

Tartu Ülikool  
Matemaatika-informaatikateaduskond  
Matemaatika ja statistika instituut  
Matemaatika eriala

Märt Riiner  
**Singulaarse spektraalanalüüsi meetod**  
Bakalaureusetöö (6 EAP)

Juhendaja: Peep Miidla

Tartu 2017

# Resümee/Abstract

## Singulaarse spektraalanalüüsi meetod

Bakalaureusetöö  
Märt Riiner

**Lühikokkuvõte.** Käesolevas bakalaureusetöös antakse ülevaade singulaarse spektraalanalüüsi meetodist ja analüüsitakse aegrida, kasutades singulaarse spektraalanalüüsi meetodit (edaspidi SSA meetod). Selles bakalaureusetöös SSA meetodi numbriliseks rakendamiseks kasutatakse Tartu Ülikooli füüsika instituudi poolt kogutud ilmavaatluse andmeid ja neid analüüsitakse programmidega R ja Matlab.

**CERCS teaduseriala:** P140 Read, Fourier analüüs, funktsionaalanalüüs

**Märksõnad:** Singulaarne spektraalanalüüs, aegrida, R, Matlab.

## Singular spectrum analysis method

Bachelors thesis  
Märt Riiner

**Abstract.** The thesis gives overview of singular spectrum analysis method overview of timeseries analysis (named also SSA method). In this thesis for numerical implementation of SSA method will be used University of Tartu Institute of Physics collected data of weather observation and it will be analyzed by programs R and Matlab.

**CERCS research specialisation:** P140 Series, Fourier analysis, functional analysis.

**Keywords:** Singular spectrum analysis, time series, R, Matlab.

# Sisukord

<b>Resümee/Abstract</b>	<b>2</b>
<b>Sissejuhatus</b>	<b>5</b>
<b>1 SSA meetodi ülevaade</b>	<b>6</b>
1.1 Ülesande püstitus . . . . .	6
1.2 Trajektoormatriksi arvutamine . . . . .	7
1.3 Trajektoormatriksi lahtus singulaarsete väärtuste järgi . . . . .	8
1.4 Rekonstruktsioon . . . . .	9
1.4.1 Gruppidesse jaotamine . . . . .	9
1.4.2 Diagonaalne keskmistamine . . . . .	10
<b>2 SSA parameetrite mõjudest</b>	<b>11</b>
2.1 Akna suuruse mõju . . . . .	11
2.2 Gruppidesse jaotuste mõju . . . . .	12
2.3 Eritingimused: Eraldatuse mõju . . . . .	12
<b>3 Numbriline näide</b>	<b>14</b>
3.1 Samm 1: trendide leidmine . . . . .	15
3.1.1 Akna suuruse määramine . . . . .	16
3.1.2 Logaritmi meetod . . . . .	18
3.1.3 Omaväärtuste graafik . . . . .	19
3.2 Samm 2: harmoonilise komponendi eraldamine . . . . .	24
3.2.1 Akna suuruse otsustamine . . . . .	25
3.2.2 Harmoonilise komponendi eraldamine . . . . .	26
3.2.3 Jääkliikme hinnang . . . . .	31
3.3 SSA meetodi rakendus MATLAB-is . . . . .	33
<b>Summary</b>	<b>39</b>
<b>Kirjandus</b>	<b>40</b>

A	Matlabi kood	41
B	R-i kood	44

# Sissejuhatus

Meie igapäeva elus mängivad suurt rolli ajas toimuvad ja teatud juhuslikkuse komponenti sisaldavad sündmused, mille efekt on sageli väljendatav erinevatele ajahetkedele vastavate numbrite jadadena ehk aegridadena. Näitena võib tuua temperatuurid, kuised sademed, toodangu mahud ajavahemikus jms. Sellistes valdkondades juhuslikkus toob kaasa riske, mille haldamiseks on väga tähtis osata juhuslikkuse iseloomu kindlaks teha, mineviku andmete põhjal võimalikult täpseid prognoose leida ning mõningatel juhtudel ka vajadusel sekkuda. Kõike seda võimaldavad aegridade numbriliste töötlemise algoritmid. Ajalooliselt on praktikute poolt kasutusele võetud mitmeid meetodeid aegridadega seotud ülesannete (nt. trendi leidmine, tulevikuväärtuste prognoosimine) lahendamiseks. Meetodi all mõistame arvutuseeskirja, mille rakendamine peaks andma soovitud tulemuse. Meetodid tuginevad sageli tervel mõistusel ja intuitsioonil ning neid võib rakendada suvalisele ajas järjestatud andmete kogumile, kuid lahtiseks jääb küsimus tulemuste tegelikkusele vastavuse ja usaldusväärsuse osas.

Käesolevas bakalaureusetöös keskendume singulaarsele spektraalanalüüsi (ingl. *singular spectrum analysis*) meetodile. Edaspidi nimetame seda SSA meetodiks. SSA meetodit on arendatud alates 1970-ndatest kui uue aegrea analüüsi meetodina ja see leiab järjest rohkem kasutust. Töös anname ülevaate SSA meetodist ja näitame ka numbriliselt, kuidas meetodit rakendada. Andmed on saadud Tartu Ülikooli Füüsika instituudist, mida on kogutud ajavahemikus 1995-2011.

Bakalaureusetöö koosneb 4 peatükist. Esimeses kuni kolmandas peatükis anname ülevaate SSA meetodist, 4 peatükis aga analüüsime ühte aegrida. Selleks kasutame programme Matlab ja R, mille koodid ja paketid on lisatud ka lisade alla (vt. [8],[9],[10],[11]). Väga hea lühiülevaate annab SSA meetodist koos näite ja selgitustega [1].

Töö aluseks on võetud SSA meetodite põhjal koostatud artiklid ja raamatud, mida on loetud kasutatud kirjanduses. Aegridadega seotud mõisteid ja definitsioone on kasutatud tööst [7].

# PEATÜKK 1

## SSA meetodi ülevaade

### 1.1 Ülesande püstitus

Selles peatükis sõnastame ülesande ja anname ülevaate SSA meetodist sammude kaupa. Esmalt aga defineerime aegrea mõiste ja kirjeldame sündmuste esinemise tõenäosusi aegreas.

Aegreaks(kronoloogiliseks reaks) nimetatakse arvandmete rida, mis kirjeldab suuruse ajalist muutumist. Aegrida saadakse korduva vaatluse kasutamisel. Harilikult esitatakse aegrida ajamomentide (kuupäev, kellaaeg) või ajaperioodide (kuu, kvartal, aasta) ja neile vastavate suuruse väärtuste kogumina.(vt [2])

Olgu meil antud aegrida  $X = x_1, \dots, x_N$ , kus  $N > 2$ , mis koosneb juhuslikest reaalarvulistest väärtustest. See tähendab, et kõigi väärtuste esinemine on võrdne tõenäosusega  $\frac{1}{N}$ . Eeldame, et  $X$  ei ole nullrida ehk eksisteerib vähemalt üks  $i > 0$  nii, et  $x_i \neq 0$ . Siin võib eeldada, et  $X$  tähendab mingit kindlat ajavahemikku väärtust ja  $i$  on mingi kindel periood, kuid ajavahemikku tegelikult ei ole oluline fikseerida meie meetodi jaoks. Arvu  $N$  saame siin mõista kui ka diskreetset ajahetke, aga samamoodi võib see tähistada ka ükskõik millise lineaarse järjestusstruktuuri(mingi elementide järjestuse) märgendit.

Meie ülesanne on eraldada aegreast  $X$  trend ja sessoonne komponent ning müra. Trendina mõistame antud töös sellist funktsiooni, mille väärtused muutuvad aeglaselt, ühtlaselt ja süstemaatiliselt väga pika aja jooksul(pikemat kirjeldust võib leida [?] tööst(sektsioon 45.12). Siin sõltub trendi funktsioon aegrea pikkusest ehk aegrea pikkus on trendi funktsiooni etteantud ajateljeks. Sessoonne komponent on perioodiline funktsioon ja müra on sisult ülejääv osa, mis pole ei trendis ega perioodilises funktsioonis.

Trendi ja muude komponentide eraldamiseks aegreast kasutame SSA meetodit. SSA on mitteparametiline meetod, millega tuvastatakse võtmeväärtusi aegreas. Analüüsi tulemused aitavad meil mõista aegrea käitumist ja ennustada tuleviku väärtusi. Antud töös keskendume aegreast trendi ja sessoonse komponendi ning müra eraldamisele. SSA-ga on võimalik ka leida tuleviku väärtusi aegreas, aga

antud juhul töös me sellele ei keskendu. SSA meetodi võrdlusest teiste tuntud meetoditega ja tuleviku väärtuste leidmisest võib leida [3],[4].

SSA meetodis on kaks põhietappi: komponentideks lahtivõtmine e. dekompositsioon ja uuesti kokkupanek e. rekonstruktsioon, mis omakorda jaotuvad eraldi etappideks. Etappide kirjeldused on järgnevad.

- Dekompositsiooni etapid:

1. Trajektoormatriksi arvutamine. Siin moodustatakse aegreast trajektoormatriksi. Selleks esmalt jaotatakse aegrida etteantud koordinaatide arvuga vektoriks ja seejärel moodustatakse neist vektoritest matriks, kus veergudes on moodustatud vektorite koordinaadid.
2. Singulaarsete väärtuste lahtus(ingl. k. *singular value decomposition*, edaspidi ka SVD). Siin esmalt konstrueerime trajektoormatriksist ruutmatriksi ja seejärel leiame saadud ruutmatriksi omaväärtused ja omavektorid. Seejärel esitame esialgse trajektoormatriksi elementaaratriksi summana.

- Rekonstruktsiooni etapid:

1. Gruppidesse jaotamine. Siin jaotatakse elementaaratriksid, mida saadi SVD etapis, mitmeks erinevaks rühmaks ja liidetakse need elementaaratriksid igas rühmas eraldi kokku. Selle etapi tulemusena saadakse igas rühmas üks matriks, mille kokkuliitmisel saadakse trajektoormatriks.
2. Diagonaalide keskmistamine. Siin etapis moodustatakse gruppidesse jaotamise etapis saadud matriksitest aegread, mis on sama pikad kui esialgne rida. Selleks kasutatakse lineaarset kujutust, mille tulemusena saadakse algse rea trajektoormatriksist uus lähendus algsele aegreale.

## 1.2 Trajektoormatriksi arvutamine

Meie eesmärgiks on koostada algsest aegreast uus matriks, mille veergudes on vektorid, mida edaspidi nimetame ka trajektoormatriksiks.

Olgu  $L, N$  naturaalarvud. Arvu  $L$  nimetame edaspidi ka akna suuruseks nii, et  $1 < L < N$ . Me jaotame aegrea  $X$ -i  $L$  koordinaatidega vektoriks  $X_i = \{x_i, \dots, x_{i+L-1}\}$  iga  $i = 1, \dots, K$  korral, kus  $K = N - L + 1$ . Koostame trajektoormatriksi  $T$ , kus igas veerus on vektorite  $X_i, i = 1, \dots, K$  koordinaadid. Matriks  $T$  mõõtmeks on  $L * K$  ja tema diagonaalide elementide kohta kehtib seos  $x_{j+i-1} = x_{i+j-1}$ . Sellist

kuju omavat maatriksit nimetatakse ka Hankeli maatriksiks.

Trajektoormatriks  $T$  on kujul

$$T = [X_1, X_2, \dots, X_K] = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \vdots & x_N \end{pmatrix} \quad (1.1)$$

Maatriksis  $T$  on esindatud kõik aegrea  $X$  elemendid.

### 1.3 Trajektoormatriksi lahutus singulaarsete väärtuste järgi

Trajektoormatriksi lahutuseks singulaarsete väärtuste järgi (edaspidi ka SVD, ingl. *singular value decomposition*) on teine etapp SSA-s. SVD-d saab nimetada erinevate nimedega ja seda saab kasutada erinevateks otstarveteks. Täpsemalt on saab selle kohta uurida [5]-st.

