ se funkce ACF a PACF shodují. Platí tedy vztah 2.31.

$$\rho_{11} = \rho_1 \quad (2.31)$$

Pro $k = 2$ platí vztah 2.32.[5] [9]

$$\rho_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \quad (2.32)$$



Obr. 2.4 Ukázka korelogramu parciální autokorelační funkce

2.5.1.4 Lineární proces

Lineární proces je definován jako nekonečná řada pomocí vztahu 2.33.

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \quad (2.33)$$

V tomto vztahu představuje ε_t bílý šum s rozptylem σ_ε^2 a ψ_j jsou parametry.

Obvykle se pro definici lineárního procesu používá operátor zpětného posunutí, který je definován vztahem 2.34.

$$By_t = y_{t-1} \quad (2.34)$$

Při opakovaném použití se dostane vztah 2.35.

$$B^j y_t = y_{t-j} \quad (2.35)$$

Lineární proces lze tedy zapsat pomocí vztahu 2.36.

$$y_t = \psi(B)\varepsilon_t; \text{ kde } \psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots \quad (2.36)$$

Postačující podmínka pro existenci lineárního procesu je, že řada popsaná vztahem 2.36 konverguje podle kvadratického středu ($\psi(B)$ konverguje pro $|B| \leq 0$).

Praktický význam mají procesy, které se získají vynulováním všech parametrů lineárního procesu až na konečný počet. Jsou to následující tři procesy.

- a) autoregresní proces řádu p (AR(p))
- b) proces klouzavých součtů řádu q (MA(q))
- c) smíšený proces řádů p a q (ARMA(p, q)) [5] [9]

2.5.2 Autoregresní proces

Autoregresní proces je popsán vztahem 2.37, kde ξ označuje konstantu a $\varphi_1, \varphi_2, \varphi_3, \dots$ označují parametry autoregresního modelu.

$$y_t = \xi + \varepsilon_t + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots \quad (2.37)$$

Z toho vyplývá, že každé pozorování je složeno z konstanty ξ , náhodného chybového členu ε a lineární kombinace předešlých pozorování.

Na autoregresní proces je kladen požadavek stacionarity procesu. Autoregresní proces je stacionární, pokud jeho parametry leží uvnitř určitého rozsahu.

Například pro proces AR(1), který je popsán vztahem 2.38, musí platit, že jediný autoregresní parametr φ_1 (procesu AR(1)) musí náležet do intervalu $\varphi \in (-1, 1)$. Pokud by tento požadavek nebyl splněn, pak se dřívější efekty nakumulují a následující hodnoty řady y_t budou konvergovat k nekonečnu a daná řada nebude stacionární.

$$y_t = \xi + \varepsilon_t + \varphi_1 y_{t-1} \quad (2.38)$$

Pro modely obsahující více než jeden parametr lze definovat podobná pravidla. [2] [5]

2.5.3 Proces klouzavých součtů

U procesu klouzavých součtů (na rozdíl od autoregresního procesu) může být každý prvek řady ovlivněn nějakou dřívější chybou. Hodnota časové řady může být vyjádřena pomocí vztahu 2.39, kde μ je konstanta a $\theta_1, \theta_2, \theta_3$ jsou parametry.

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots \quad (2.39)$$

Každé pozorování je složeno z náhodného chybového členu ε_t , konstanty μ a lineární kombinace předchozích chybových členů $\theta_1, \theta_2, \theta_3, \dots$

Mezi procesem klouzavých součtů a autoregresním procesem existuje úzký vztah. Proces klouzavých součtů může být invertován do tvaru autoregresního procesu nekonečného řádu. To se dá uskutečnit pouze tehdy, pokud parametry procesu klouzavých součtů splňují podmínku invertibility. [2] [5]

2.5.4 ARIMA metodologie

Tento model obsahuje kombinaci obou výše popsaných procesů (autoregresní proces a proces klouzavých součtů). Na zkoumanou časovou řadu je kladen požadavek stacionarity. To znamená, že časová řada musí mít konstantní střední hodnotu, rozptyl a autokorelační funkci. Aby bylo dosaženo stacionarity, je často nutné časovou řadu nejprve diferencovat.

Model lze obecně popsat jako ARIMA(p , d , q). Parametr d udává, kolikrát je daná časová řada diferencována. Například model ARIMA(0, 1, 1) udává, že model neobsahuje žádný autoregresní parametr a obsahuje jeden parametr pro klouzavé součty, který je počítán pro jednu diferencovanou řadu. [2] [5]

2.5.4.1 Identifikace modelu

Pro určení potřebného řádu diferencování je nutné zkoumat graf časové řady a graf autokorelogramu. Významné změny na úrovni časové řady většinou vyžadují nesezónní diferencování prvního řádu a velké změny ve sklonu diferencování druhého řádu. Pokud řada obsahuje sezónní vzory, je potřeba diferencovat sezónně.

Pokud při sledování autokorelogramu klesají odhadnuté autokorelační koeficienty pomalu pro delší zpoždění, pak je většinou nutné použít diferencování prvního řádu.

Některé časové řady vyžadují být diferencovány jen trochu nebo dokonce vůbec. Přediferencovaná řada má méně stabilní odhady koeficientů.

Dále je nutné určit, kolik parametrů klouzavých součtů q a kolik autoregresních parametrů p je potřeba pro získání efektivního, ale šetrného modelu procesu. Šetrný model znamená, že obsahuje nejmenší počet parametrů a zároveň největší počet stupňů volnosti mezi všemi modely, které připadají v úvahu pro daná data. Parametry p a q málokdy překročí hodnotu 2.

Pro určení typů a počtů ARIMA parametrů jsou hlavními pomocnými nástroji graf časové řady a grafy autokorelační a parciální autokorelační funkce. Toto rozhodnutí není vždy snadné a často vyžaduje zkušenosti a zkoušení alternativních modelů.

Většina empirických vzorů časových řad může být aproximována za pomoci některého z pěti konkrétních modelů uvedených v tab. 2.1. Pro který z těchto základních modelů se rozhodnout pomůže graf autokorelační a parciální autokorelační funkce. [2] [5]

Tab. 2.1 Základní modely a charakteristiky ACF a PACF [2]

Typ modelu	Charakteristika ACF	Charakteristika PACF
AR(1)	exponenciální pokles	hrot u zpoždění 1, žádné jiné korelace u ostatních zpoždění
AR(2)	sinusový tvar nebo exponenciální pokles	hroty u zpoždění 1 a 2, žádné jiné korelace u ostatních zpoždění
MA(1)	hrot u zpoždění 1, žádné jiné korelace u ostatních zpoždění	exponenciální pokles
MA(2)	hroty u zpoždění 1 a 2, žádné jiné korelace u ostatních zpoždění	sinusový tvar nebo exponenciální pokles
ARMA(1, 1)	exponenciální pokles začínající od zpoždění 1	exponenciální pokles začínající od zpoždění 1

2.5.4.2 Odhad parametrů modelu

Odhad parametrů může být proveden pomocí několika metod. Všechny tyto metody vedou k velmi podobným odhadům, ale pro některé modely mohou být více nebo méně efektivní.

Jedna z metod pro odhad parametrů je pomocí algoritmů pro minimalizaci součtu čtverců (kvazi Newtonova metoda). Odhadování je prováděno na diferencovaných datech a před předpovídáním je nutno mít časovou řadu integrovanou tak, že předpovědi jsou vyjádřeny v kompatibilních hodnotách se vstupními daty. Odhady parametrů se použijí také pro spočtení nových hodnot časové řady tam, kde ve vstupní množině dat už nejsou data.

ARIMA model může obsahovat kromě parametrů klouzavých součtů a autoregresních parametrů také konstantu. Interpretace významně nenulové konstanty závisí na modelu, který je proložen daty. Hlavně u modelů bez autoregresních parametrů se předpokládá hodnota konstanty μ , která je rovna průměru řady. Konstanta ζ u modelů s autoregresními parametry představuje intercept. U diferencované řady konstanta představuje intercept nebo průměr diferencované řady. Pokud je časová řada jednou diferencována a její model neobsahuje žádný autoregresní parametr, pak konstanta reprezentuje průměr diferencované řady a tedy také sklon lineárního trendu původní nediferencované řady. [2] [5]

2.5.4.3 Ověřování modelu

Pro ověřování modelu existuje několik způsobů. Prvním a nejjednodušším způsobem je vizuální kontrola grafu předpovědi. Již na první pohled lze většinou vidět, jestli zvolený model je správný.

Druhým způsobem je kontrola reziduí. Na ně je kladen požadavek, aby měla normální rozdělení a nebyla navzájem korelovaná. Požadavek na normální rozdělení lze například zkontrolovat pomocí grafu normálního rozdělení reziduí a histogramu reziduí. Korelovanost reziduí lze zkoumat pomocí autokorelační a parciální autokorelační funkce. [5]

2.5.5 Sezónní modely

Sezónní ARIMA model (někdy označován také jako SARIMA) je zobecněnou a rozšířenou metodou, pro řady, které obsahují sezónně se opakující vzor. Kromě nesezónních parametrů je nutné také odhadnout sezónní parametry pro specifická zpoždění. Tato metoda je tedy rozšířena o sezónní autoregresní parametry ps , sezónní parametry klouzavých součtů qs a sezónní diferencování ds .

Příkladem je sezónní model $ARIMA(0, 1, 2)(0, 1, 1)$, který neobsahuje žádný autoregresní parametr a obsahuje 2 parametry klouzavých součtů a 1 parametr sezónních klouzavých součtů, přičemž tyto parametry jsou počítány pro řadu, která je jednou diferencována se zpožděním 1 a jednou diferencována sezónně.

Obecná doporučení pro výběr parametrů podle ACF a PACF funkcí a tabulky 2.1 mohou být použita také na tyto sezónní modely. ACF a PACF pro sezónní řady obsahují poměrně velké hodnoty koeficientů u násobků sezónních zpoždění. [2]

3 Porovnání statistických SW STATISTICA a NCSS v oblasti analýzy časových řad

V této kapitole jsem se zabýval porovnáním možností obou statistických SW v oblasti analýzy časových řad. Cílem bylo určit, jaké metody analýzy jednotlivé SW umožňují a jaké možnosti tyto metody nabízí. Kromě samotných metod jsem porovnával také možnosti zobrazení výsledků.

Při porovnávání jsem vycházel z dokumentace a manuálů (obsažených přímo v SW a ze zdrojů: [10] [11]) od výrobců jednotlivých SW a z vlastních zkušeností získaných prací v těchto SW.

3.1 SW STATISTICA

SW STATISTICA umožňuje k analýze časových řad využít několik metod: ARIMA metodologie, sezónní a nesezónní exponenciální vyhlazování, dekompozice časových řad (metoda Census I a její vylepšení: metoda Census II) a spektrální analýza.

3.1.1 ARIMA metodologie

Tato metoda kromě samotného výpočtu umožňuje diferencování časové řady a zobrazení ACF a PACF. Kromě těchto základních možností je také možnost volby různých transformací: detrendování řady, odečtení průměru řady, přičtení konstanty, transformace řady pomocí různých funkcí (logaritmická, exponenciální, ...). Zároveň je součástí této metody možnost vyhlazování pomocí klouzavého průměru, klouzavého mediánu, jednoduchého exponenciálního vyhlazování a 4253H filtru. Dále je součástí také možnost tvorby různých grafů (histogram, normální rozdělení, ...). Zařazení všech těchto úprav a možností přímo v této metodě zjednodušuje a zrychluje samotné analyzování časových řad.

Problematickou částí je požadavek na časové řady, že musí obsahovat alespoň 8 kompletních cyklů, jinak nelze u řad, kde se vyskytuje sezónní složka, nastavit sezónní diferencování se zpožděním rovném délce sezóny.

SW STATISTICA umožňuje také pomocí ARIMA metodologie analyzovat přerušené časové řady. To jsou řady, kde došlo k náhlému a významnému poklesu nebo navýšení hodnot.

3.1.2 Dekompozice časových řad

SW STATISTICA nabízí možnost analýzy pomocí metody Census I nebo pomocí rozšířené metody Census II.

Při volbě metody Census I lze zvolit délku sezóny a určit typ modelu řady (aditivní nebo multiplikativní). Poté lze spočítat klouzavé průměry, podíly nebo diference, sezónní faktory, sezónně upravené řady, komponenty vyhlazeného trendového cyklu a nepravidelnou složku.

Volba metody Census II je vylepšením metody Census I a nabízí možnost měsíční a čtvrtletní sezónní dekompozice. Je možné zvolit očištění od obchodních dnů, sezónní přizpůsobení a další možnosti spadající do této metody.

SW STATISTICA neumožňuje predikci časové řady, ale pouze spočtení výše uvedených složek. Další možností je tvorba grafů (2D graf, histogram, normální rozdělení,...) z těchto složek.

3.1.3 Sezónní a nesezónní exponenciální vyhlazování

SW STATISTICA nabízí 12 exponenciálních vyhlazovacích funkcí. Lze zvolit typ modelu (aditivní nebo multiplikativní) a určit charakter trendu (bez trendu, lineární trend, exponenciální trend a tlumený trend). Pro sezónní variantu lze nastavit délku sezóny.

Je možné zadat počáteční hodnotu pro vyhlazovací transformaci, počáteční hodnotu trendu a sezónní faktory.

Identifikaci nejlepších parametrů lze nastavit průzkumem mřížky parametrů (tabulka výsledků obsahující kombinace hodnot parametrů: směrodatná odchylka, absolutní odchylka, součet čtverců chyb, rozptyl, procentuální odchylka a absolutní procentuální odchylka) nebo pomocí automatického vyhledání nejlepších parametrů.

Po provedení analýzy lze zobrazit výsledky příslušné exponenciální transformace, rezidua a předpověď.

3.1.4 Spektrální analýza

Tato metoda umožňuje doplnit nebo zkrátit řady ještě před provedením analýzy. Dále nabízí možnost různých předběžných transformací (zúžení, odečtení průměru, centrování,...).

Spektrální analýza počítá frekvence, periody, koeficienty sinu a cosinu, hodnoty pro frekvenční graf a odhady spektrální hustoty.

Odhady spektrální hustoty lze počítat pomocí různých váhových koeficientů (Daniellovo, Hammingovo, Bartlettovo, Tukeyovo, Parzenovo nebo pomocí koeficientu definovaného uživatelem).

Další možností je zobrazit pouze definovaný počet hodnot z frekvenčního grafu (periodogramu) nebo grafu spektrální hustoty.

Je možné znázornit koeficienty sinu a cosinu, hodnoty frekvenčního grafu a hodnoty spektrální hustoty oproti frekvenci nebo periodě. Také je možné zvolit rozsah, pro který bude znázorněn frekvenční graf a graf spektrální hustoty, a tak zvýšit rozlišení a vypovídací hodnotu grafu.

Další volbou je křížová spektrální analýza. Je to analýza dvou řad, kde jedna řada je nastavena jako závislá proměnná a druhá jako nezávislá.

3.1.5 Zobrazení výsledků

SW STATISTICA umožňuje jak grafické, tak i číselné zobrazení výsledků (i průběžných výsledků). Číselné zobrazení je ve speciální tabulce a tyto hodnoty lze okamžitě použít k dalším analýzám. Výsledky lze zobrazit také pomocí nejrůznějších grafů. Zároveň je možné tyto grafy uložit a také jednoduchým způsobem formátovat.

3.2 SW NCSS

SW NCSS umožňuje k analýze časových řad využít několik metod: ARIMA metodologie, exponenciální vyhlazování, dekompozice a předpověď časových řad a spektrální analýza.

3.2.1 ARIMA metodologie

SW NCSS umožňuje analýzu časových řad pomocí ARIMA metodologie. Kromě klasického nastavení modelu (volba parametrů modelu, parametry diferencování, délka sezóny, počet hodnot předpovědi) umožňuje také odečtení průměru řady, její detrendování a logaritmování. ACF a PACF nejdou zobrazit přímo v této metodě, ale musí se otevřít speciální modul, který tuto možnost nabízí. Nejdůležitějšími výsledky této metody jsou: předpověď pro počet požadovaných hodnot, rezidua a ACF reziduí (grafické i číselné zobrazení). Pro zobrazení například histogramu reziduí je nutné hodnoty reziduí zkopírovat a vložit do datové tabulky a spustit speciální modul, který je určený k tvorbě histogramu.

Speciální možností je teoretická ARMA. Ta nevyužívá žádná data, ale slouží ke studiu různých modelů a ukazuje například jejich ACF nebo PACF.

3.2.2 Dekompozice a předpověď časových řad

Tato metoda nabízí volbu počtu hodnot předpovědi a volbu délky sezóny. Výsledkem této analýzy je předpověď a jednotlivé složky řady (cyklická složka, sezónní složka, trendová složka a nepravidelná složka). Tyto výsledky jsou zobrazeny jak v grafické, tak i číselné podobě.

3.2.3 Exponenciální vyhlazování

SW NCSS nabízí tři druhy exponenciálního vyhlazování (horizontální, trendové a trendové/sezónní).

Horizontální exponenciální vyhlazování je určeno pro řady, které nevykazují trend ani sezónnost. Výsledkem je horizontální přímka. Tato metoda je vhodná pro krátkodobé předpovědi.

Trendové exponenciální vyhlazování umožňuje analýzu časových řad vykazujících stoupající nebo klesající trend.

