



TAMPEREEN TEKNILLINEN YLIOPISTO

**SAMULI HAVERINEN**

**Epälineaarinen aikasarja-analyysi ja sykevälivaihtelu**

Diplomityö

Tarkastaja: professori Esa Räsänen  
Jätetty tarkastettavaksi 17.11.2017

# TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Teknis-luonnontieteellinen koulutusohjelma

**Samuli Haverinen: Epälineaarinen aikasarja-analyysi ja sykevälivaihtelu**

Diplomityö, 52 sivua

Marraskuu 2017

Pääaine: teknillinen fysiikka

Tarkastaja: professori Esa Räsänen

Avainsanat: pitkän kantaman korrelaatiot, fraktaaliset aikasarjat, Brownin liike,  $1/f$ -kohina, DFA, sykevälivaihtelu, RR-sarjat

Monissa kiinnostavissa aikasarjoissa on havaittu pitkän kantaman korrelaatioita tai fraktaalisuutta. Näitä ominaisuuksia on löydetty esimerkiksi erilaisista geofysikaalisissa aikasarjoista, DNA-jaksojen järjestymisestä, eräiden kvanttisysteemien energiatilojen jakaumista sekä sydämen sykevälivaihtelusta. Pitkän kantaman korrelaatioita tai fraktaalisuutta ei voida määrittää tarkasti tavallisimmilla tilastollisilla työkaluilla, joten tehtävään on kehitetty erityisiä laskennallisia menetelmiä.

Tässä tutkielmassa perehdytään trendit poistavaan fluktuaatioanalyysiin (*detrended fluctuation analysis*, DFA), joka on suosittu menetelmä pitkän kantaman korrelaatioiden ja fraktaalisuuden kvantitatiiviseen määrittämiseen. DFA:n todetaan olevan täsmällinen menetelmä, jos tutkitut aikasarjat noudattavat tiettyjä teoreettisia taustaoletuksia. Toisaalta havaittiin, että erilaiset poikkeamat taustaoletuksista vaikuttavat yleensä selvästi lopputulokseen ja näin vaikeuttavat myös tulosten tulkintaa.

Menetelmän ominaisuuksien lisäksi tässä työssä on tutkittu sen soveltamista sydämen sykevälivaihteluun. Sykevälivaihtelua pidetään tärkeänä mittarina autonomisen hermoston tilasta. Sykevälivaihtelu sisältää pitkän kantaman korrelaatioita, joihin esimerkiksi joidenkin patologisten tilojen on havaittu vaikuttavan merkittävästi. Tutkielmassa kuitenkin havaitaan, että patologisten tilojen lisäksi DFA:lla määritetyt korrelaatio-ominaisuudet vaihtelevat merkittävästi tavallisissakin oloissa. Tämä on seurausta aineiston luonteesta; sykevälivaihteluun voivat potentiaalisesti vaikuttaa monet tekijät, eikä se siten noudata täsmällisiä taustaoletuksia.

Lopuksi DFA:lla on tutkittu esimerkinomaisesti avantouimarien sykevälivaihtelua. Yksittäisille koehenkilöille tehtiin kolme pitkää sydänsähkökäyrä- eli EKG-mittausta, joista yksi sisälsi saunomista, toinen saunomista ja avantouintia ja kolmas eli referenssimittaus ei kumpaakaan. DFA:lla määritetyistä korrelaatio-ominaisuuksista löydettiin samojen henkilöiden sauna- ja referenssimittausten väliltä tilastollisesti merkittäviä eroavaisuuksia, mutta muiden mittausten väliltä ei. Havaintojen syytä ei pystytty tarkasti selvittämään, mutta hypoteeseja ja ehdotuksia jatkotutkimuksista esitettiin.

## ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Science and Engineering

**Samuli Haverinen: Nonlinear time series analysis and heart rate variability**

Master of Science Thesis, 52 pages

November 2017

Major: Engineering physics

Examiners: Professor Esa Räsänen

Keywords: long-term correlations, fractal time series, Brownian motion,  $1/f$ -noise, DFA, heart rate variability, RR-series

Long-term correlations and fractality have been observed in many interesting time series. These properties have been found, for example, in various geophysical time series, in DNA sequences, in distributions of energy levels of some quantum systems, and in the human heart rate variability. Long-term correlations or fractality cannot be estimated accurately with the most common statistical tools. Hence, specific computational methods have been developed.