Rakendame trajektoormatriksile  $T$  singulaarsete väärtuste lahutust ja saame trajektoormatriksid  $T_i$ , iga  $i = 1, \dots, L$  korral, mis omakorda peegeldab meile trende, sestooneid komponente ja müra vastavalt nende singulaarsetele väärtustele. Singulaarsete väärtuste all mõistame omaväärtuste ruutjuurt. Meenutame, et skalaarit  $\lambda$  nimetatakse ruutmaatriksi  $A$  omaväärtuseks, kui kehtib võrdus  $A\vec{x} = \lambda\vec{x}$ , kus vektor  $\vec{x} \neq \vec{0}$  on üheveerulise maatriksi kujul. Vektori  $\vec{x}$  kohta öeldakse, et see on omaväärtusele  $\lambda$  vastav omavektor.

Meetodi rakendamiseks tuleb esmalt arvutada maatriks  $S = TT^T$  ja tulemuseks saame sümmeetrilise ruutmaatriksi mõõtmetega  $L * L$ .

Singulaarsete väärtuste leidmiseks on vajalik esmalt arvutada maatriksi  $S$  omaväärtused ja omavektorid. Tähistagu  $\lambda_1, \lambda_2, \dots, \lambda_L$  maatriksi  $S$  omaväärtusi kahe-nevas järjekorras niiviisi  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$ . Olgu  $\{U_1, U_2, \dots, U_L\}$  maatriksi  $S$  ortonormeeritud omavektorite süsteem (st.  $(U_i, U_j) = 0$ , iga  $i, j = 1, \dots, L$  korral), mis vastavad  $\lambda_1, \lambda_2, \dots, \lambda_L$  omaväärtustele. Maatriksit  $S$  omaväärtusi on võimalik nii järjestada, sest nad on üheselt määratud ja tulenevalt maatriksi  $S$  konstruktsioonist.

Olgu  $V_i = T^T U_i / \sqrt{\lambda_i}$  ( $i = 1, 2, \dots, d$ ), kus  $d = \max(i : \lambda_i > 0)$  on maatriksi  $T$  astak. Tavaliselt  $d$  väärtus on minimaalne väärtus  $L$ -st ja  $K$ -st. Me saame trajektoormatriksi lahti kirjutada kujul



$$T = T_1 + T_2 + \dots + T_d = \sum_{i \in d} \sqrt{\lambda_i} U_i V_i^T, \quad (1.2)$$

kus  $U_i$  ja  $V_i$  on  $L * L$  ja  $K * K$  matriksid. (1.2) tõestuse leiab [5] peatükk 4-st. Kolmikut  $(U_i, \sqrt{\lambda_i}, V_i^T)$  nimetatakse (1.2) SVD  $i$ -ndaks omafunktsiooni kolmikuks (ingl. k. *ith eigentriple of the SVD*).

## 1.4 Rekonstruktsioon

### 1.4.1 Gruppidesse jaotamine

Siin jaotame SVD etapis lahtivõetud matriksi  $T_i$  laiali. Selle eesmärgiks on eraldada meid huvitavad komponendid. Täpsemalt arutame gruppidesse jaotamise mõjust (2.2).

Põhimõtteliselt jaotame indeksid  $\{1, 2, \dots, d\}$ , mis tähistavad  $T = T_1 + T_2 + \dots + T_d$  matriksite  $T$ - indekseid, ühisosata alamgruppidesse  $I = \{I_1, \dots, I_m\}$ . Seega  $T_I$ -le vastab hulk  $I = \{I_1, \dots, I_m\}$ . Matriks  $T_I$  on summa  $T_j$ -st, kus  $j \in I_i$ . Trajektoormatriksi  $T$  saab lahti kirjutada kujul

$$\begin{aligned} T &= \overbrace{T_1 + \dots + T_L}^{\text{SVD}} \\ &= \overbrace{T_{I_1} + \dots + T_{I_m}}^{\text{Gruppideks jaotamine}} \end{aligned} \quad (1.3)$$

Selleks, et mõista paremini gruppidesse jagamist, selgitame protsessi veidi detailsemalt. Oletame, et meil on kahe grupi trajektoormatriksi  $T$  omafunktsiooni kolmikuid. Olgu need  $T_L$  ja  $T_R$ . Olgu hulk  $I = \{1, \dots, d\}$ ,  $R \cup L = I$ , kus  $R$  ei ole  $L$  alamhulk.  $T_I$  on siis

$$T_I = \sum_{i \in I} \sqrt{\lambda_i} U_i V_i^T \quad (1.4)$$

ja saame arvutada  $T_L = T_I - T_R$ , eeldusel, et nende seos omavahel on nõrk ehk nad on eraldatavad. See tähendab sisuliselt seda, et kaks rida  $X^{(1)}$  ja  $X^{(2)}$  on eraldatavad trajektoormatriksi lahutuses (1.2) parajasti siis, kui eksisteerib selline indeksite hulk  $I \subset \{1, \dots, d\}$  nii, et  $X^{(1)} = \sum_{i \in I} X_i$  ja  $X^{(2)} = \sum_{i \notin I} X_i$ .

Seega saame  $T_L$  kirjutada kujul

$$T_L = \sum_{i \in L} \sqrt{\lambda_i} U_i V_i^T \quad (1.5)$$

Eraldatus on vajalik selleks, et paremini saaks teha grupeerimist ja diagonaalsest keskmistamist. Eraldatuse mõjudest räägime pikemalt alapunktis (2.3).

## 1.4.2 Diagonaalne keskmistamine

SSA algoritmis mõeldakse diagonaalne keskmistamise all sammu, kus me moodustame grupeeritud maatriksitest  $T_{I_i}$  uue rea pikkusega  $N$ .

Olgu  $Y$   $L * K$  mõõtmetega maatriks, mille elementideks on  $y_{ij}$ ,  $1 \leq i \leq L$ ,  $1 \leq j \leq K$ . Fikseerime  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$  ja  $N = L + K - 1$ . Olgu  $y_{ij}^* = y_{ij}$ , kui  $L < K$ , muudel juhtudel  $y_{ij}^* = y_{ji}$ . Diagonaalne keskmistamine muundab maatriksi  $Y$  reaks  $g_0, \dots, g_{N-1}$  järgneva skeemi abil:

$$g_k = \left\{ \begin{array}{ll} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^* & \text{iga } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^* & \text{iga } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2}^* & \text{iga } K^* \leq k < N \end{array} \right\} \quad (1.6)$$

Skeem (1.6) teostab keskmistamist üle maatriksi elementide üle "diagonaalide"  $i + j = k + 2$ , kui  $k = 0$ , siis  $g_0 = y_{11}$ , iga  $k = 1$  korral saame  $g_1 = (y_{12} + y_{21})/2$  jne. Paneme tähele, et kui maatriks  $Y$  on trajektoormaatrisi mõnele reale  $h_0, \dots, h_{N-1}$ , teiste sõnadega, kui  $Y$  on Hankeli maatriks, siis  $g_i = h_i$  iga  $i$  korral. Diagonaalne keskmistamine (1.6) rakendatuna  $T_{i_k}$ -le annab meile rea  $\tilde{X}^{(i)} = \{\tilde{X}^{(i)} + \tilde{X}_0^{(k)} + \dots + \tilde{X}_N^{(k)}\}$  ja seega aegrida  $X = \{x_1, \dots, x_N\}$  dekomponeeritakse  $m$ -rea summaks eeldusel, et kehtib nõrk seos.

# PEATÜKK 2

## SSA parameetrite mõjudest

### 2.1 Akna suuruse mõju

SSA meetodi alguses määrasime akna  $L$  suuruse, mis määrab ka trajektoormaat-  
riks  $T$  mõõtme. Vaatame meie algset aegrida  $X = \{x_1, \dots, x_N\}$ . Arv  $L$  jaotab  
aegrea  $X$   $L$  koordinaadiga vektoriks  $X_i = \{x_i, \dots, x_{i+L-1}\}$  iga  $i = 1, \dots, K$  korral  
ja kus  $K = N - L + 1$ . Akna suurus  $L$  rahuldab võrratust  $2 < L < N/2$ . Kuna  
akna suurus  $L$  rahuldab võrdust  $K = N - L + 1$  trajektoormaatrisis  $T$ , siis meile  
pole vaja suuremat  $L$  kui  $N/2$ .

Oletame, et meil on mingi aegrida kolme aasta kohta. Kui akna suurus  $L$  läheneb  
 $N/2$ -le, siis algse rea jaotamine muutub detailsemaks. Maatriksi  $T$  maksimaalne  
astak saab olla  $N/2$ . Me peame olema samuti ettevaatlik akna suuruse valikuga,  
sest kui nt aegrida on perioodiline, siis akna  $L$  suurus ei saa olla akna  $L$  maksimum  
ega ka miinimum. Kui  $L$  on väga lähedal väärtusele  $N/2$ , siis kipub trajektoori  
maatriks  $T$  ületama oma ridu ja veerge. Kui aga  $L$  on väga väike, siis võib juhtuda  
nii, et lahti võtmise käigus kaovad ära perioodilised komponendid.

Allolev tabel (2.1) näitab, mis juhtub suurte või väiksete  $L$ -ide korral SSA meeto-  
dis.

Seega, mida täpsema akna suuruse  $L$  me valime, saab trendiks sile või regulaarne  
muster. Kui me vähendame akna väärtust  $L$ , trend muutub "karedamaks" ja näitab  
rohkem ebaregulaarusi, sest ta sisaldab ka teisi komponente.

Tabel 2.1: Akna suuruse mõju

Akna suurus	$T$ aste	$\sqrt{\lambda_1}$	Peamine trend	
$L \uparrow$	$\uparrow$	$\downarrow$	sile	tavaline
$L \downarrow$	$\downarrow$	$\uparrow$	ebatasane	ebatavaline

Kui me teame aegrea perioodi  $T$ , siis soovitav on võtta selline akna suurus  $L$ ,

et  $T/L$  oleks täisarv. Näiteks kui meil on aegrea andmed sessoonsed ja perioodiks on 5, siis on mõistlik võtta  $L$ -ks 5-kordsed 5, 10, 15, jne. Akna  $L$  väärtus ei tohiks olla ka liiga suur, sest see võib lüüa segamini perioodid teiste väärtustega.

## 2.2 Gruppidesse jaotuste mõju

Kokku panemisel mängib olulist rolli grupeerimine, mis määrab ära trendid, sileduse, poolaasta, kvartali, igakuise arverea ja müra, mis kajastub graafikutel. Kui lahtivõtmise etapid on läbitud, siis grupeerime valitud akna suurusega maatriksid  $T_{L_i}$ , mille abil saame võrrelda algset rida ja uut koostatud rida.

Olulist rolli mängib, kuidas leida grupeerimise sammus õiged alamgrupid. SVD sammust saame omafunktsiooni kolmikud  $L$ -tükki ja diagonaalsest keskmistamisest saame  $\tilde{X}^{(i)}$  iga  $i = 1, \dots, d$  diagonaalsest keskmistamisest saame  $\tilde{X}^{(i)}$ . Erinevate omafunktsiooni kolmikute graafikute võrdlemine aitab meil määrata trende või silumist või teiste komponentide sobivust.

Kahe komponendi vahelist vastastikust seost saab arvutada kasutades selleks nn  $w$ -korrelatsiooni (ingl.k. *w-correlation*). Selle all mõistame seost (2.3). Komponentide  $w$ -korrelatsiooni maatriks võib olla kasulik grupeerimise jaoks.  $W$ -korrelatsiooni väärtused, mis on lähedal 1-le, on omavahel kõrges korrelatsioonis ja peaksid seega ühte gruppi kuuluma, samal ajal väärtused, mis on lähedal 0-le viitavad, et need pole omavahel korrelatsioonis.

Veel üks võimalus grupeerimiseks on liita omavahel kokku komponendid, millel on ligilähedaselt sama periood. See on efektiivne eriti siis, kui kui lahtivõetud komponendid sisaldavad palju harmoonilisi (perioodilisi) komponente. Perioode on võimalik leida nii, et teha iga komponendi kohta periodogrammi ja leida sealt domineeriv periood.