Naopak trendové/sezónní vyhlazování si poradí jak s trendovou, tak i sezónní složkou. Je možné zvolit typ modelu (aditivní nebo multiplikativní) a nastavit počet hodnot předpovědi. Výsledkem je předpověď a rezidua.

3.2.4 Spektrální analýza

Spektrální analýza počítá frekvence, periody, hodnoty pro frekvenční graf (periodogram) a odhady spektrální hustoty. Lze nastavit odečtení průměru řady a její detrendování.

Je také možné zobrazit hodnoty frekvenčního grafu a hodnoty spektrální hustoty proti frekvenci nebo periodě. Také lze zvolit rozsah, pro který bude znázorněn periodogram nebo graf spektrální hustoty, a tak zvýšit rozlišení a vypovídací hodnotu grafu.

Další volbou je křížová spektrální analýza. Je to analýza dvou řad, kde jedna řada je nastavena jako závislá proměnná a druhá jako nezávislá.

3.2.5 Zobrazení výsledků

SW NCSS zobrazuje výsledky jak v podobě grafické, tak i číselné. Nevýhodou grafů je, že jejich formátování (název, popisky os, rozsahy hodnot na osách,...) musí být nastaveno ještě před spuštěním výpočtu konkrétní analýzy. Číselné zobrazení výsledků je "jakoby v textovém souboru, který je připraven k tisku – na začátku každé nové strany je nadpis hodnot". Pokud chceme výsledné hodnoty použít k další analýze, je nutné je kopírovat a nadpisy vložené mezi nimi filtrovat.

3.2.6 Porovnání SW STATISTICA a NCSS

SW STATISTICA celkově dovoluje uživateli nastavit více parametrů a možností. Také výsledky (i mezivýsledky) jsou jednoduchým způsobem připraveny k dalším analýzám.

SW NCSS požaduje nastavení formátování grafů ještě před spuštěním výpočtu dané analýzy a také data z výsledků analýzy nejsou okamžitě připravena pro další analyzování pomocí jiné požadované metody.

Celkově oba dva SW dokážou analyzovat časové řady dostatečně, ale SW STATISTICA umožňuje více voleb v jednotlivých metodách, má lepší řešení pro výstup hodnot provedených analýz a dle mého subjektivního názoru je i lépe strukturován (lépe se v něm pracuje a je více intuitivní).

4 Analýza časových řad z výsledků materiálových zkoušek

V této kapitole jsem analyzoval časové řady z výsledků materiálových zkoušek tří různých senzorů: textilní senzor, senzor - flexibilní substrát + PEDOT:PSS a senzor - BI2 + sítotisk. Tyto senzory jsou vyvíjeny v rámci materiálového výzkumu na Západočeské univerzitě v Plzni (Fakulta elektrotechnická – Katedra technologií a měření).

Analýzu časových řad jsem provedl v obou statistických SW, za použití ARIMA modelů. Cílem bylo provést potřebné transformace časové řady (vyhlazení, diferencování, stacionarizace,...), pokud byly zapotřebí. Dále nalézt vhodný model, který nejlépe odpovídá dané časové řadě a získat předpověď pro další hodnoty.

Data časových řad jsou velmi rozsáhlá, a proto jsem je umístil na CD, které je přiloženo k práci. Data pro textilní senzor jsou umístěna v souboru *data1.xls*, data senzoru - flexibilní substrát + PEDOT:PSS v souboru *data2.xls* a data senzoru - BI2 + sítotisk v souboru *data3.xls*.

Základní specifikace senzorů:

a) textilní senzor

Jedná se o senzor, který je vsítý pomocí nití do textilie.

Spodní nit je vodivý drát o průměru 0,002 mm. (MONEL 1.4301)

Horní nit je nevodivá. (PES)

b) senzor – flexibilní substrát + PEDOT:PSS

Je to senzor, který je tvořen z flexibilního substrátu a materiálu PEDOT:PSS.

PEDOT (poly-ethylenedioxythiophene) je vodivý polymer, který je použit v kombinaci s PSS (poly-styrenesulphonic acid).

Senzor se skládá z elektrodového systému a senzitivní vrstvy. Elektrodový systém má jednu vrstvu, elektroda je uhlíková a substrát z materiálu NFC/MFC 75/25. Senzitivní vrstva je z materiálu NFC/MFC 75/25. NFC je nanokrystalická celulóza a MFC je mikrokrytalická celulóza.

c) senzor – BI2 + sítotisk

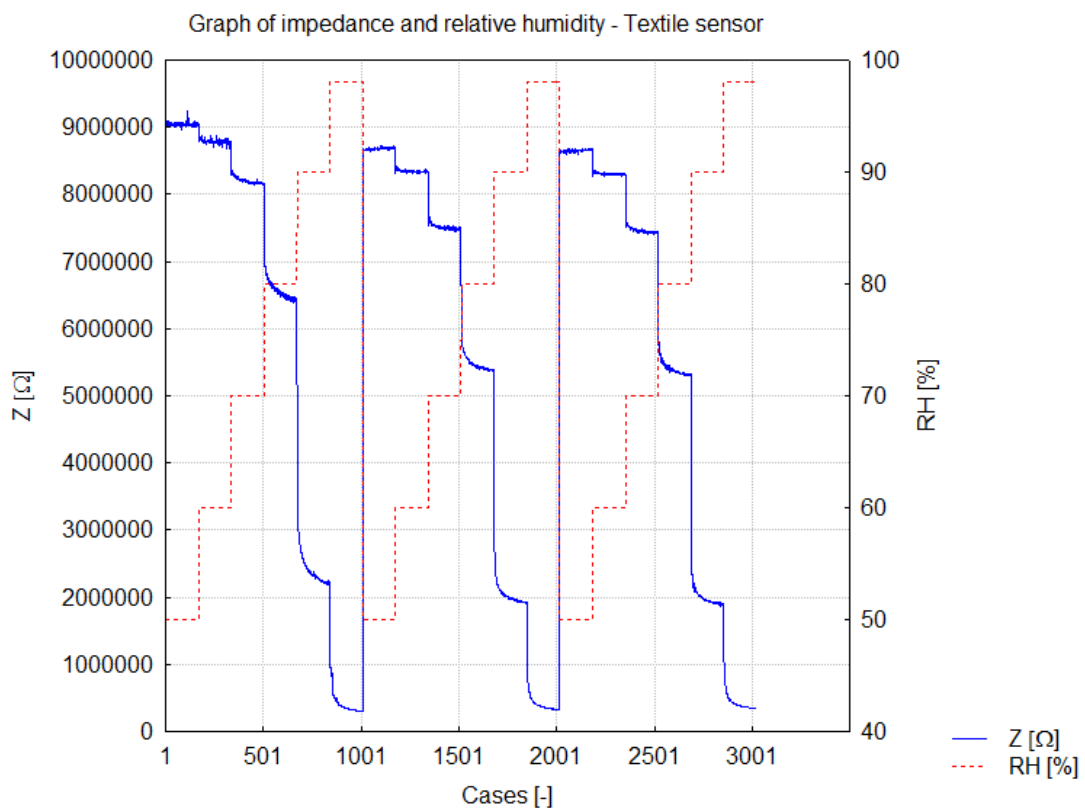
Pro depozici organických senzorových vrstev byl použit substrát s elektrodovým systémem BI2. Jedná se o jemnozrný 96 % korundový substrát tloušťky 0,64 mm s rozměry 9 x 6 mm. Na substrátu je dvojice zlatých interdigitálních elektrod. Jako adhezní mezivrstva je použit NiCr a Ni. Šířka elektrod je 25 μm , šířka mezery 25 μm . Tloušťka elektrod je

0,4 μm . Aktivní plocha interdigitálních elektrod je 4,1 x 4,1 mm. Senzitivní vrstva je z materiálu FTC (ftalocyanin) a krycí vrstva je z materiálu PEEC-4 (polyetheretherketon).

4.1 Textilní senzor

V této kapitole jsem analyzoval data časové řady z měření textilního senzoru. Jedná se o hodnoty impedance senzoru v závislosti na změně vlhkosti při konstantní teplotě 30 °C. Hodnota relativní vlhkosti se mění skokově: 50%, 60%, 70%, 80%, 90% a 98%. Celkově tato řada obsahuje 3 celé cykly. Graf popisující změnu impedance a vlhkosti je na obr. 4.1.

Každý cyklus a podcyklus (data pro určitou míru relativní vlhkosti) obsahují rozdílné množství naměřených hodnot. Rozdíl v počtu hodnot je minimální oproti celkovému počtu pozorování v jednotlivých cyklech, a tak jsem provedl filtraci dat. Každý podcyklus má na svém konci již stabilizované hodnoty impedance, a tak jsem mohl odfiltrovat potřebný počet dat vždy z konce jednotlivého podcyklu tak, aby to nemělo vliv na průběh časové řady. Data původní i odfiltrovaná časové řady jsou uvedena v souboru *data1.xls*.

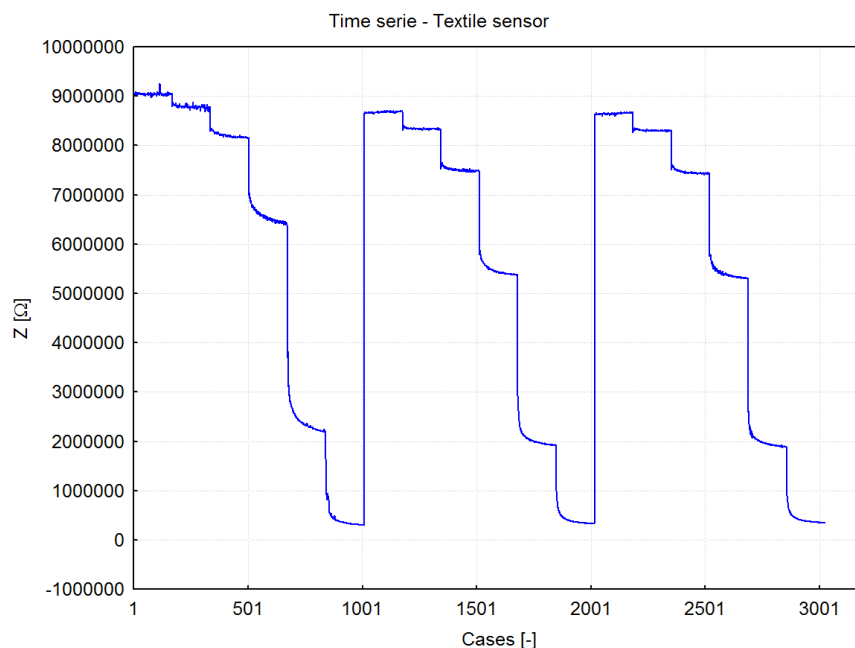


Obr. 4.1 Graf změny impedance a relativní vlhkosti - textilní senzor

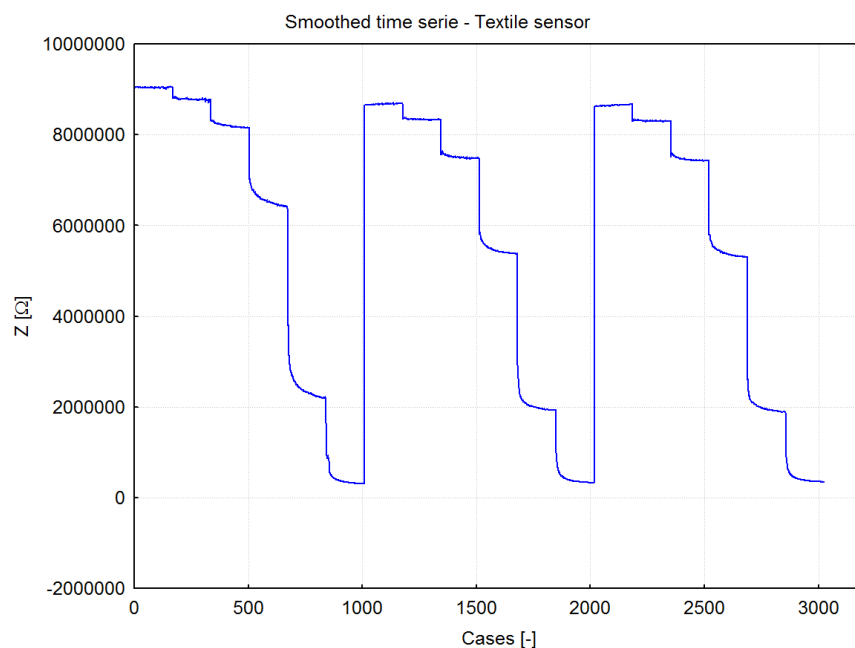
4.1.1 SW STATISTICA

Napřed jsem načetl data ze souboru *data1.xls* zobrazil je (obr. 4.2), abych měl lepší představu o průběhu časové řady.

Z grafu na obr. 4.2 je vidět, že tato časová řada potřebuje vyhladit, hlavně její první cyklus. Pro její vyrovnání jsem zvolil tří – bodový klouzavý medián, který řadu dostatečně vyhladil a přitom zachoval její významné charakteristiky (obr. 4.3).



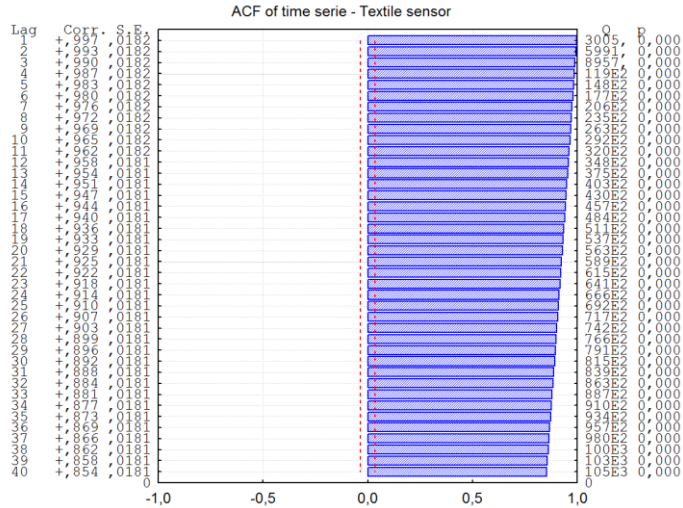
Obr. 4.2 Časová řada - textilní senzor – SW STATISTICA



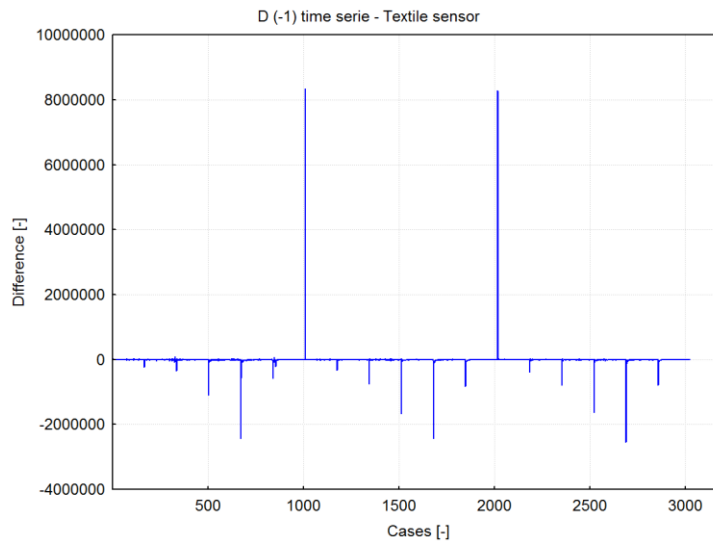
Obr. 4.3 Vyhlazená časová řada - textilní senzor – SW STATISTICA

Po vyhlazení časové řady jsem zkontroloval průběh ACF a PACF. Z grafu ACF (obr. 4.4) je patrná silná sériová závislost. Pro odstranění této závislosti je nutné řadu jednou nesezónně diferencovat. Graf takto diferencované řady je na obr. 4.5.

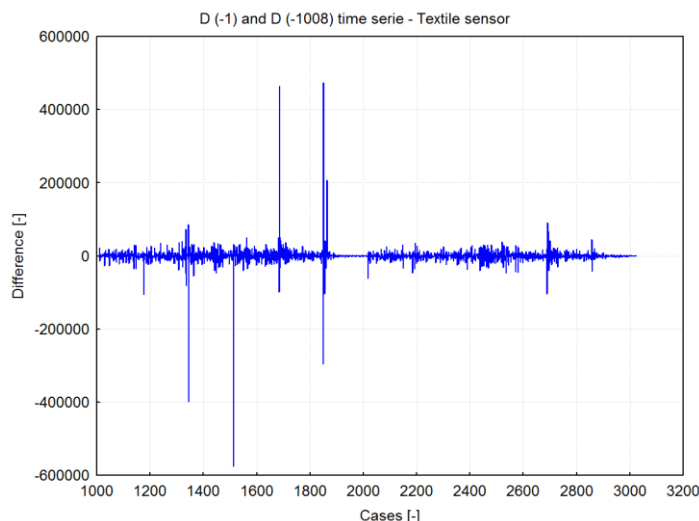
V dalším kroku je zapotřebí opět zkontrolovat průběhy ACF a PACF jednou nesezónně diferencované řady. Ty jsou v SW STATISTICA omezeny na 999 hodnot. Ze znalosti časové řady lze usuzovat, že tato časová řada bude potřebovat sezónní diferencování se zpožděním 1008. Takto diferencovaná řada je zobrazena na obr. 4.6.