In this thesis, we study detrended fluctuation analysis, DFA, which is a popular method for estimating long-term correlations and fractality in a quantitative fashion. We find out that DFA is an accurate method if time series follow certain theoretical assumptions. On the other hand, various deviations from these assumptions affect the results and make interpretations more complicated.

In addition to the properties of the method, we have studied its application to heart rate variability. Human heart rate variability is thought to be an important measure of the state of the autonomous nervous system. Heart rate variability contains long-term correlations, and it has been observed that some pathological states affect correlations significantly. However, we also observed that in addition to the pathological states, correlation properties estimated by DFA can vary significantly in normal conditions. This is a result from the nature of the data; heart rate variability can potentially be affected by many factors. Thus, heart rate variability does not follow rigorous theoretical assumptions.

Finally, we used DFA to study heart rate variability of winter swimmers as an example. There were three long electrocardiogram (ECG) measurements from each individual. The first measurement involved sauna bathing, the second measurement involved sauna bathing and winter swimming, and the third, the reference measurement, did not involve either. We found statistically significant differences between the sauna measurement and the reference measurement using correlation properties estimated by DFA in the statistical analysis. Statistical significance was not found between the other measurements. The reason for these observations was not fully clarified, but we propose some hypotheses and follow-up studies.

## ALKUSANAT

Aikasarjat ja aikasarja-analyysi eivät olleet minulle ennestään tuttuja aihepiirejä. Aikasarjat valikoituivat aiheeksi osittain oman mielenkiinnon ja osittain sattuman kautta. Sovelluskohde eli sykevälivaihtelu lähinnä sattuman kautta. Sovelluskohde olisi hyvin voinut olla jokin muukin, koska tutkielman pääpaino on menetelmäpuolella ja sen suhteessa ns. reaali maailman aineistoon.

Kiitos kaikille, jotka olivat millään tavalla, virallisesti tai epävirallisesti, osallisina tässä diplomityöprosessissa. Tiedätte itse keitä olette.

# SISÄLLYS

1. Johdanto . . . . .	1
2. Aikasarja-analyysistä ja fraktaalista aikasarjoista . . . . .	3
2.1 Stationäärisyys ja ergodisuus . . . . .	3
2.2 Aikasarjat ja fraktaalit . . . . .	4
2.3 Korrelaatiot aikasarjassa . . . . .	6
3. Trendit poistava flukтуаatioanalyysi, DFA . . . . .	9
3.1 Algoritmin kulku . . . . .	10
3.2 Skaalauseksponentti $\alpha$ . . . . .	11
3.3 Algoritmin ominaisuuksia . . . . .	12
3.3.1 Ikkunakoon vaikutus . . . . .	12
3.3.2 Aikasarjan pituus . . . . .	13
3.3.3 Aikasarjan generointimenetelmän vaikutus . . . . .	16
3.3.4 Ulkoiset trendit . . . . .	18
3.3.5 Muut ei-stationäärisyydet . . . . .	19
3.3.6 Lyhyen kantaman korrelaatiot . . . . .	20
3.3.7 Mahdolliset valesiirtymät . . . . .	21
3.3.8 Tuntemattomat aikasarjat . . . . .	22
3.4 Multifraktaalisuus ja MF-DFA . . . . .	22
4. DFA ja sykevälialyysi . . . . .	25
4.1 Taustaa . . . . .	25
4.2 Pitkän ja lyhyen kantaman skaalauseksponentit . . . . .	27
4.3 Paikalliset skaalauseksponentit . . . . .	28
4.4 Tuloksista ja niiden tulkinnasta . . . . .	30
5. Muita menetelmiä sykevälialyysissä . . . . .	32
5.1 Tehospektri . . . . .	32
5.2 Entropiat . . . . .	33
5.3 Paluukartat . . . . .	35
6. Tutkimusesimerkki: sykeväli vaihtelu avantouimareilla . . . . .	37
6.1 Aineisto . . . . .	37
6.2 Tutkimuskysymykset . . . . .	38
6.3 RR-sarjojen esiprosessointi . . . . .	38
6.4 Paikallisten skaalauseksponenttien vaihtelu . . . . .	39
6.5 Skaalauseksponenttien tilastollinen analyysi . . . . .	42
6.6 Analyysin rajoituksia . . . . .	44
7. Yhteenveto . . . . .	46
Viitteet . . . . .	49