## 2.3 Eritingimused: Eraldatuse mõju

Singulaarsete väärtuste lahtivõtmise sammus(SVD) ja diagonaalide keskmistamise sammus peame eeldama, et trendid ja teised komponendid on eraldatavad algsest aegreast. Seda eeldust kutsume nõrgaks seoseks. Sisult tähendab see järgnevat.

Olgu  $L^* = \min(L, K)$  ja  $K^* = \max(L, K)$  ning

$$w_i = \begin{cases} i & \text{iga } 0 \leq i < L^*, \\ L^* & \text{iga } L^* \leq i < K^*, \\ N - i + 1 & \text{iga } K^* \leq i < N \end{cases} \quad (2.1)$$

Defineerime kaks aegrida  $F^{(1)}$  ja  $F^{(2)}$  pikkusega  $N$  niiviisi:

$$(F^{(1)}, F^{(2)})_w \stackrel{\text{def}}{=} \sum_{i=1}^N w_i x_i^{(1)} x_i^{(2)} \quad (2.2)$$

ja nimetame read  $F^{(1)}$  ja  $F^{(2)}$   $w$ -ortogonaalseks (ingl. k. *w-orthogonal*), kui  $(F^{(1)}, F^{(2)})_w = 0$ . Järgmisena defineerime

$$\rho^w(F^{(1)}, F^{(2)}) \stackrel{\text{def}}{=} \frac{(F^{(1)}, F^{(2)})_w}{\|F^{(1)}\|_w \|F^{(2)}\|_w}, \quad (2.3)$$

kus  $\|F^{(i)}\| = \sqrt{(F^{(1)}, F^{(2)})}$ ,  $i = 1, 2$ . Me ütleme, et kaks rida  $F^{(1)}$  ja  $F^{(2)}$  on ligikaudu eraldatavad (ingl.k. *approximately separable*) kui

$$\rho^w(F^{(1)}, F^{(2)}) \simeq 0. \quad (2.4)$$

Definitsioonid ja kahe rea eraldatuse näidete ja tõestuste kohta leiab [5], peatükk 6.

Eraldatuse eelduse tõttu rakendame SSA meetodit aegreale ja eraldame trendi või teised komponendid algsest aegreast. Sellisel juhul kõik eraldised trendidest või komponentidest, mis on võetud algandmetest, tehakse eeldusel, et kehtib (2.4).

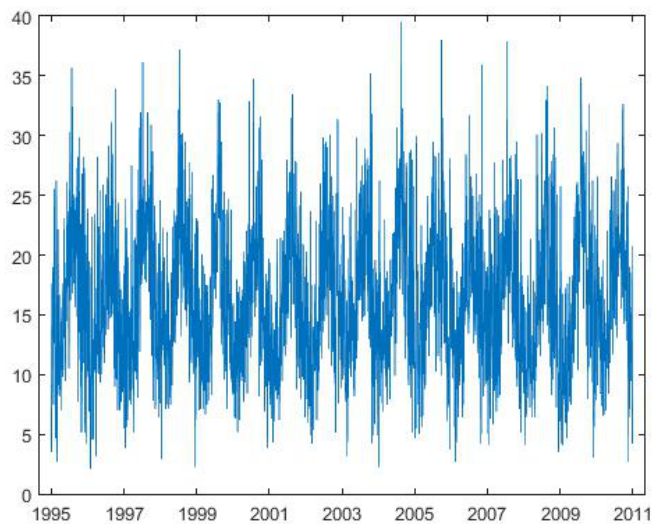
Peame ära märkima, et praktikas tegelikult on harva selline situatsioon, kus aegrea komponendid on täpselt eraldatavad ja seetõttu kasutatakse rohkem ligikaudset eraldatavust. Aegridadel, millel on väga kompleksed struktuurid, on ka keeruline teostada komponentide eraldamist. Selle jaoks modifitseeritakse SSA meetodit erinevate võtete abil nagu tsentreerimine ja Toeplitz. Neid meetodeid on kirjeldatud [5].

# PEATÜKK 3

## Numbriline näide

Selleks, et illustreerida SSA meetodit, kasutame SSA meetodit konkreetsel aegreal ja esitame graafikuid. Töös on analüüsi teostamise eeskujuks võetud [6] ja [12]. Meie andmed pärinevad Tartu Ülikooli füüsika instituudi uurimustööst, mille aluseks on atmosfääri veeauru sisaldus arvatuna GPS signaali hilistumise põhjal. Atmosfääri veeauru sisaldust kilogrammides ruutmeetri kohta (ingl.k. *GNSS IPW - Integrated Precipitable Water*) on kogutud igapäevaselt ja need andmed on jaotatud veergudeks kuupäeva, veeauru, standardhälvete kaupa. Meie analüüsime veeauru veergu, kasutades selleks Matlabi ja R programmi ning R-i pakette Rssa, gdata.

Meie algne aegrida  $X$ , milles on  $N = 4568$  väärtust, on kujul



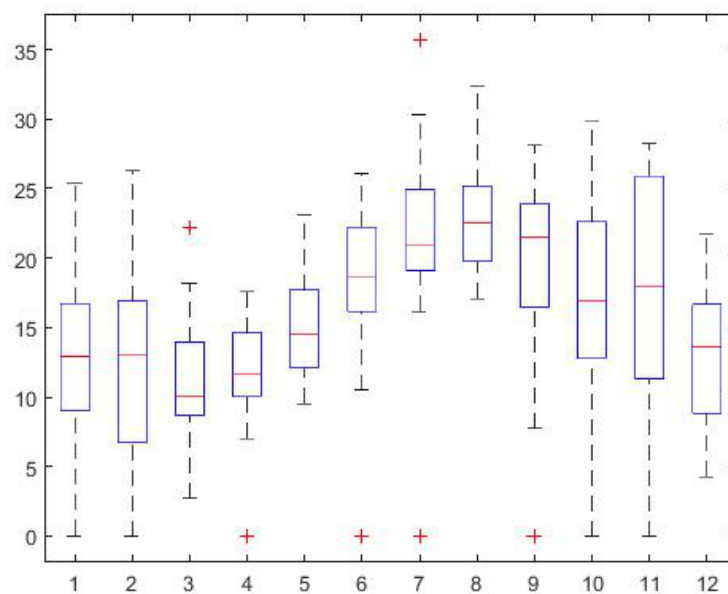
Joonis 3.1: Atmosfääri veeaurusisaldus

Joonisel (3.1) on vertikaalteljel aegrea ühik  $\text{kg}/\text{m}^2$  kohta ja horisontaalteljel kuupäevad.

Aegrea graafik näitab, et aasta alguses on atmosfääris veeauru sisaldus madalam, aasta keskel tundub olema atmosfääris veeauru sisaldus kõrgem. Mõnel aastal tundub olema väga kõrge atmosfääri veeauru sisaldus.

### 3.1 Samm 1: trendide leidmine

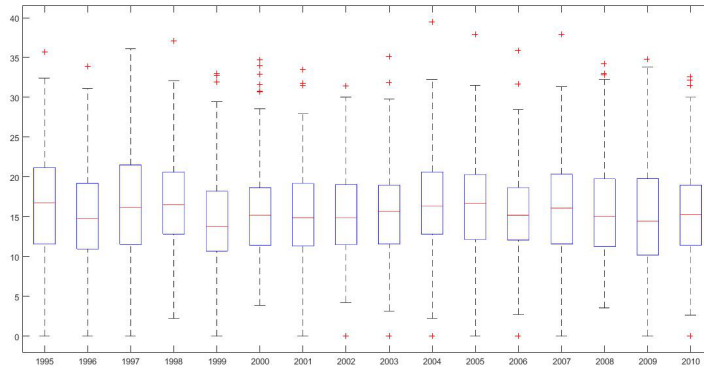
Uurime esmalt kuude kaupa, kas suudame mingeid trende leida.



Joonis 3.2: Atmosfääri veeaurusisalduse karpdiagramm 1995 aasta kohta

Kuuline karpdiagramm näitab meile, et kõrgem veeauru mediaan on suvekuudel ja kõige kõrgem on augustis. Samal ajal madalam veeauru mediaan on talvekuudel ja madalaim mediaan on märtsis. Seega on meil aastaaegadest sõltuv aegrida.

Aasta karpdiagrammi joonis (3.3) näitab meile, et üldiselt on veeaur suhteliselt sarnasel tasemel aastate jooksul, kuid siiski on ka mõned kuivemad ja mõned niiskemad aastad ehk eksisteerivad irregulaarsed tsüklid. Seega tegelikult on keeruline otsesid trende määrata, kuna veeauru sisaldus sõltub millised on aastaajad. Kõige täpsem määratlus oleks ehk aastaaegade kaupa trende otsida.



Joonis 3.3: Atmosfääri veeaurusisalduse karpdiagramm

### 3.1.1 Akna suuruse määramine

SSA meetodi esimene samm on arvutada trajektoori maatriks. Selleks peame fikseerima akna suuruse  $L$ . Akna fikseerimisel on oluline teada, et kui aegrida on lühike, siis on parem fikseerida akna suurus proportsioonides perioodiga. Näiteks, oletame et meil on aegreas aastased perioodid, siis sellisel juhul peame jälgima ka aastaste puhul trende ja kordusi, kui  $k/12$  ( $k = 1, \dots, 12$ ).

Kui meil on tegemist pika aegreaga, siis on mõistlik valida  $L$  nii suur kui võimalik, kuid väiksem kui  $T/2$ .

Sagedus on defineeritud kui tsüklite arv ajaühikus. Kuna iga üksik vaatlus tähendab üht ajaühikut, siis on sagedus arvatud kui vaatluste tsükkel. Sukses-siivseteks sagedusteks on väärtused  $K/N$  ( $K = 0, \dots, N/2$ ), kus  $N$  on andmete arv algreas. Seega näiteks sagedus  $0,0833$  tähendab, et 12 andmetega aegreast täidetakse üks tsükkel. Kui aegrida koosneb kuu andmetest, mis on kogutud üle paljude aastate, siis vastav sagedus näitab aastast tsükli.

Kuna veeauru sisaldus õhus on sõltuv kuudest ja aastaegadest, siis meie rea puhul on mõistlik eeldada, et periood  $T$  võiks olla 12 ja  $L$  oleks 12-kordsed.

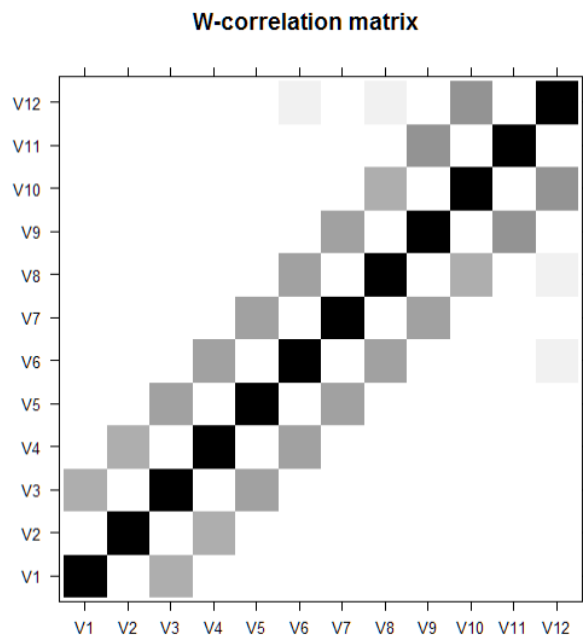
Meil on tegemist pika reaga, millel on  $N = 4568$  väärtust, siis  $L = 12 * M$ , kus  $M$  on positiivne täisarv.

$W$ -korrelatsiooni graafik aitab meil otsustada akna suuruse  $L$ -i üle.  $W$ -korrelatsiooni graafiku  $L = 12$  maatriks on nähtav joonisel (3.4).