Obr. 4.4 ACF časové řady - textilní senzor – SW STATISTICA



Obr. 4.5 Jednou nesezónně diferencovaná časová řada - textilní senzor – SW STATISTICA



Obr. 4.6 Jednou nesezónně a jednou sezónně diferencovaná časová řada - textilní senzor – SW STATISTICA

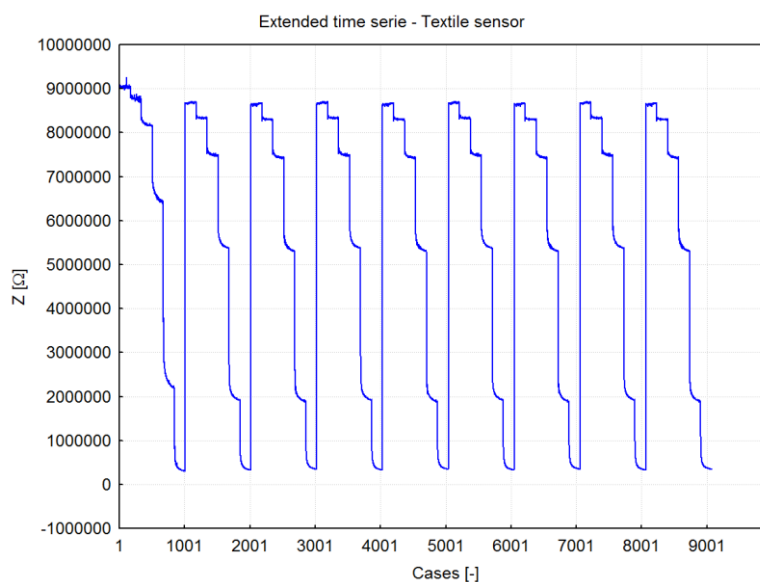
Dále jsem zkontroloval ACF a PACF jednou nesezónně a jednou sezónně (se zpožděním 1008) diferencované řady. Obě tyto funkce jsou v pořádku a lze přistoupit k volbě ARIMA modelu pro tuto časovou řadu a odhadu parametrů klouzavých součtů a autoregresních parametrů (sezónních i nesezónních). Po vyzkoušení různých modelů jsem zvolil model $ARIMA(1, 1, 1)(1, 1, 1)$. Nastavení jsem provedl podle obr. 4.7.

Obr. 4.7 Nastavení ARIMA modelu - textilní senzor – SW STATISTICA

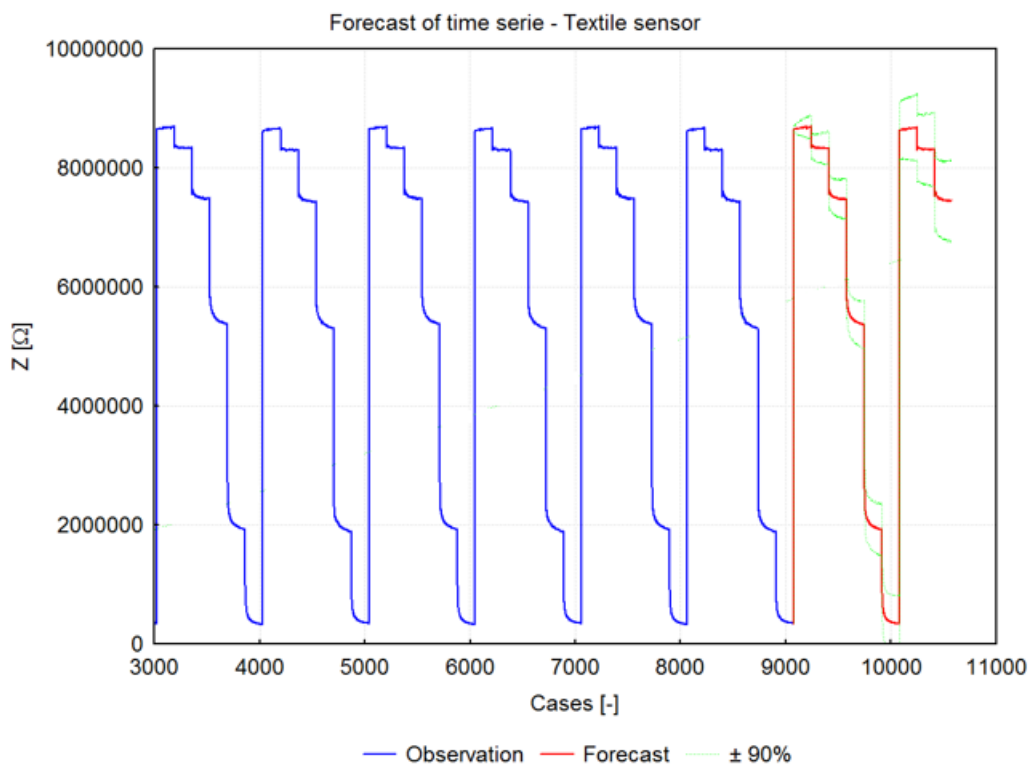
SW STATISTICA vyžaduje pro výpočet alespoň 8 kompletních cyklů, jinak nelze sezónní diferencování o délce cyklu nastavit. Tato časová řada obsahuje pouze 3 cykly. Možným řešením je tuto řadu uměle nastavit na 9 cyklů (přesně 8 nestačí – diferencováním a vyhlazením dojde ke zkrácení řady). První cyklus je oproti dalším posunut výše (stabilizační cyklus), a tak jsem k prodloužení řady využil jen druhý a třetí cyklus. V nově vytvořené řadě

jsou jednotlivé cykly seřazeny v pořadí: 1-2-3-2-3-2-3-2-3 (obr. 4.8). Tuto řadu jsem musel nejprve znovu vyhladit pomocí klouzavého mediánu a až poté spustit výpočet.

Po skončení výpočtu lze zobrazit výsledky této analýzy a zároveň otestovat vhodnost zvoleného modelu. Provedl jsem předpověď pro 1500 dalších hodnot, kde je vidět, že tento model zachytil všechny charakteristické aspekty této řady a předpověď vypadá věrohodně (obr. 4.9). Navíc oblast, která předpovídá s 90% pravděpodobností, je velmi malá.

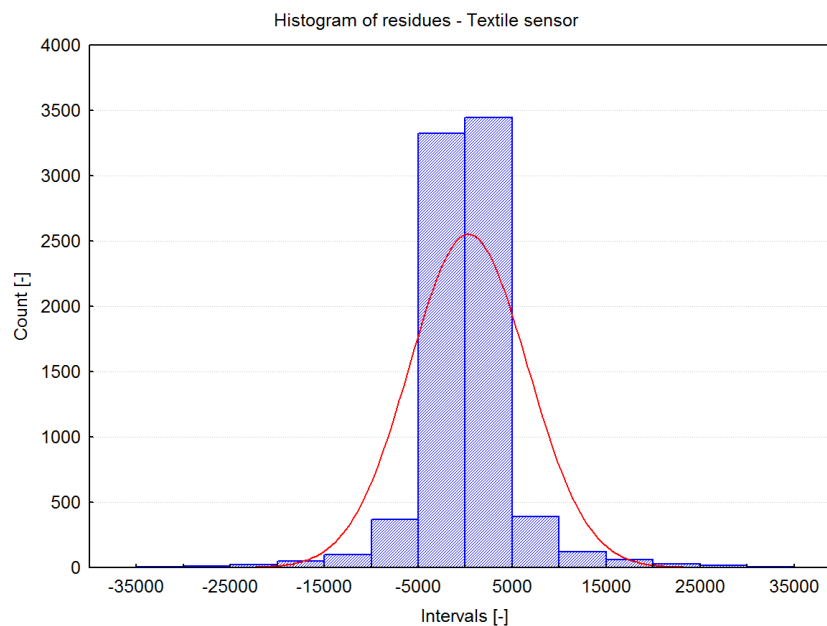


Obr. 4.8 Prodloužená časová řada - textilní senzor – SW STATISTICA

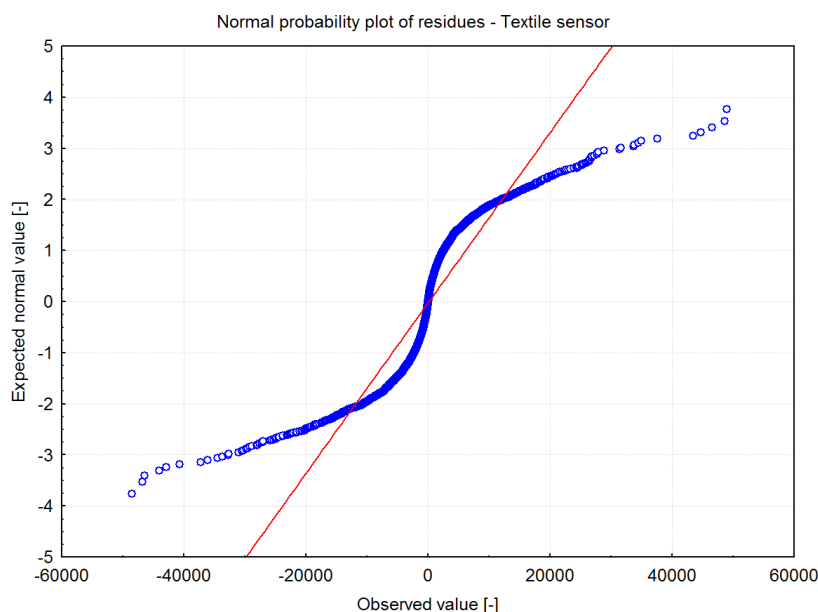


Obr. 4.9 Předpověď časové řady - textilní senzor - SW STATISTICA

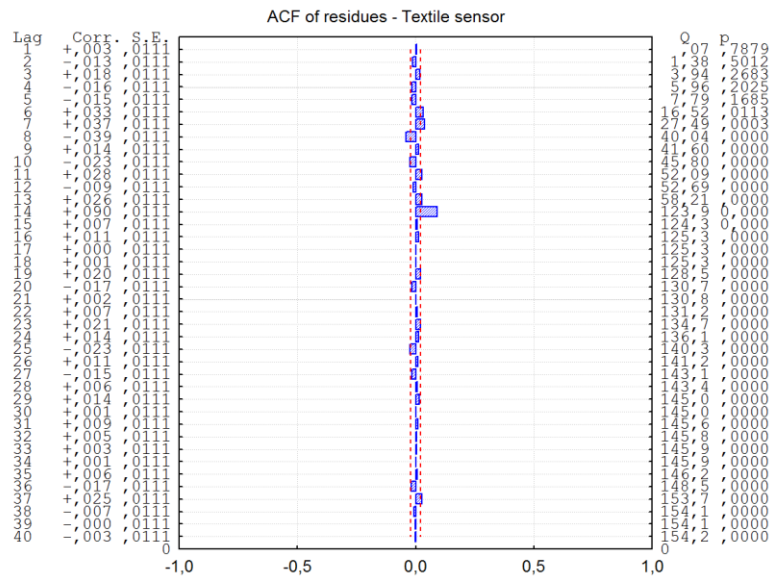
Dále jsem pro ověření zvoleného modelu zobrazil graf histogramu reziduí (obr. 4.10), graf normálního rozdělení reziduí (obr. 4.11), ACF (obr. 4.12) a PACF (obr. 4.13) reziduí. Z histogramu a grafu normálního rozdělení je patrné, že rozdělení reziduí není přesně normální, ale hodně se mu přibližuje. ACF a PACF nevykazují korelaci mezi reziduí. Vizuální kontrola grafu předpovědi je také v pořádku. Z toho lze usuzovat, že zvolený model je správný, a získaná předpověď je věrohodná.



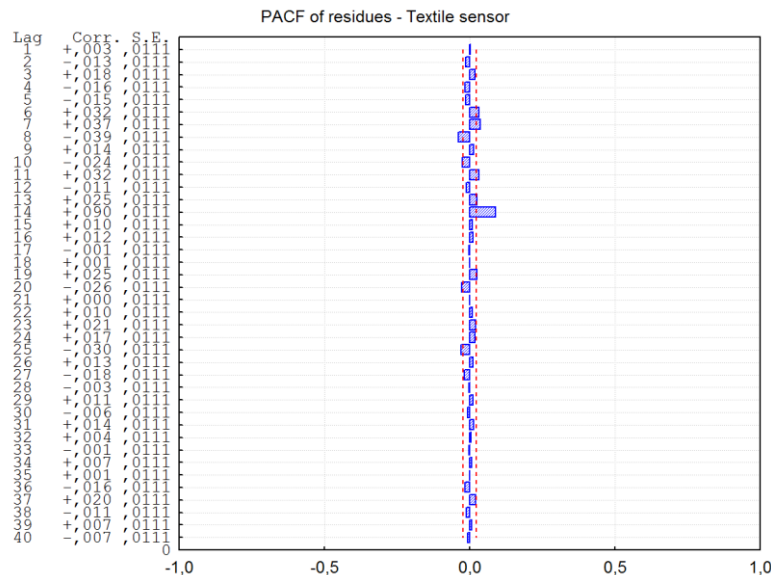
Obr. 4.10 Histogram reziduí - textilní senzor – SW STATISTICA



Obr. 4.11 Normální rozdělení reziduí - textilní senzor – SW STATISTICA



Obr. 4.12 ACF reziduí - textilní senzor – SW STATISTICA



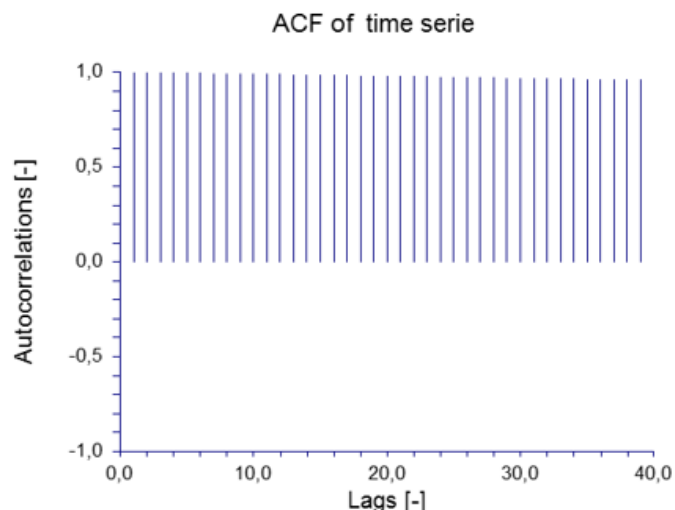
Obr. 4.13 PACF reziduí - textilní senzor – SW STATISTICA

4.1.2 SW NCSS

Napřed jsem načítal data ze souboru *data1.xls*. Použitá verze SW NCSS neobsahovala možnost vytvoření klasického 2D grafu, takže jsem vycházel z průběhu zobrazeného na obr. 4.2 pomocí SW STATISTICA (lze využít i jiný SW).

Z průběhu časové řady lze vidět, že je zapotřebí ji vyhladit. Pro vyhlazení jsem zvolil 3 – bodový klouzavý medián. Ten řadu dostatečně vyhladil a zachoval její charakteristický průběh.

Poté jsem zkontroloval průběhy ACF (obr. 4.14) a PACF vyhlazené řady.

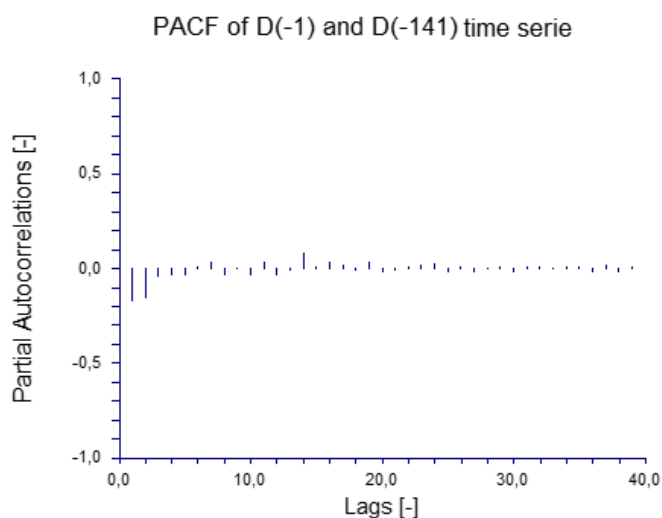


Obr. 4.14 ACF časové řady - textilní senzor - SW NCSS

Na obr. 4.14 je vidět silná sériová závislost. Pro její odstranění jsem řadu jednou nesezónně diferencoval.

Dalším krokem je opět kontrola ACF a PACF takto diferencované řady. Ze znalosti časové řady lze usuzovat, že tato časová řada bude potřebovat sezónní diferencování se zpožděním 1008. Průběhy ACF a PACF pro tak vysoká zpoždění nebylo možné zobrazit. Na základě dosud získaných zkušeností a znalosti časové řady jsem provedl sezónní diferencování se zpožděním 1008.

Poté jsem opět zkontroloval ACF a PACF (obr. 4.15) jednou nesezónně a jednou sezónně (se zpožděním 1008) diferencované řady. Tyto funkce již neukazují na žádné další významné korelace.