## SYMBOLIT

$X$	Stokastinen prosessi
$X_i$	Satunnaismuuttuja
$x_i$	Realisoitunut satunnaismuuttujan arvo
$(x_i)$	Aikasarja, jono realisoituneita satunnaismuuttujia
$H$	Hurstin eksponentti
$\alpha$	DFA:lla määritettävä skaalauseksponentti
$\langle x \rangle, \mu(x)$	$x$ :n keskiarvo
$\sigma(x)$	$x$ :n standardipoikkeama

# 1. JOHDANTO

Aikasarja on jono mittauksia tai havaintoja aikajärjestyksessä. Aikasarjat ovat yleisiä monilla tieteen ja tekniikan aloilla. Aikasarja voi olla esimerkiksi päivittäinen lämpötila, tietoliikenteen määrä verkossa, osakekurssien hinta tai vaikkapa laserin teho. Usein aikasarja kuvaa monimutkaisemman systeemin tai prosessin tilaa ja aikakehitystä dimensionaalisesti alennetussa esitysmuodossa.

Aikasarja-analyysi on mielekkäiden suureiden määrittämistä taustalla olevasta prosessista aikasarjan kautta, sekä ennustamista ja mallintamista. Menetelmien ja työkalujen kirjo on laaja ja yleensä ne perustuvat joihinkin oletuksiin taustalla olevasta prosessista. Prosessi voi olla esimerkiksi kaoottinen, lineaarinen tai epälineaarinen, deterministinen tai satunnainen tai jokin yhdistelmä näistä. Monesti prosessin luonne ei ole edes selvä eikä sen selvittäminen ja siten oikeiden työkalujen valinta ole välttämättä yksinkertaista. Huomautettakoon myös, että muitakin sarjoja kuin aikasarjoja voidaan käsitellä aikasarja-analyysin työkaluilla.

Tämä tutkielma käsittelee erästä niin sanottujen kompleksisten aikasarjojen tutkimuksessa käytettyä menetelmää ja yhtä sen sovelluskohteista; trendit poistavaa fluktuaatioanalyysia (DFA, *detrended fluctuation analysis*) ja sykevälianalyysiä. Termillä ”kompleksinen aikasarja” tarkoitetaan aikasarjaa, jonka tulevat arvot eivät riipu vain lähimenneisyydestä, vaan joka sisältää pitkän kantaman korrelaatioita<sup>1</sup>. Sykeväleillä tarkoitetaan ihmisen sydämenlyöntien välistä aikaa. Sykeväliaikasarjojen on näytetty olevan kompleksisia yllä esitetyssä merkityksessä.

Tutkielma voidaan jakaa kahteen pääaiheeseen. Ensimmäinen on DFA:n tarkastelu ja tutkiminen menetelmänä. Luvussa 2 selvitetään tarvittavia taustaoletuksia ja joitain teoreettisia ominaisuuksia; millaisia aikasarjoja halutaan tutkia ja mistä niissä ollaan kiinnostuneita. Luvussa 3 esitellään tutkielman ytimessä oleva menetelmä, DFA, sekä paneudutaan laajasti sen ominaisuuksiin eri tilanteissa, osittain kirjallisuuden kautta ja osittain omien simulaatioiden avulla. Tämä antaa kokonaiskuvan DFA:n heikkouksista, vahvuuksista ja sovellusalueesta.

Toinen pääaihe on DFA:n soveltaminen sydämen sykevälianalyysiin. Menetelmän soveltaminen ei ole suoraviivainen rutiinitoimenpide, koska sykeväliaikasarjoilla on kiusallinen ominaisuus, joka aiheuttaa valtaosan haasteista aikasarja-analyysin sovelluksissa yleisestikin. Sykeväliaikasarjat eivät välttämättä ole stationäärisiä, mikä

---

<sup>1</sup>Termi ”kompleksinen aikasarja” esiintyy usein kirjallisuudessa, mutta sitä käytetään vaihtelevissa merkityksissä. Usein sille ei edes anneta mitään täsmällistä määritelmää.