$W$ -korrelatsioon maatriksi kujul annab meile aimu, kuidas teha õiget grupeerimist. Joonisel (3.4) iga ruudu tumedus näitab  $w$ -korrelatsiooni tugevust kahe komponendi vahel. Mida tumedam, seda tugevam on  $w$ -korrelatsioon.

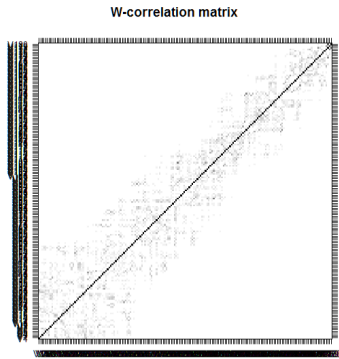
Joonis (3.4) näitab meile antud juhul, et esimene element on nõrgas  $w$ -korrelatsioonis kolmanda komponendiga. Teine element on nõrgas  $w$ -korrelatsioonis neljanda komponendiga. Kolmas element on nõrgas  $w$ -korrelatsioonis kolmanda ja viienda ele-



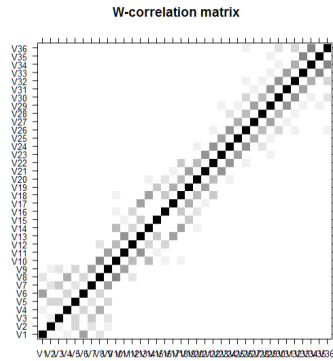


Joonis 3.4: W-korrelatsiooni maatriks  $L = 12$

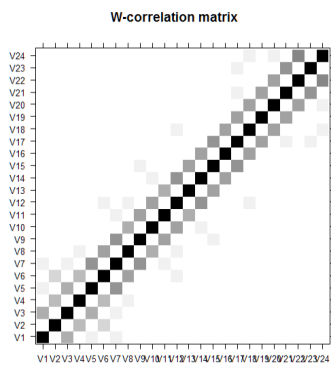
mendiga. Samas võib olla ka tegemist lihtsalt müraga. Ühe kuni kümne puhul võime eeldada, et need kirjeldavad trende.



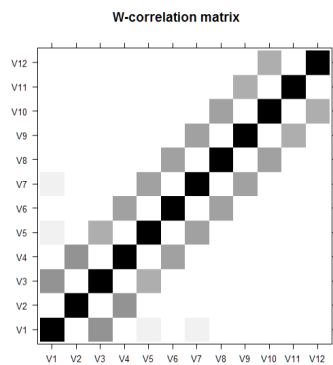
(a)  $L = 180$   $w$ -korrelatsiooni maatriks



(b)  $L = 36$   $w$ -korrelatsiooni maatriks



(c)  $L = 24$   $w$ -korrelatsiooni maatriks



(d)  $L = 12$   $w$ -korrelatsiooni maatriks

Joonis 3.5:  $W$ -korrelatsiooni maatriksid

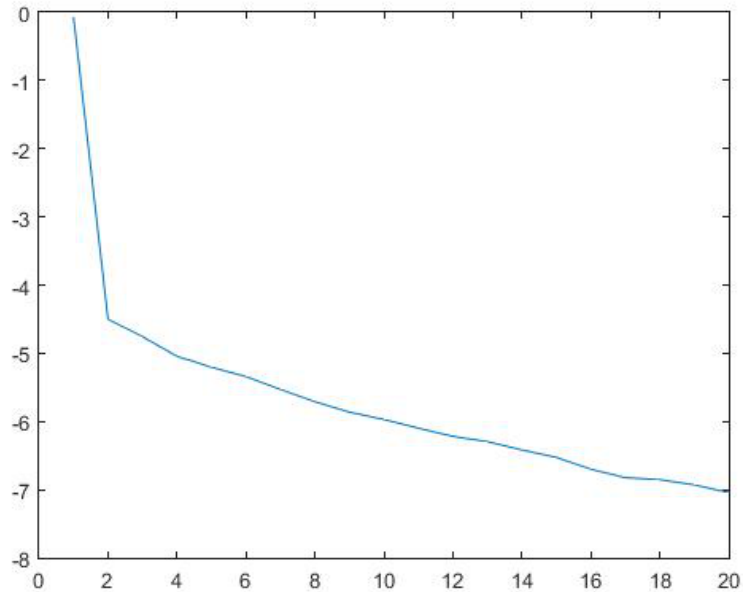
Need neli akna  $L$  suurust joonisel (3.5) näitavad meile, et müra on rohkem väiksemate akende juures ja müra väheneb 8-ndast komponendist alates.

### 3.1.2 Logaritmi meetod

Üks meetod müra eraldamiseks on otsida erisusi omaväärtuste hulgas. Müra tunnuseks on aeglaselt kahaneva singulaarsete väärtuste jada. Kahjuks ei ole mingit

kindlat eeskirja selle tuvastamiseks, vaid peabki uurima graafikutelt. Sisult otsitakse sellist kohta, kus logaritmilised väärtused hakkavad aeglaselt ja ühtlaselt kahanema ja ei esine nn hüppeid graafikus.

Arvutame 20 omaväärtust, mis on järjestatud vastavalt lahtivõtmise osadele. Järgnevas graafikus on näha omaväärtuste logaritmilisi väärtusi. Jooniselt (3.6)

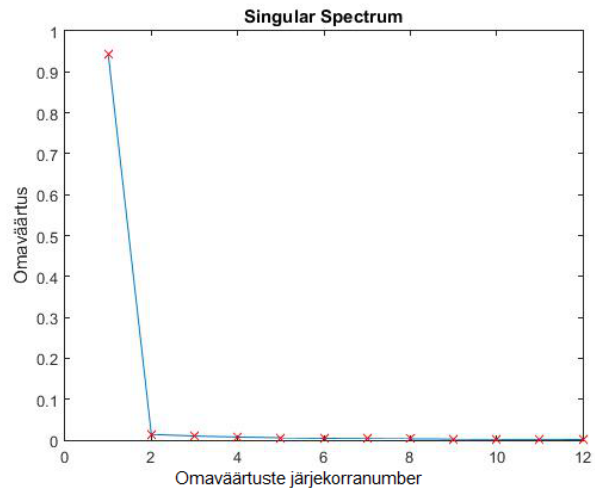


Joonis 3.6: Omaväärtuste logaritmid

näeme, et kiire langus toimub teise omaväärtuste juures ja saame tegelikult oletada, et siit alates algab müra suurenemine.

### 3.1.3 Omaväärtuste graafik

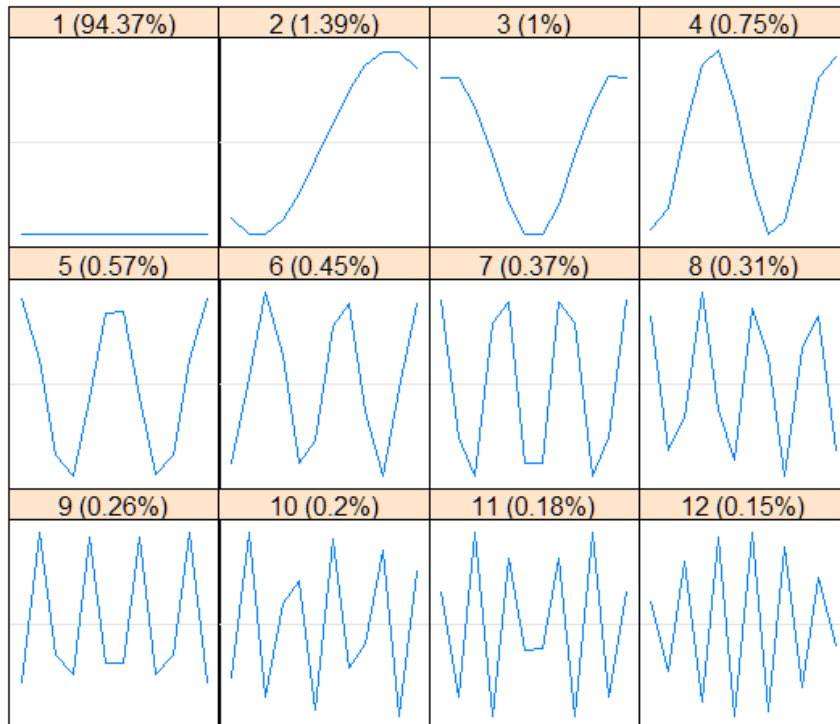
Omaväärtuste graafikult, millel on omaväärtuste korrelatsiooni maatriks kahanevalt, on oluline jälgida langeva graafiku murdepunkti. Olulised komponendid on meie jaoks enne murdepunkti.



Joonis 3.7: Omaväärtuste graafik

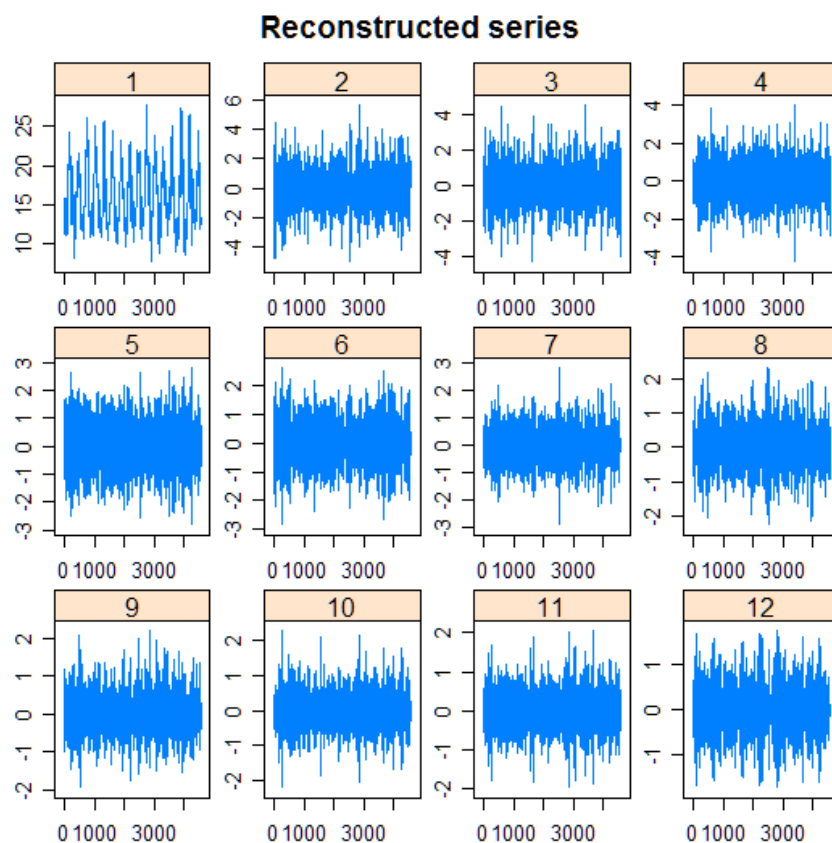
Jooniselt (3.7) on näha, et murdepunkt on number 2 omaväärtuse juures. See tähendab, et oluline komponent on esimene.