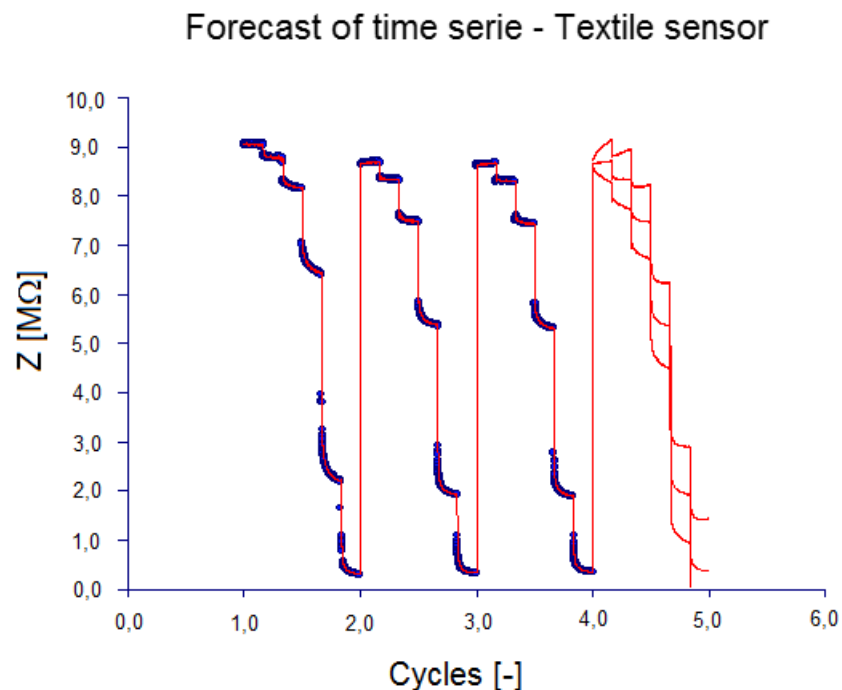


Obr. 4.15 PACF jednou nesezónně a jednou sezónně diferencované časové řady - textilní senzor - SW NCSS

Dalším krokem je volba ARIMA modelu pro tuto časovou řadu a odhad parametrů klouzavých součtů a autoregresních parametrů (sezónních i nesezónních). Po vyzkoušení různých modelů a po přihlédnutí k průběhům ACF a PACF jsem zvolil model ARIMA(1, 1, 1)(1, 1, 0). Nastavení jsem provedl podle obr. 4.16.

Obr. 4.16 Nastavení ARIMA modelu - textilní senzor – SW NCSS

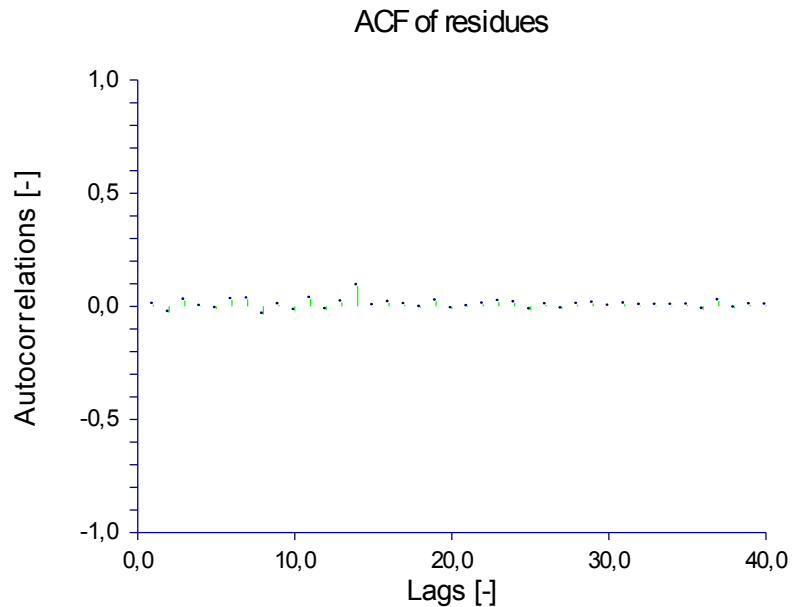
Dále stačí spustit výpočet, po jehož skončení lze zobrazit výsledky této analýzy a zároveň otestovat vhodnost zvoleného modelu. Provedl jsem předpověď pro jeden další cyklus. Graf předpovědi je zobrazen na obr. 4.17.



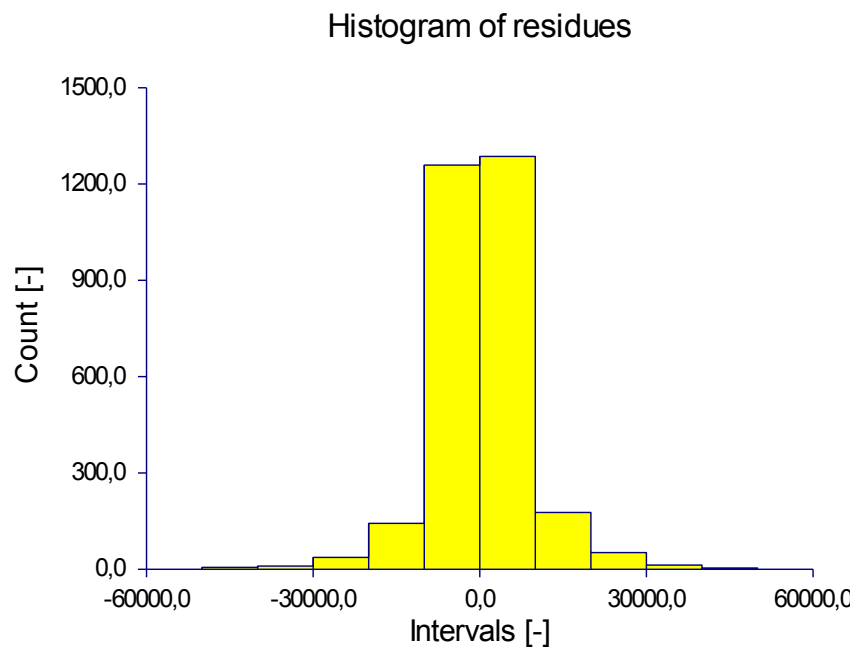
Obr. 4.17 Předpověď časové řady - textilní senzor - SW NCSS

Na obr. 4.17 je vidět, že tento model zachytil všechny charakteristické aspekty této řady a předpověď vypadá věrohodně.

Pro ověření modelu jsem zkontroloval ACF reziduí (obr. 4.18) a histogram reziduí (obr. 4.19). Jak ACF reziduí, tak jejich histogram jsou v pořádku. Zvolený model lze tedy považovat za vhodný.



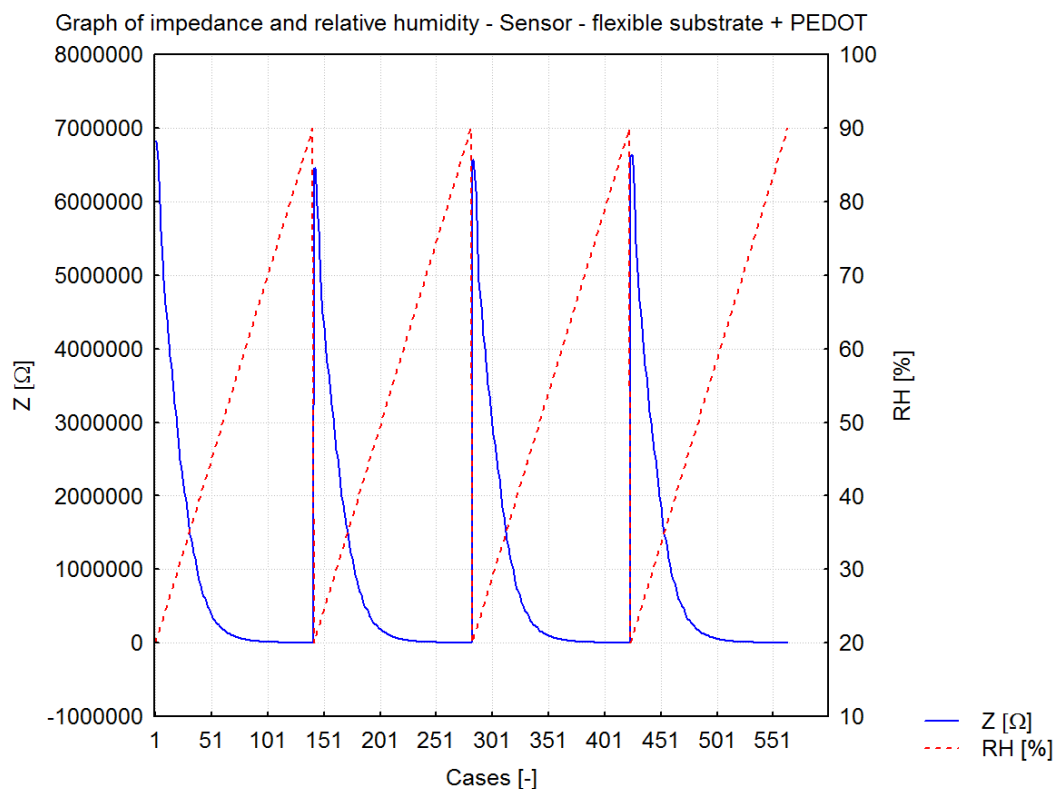
Obr. 4.18 ACF reziduí - textilní senzor - SW NCSS



Obr. 4.19 Histogram reziduí - textilní senzor - SW NCSS

4.2 Senzor - flexibilní substrát + PEDOT:PSS

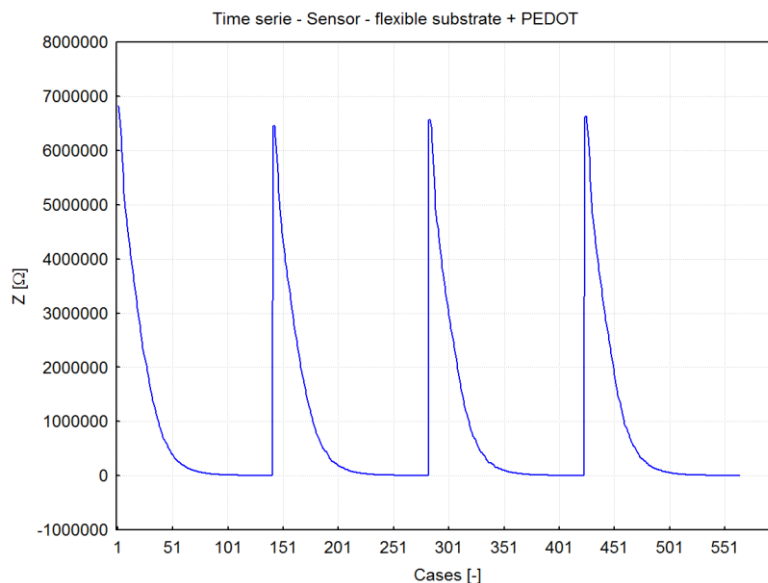
V této kapitole jsem analyzoval data časové řady z měření senzoru: Senzor - flexibilní substrát + PEDOT:PSS. Jedná se o hodnoty impedance substrátu v závislosti na změně vlhkosti při konstantní teplotě 30°C a napětí 1V při frekvenci 1000 Hz. Relativní vlhkost se měnila od hodnoty 20% do hodnoty 90% s krokem 0,5%. Celkově tato řada obsahuje 4 celé cykly a každý cyklus se skládá ze 141 hodnot. Graf popisující změnu impedance a vlhkosti je na obr. 4.20. Na tomto obrázku je vidět, že hodnoty impedance klesají napřed rychleji a poté pomaleji, až se ustálí kolem své minimální hodnoty. Data této časové řady jsou uvedena v souboru *data2.xls*.



Obr. 4.20 Graf změny impedance a relativní vlhkosti - senzor - flexibilní substrát + PEDOT:PSS

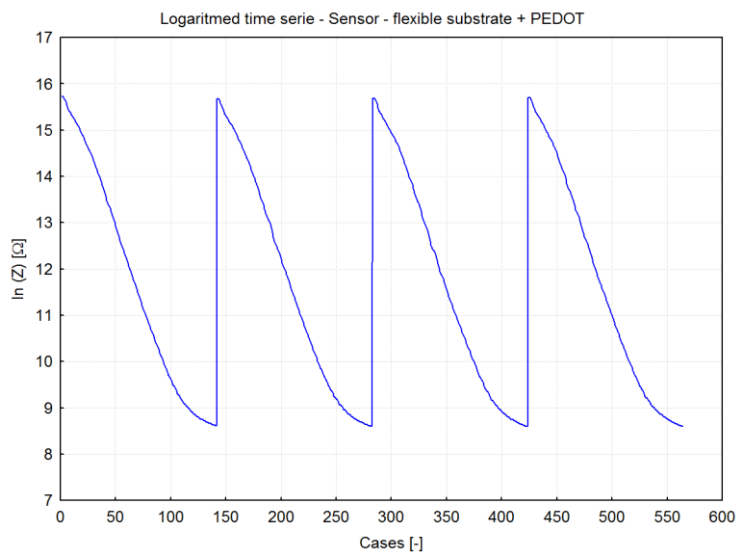
4.2.1 SW STATISTICA

Nejprve jsem načel data této časové řady ze souboru *data2.xls* a zobrazil je pomocí 2D grafu (obr. 4.21).



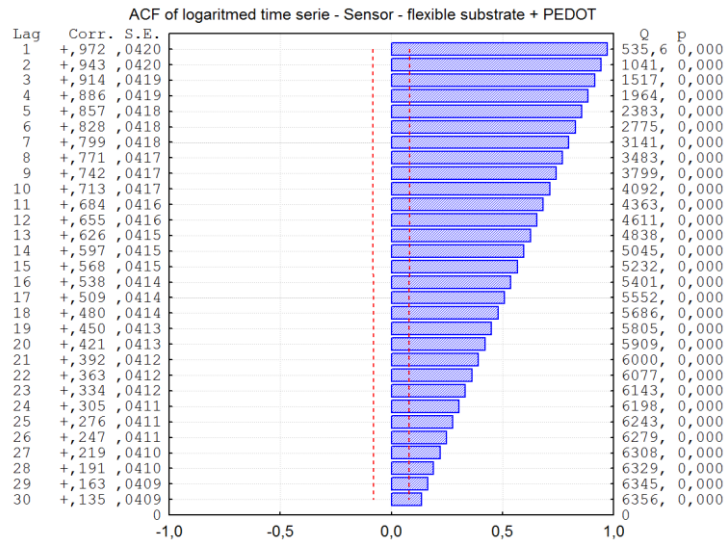
Obr. 4.21 Časová řada - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

Tato časová řada nepotřebuje vyhladit, ale objevují se zde mírné rozdíly v její amplitudě. To by mohlo ovlivnit hodnoty autokorelací, a proto jsem řadu stabilizoval pomocí logaritmizace přirozeným logaritmem (obr. 4.22).



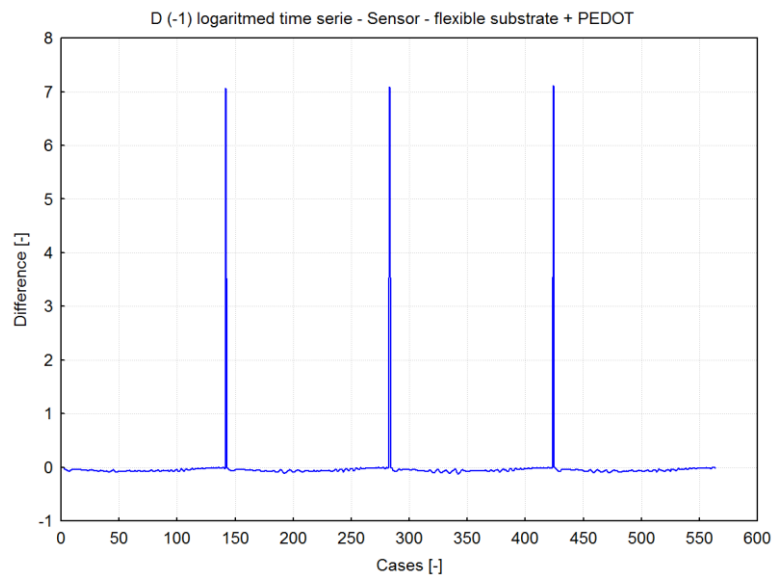
Obr. 4.22 Logaritmovaná časová řada - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

Z grafu na obr. 4.22 je patrné, že hodnoty amplitud jsou nyní více vyrovnané. Dalším krokem je kontrola ACF a PACF. Graf ACF pro prvních 30 hodnot je zobrazen na obr. 4.23.