tarkoittaa että taustalla olevan prosessin ei voida olettaa pysyvän muuttumattomana. Stationäärisyys on taustaoletus useimmille aikasarja-analyysin menetelmille. Luvussa 4 käsitellään DFA:n soveltamista sykevälialianalyysiin yleisesti, osittain jälleen kirjallisuuden kautta ja osittain itse kokeilemalla. Lopuksi luvussa 6 selvitetään mitä kaiken aikaisemman perusteella voidaan saada irti todellisesta aineistosta. Luvussa tarkastellaan avantouimarien sykeväliaikasarjoja, jotka ovat peräisin aikaisemmin Tampereella tehdyistä tutkimuksista, joissa selvitettiin avantouinnin mahdollisia terveysvaikutuksia. Näiden pääaiheiden lisäksi luvuissa 3.4 ja 5 tehdään pieni katsaus muihin sykevälialianalyysissä käytettyihin aikasarja-analyysin menetelmiin.

Tutkielman otsikon “epälineaarinen aikasarja-analyysi” on osittain harhaanjohdava. Kirjallisuudessa DFA:ta ja muita tutkielmassa esille tuotuja menetelmiä kutsutaan usein epälineaariseksi. Menetelmät itsessään eivät kuitenkaan ole missään yleisessä merkityksessä epälineaarisia. Tämä epäjohdonmukaisuus juontaa juurensa siitä, että näiden menetelmien sovelluskohteina olevat aikasarjat usein ovat tai niiden oletetaan olevan peräisin epälineaarista systeemeistä.



## 2. AIKASARJA-ANALYYSISTÄ JA FRAKTAALISISTA AIKASARJOISTA

Aikasarja-analyysi on laaja tieteenala. Erilaisia aikasarjoja, menetelmiä ja sovelluskohteita on valtavan paljon. Tässä luvussa ei käsitellä aikasarja-analyysia kattavasti, vaan tutustutaan ohueen siivuun, jota tutkielmassa tarvitaan.

Jotta yksittäisestä aikasarjasta voitaisiin määrittää koko prosessin ominaisuuksia, tarvitaan jonkinlainen käsitys prosessin muuttumattomuudesta ja toistettavuudesta. Tämän vuoksi määritellään aluksi stationäärisyyden ja ergodisuuden käsitteet, jotka luovat pohjan tilastollisten menetelmien soveltamiselle. Tämän jälkeen tutustutaan aikasarjoihin liittyviin fraktaalisiin ilmiöihin ja korrelaatioihin aikasarjoissa, jotka ovat keskeisiä jatkossa käytettävien menetelmien kannalta.

### 2.1 Stationäärisuus ja ergodisuus

Tässä tutkielmassa aikasarjalla tarkoitetaan lukujonoa

$$(x_i), \quad i = 1, 2, 3, \dots, N, \quad (2.1)$$

missä lukujonon alkiot ovat skalaariarvoisia ja tasavälein tehtyjä mittauksia tai havaintoja. Aikasarja on realisaatio taustalla olevasta stokastisesta prosessista  $X$ . Huomautettakoon, että myös deterministiset dynaamiset systeemit voidaan nähdä stokastisina prosesseina, joten tämä määritelmä ei rajoita tarkastelua pelkästään satunnaisuusluonteisiin prosesseihin.

Stokastinen prosessi on stationäärinen<sup>1</sup>, jos yhteisjakaumat  $(X_1, X_2, \dots, X_N)$  ja  $(X_{1+\tau}, X_{2+\tau}, \dots, X_{N+\tau})$  ovat samat eli

$$(X_1, X_2, \dots, X_N) = (X_{1+\tau}, X_{2+\tau}, \dots, X_{N+\tau}) \quad (2.2)$$

kaikilla valinnoilla  $N$  ja  $\tau$ , kun  $X_i$  on stokastisen prosessin satunnaismuuttuja indeksillä  $i$  [1]. Aikasarjan kannalta tämä tarkoittaa sitä, että todennäköisyys tietyille realisoituneille arvoille ei riipu aloitusajankohdasta. Täten stationäärisestä prosessista peräisin olevaa aikasarjaa voidaan pitää satunnaisotoksena.