### Eigenvectors



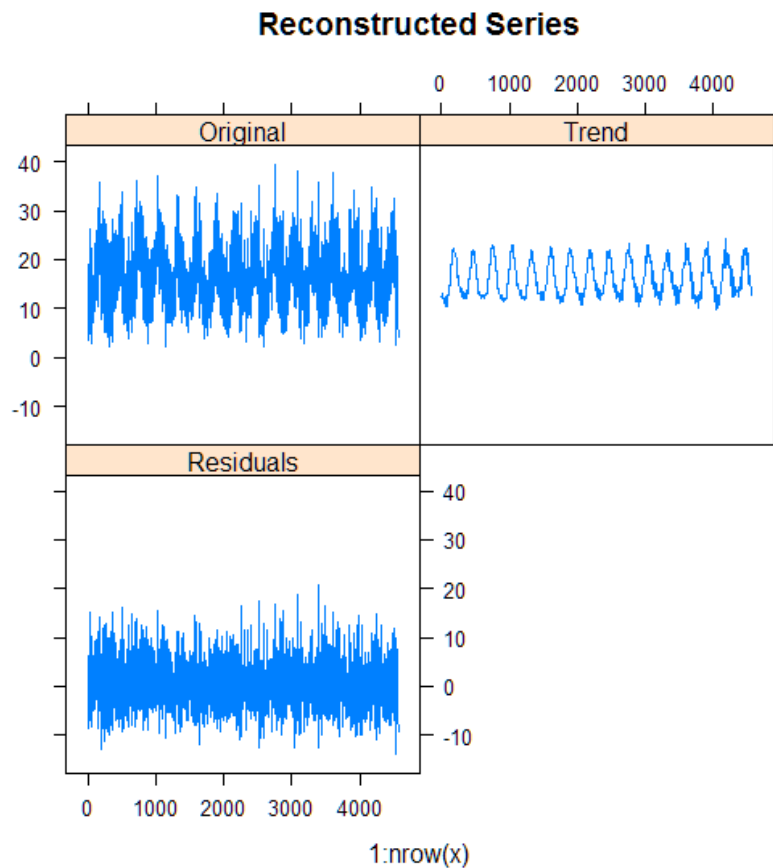
Joonis 3.8: Omavektorid,  $L = 12$

Jooniselt (3.8) on et esimesel omavektoril on peaaegu konstantsed koordinaadid. Sellist omavektori käitumist saab tõlgendada trendina. Jooniselt (3.8) 94.37% ütleb meile seda, et see omavektor kirjeldab 94.37% reast. Teised omavektorid kirjeldavad meile mu"ra.



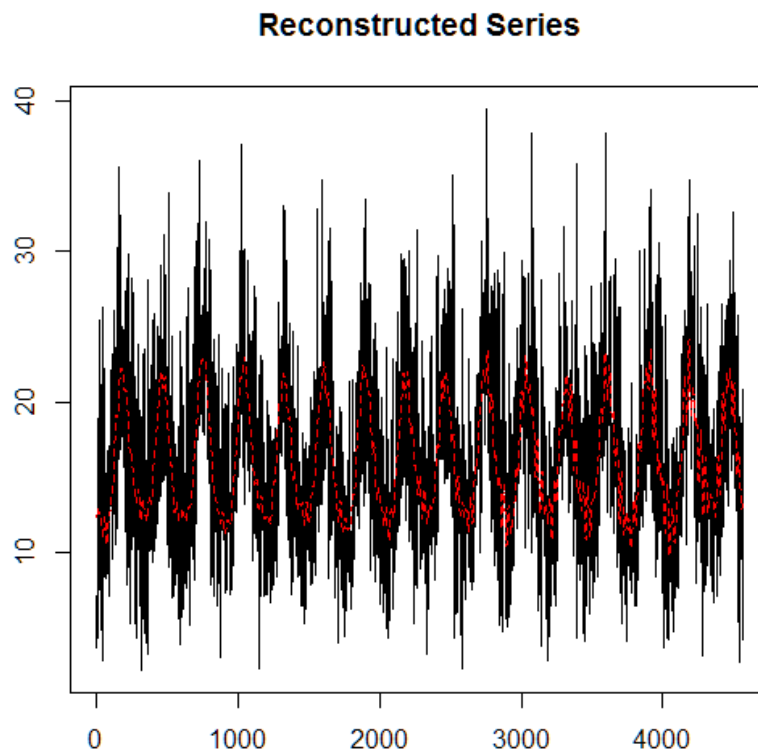
Joonis 3.9: Rekonstrueeritud rida,  $L = 12$

Joonisel (3.9) on kujutatud rekonstrueeritud rea graafikut. Esimene kast rekonstrueeritud rea joonisel kujutab trendi ja ülejäänud kastid näitavad kõrge sagedusega komponente, mis ei ole seotud trendiga. Küll aga võime leida sarnasust karpdiagrammi joonisega (3.3).



Joonis 3.10: Trend, jääkliige

Joonisel (3.10) on kujutatud rekonstrueeritud rea trendi, originaalrida ja jääkliiget. Joonis näitab, et trend on tegelikult suhteliselt regulaarne ja tsüklid on suhteliselt sarnased. Joonisel (3.11) kujutatud trend on nüüd R-i SSA meetodi tuvastatud. Punane on trend ja must on originaalrida.



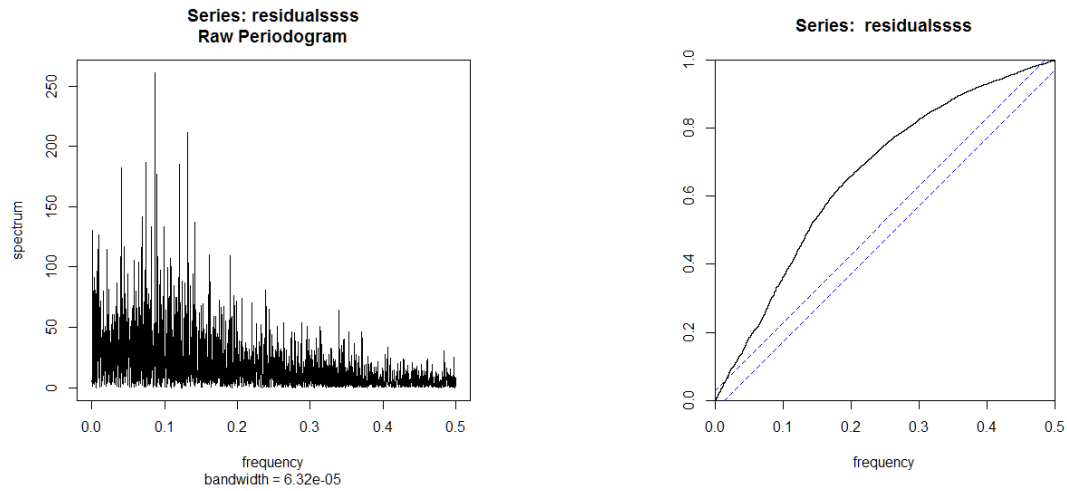
Joonis 3.11: Rekonstrueeritud rida

## 3.2 Samm 2: harmoonilise komponendi eraldamine

Jääkliiget arvutatakse nii, et lahutatakse trendi väärtused ära originaalreast. Teises sammus tuvastame harmoonilise komponendi jääkliikme aegreas.

Alge rea periodogramm oli niisugune





Joonis 3.12: Jääkrea periodogramm

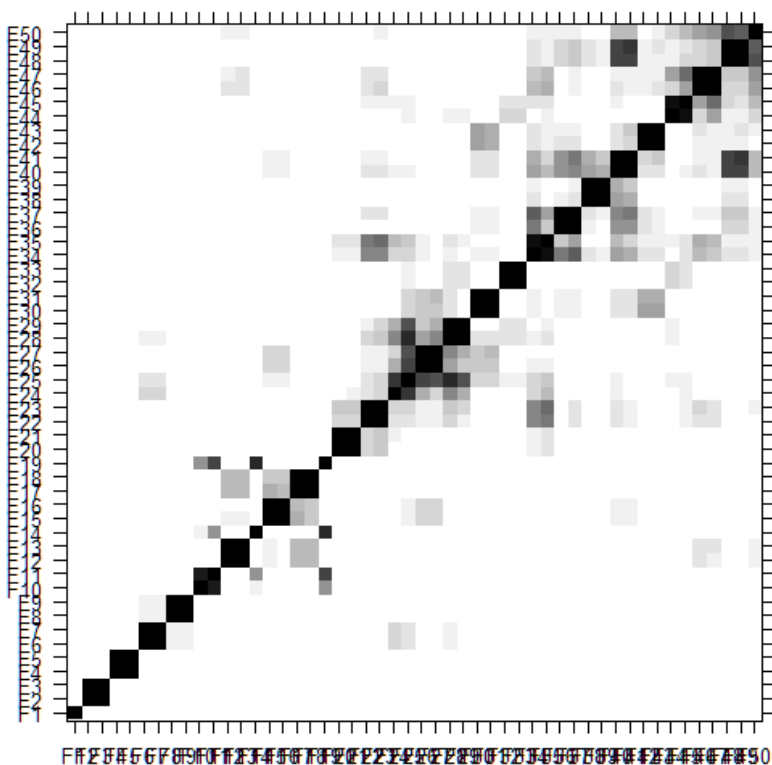
Jääkrea periodogrammist on maha lahutatud trend  $1/12$  on ühtlustanud väärtusi. Need väärtused, mis eemaldati, saame seostada, et nende tsüklid olid pikemad kui aasta.

Vasak graafik joonisel (3.12) ütleb meile, et meil on tipp sageduse 0.1 juures ja veidi väiksem 0.5 juures. Nende sagedused on võrdsed vastavalt  $0.01/12$  ja  $0.5/12$ . Parempoolne graafik joonisel (3.12) näitab meile seda, et kui jääkliige on lineaarselt tõusev graafik, siis on tegemist meil mu"raga. Joonisel on näha see punktiirjoone-na. Jääkliikme rea graafikut ennast kujutab aga kumer joon. X-telg kujutab endast esinemise sagedusi ja y-telg nende esinemise tõenäosusi.

### 3.2.1 Akna suuruse otsustamine

Akna suuruse määramine on esimene samm, mida tuleb otsustada enne SSA analüüsi teostamist. Paremaks eraldumiseks peab  $L$  olema võimalikult suur ehk siis maksimaalselt  $N/2$ . Seega peaks  $L = 2274$ .

**W-correlation matrix**

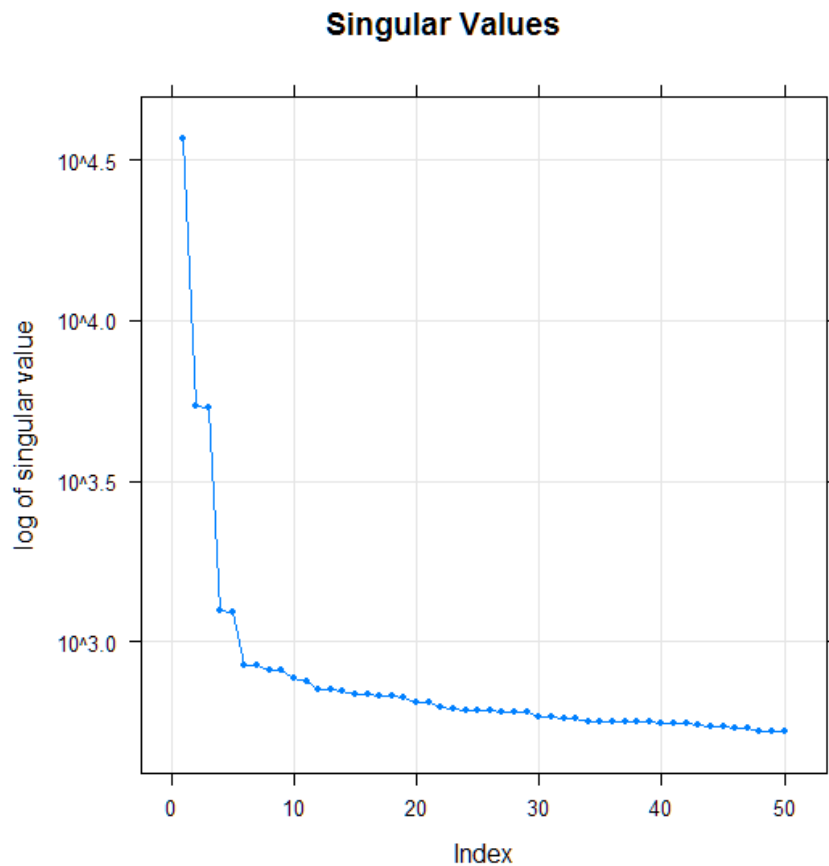


Joonis 3.13:  $W$ -korrelatsiooni maatriks,  $L = 2274$

Joonis (3.13) jaotab esimesed 50 omafunktsiooni kolmikud kahte peamisesse gruppi: esimesest kuni 14-ndani ja ülejäänud. 19. omafunktsiooni kolmik tundub ka olema eraldiseisev. Kuigi 15. ja 16. on omavahel tihedas korrelatsioonis, siis on nad ka paljude teistega tihedas korrelatsioonis. 15. ja 16. omafunktsiooni kolmikud vastavad mürale.