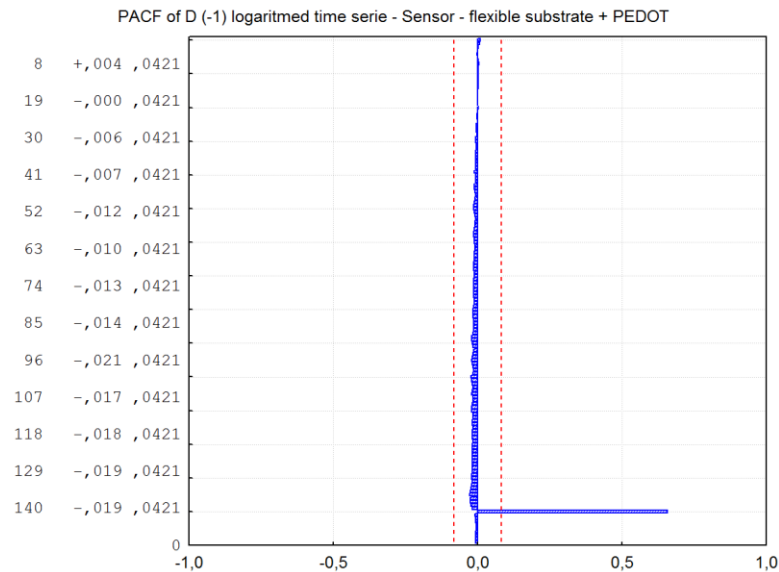


Obr. 4.23 ACF logaritmované časové řady - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

Na obr. 4.23 je vidět silná sériová závislost. Pro její odstranění jsem řadu jednou nesezónně diferencoval (obr. 4.24). Jelikož první pozorování nelze diferencovat, tak bylo vypuštěno. Dále je opět nutné zkontrolovat ACF a PACF. Graf PACF je zobrazen na obr. 4.25.

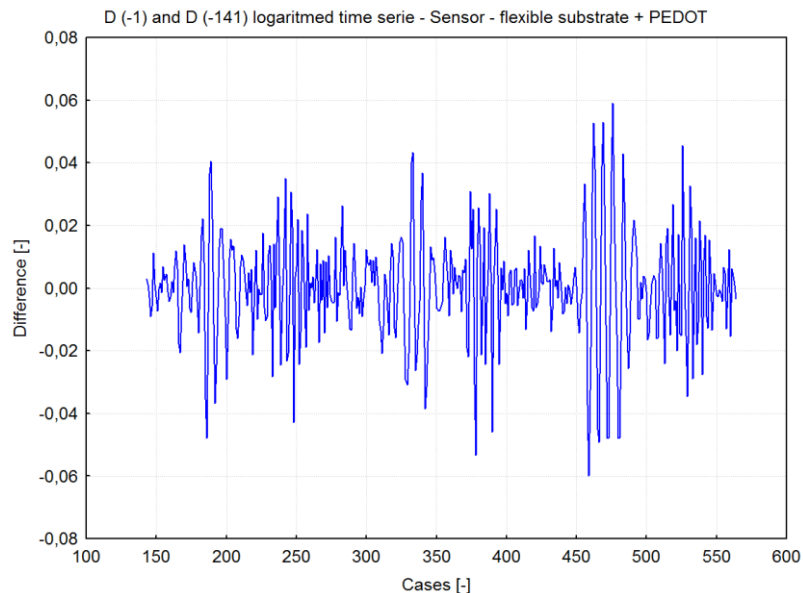


Obr. 4.24 Jednou nesezónně diferencovaná logaritmovaná časová řada - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA



Obr. 4.25 PACF jednou nesezónně diferencované logaritmované časové řady - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

Po zkontrolování ACF a PACF (obr. 4.25) jednou nesezónně diferencované řady jsem objevil významné hroty pro zpoždění o násobcích 141. To napovídá, že v této řadě je obsažena také významná sezónní složka. Pro její odstranění jsem řadu jednou sezónně diferencoval se zpožděním 141 (obr. 4.26). Poté jsem opět zkontroloval ACF a PACF takto diferencované řady, kde jsem už neobjevil žádné další významné autokorelace.

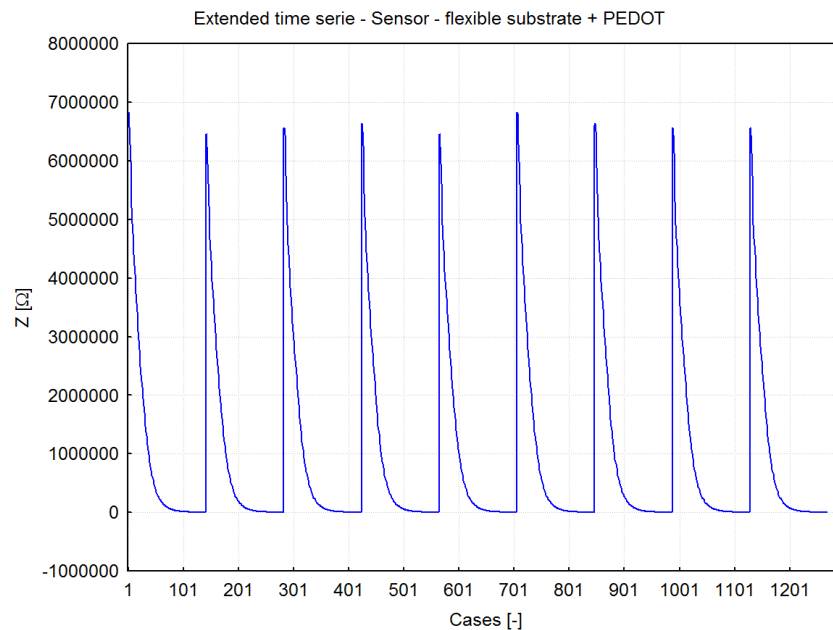


Obr. 4.26 Jednou nesezónně a jednou sezónně diferencovaná logaritmovaná časová řada - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

Dalším krokem je volba ARIMA modelu pro tuto časovou řadu a odhad parametrů klouzavých součtů a autoregresních parametrů (sezónních i nesezónních). Po vyzkoušení různých modelů a po přihlédnutí k průběhům ACF a PACF jsem zvolil model ARIMA(1, 1, 0)(1, 1, 0). Nastavení jsem provedl podle obr. 4.27.

Obr. 4.27 Nastavení ARIMA modelu - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

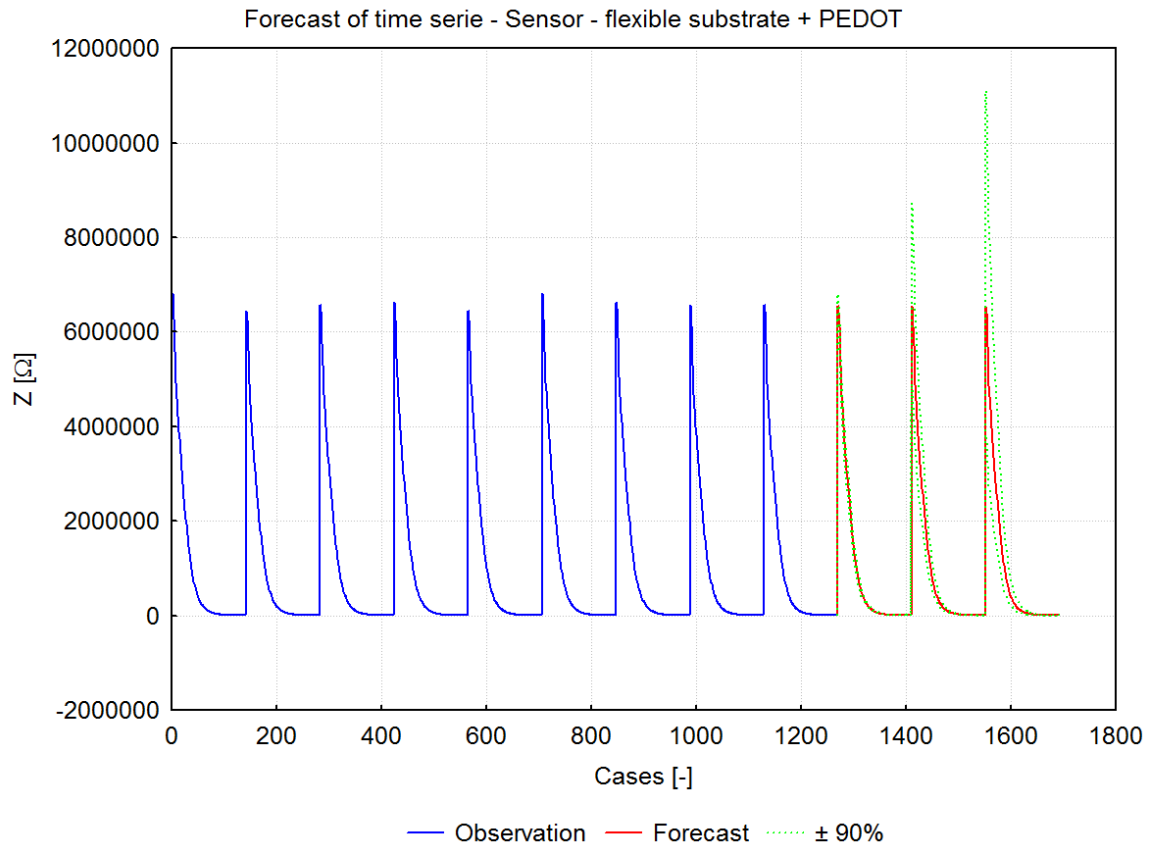
SW STATISTICA vyžaduje pro výpočet alespoň 8 kompletních cyklů, jinak nelze sezónní diferencování nastavit. Tato časová řada obsahuje pouze 4 cykly. Možným řešením je tuto řadu uměle nastavit, například na 9 cyklů. Jelikož jednotlivé cykly si jsou velmi podobné, tak jsem k prvním čtyřem zadaným doplnil dalších pět. V nově vytvořené řadě jsou jednotlivé cykly seřazeny v pořadí: 1-2-3-4-3-2-4-3-3 (obr. 4.28).



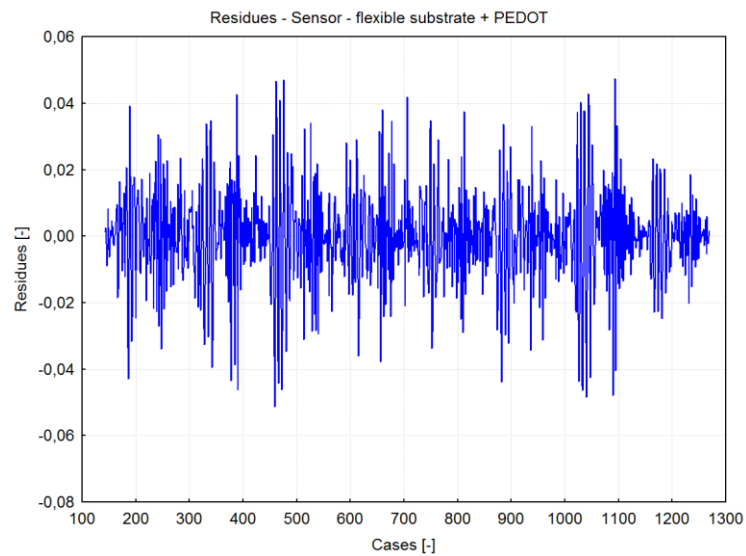
Obr. 4.28 Prodloužená časová řada - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

V dalším kroku jsem spustil výpočet pro takto vytvořenou řadu. Poté lze zobrazit výsledky této analýzy a zároveň otestovat vhodnost zvoleného modelu. Provedl jsem předpověď dalších tří cyklů (obr. 4.29), kde lze vidět, že tento model zachytil všechny charakteristické aspekty této řady. Navíc oblast, která předpovídá s 90% pravděpodobností, je velmi malá.

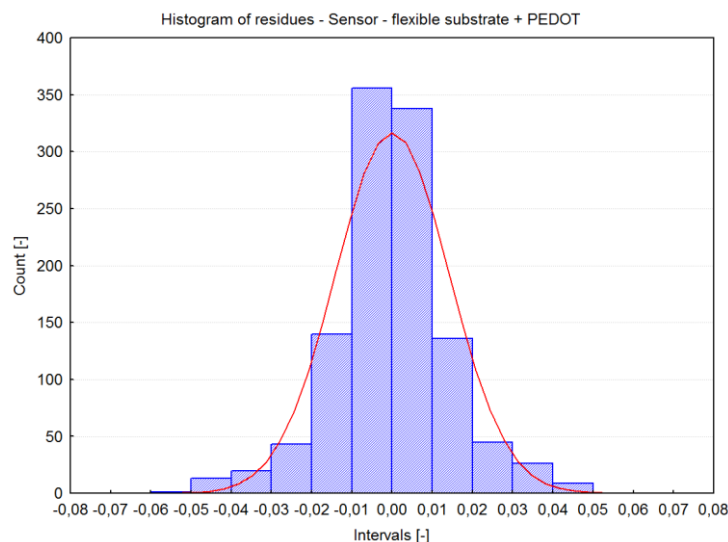
Dále jsem pro kontrolu modelu zobrazil graf reziduí (obr. 4.30), graf histogramu reziduí (obr. 4.31) a graf normálního rozdělení (obr. 4.32) reziduí. Z grafů je patrné, že rezidua mají očekávané rozdělení a vizuální kontrola grafu předpovědi je také podle očekávání. Kontrola ACF a PACF reziduí byla také v pořádku. Z toho lze usuzovat, že zvolený model je správný a získaná předpověď je věrohodná.



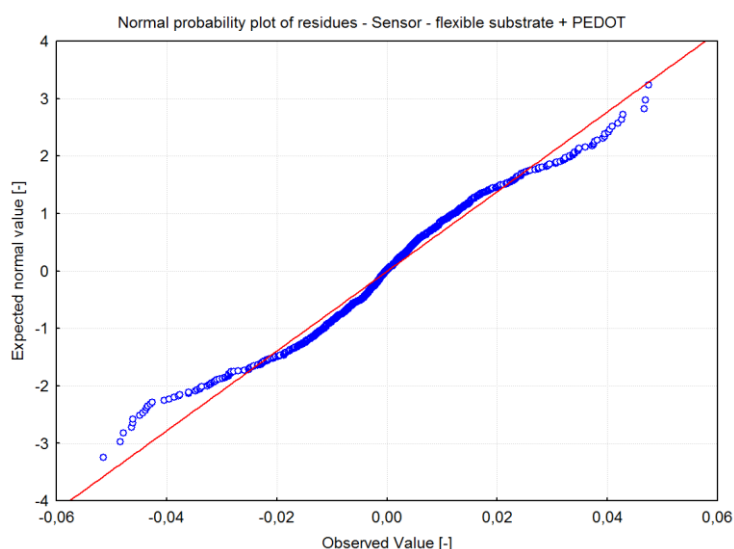
Obr. 4.29 Předpověď časové řady - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA



Obr. 4.30 Graf reziduí - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA



Obr. 4.31 Histogram reziduí -senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA



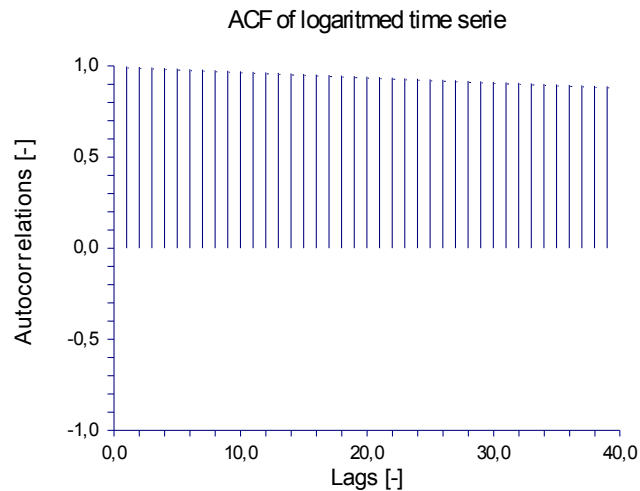
Obr. 4.32 Normální rozdělení reziduí - senzor - flexibilní substrát + PEDOT:PSS – SW STATISTICA

4.2.2 SW NCSS

Nejprve jsem načel data ze souboru *data2.xls*. Použitá verze SW NCSS neobsahovala možnost vytvoření klasického 2D grafu, takže jsem vycházel z průběhu zobrazeného na obr. 4.21 pomocí SW STATISTICA (lze využít i jiný SW).

Tato časová řada nepotřebuje vyhladit, ale objevují se zde mírné rozdíly v její amplitudě. Tyto rozdíly by mohly ovlivnit hodnoty autokorelací, a proto jsem řadu stabilizoval pomocí logaritmizace.

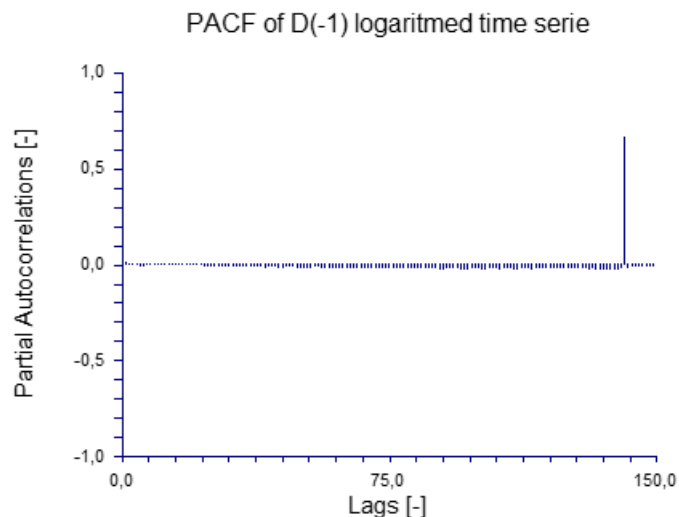
Dalším krokem je kontrola ACF a PACF. Graf ACF pro prvních 40 hodnot je zobrazen na obr. 4.33.