### 3.2.2 Harmoonilise komponendi eraldamine

Joonis (3.14) näitab meile õlga 6-nda omaväärtuse juures. Jooniselt (3.14) näeme veel, et meil osad väärtused on peaaegu harmoonilised: 2 ja 3, 4 ja 5, 6 ja 7, 8 ja 9, 10, 11, 12 ja 13.



Joonis 3.14: Singulaarväärtuste logaritmid

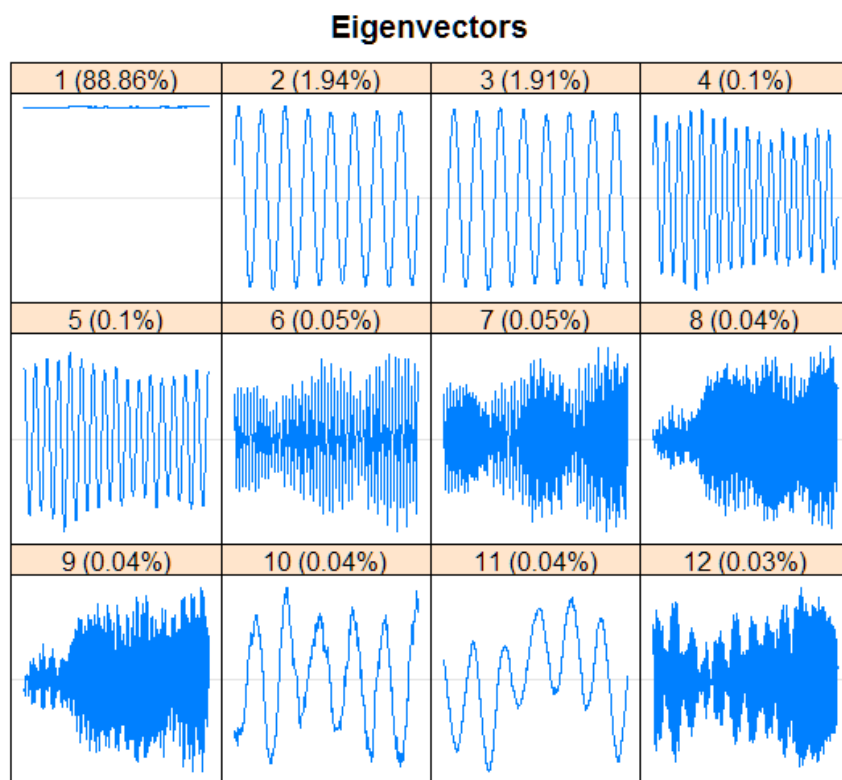
Joonis (3.15) näitab, et esimene komponent moodustab 88,86% jääkreast samm 1-s.

Praktikas selgub, et harmoonilise rea kahe omavektorsiooni kolmiku singulaarväärtused on sageli väga lähedased üksteisele. Omavektorite paaride graafik näitab meile visuaalselt ära need omavektorsiooni kolmikud, mis vastavad harmoonilistele komponentidele reas, eeldusel, et need komponendid on eraldatavad jääkliikmete hulgast.

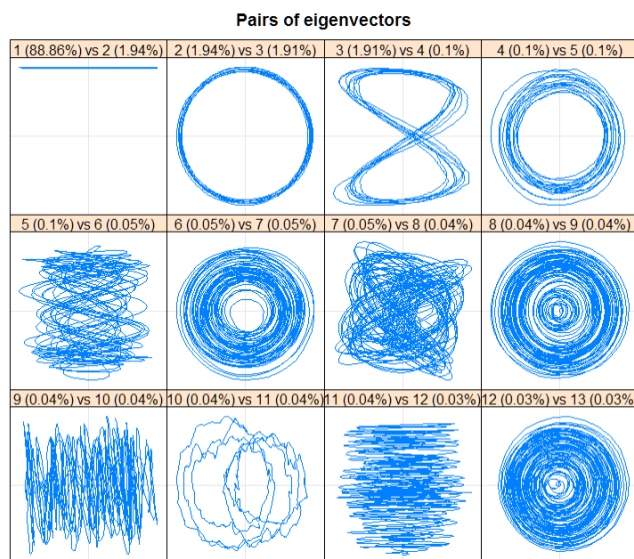
Oletame, et puhas harmooniline komponent sagedusega  $w$ , millel on kindel faas, amplituud ja ideaalses situatsioonis, kus  $P = 1/w$  on täisarv, mis jagab akna suurust  $L$  ja  $K$ . Kuna  $P$  on täisarv, on ta harmoonilise rea periood. Ideaalses situatsioonis on vasakul omavektoril ja põhikomponendil siinuse ja koosinuse jada kuju, millel on sama  $P$  ja sama faas. Seega, nende komponentide tuvastamiseks, mis genereeritakse harmoonilisusest, taandatakse sellele, et need paarid tuleb tuvastada.

Puhas siinus ja koosinus, millel on võrdsed sagedused, amplituudid ja faasid, joo-

nestuvad graafiku ringi kujulisena.



Joonis 3.15: Omavektorid sammus 2



Joonis 3.16: Omavektorite paarid samm 2

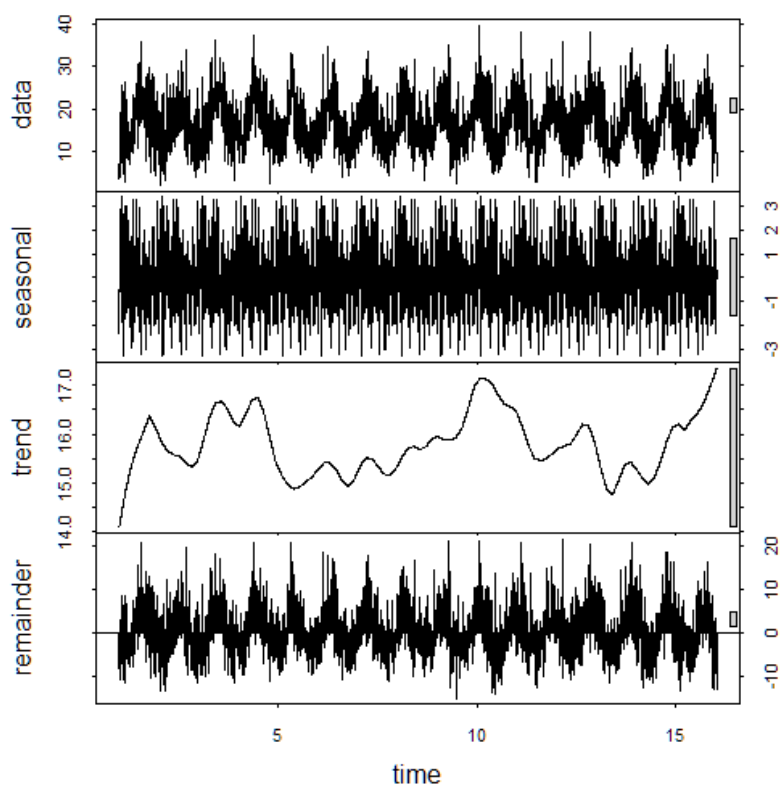
Joonisel (3.16) näeme, 2 ja 3, 4 ja 5, 6 ja 7, 8 ja 9, 12 ja 13 näitavad p-tipu polügoone, mis tähendab, et nad on siinuse/koosinuse jada paarid koos nullfaasiga ja sama amplituudiga. Teiste sõnadega, nad moodustavad harmoonilisi komponente erinevad perioodiga.

Tabel (3.1) näitab, et 2 ja 3 moodustavad perioodi 238, 4 ja 5 perioodi 106, 6 ja 7

Tabel 3.1

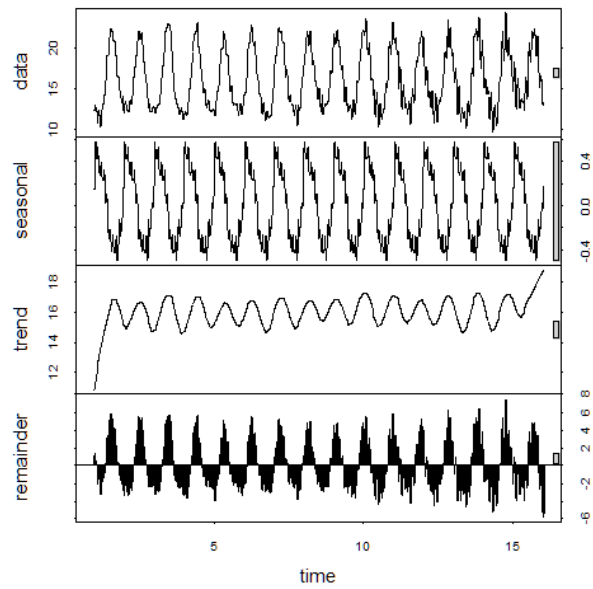
Komponent A ja B	Periood	Määr	Mod	Arg	Re	Im
A=2 ja B=3	238,278	0	1	0,03	0,99965	0,02637
A=4 ja B=5	105,611	0	1	0,06	0,99823	0,05946
A=6 ja B=7	38,979	0	1	0,16	0,98704	0,1605
A=8 ja B=9	29,81	0	1	0,21	0,97787	0,20922
A=12 ja B=13	26,044	0	1	0,24	0,97104	0,23892

perioodi 39, 8 ja 9 perioodi 30, 12 ja 13 perioodi 26, mis tähendab, et moodustavad pika tsükli. Pikad tsüklid ei ole sessoonsed. Hiljem lisame need trendile.



Joonis 3.17: Trendide ja sessoonide graafik algsel real

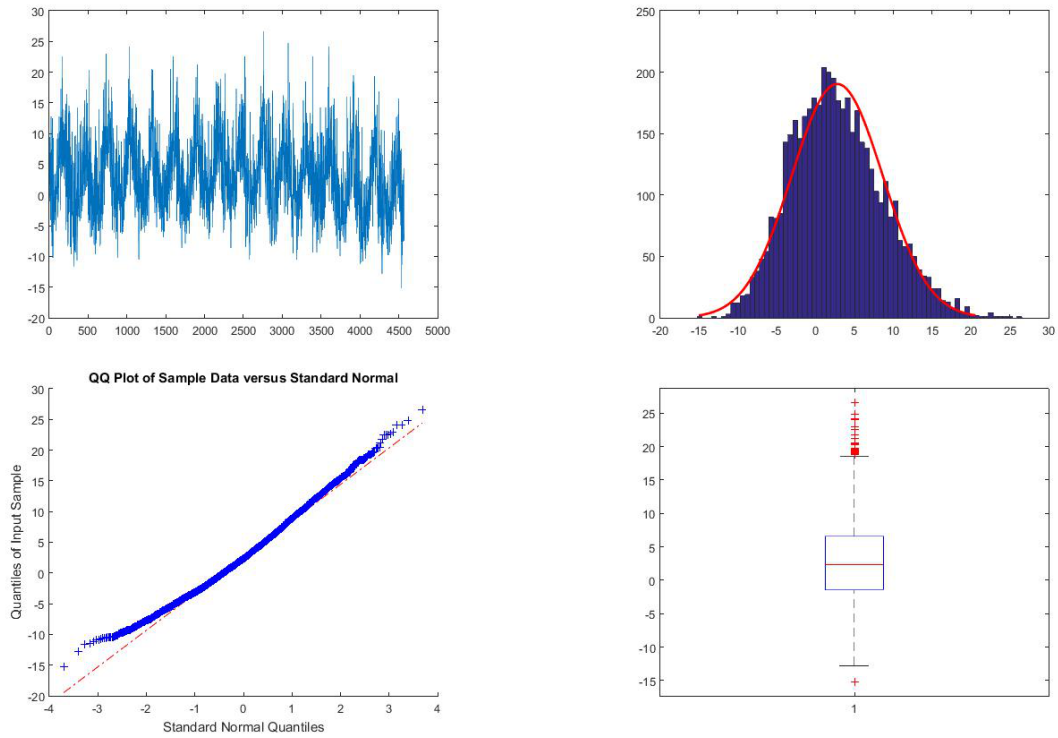
Jooniselt (3.17) näeme, et tegelikult on keeruline leida mingit sessoonsust, sest ka trendid on väga erinevad aastate lõikes.