Obr. 4.33 ACF logaritmované časové řady - senzor - flexibilní substrát + PEDOT:PSS – SW NCSS

Na obr. 4.33 je vidět silná sériová závislost. Pro její odstranění jsem řadu jednou nesezónně diferencoval.

Poté jsem zkontroloval ACF a PACF (obr. 4.34) takto diferencované řady.



Obr. 4.34 PACF logaritmované časové řady - senzor - flexibilní substrát + PEDOT:PSS – SW NCSS

Po zkontrolování ACF a PACF (obr. 4.34) jednou nesezónně diferencované řady jsem objevil významné hroty pro zpoždění o násobcích 141. To ukazuje, že v časové řadě je obsažena významná sezónní složka. Pro její odstranění jsem řadu jednou sezónně diferencoval se zpožděním 141.

Poté jsem opět zkontroloval ACF a PACF takto diferencované řady, kde jsem už neobjevil žádné další významné autokorelace.

Dalším bodem je volba ARIMA modelu pro tuto časovou řadu a odhad parametrů klouzavých součtů a autoregresních parametrů (sezónních i nesezónních). Po vyzkoušení

různých modelů a po přihlednutí k průběhům ACF a PACF jsem zvolil model ARIMA(1, 1, 0)(1, 1, 0). Nastavení jsem provedl podle obr. 4.35.

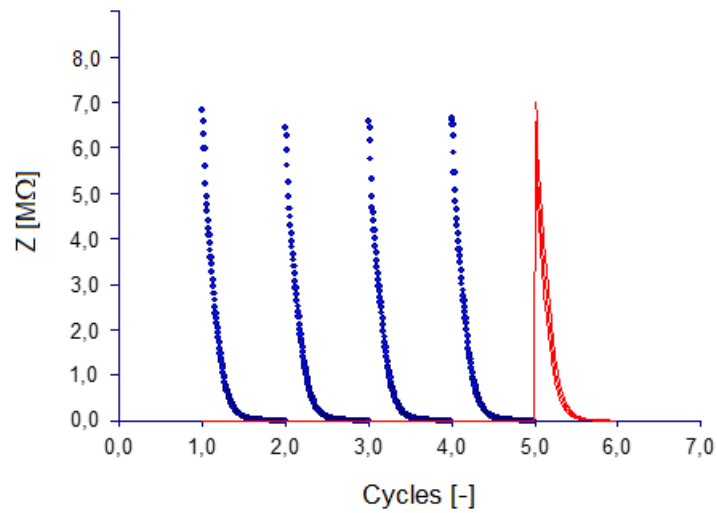
The screenshot shows the 'ARIMA Model Options' dialog box in STATISTICA. The 'Time Series Variable' is 'C1'. Under 'Forecasting Options', 'Number of Forecasts' is 141. Under 'Data Adjustment Options', 'Remove Mean' is checked. Under 'Seasonality Options', 'Seasons' is 141 and 'First Season' is 1. Under 'ARIMA Model Options', 'Regular AR' is 1, 'Regular MA' is 0, 'Seasonal AR' is 1, and 'Seasonal MA' is 0. Under 'Estimation Options', 'Max Iterations' is 20, 'Convergence' is 0.00001, and 'Lambda' is 0.1.

Obr. 4.35 Nastavení ARIMA modelu - senzor - flexibilní substrát + PEDOT:PSS – SW NCSS

Dále jsem spustil výpočet pro takto zvolený model. Nastavil jsem předpověď pro 141 hodnot (jeden cyklus). Po skončení výpočtu se zobrazí výsledky. Graf předpovědi je na obr. 4.36. Na tomto obrázku je vidět, že zvolený model očekávaně předpovídá budoucí průběh časové řady.

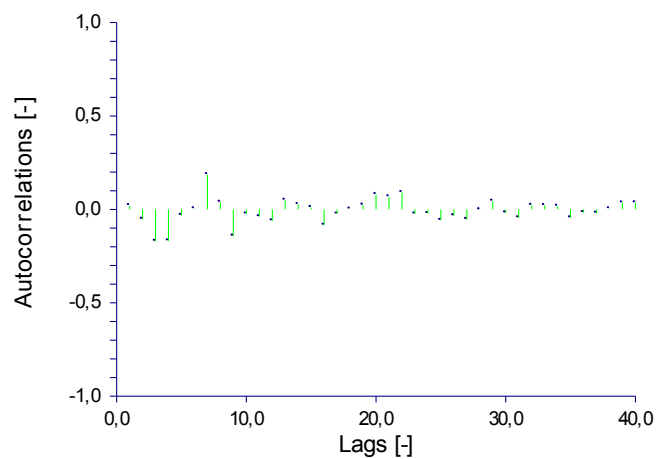
Pro ověření modelu jsem zkontroloval ACF reziduí (obr. 4.37) a histogram reziduí (obr. 4.38). Histogram reziduí nemá nejlepší rozdělení, ale požadovanému normálnímu rozdělení se blíží. Zvolený model lze tedy považovat za vhodný.

Forecast of time serie - Sensor-flexible substrate+PEDOT



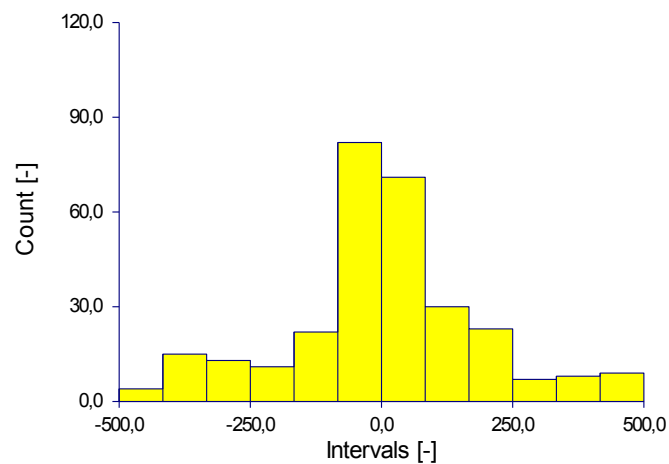
Obr. 4.36 Předpověď časové řady - senzor - flexibilní substrát + PEDOT:PSS – SW NCSS

ACF of residues



Obr. 4.37 ACF reziduí - senzor - flexibilní substrát + PEDOT:PSS – SW NCSS

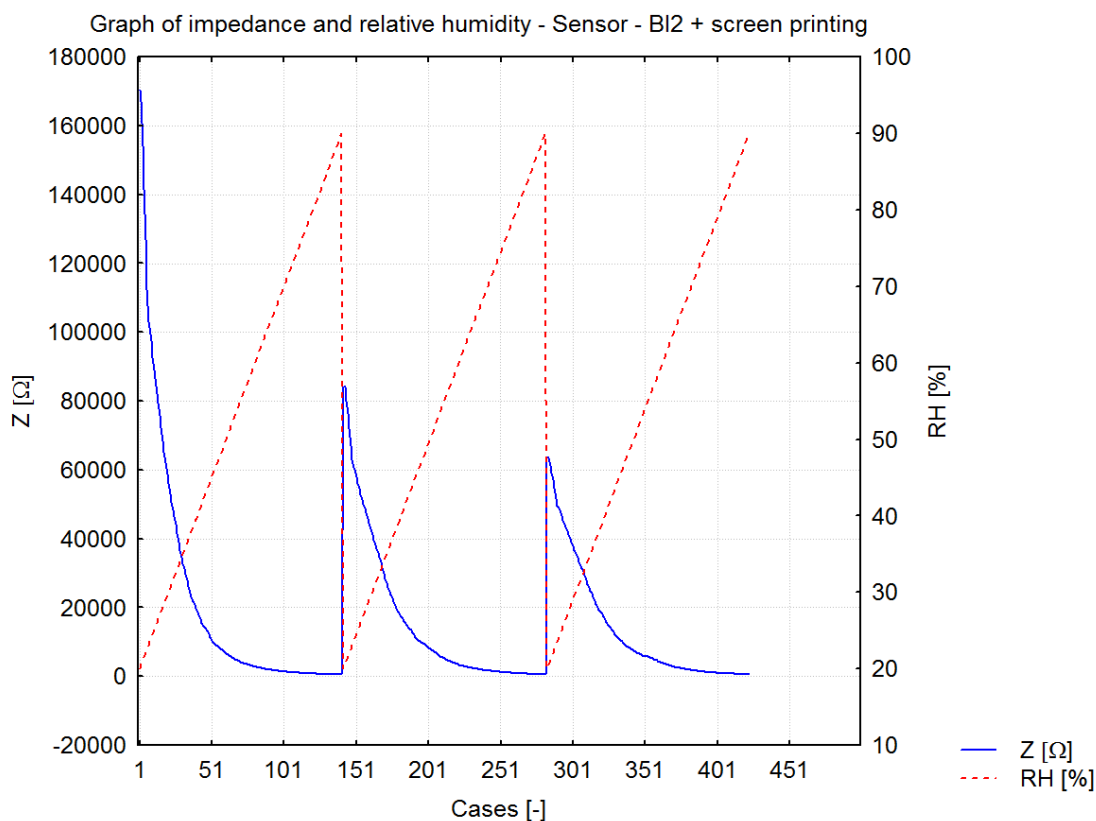
Histogram of residues



Obr. 4.38 Histogram reziduí - senzor - flexibilní substrát + PEDOT:PSS – SW NCSS

4.3 Senzor - BI2 + sítotisk

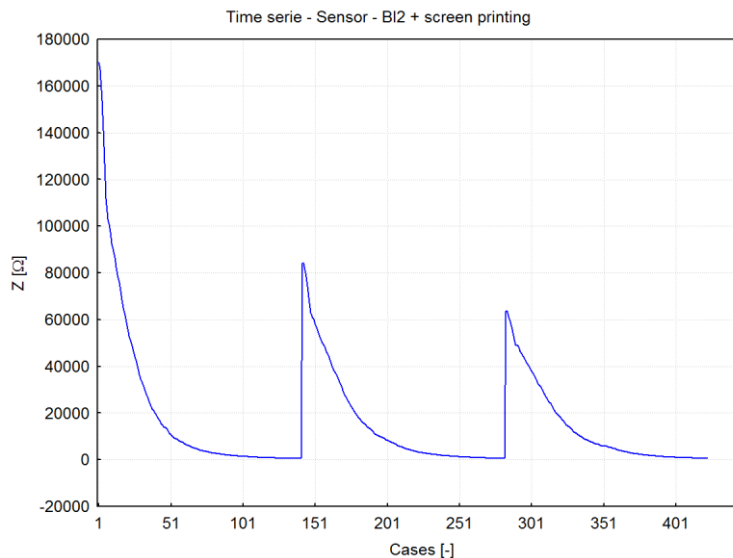
V této podkapitole jsem analyzoval data časové řady z měření senzoru: Senzor – BI2 + sítotisk. Tato časová řada popisuje hodnoty impedance substrátu v závislosti na změně vlhkosti při konstantní teplotě 30°C a napětím 1 V při frekvenci 1000 Hz. Relativní vlhkost se měnila od hodnoty 20% do hodnoty 90% s krokem 0,5%. Celkově tato řada obsahuje 3 celé cykly a každý cyklus se skládá ze 141 hodnot. Graf popisující změnu impedance a relativní vlhkosti je na obr. 4.39. Je zde vidět, že hodnoty impedance klesají napřed rychleji a poté pomaleji (odhadem exponenciálně), až se ustálí kolem své minimální hodnoty. Je tedy patrné, že s rostoucí vlhkostí se zmenšuje hodnota impedance. Navíc každý nový cyklus začíná na nižší hodnotě impedance. Data této časové řady jsou uvedena v souboru *data3.xls*.



Obr. 4.39 Graf změny impedance a relativní vlhkosti - senzor - BI2 + sítotisk

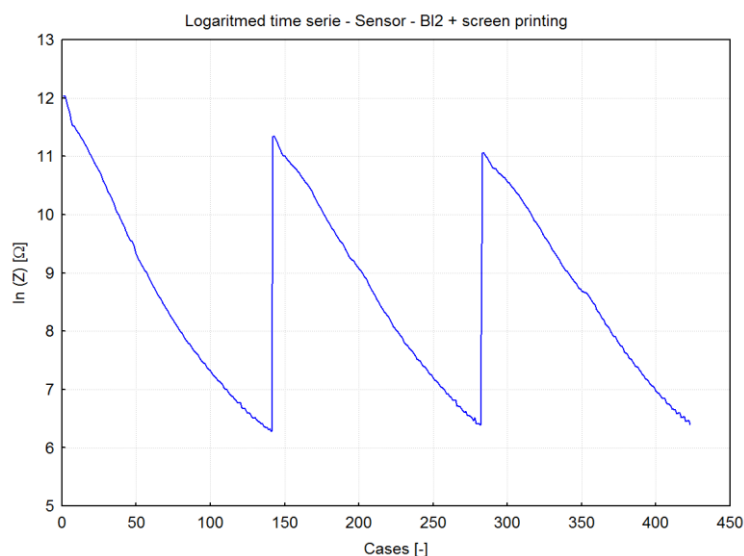
4.3.1 SW Statistica

V prvním kroku jsem načel data této řady a zobrazil je pomocí 2D grafu (obr. 4.40).



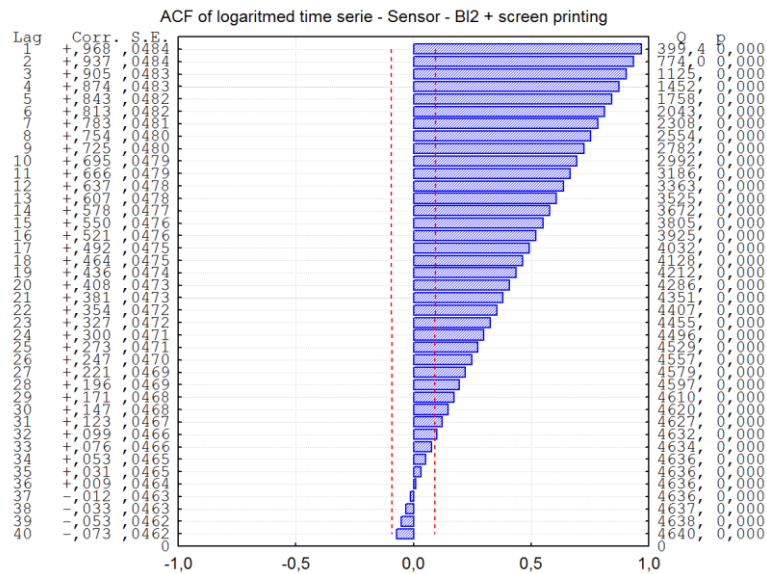
Obr. 4.40 Časová řada - senzor - BI2 + sítotisk – SW STATISTICA

Tato časová řada nepotřebuje vyhladit, ale objevují se zde velké rozdíly v její amplitudě. To by mohlo ovlivnit hodnoty autokorelací, a tak jsem řadu stabilizoval pomocí logaritmizace přirozeným logaritmem (obr. 4.41).



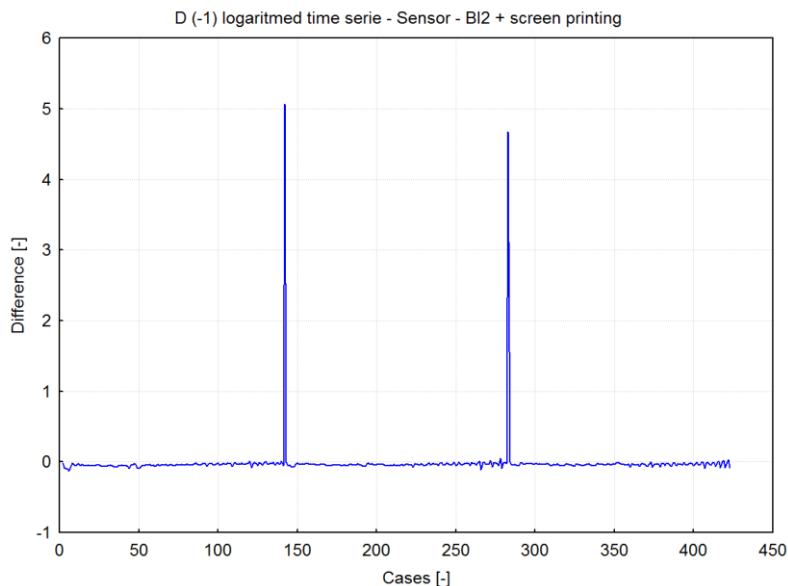
Obr. 4.41 Logaritmovaná časová řada - senzor - BI2 + sítotisk – SW STATISTICA

Z grafu na obr. 4.41 je zřejmé, že hodnoty amplitud jsou nyní více vyrovnané. Dalším krokem je kontrola ACF a PACF. Graf ACF pro prvních 40 hodnot je zobrazen na obr. 4.42.

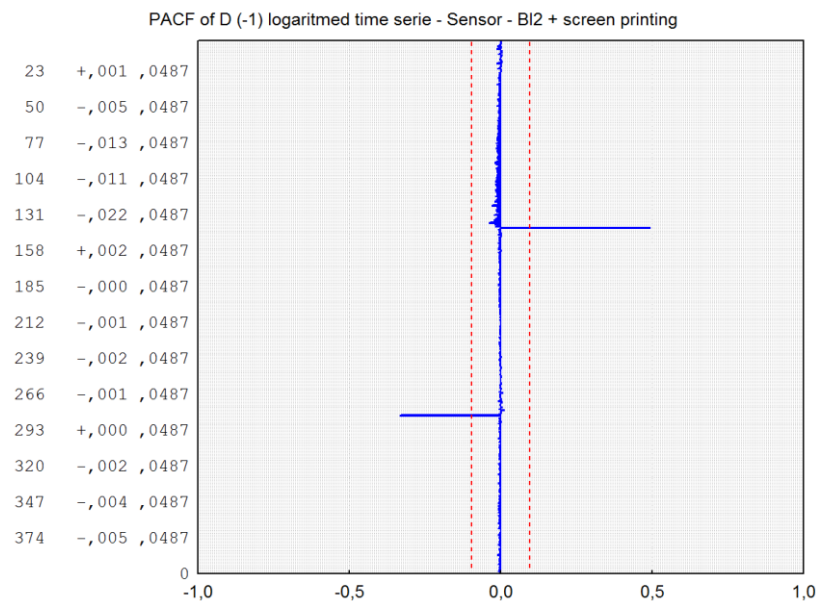


Obr. 4.42 ACF logaritmované časové řady - senzor - BI2 + sitotisk – SW STATISTICA

Na obrázku 4.42 je vidět silná sériová závislost. Pro její odstranění jsem řadu jednou nesezónně diferencoval (obr. 4.43). Protože první pozorování nemůže být diferencováno, tak bylo vypuštěno. Poté je opět nutné zkontrolovat ACF a PACF. Graf PACF je zobrazen na obr. 4.44.



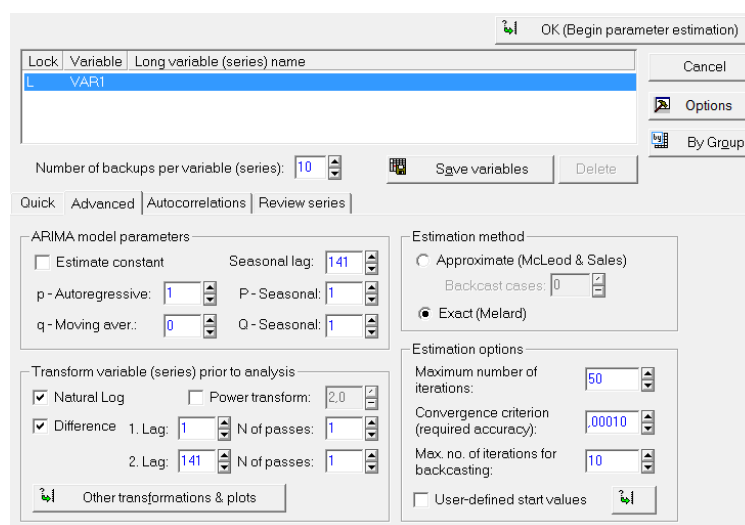
Obr. 4.43 Jednou nesezónně diferencovaná časová řada - senzor - BI2 + sitotisk – SW STATISTICA



Obr. 4.44 PACF jednou nesezónně diferencované časové řady - senzor - BI2 + sítotisk – SW STATISTICA

Po zkontrolování ACF a PACF (obr. 4.44) jednou nesezónně diferencované řady jsem objevil významné hroty pro zpoždění o násobcích 141. To ukazuje, že v časové řadě je obsažena významná sezónní složka. Pro její odstranění jsem řadu jednou sezónně diferencoval se zpožděním 141. Poté jsem opět zkontroloval ACF a PACF takto diferencované řady, kde jsem už neobjevil žádné další významné autokorelace.

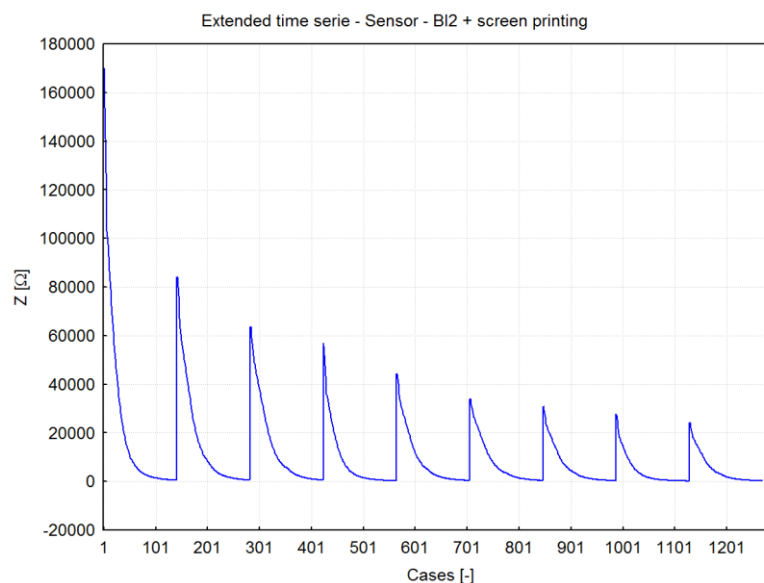
Dalším krokem je určení ARIMA modelu pro tuto časovou řadu a odhad parametrů klouzavých součtů a autoregresních parametrů (sezónních i nesezónních). Po vyzkoušení různých modelů jsem zvolil model ARIMA(1, 1, 0)(1, 1, 1). Nastavení jsem provedl podle obr. 4.45.



Obr. 4.45 Nastavení ARIMA modelu - senzor - BI2 + sítotisk – SW STATISTICA

SW STATISTICA vyžaduje pro výpočet alespoň 8 kompletních cyklů, jinak nelze sezónní diferencování nastavit. Tato časová řada obsahuje pouze 3 cykly. Možným řešením je tuto řadu uměle nastavit na alespoň 8 cyklů. Jelikož diferencováním řady dojde k jejímu zkrácení, tak ke 3 zadaným cyklům jsem doplnil dalších 6 cyklů (přesně 8 nestačí). Protože zadané cykly vykazují výrazné rozdíly v amplitudě, tak není možné tuto řadu nastavit pomocí zkopírování jednotlivých cyklů, ale je nutné jejich průběh odhadnout.

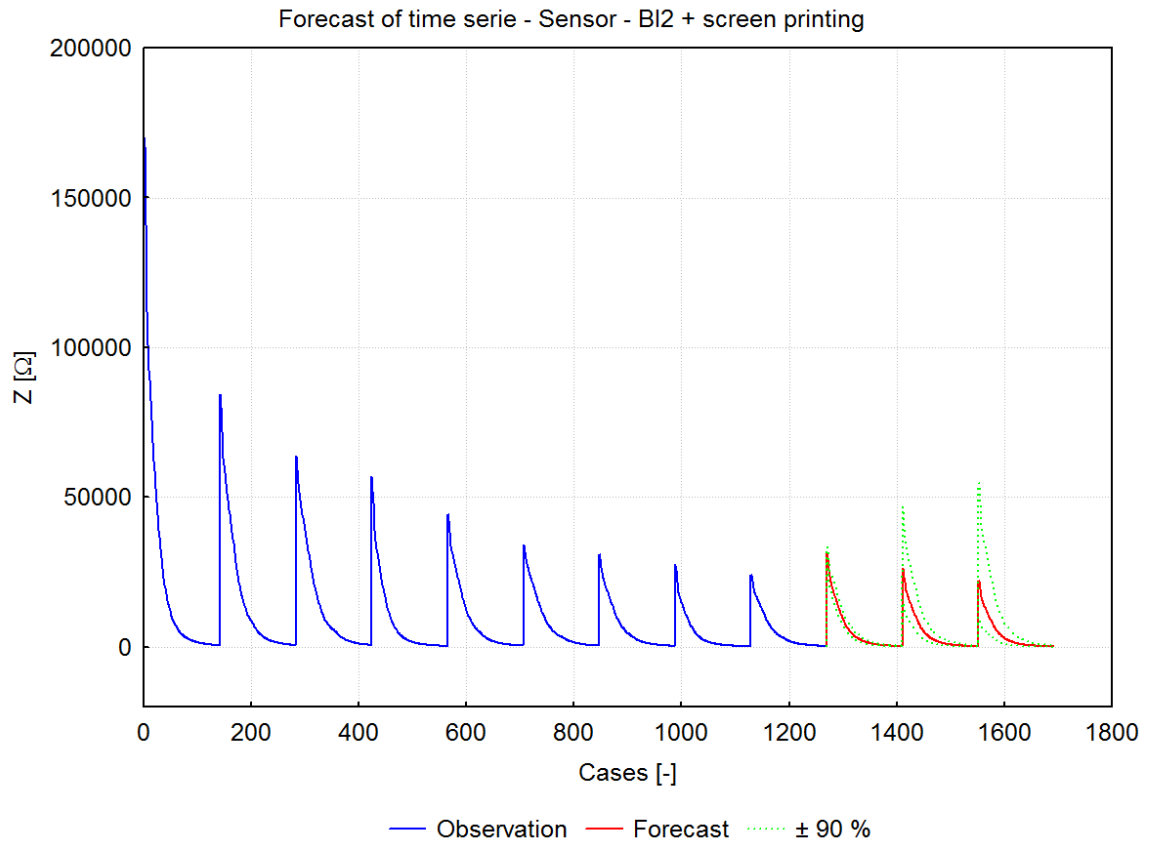
Podle obr. 4.40 jsem odhadl, že velikosti amplitud jednotlivých cyklů klesají exponenciálně. Proto jsem provedl transformaci řady jejím logaritmováním. Tím jsem převedl problém exponenciální na problém lineární. Touto úpravou se odhad významně zjednodušil. Poté jsem odhadl průběh dalších cyklů a transformoval je pomocí inverzní transformace zpět. Data této prodloužené řady jsou uvedena v souboru *data3.xls*. Výsledný průběh prodloužené řady je na obr. 4.46.



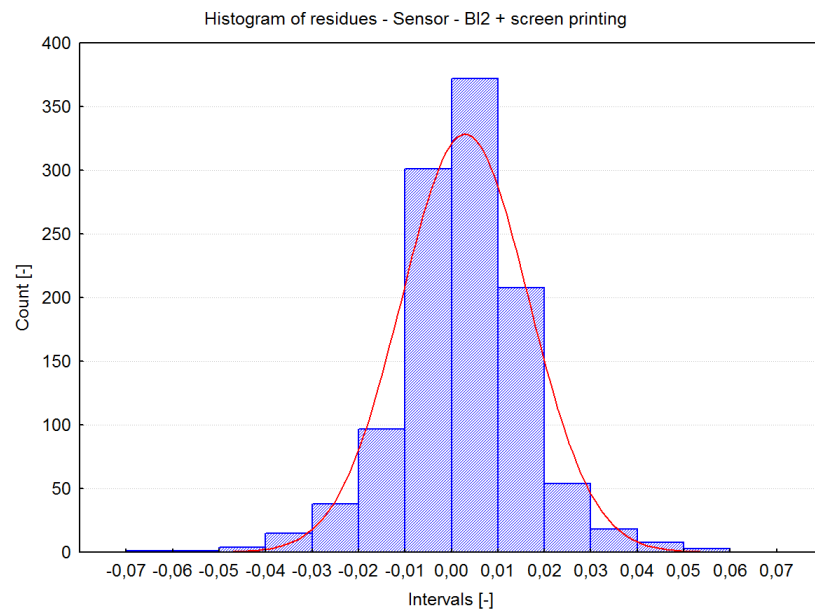
Obr. 4.46 Prodloužená časová řada - senzor - BI2 + sítotisk – SW STATISTICA

V dalším kroku jsem spustil výpočet pro takto vytvořenou řadu. Také lze zobrazit výsledky této analýzy a otestovat vhodnost zvoleného modelu. Provedl jsem předpověď dalších tří cyklů (obr. 4.47). Tento model zachytil všechny charakteristické vlastnosti této řady. Oblast, která předpovídá s 90% pravděpodobností, je velmi malá.

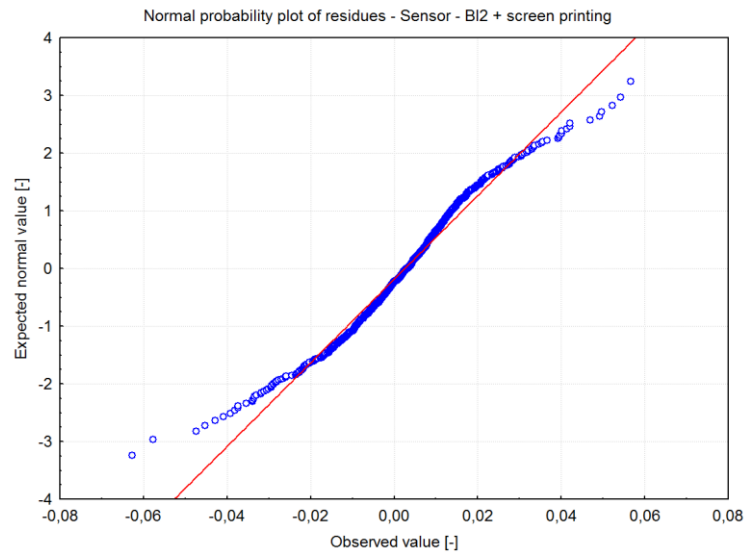
Také jsem pro kontrolu modelu zobrazil graf reziduí, graf histogramu reziduí (obr. 4.48) a graf normálního rozdělení reziduí (obr. 4.49). Z grafů je patrné, že rezidua mají očekávané rozdělení. Dále jsem zobrazil ACF (obr. 4.50) a PACF (obr. 4.51) reziduí. ACF a PACF nevykazují korelaci mezi rezidui. Z kontroly reziduí lze usuzovat, že zvolený model je správný, a tedy získaná předpověď je věrohodná.



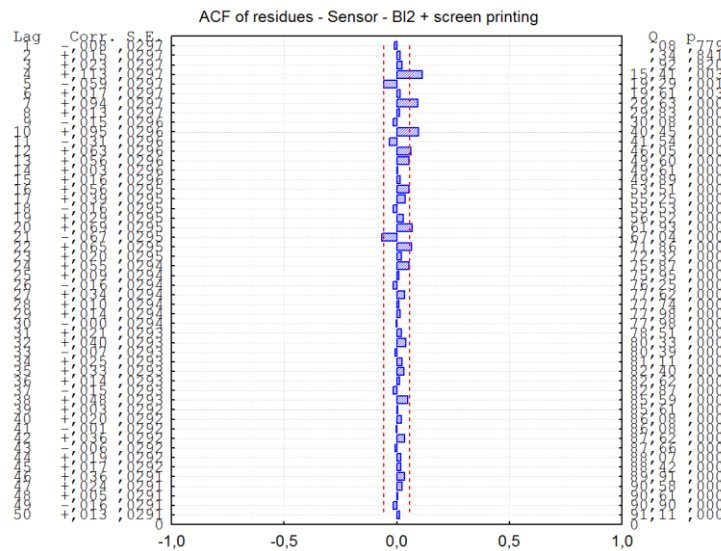
Obr. 4.47 Předpověď časové řady - senzor - BI2 + sítotisk – SW STATISTICA



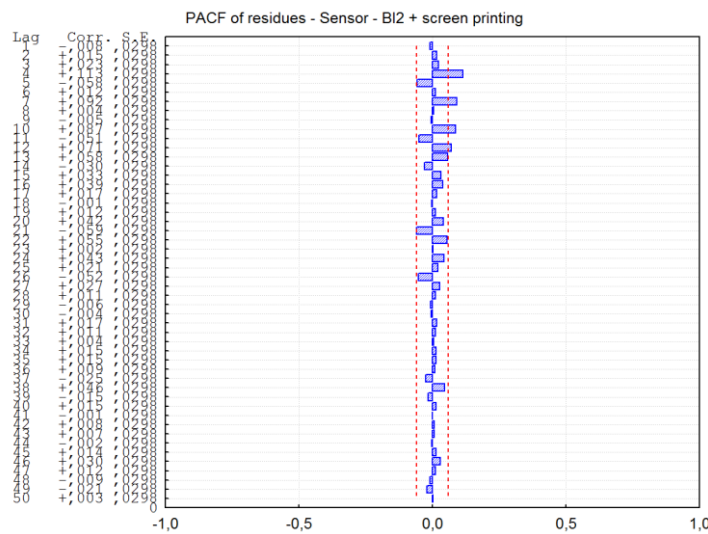
Obr. 4.48 Histogram reziduí - senzor - BI2 + sítotisk – SW STATISTICA



Obr. 4.49 Normální rozdělení reziduí - senzor - BI2 + sítotisk – SW STATISTICA



Obr. 4.50 ACF reziduí - senzor - BI2 + sítotisk – SW STATISTICA



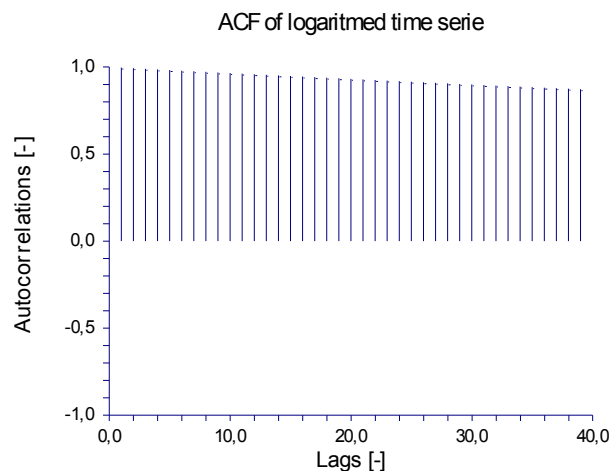
Obr. 4.51 PACF reziduí - senzor - BI2 + sítotisk – SW STATISTICA

4.3.2 SW NCSS

Nejprve jsem načel data ze souboru *data3.xls*. Použitá verze SW NCSS neobsahovala možnost vytvoření klasického 2D grafu, takže jsem vycházel z průběhu zobrazeného na obr. 4.40 pomocí SW STATISTICA (lze využít i jiný SW).