Joonis 3.18: Rekonstrueeritud rea trendid

### 3.2.3 Jääkliikme hinnang

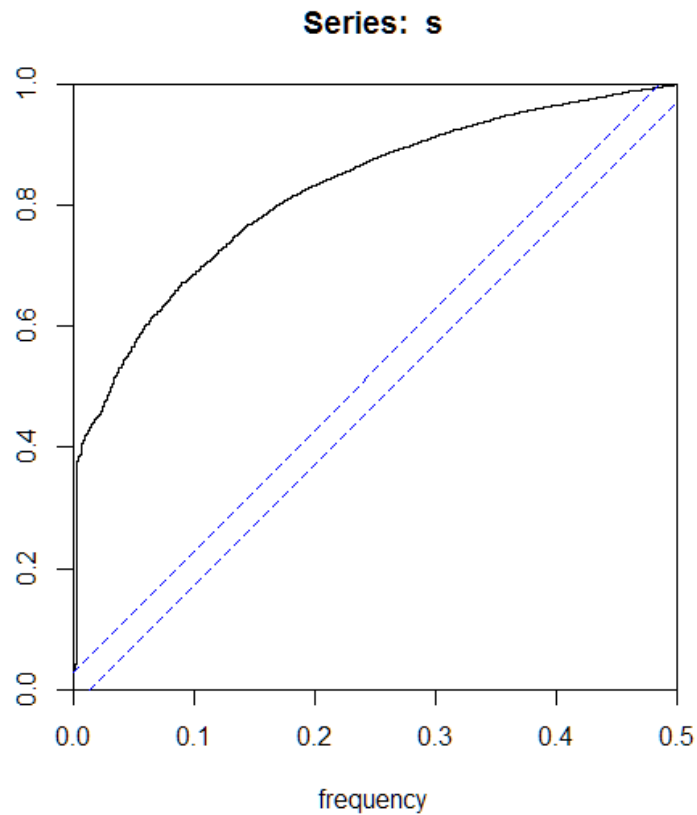
Jääkliige on, mis jääb üle reast, kui sealt on välja võetud trend, ja sestoone osa. Vaatleme jääkliikme graafikute kujusid.



Joonis 3.19: Jääkliikme graafikud  $L=50$

Histogrammil on näha, et kerge kalle vasakule, mis on tingitud mõnest väärtusest, mis kalduvad normaalsusest kõrvale. Karpdiagramm näitab meile, et on mõned ebanormaalsused, kuid peamine kast on sümmeetriline. Samas on ta nullpunktist veidi kõrgemal. Positiivsed väärtused näitavad meile suuremat veeauru sisaldust, kui mida võiks eeldada trendide ja sessoonide põhjal. Kvantiilide graafik näitab meile, et väärtused  $-2, -1, 2$  ei ole kaugel normaaldistributsioonist. Siiski on ka suuremaid väärtusi, mis näitavad kõrvalekallet normaalsusest. Shapiro-Wilk test näitab meile, et  $p - value = 5.493e - 16$ . See tähendab, et jääkliikmeid ei saa käsitleda normaalsetena. Kuna meil on 4568 väärtust, siis iga kõrvalekalle ei pruugi eriti oluline olla.





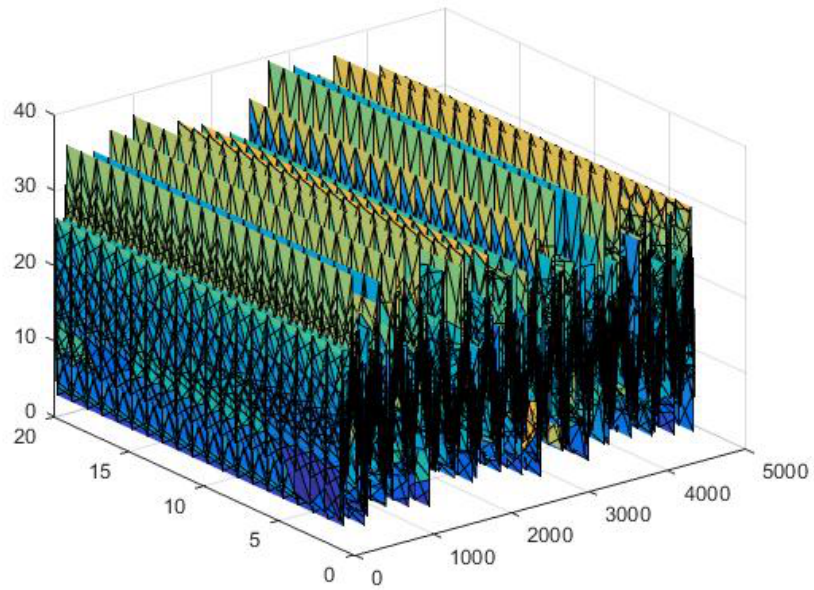
Joonis 3.20: Kumulatiivne jääkliikme funktsioon

Joonis (3.20) näitab, et jääkliikme puhul pole tegemist lihtsalt müraga.

### 3.3 SSA meetodi rakendus MATLAB-is

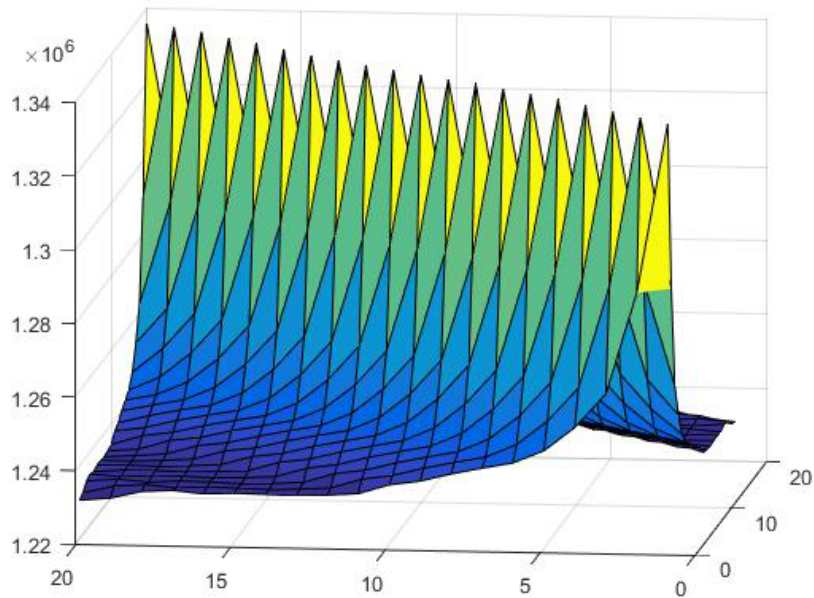
Valisime antud juhul  $L = 20$ .

Trajektoori maatriks  $T$  on mõõtmetega  $20 \times 4549$  (20 rida ja 4549 veergu). Trajektoormaatricsi  $T$  diagonaalis on algse rea  $X$  väärtused üle kahe ehk siis diagonaalis on algrea elementidest  $1, 3, 5, 7, \dots, 4549$ . Trajektoormaatricsi  $T$  graafik on selline:



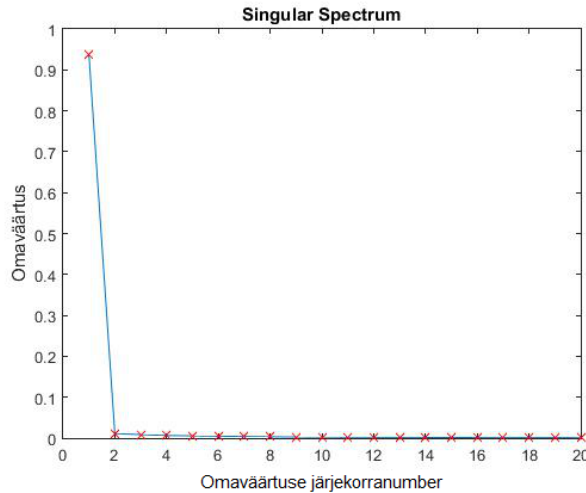
Joonis 3.21: Trajektoormatriksi elemendid

Järgmisena arvutame matriksi  $T * T^T$ , mille tulemusena saame 20\*20 ruutmatriksi:



Joonis 3.22: Maatriksi  $T * T^T$  väärtused

Kolmas samm on leida SVD. Iga omaväärtuste suhe  $\lambda_i / \sum_{i=1}^d \lambda_i$  on siis ka mõistetav kui  $T_i$  osa  $T$ -st, mis on graafikul järgnev

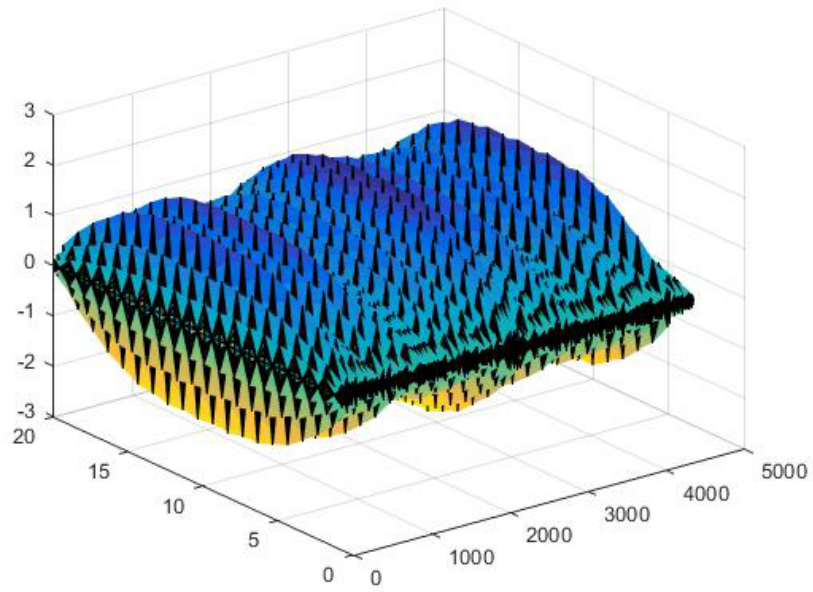


Joonis 3.23: Omaväärtused

Järgmisena valime  $l$  ( $1 \leq l \leq L$ ) suuruse grupi omavektoreid  $P_{i_1}, P_{i_2}, \dots, P_{i_l}$ . Olgu  $l = 20$ . Olgu hulk  $I = \{i_1, \dots, i_l\}$  indekse  $i_1, \dots, i_l$  grupp. Siis maatriks  $T_I$  vastavalt grupeerimisele on defineeritud niiviisi:

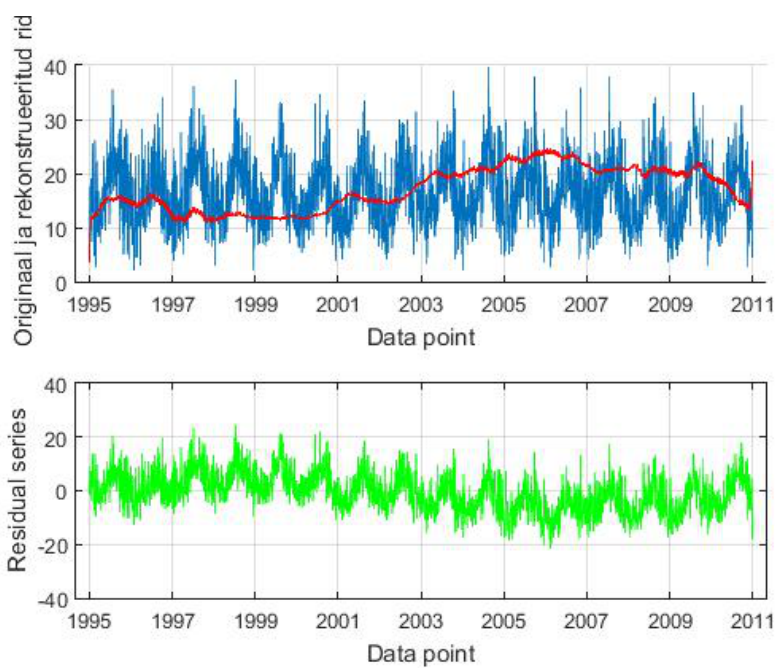
$$T_I = T_{i_1} + \dots + T_{i_l}.$$

Viimane samm on arvutada aegrida  $\tilde{X}$ , mis on lähendiks algele aegreale  $X$ . Selleks rakendame diagonaalset keskmistamist eelmises sammus arvutatud maatriksile  $T_{I_i}$  ja saame esialgse pikkusega aegrea  $\tilde{X}^{(i)}$ . Rakendades diagonaalset keskmistamist kõigile eelmises sammus saadud maatriksitele, saame  $T = \tilde{X}_{I_1} + \dots + \tilde{X}_{I_l}$ , kus  $\tilde{X}_{I_1} = \varkappa \tilde{X}_{I_1}$  ja seega esitatakse algne rida  $X$  maatriksitest  $\tilde{X}_{I_i}$  diagonaalse keskmistamise abil saadud ridade  $\tilde{X}^{(i)}$  summana. Meie andmete puhul on rida  $\tilde{X}^{(i)}$  kujul:



Joonis 3.24: Maatriks  $\tilde{X}^{(i)}$  väärtused

Nagu näha, ei olnud meie valitud tükeldus parim. Jaotades indeksid 1,...,20 saame:



Joonis 3.25: Algse rea silumine ja jääkliige, kui akna suurus  $L = 20$

Joonise (3.25) alumisel graafikul on näha jääkliige. Aegrida saab kirjeldada kolme komponendi kaudu sisuliselt: Aegrida=trendid+tsüklid+müra(ebaregulaarsed muutused). Seega, mida suurem on jääkliige, seda rohkem on jääkliikmete reas väärtusi.

# Singular spectrum analysis method

Märt Riiner

## Summary

This thesis is motivated by the difficulty of time series analysis. The longer the time series are, the harder is to analyse by computers the trends and future values. It needs faster computers to analyse the series. So for this reasons we are investigating the SSA method for analysing and smoothing time series.

This bachelor thesis consists of four chapters. The first, second and third chapter give the overview of SSA method. The fourth chapter gives us numerical example of SSA method. In the fourth chapter, we use University of Tartu Institute of Physics collected data of weather observation and give overview of choosing window size, analyzing graphics and deciding over window length.

In this thesis we use computer programs Matlab and R.

# Kirjandus

- [1] A BEGINNER'S GUIDE TO SSA - CERES [environnement.ens.fr/IMG/file/DavidPDF/SSA\\_beginners\\_guide\\_v9.pdf](http://environnement.ens.fr/IMG/file/DavidPDF/SSA_beginners_guide_v9.pdf)
- [2] AEGREA MÕISTE [http://www.lvrkk.ee/kristiina/Heli\\_Freienthal/aegread/](http://www.lvrkk.ee/kristiina/Heli_Freienthal/aegread/)
- [3] ANALYSIS OF PROCESS DATA WITH SINGULAR SPECTRUM METHODS [https://scholar.sun.ac.za/bitstream/handle/10019.1/16252/barkhuizen\\_analysis\\_2003.pdf;sequence=1](https://scholar.sun.ac.za/bitstream/handle/10019.1/16252/barkhuizen_analysis_2003.pdf;sequence=1)
- [4] BASIC SINGULAR SPECTRUM ANALYSIS AND FORECASTING WITH R <https://arxiv.org/pdf/1206.6910.pdf>
- [5] GOLYANDINA, N., NEKRUTKIN, V., ZHIGLJAVSKY, A., *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman & Hall, 2001.
- [6] HASSANI, HOSSEIN, *Singular Spectrum Analysis: Methodology and Comparison*, Journal of Data Science 5(2007),p. 239-257
- [7] KANGRO, RAUL, *Aegridade analüüs*, <http://kodu.ut.ee/~rkangro/aegread/2011/aegread.pdf>
- [8] MATLAB-I KOOD <https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/8115/versions/1/previews/ssa.m/index.html>
- [9] PACKAGE RSSA <https://cran.r-project.org/web/packages/Rssa/Rssa.pdf>
- [10] R-I PAKETID <http://www.inside-r.org/>
- [11] SINGULAR SPECTRUM ANALYSIS WITH RSSA <http://www.milanor.net/blog/wp-content/uploads/2014/07/SingularSpectrumAnalysisWithRssa.pdf>



- [12] TIME SERIES DECOMPOSITION USING SINGULAR SPECTRUM ANALYSIS <http://dc.etsu.edu/cgi/viewcontent.cgi?article=3725&context=etd>

# LISA A

## Matlabi kood

Allikas: <https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/8115/versions/1/previews/ssa.m/index.html>

```
load albh.txt; % Loeme andmed sisse failist albh.txt
x1 = albh(:,3); % Loeme failist albh.txt 3 veeru
L=4; % M"argime akna suuruse
```

```
% Step1 : Build trayjectory matrix
```

```
N=length(x1);
if L>N/2;L=N-L;end
K=N-L+1;
X=zeros(L,K);
for i=1:K
    X(1:L,i)=x1(i:L+i-1);
end
```

```
% Step 2: Constructing a matrix for applying SVD
```

```
S=X*X';
```

```
%Step 3: SVD of the matrix XXT
```

```
[U,autoval]=eig(S);
[d,i]=sort(-diag(autoval));
d=-d;
U=U(:,i);sev=sum(d);
plot((d./sev),hold on,plot((d./sev),'rx');
title('Singular Spectrum');xlabel('0mav"a"artuste arv');
ylabel('0mav"a"artus')
```

```

V=(X')*U;
rc=U*V';

% Step 4: Selection of eigen(vectors):

I=input('Valime t"ukelduste suurused I vormis [i1,i2:ik,...,iL]')
Vt=V';
rca=U(:,I)*Vt(I,:);

% Step 5: Reconstruction

y=zeros(N,1);
Lp=min(L,K);
Kp=max(L,K);

for k=0:Lp-2
    for m=1:k+1;
        y(k+1)=y(k+1)+(1/(k+1))*rca(k-m+2);
    end
end

for k=Lp-1:Kp-1
    for m=1:Lp;
        y(k+1)=y(k+1)+(1/(Lp))*rca(k-m+2);
    end
end

for k=Kp:N
    for m=k-Kp+2:N-Kp+1;
        y(k+1)=y(k+1)+(1/(N-k))*rca(k-m+2);
    end
end

figure;subplot(2,1,1);hold on;xlabel('Data point');
ylabel('Originaal ja rekonstrueeritud rida')
plot(x1);grid on;plot(y,'r')

r=x1-y;
subplot(2,1,2);plot(r,'g');

```

```
xlabel('Data point');ylabel('Residual series');grid on  
vr=(sum(d(I))/sev);
```

# LISA B

## R-i kood

```
#Gdata, Rssa, TSA paketid, install.packages("gtools", dependencies = T)

#w-korrelatsioon
s <-read.xls("C:/algerida.xls",
             sheet=1, header =FALSE,
             perl="C:/Perl64/bin/perl.exe")
ss <- ssa(s, L=2274)
w <- wcor(ss)
print(w)
plot(w)
plot(ss)
plot(ss, type = "vectors",idx = 1:12) # omavektorid
plot(ss, type = "paired", idx = 1:12) #omavektorite paarid
print(parestimate(ss,groups=list(c(2,3),c(4,5),c(6,7),c(8,9),c(12,13))))

#####kood
# plot
coo2 <-read.xls("C:/algerida.xls",
               sheet=1, header =FALSE,
               perl="C:/Perl64/bin/perl.exe")
s<- as.numeric(unlist(coo2))
ss <- ssa(s, L=12)
plot(ss, type = "vectors", idx = 1:12)
plot(ss, type = "series", groups = 1:12)

#Trend
coo2 <-read.xls("C:/algerida.xls",sheet=1,
               header =FALSE,perl="C:/Perl64/bin/perl.exe")
s<- as.numeric(unlist(coo2))
```

```

ss <- ssa(s)
r <- reconstruct(ss, groups = list(Trend = c(1:12)))
plot(r, plot.method = "xyplot", add.residuals = TRUE,
add.original = TRUE,superpose = TRUE, auto.key = list(columns = 2), col="red")
#plot(r, add.residuals = FALSE,add.original = TRUE);
#legend(4000,40,legend="series", fill="black")
#typeof(s)

#####
plot(ss, type = "vectors", vectors = "factor", idx = 1:12)

plot(ss, type = "wcor", groups = 1:10)

plot(ss, plot.method = "xyplot", superpose = TRUE,
auto.key = list(columns = 1),
col = c("blue"),
lty = c(rep(1)))

lst <- grouping.auto(ss, grouping.method = "wcor",
groups = 1:6, nclust=3)

g1 <- grouping.auto(ss, base = "series", freq.bins = list(0.005),
threshold = 0.95)

#RAW periodogram, periodogramm
spec.pgram(coo2, taper=0, fast=FALSE, detrend=FALSE, log="no")
cpgram(coo2)

#Residuals
coo2 <-read.xls("C:/algnerida.xls",
sheet=1, header =FALSE,
perl="C:/Perl64/bin/perl.exe")
s<- as.numeric(unlist(coo2))
ss <- ssa(s, L=2274)
residualssss<- residuals(ss)
spec.pgram(residualssss, taper=0.1, fast=FALSE, detrend=FALSE, log="no")

```

```

cpgram(residualsssss)

#Seasonality
ss<-read.xls("C:/algnnerida.xls",
             sheet=1, header =FALSE,
             perl="C:/Perl64/bin/perl.exe")
s<- as.numeric(unlist(ss))
plot(stl(ts(s,freq=304), t.window=365, s.window="per", robust=TRUE))
plot(stl(ts(ss,freq=304),s.window="per", robust=TRUE))
#trendide graafik aasta kohta

# reconstructed
coo2 <-read.xls("C:/algnnerida.xls",
               sheet=1, header =FALSE,
               perl="C:/Perl64/bin/perl.exe")
s<- as.numeric(unlist(coo2))
ss <- ssa(s)
r <- reconstruct(ss, groups = list(Trend = c(2,3),c(4,5),c(6,7),c(8,9),c(12,13)))
s<- as.numeric(unlist(r))
plot(stl(ts(s,freq=2280),t.window=365,s.window="per", robust=TRUE))

#Periodogram
s <-read.xls("C:/algnnerida.xls",
             sheet=1, header =FALSE,
             perl="C:/Perl64/bin/perl.exe")
ss <- ssa(s, L=2274)
periodogram(ss,groups=list(c(2,3),c(4,5),c(6,7),c(8,9),c(12,13)),
           ylab='Variable Star Periodogram'); abline(h=0)

#Smoothed Periodogram
s <-read.xls("C:/algnnerida.xls",
             sheet=1, header =FALSE,
             perl="C:/Perl64/bin/perl.exe")
spectrum(s, spans = c(2,3))
spectrum(s, c(2),c(3))

#

```

```
s <-read.xls("C:/Vihik1.xlsx",
            sheet=1, header =FALSE,
            perl="C:/Perl64/bin/perl.exe")
> ss<- as.numeric(unlist(s))
> shapiro.test(ss)

#cummulative program
cpgram(s)
```



## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Märt Riiner (sünnikuupäev 25.02.1985),

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Singulaarse spektraalanalüüsi meetod", mille juhendaja on Peep Miidla,
  - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace'is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **11.05.2017**