Tato časová řada nepotřebuje vyhladit, ale objevují se zde výrazné rozdíly v její amplitudě. Tyto rozdíly by mohly ovlivnit hodnoty autokorelací, a proto jsem řadu stabilizoval pomocí logaritmizace.

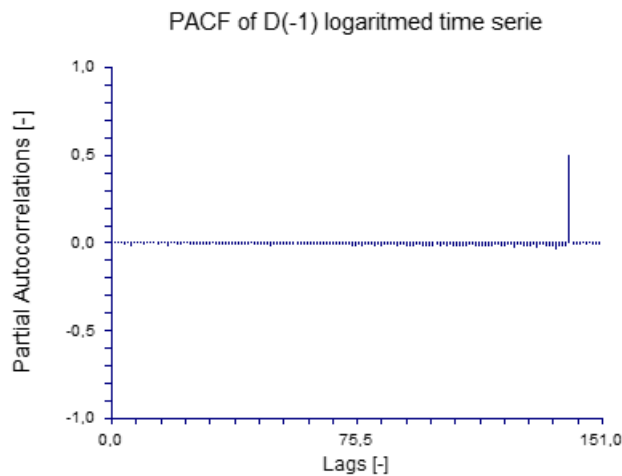
Dalším krokem je kontrola ACF a PACF. Graf ACF pro prvních 40 hodnot je zobrazen na obr. 4.52.



Obr. 4.52 ACF logaritmované časové řady - senzor - BI2 + sítotisk – SW NCSS

Na obr. 4.52 je patrná silná sériová závislost. Pro její odstranění jsem řadu jednou nesezónně diferencoval.

Poté jsem zobrazil ACF a PACF (obr. 4.53) takto diferencované řady.



Obr. 4.53 PACF jednou nesezónně diferencované logaritmované časové řady - senzor - BI2 + sítotisk – SW NCSS

Po zkontrolování ACF a PACF (obr. 4.53) jednou nesezónně diferencované řady jsem objevil významné hroty pro zpoždění o násobcích 141. To naznačuje, že v časové řadě je obsažena významná sezónní složka. Pro její odstranění jsem řadu jednou sezónně diferencoval se zpožděním 141.

V dalším kroku jsem opět zkontroloval ACF a PACF takto diferencované řady, kde jsem už neobjevil žádné další významné autokorelace.

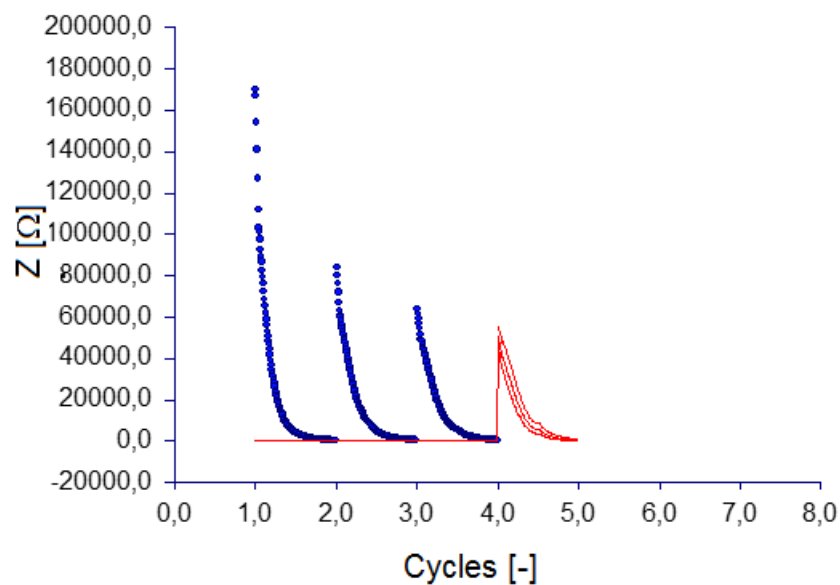
Následně jsem přistoupil k volbě ARIMA modelu pro tuto časovou řadu a odhadu parametrů klouzavých součtů a autoregresních parametrů (sezónních i nesezónních). Po vyzkoušení různých modelů a po přihlédnutí k průběhům ACF a PACF jsem zvolil model $ARIMA(1, 1, 0)(1, 1, 0)$. Nastavení jsem provedl podle obr. 4.54.

Obr. 4.54 Nastavení ARIMA modelu - senzor - BI2 + sítotisk – SW NCSS

Dále jsem spustil výpočet pro takto zvolený model. Nastavil jsem předpověď pro 141 hodnot (jeden cyklus). Po skončení výpočtu lze zobrazit výsledky. Graf předpovědi je na obr. 4.55. Na tomto obrázku je vidět, že zvolený model očekávaně předpovídá budoucí průběh časové řady.

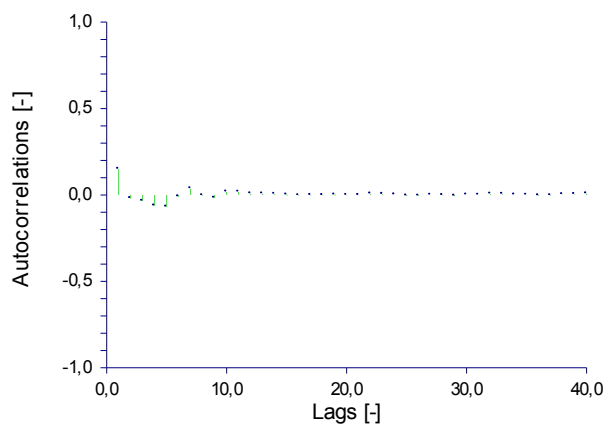
Pro ověření modelu jsem zkontroloval ACF reziduí (obr. 4.56) a histogram reziduí (obr. 4.57). Histogram reziduí nemá ideální rozdělení, ale požadovanému normálnímu rozdělení se blíží. Zvolený model lze tedy považovat za vhodný.

Forecast of time serie - Sensor - BI2 + screen printing



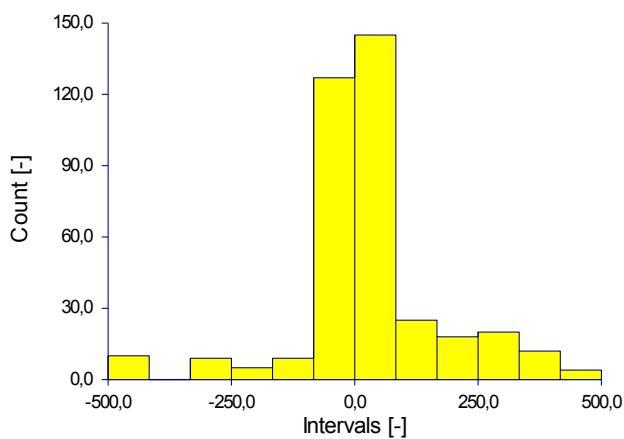
Obr. 4.55 Předpověď časové řady - senzor - BI2 + sítotisk – SW NCSS

ACF of residues



Obr. 4.56 ACF reziduí - senzor - BI2 + sítotisk – SW NCSS

Histogram of residues



Obr. 4.57 Histogram reziduí - senzor - BI2 + sítotisk – SW NCSS

4.4 Výsledky a porovnání provedených analýz

Pro analýzu časových řad jsem vybral ARIMA metodologii. Ta je sice nejnáročnější na provedení a rozhodování, ale poskytuje velmi dobré výsledky.

Pomocí této metodologie jsem analyzoval časové řady z výsledků materiálových zkoušek tří různých senzorů: textilní senzor, senzor - flexibilní substrát + PEDOT:PSS a senzor - BI2 + sítotisk. Analýzu jsem provedl pomocí SW STATISTICA (verze 7.1) a NCSS (verze 7).

SW STATISTICA vyžadoval pro nastavení sezónního diferencování alespoň 8 kompletních cyklů. Ani jedna časová řada neobsahovala požadovaný počet cyklů, a tak jsem musel časové řady uměle nastavit kombinací zadaných cyklů nebo odhadem dalších cyklů. Tím mohlo dojít (došlo) k zanesení mírné chyby, protože některé průběhy se opakovaly a staly se tak více významnými. Byl to ale jediný způsob, jak analýzu provést.

Použitá verze SW NCSS neobsahovala možnost tvorby obyčejných 2D grafů, a tak jsem pro zobrazení průběhu časových řad musel využít SW STATISTICA. Dále tato verze neumožňovala detailní formátování grafů, takže některé grafy nejsou vizuálně nejlepší. Vyzkoušel jsem i trial verzi SW NCSS 10 a tam jsou už tyto nedostatky odstraněny.

Výsledky provedených analýz:

a) textilní senzor

Data této časové řady potřebovala napřed filtrovat, aby se dosáhlo stejné délky jednotlivých cyklů. Poté jsem ještě řadu vyhladil pomocí tříbodového klouzavého mediánu. Dále jsem řadu diferencoval jednou nesezónně a jednou sezónně se zpožděním 1008.

V SW STATISTICA mi nejlépe vyšel model $ARIMA(1, 1, 1)(1, 1, 1)$ a v SW NCSS model $ARIMA(1, 1, 1)(1, 1, 0)$. To, že vyšel v každém SW jiný model, může být způsobeno tím, že v SW STATISTICA jsem danou řadu uměle nastavoval a tím, že oba dva modely dostatečně odpovídají této řadě. Také jsem provedl předpověď časové řady a ověřil správnost zvoleného modelu.

b) senzor - flexibilní substrát + PEDOT:PSS

Data této časové řady nepotřebovala filtrovat ani vyhladit, ale objevovali se zde malé rozdíly v amplitudách jednotlivých cyklů, takže jsem tuto řadu stabilizoval pomocí transformace logaritmováním. Dále jsem řadu diferencoval jednou nesezónně a jednou sezónně se zpožděním 141.

V SW STATISTICA mi nejlépe vyšel model $ARIMA(1, 1, 0)(1, 1, 0)$ a v SW NCSS model $ARIMA(1, 1, 0)(1, 1, 0)$. V obou dvou SW mi vyšel stejný model. Také jsem provedl předpověď časové řady a ověřil správnost zvoleného modelu.

c) senzor - BI2 + sítotisk

Data této časové řady nepotřebovala filtrovat ani vyhladit, ale objevovali se zde rozdíly v amplitudách jednotlivých cyklů, takže jsem tuto řadu stabilizoval pomocí transformace logaritmováním. Dále jsem řadu diferencoval jednou nesezónně a jednou sezónně se zpožděním 141.

V SW STATISTICA mi nejlépe vyšel model $ARIMA(1, 1, 0)(1, 1, 1)$ a v SW NCSS model $ARIMA(1, 1, 0)(1, 1, 0)$. To, že vyšel v každém SW jiný model, může být způsobeno tím, že v SW STATISTICA jsem danou řadu uměle nastavoval a tím, že oba dva modely dostatečně odpovídají této řadě. Také jsem provedl předpověď časové řady a ověřil správnost zvoleného modelu.

Závěr

V této práci jsem se zabýval analýzou časových řad a její text jsem tematicky rozdělil do čtyř hlavních částí.

V první kapitole jsem vysvětlil, co jsou to časové řady a definoval základní cíle jejich analýzy. Dále jsem uvedl specifické problémy, které se vyskytují při jejich analýze.

Ve druhé kapitole jsem se zabýval metodami pro statistické vyhodnocování časových řad. Popsal jsem dekompoziční metodu, metodu Census I, metodu Census II, spektrální analýzu časových řad a Box-Jenkinsovu metodologii (ARIMA modely).

Ve třetí kapitole jsem se zabýval možnostmi SW STATISTICA a NCSS v oblasti analýzy časových řad. Při analýze jejich možností jsem vycházel z dokumentace a manuálů od obou SW a také z vlastních zkušeností, které jsem získal prací v těchto SW. Uvedl jsem, jaké metody tyto SW nabízejí a jaké jsou možnosti těchto metod v obou SW. Nakonec jsem oba statistické programy porovnal a uvedl jejich hlavní výhody a nevýhody.

Ve čtvrté kapitole jsem analyzoval časové řady z výsledků materiálových zkoušek tří různých senzorů (textilní senzor, senzor – flexibilní substrát + PEDOT:PSS a senzor – BI2 + sítotisk). Analýzu jsem provedl v obou SW pomocí Box-Jenkinsovy metodologie (ARIMA modely). Tuto metodu jsem zvolil, protože lze pomocí ní dosáhnout nejlepších výsledků, přestože je nejnáročnější na provedení a rozhodování.

Ani jedna časová řada neobsahovala potřebný počet cyklů, aby bylo možné v SW STATISTICA tuto analýzu provést. Z toho důvodu jsem musel časové řady uměle prodloužit na požadovaný počet cyklů, abych mohl analýzu v tomto SW provést. Prodloužení řad jsem provedl kopírováním jednotlivých cyklů, pokud to bylo možné. U vzorku senzor – BI2 + sítotisk se vyskytoval významný pokles v amplitudě jednotlivých cyklů. Z toho důvodu jsem nemohl řadu prodloužit kopírováním zadaných cyklů, ale musel jsem jejich průběh sám odhadnout. Tím sice došlo k jisté míře nepřesnosti této analýzy, ale bylo to nejlepší řešení, na které jsem přišel.

Pro všechny tři senzory jsem identifikoval a ověřil model, který nejlépe popisuje danou časovou řadu. Po identifikaci modelu jsem následně provedl předpovědi jednotlivých časových řad. Analýzu časových řad všech tří senzorů jsem provedl s využitím obou SW. Nakonec jsem také porovnal výsledky získané pomocí SW STATISTICA a NCSS.

Seznam použité literatury

- [1] LITSCHMANNOVÁ, . 2010. Úvod do analýzy časových řad. [Http://homel.vsb.cz/~lit40/SMAD/Casove_rady.pdf](http://homel.vsb.cz/~lit40/SMAD/Casove_rady.pdf) [online]. [cit. 2015-05-01]. Dostupné z: http://homel.vsb.cz/~lit40/SMAD/Casove_rady.pdf
- [2] STATSOFT. ČASOVÉ ŘADY S PROGRAMEM STATISTICA.
- [3] How To Identify Patterns in Time Series Data: Time Series Analysis. StatSoft [online]. [cit. 2015-05-01]. Dostupné z: <http://www.statsoft.com/textbook/time-series-analysis>
- [4] ALEŠ, KRIŠTOF. 2006. Nové metody a přístupy k analýze a prognóze ekonomických časových řad [online]. Praha [cit. 2015-05-10]. Dostupné z: www.pef.czu.cz/cs/?dl=1&f=12920. disertační práce. ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE.
- [5] MONTGOMERY, Douglas C, Cheryl L JENNINGS a Murat KULAHCI. 2008. Introduction to time series analysis and forecasting. Hoboken, N.J.: Wiley-Interscience, xi, 445 p. ISBN 04-716-5397-7.
- [6] Transformations of Variables - Smoothing Tab. STATISTICA Help [online]. [cit. 2015-05-01]. Dostupné z: <http://documentation.statsoft.com/STATISTICAHelp.aspx?path=TimeSeries/TimeSeries/Dialogs/TransformationofVariables/TransformationofVariablesSmoothingTab>
- [7] HAMILTON, James D. 1994. Time series analysis. Princeton: Princeton University Press, xiv, 799 s. ISBN 06-910-4289-6.
- [8] Spectrum Analysis Basic Notation and Principles - Data Windows and Spectral Density Estimates. STATISTICA Help [online]. [cit. 2015-05-01]. Dostupné z: <http://documentation.statsoft.com/STATISTICAHelp.aspx?path=TimeSeries/TimeSeries/Overview/SpectrumAnalysisBasicNotationandPrinciples/SpectrumAnalysisBasicNotationandPrinciplesDataWindowsandSpectralDensityEstimates>

- [9] VLADIMÍRA, Petrášková. 2006. Prognostické modely v oblasti modelování finančních časových řad [online]. Praha [cit. 2015-05-01]. Dostupné z: www.pef.czu.cz/cs/?dl=1&f=12935. disertační práce. ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE.
- [10] STATISTICA Help. StatSoft [online]. [cit. 2015-05-01]. Dostupné z: <http://documentation.statsoft.com/>
- [11] NCSS Documentation. NCSS Statistical Software [online]. [cit. 2015-05-01]. Dostupné z: <http://www.ncss.com/software/ncss/ncss-documentation/>