---

<sup>1</sup>Kyseessä on ns. vahva stationäärisuus. Kirjallisuudessa esiintyy usein myös toinen stationäärisyyden määritelmä, heikko stationäärisuus. Vahva stationäärisuus ei implikoi heikkoa stationäärisyyttä tai päinvastoin.

Aikasarja-analyysin haasteena ovat otoksen koko sekä aikasarjan pituus. Monesti käytössä saattaa olla vain yksi aikasarja, jolloin otoksen koko on 1. Toisaalta aikasarjat ovat myös äärellisen pituisia eivätkä siten välttämättä edusta koko prosessia. Stokastinen prosessi  $X$  on ergodinen, jos sen tilastolliset ominaisuudet voidaan määrittää yhdestä riittävän pitkästä aikasarjasta  $(x_i)$  [1]. Erityisesti voidaan käsitellä tietyn suureen ergodisuutta, esimerkiksi keskiarvon. Stokastinen prosessi  $X$  on keskiarvoergodinen, jos aikasarjan  $(x_i)$  keskiarvo lähestyy prosessin otoskeskiarvoa  $E[X]$ , kun aikasarjan pituus lähestyy ääretöntä:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i = E[X]. \quad (2.3)$$

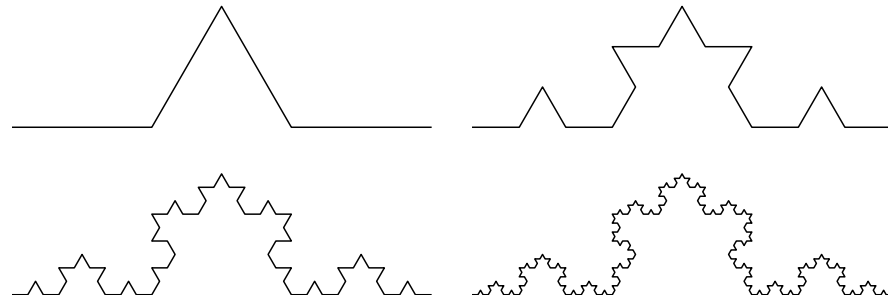
Vastaavasti voidaan määritellä muille suureille.

Stationäärisuus ja ergodisuus mahdollistavat prosessia koskevien suureiden määrittämisen yksittäisestä aikasarjasta. Valtaosa aikasarja-analyysistä nojaa tähän. Valitettavasti stationäärisyyden määritelmä ei kerro, miten stationäärisyyden voi todentaa aikasarjasta eikä ergodisuuden määritelmä anna viitteitä riittävästä aikasarjan pituudesta. Molemmat kriteerit pitää tutkia tapauskohtaisesti. Intuitiivisesti stationäärin prosessin pitää säilyä muuttumattomana (parametrit pysyvät vakioina) ja ergodisuuden kannalta riittävän pitkässä aikasarjassa prosessin eri ilmiöiden tulee olla edustettuna riittävästi. Jos prosessi on esimerkiksi jaksollinen, tulee aikasarjan sisältää riittävän monta jaksoa. Jos taas aikasarja sisältää esimerkiksi yhden epäsäännöllisen osan, prosessia täytyy pitää ei-stationäärinä tai ergodisuuden kannalta liian lyhyenä.

Nyt määritellyt käsitteet, stationäärisuus ja ergodisuus, ovat prosessin ominaisuuksia. Ilmaisun yksinkertaistamiseksi jatkossa saatetaan kuitenkin puhua stationäärisestä aikasarjasta. Tällä tarkoitetaan että aikasarja on peräisin prosessista, jota voidaan pitää stationäärinä.

## 2.2 Aikasarjat ja fraktaalit

Yleisimmin fraktaalit tunnetaan ei-triviaaleina geometrisina kuvioina tai muotoina, jotka voidaan jakaa yhä pienempiin ja pienempiin aliyksiköihin, jotka muistuttavat kokonaisuutta, kuten esimerkiksi kuvan 2.1 Kochin käyrä. Tämä ominaisuus tunnetaan itsesimilaarisuutena, mikä on fraktaalien yleinen ominaisuus. Huomautettakoon, että itsesimilaarisuus ei kuitenkaan ole riittävä ominaisuus fraktaalisuudelle, koska triviaalit geometriset oliot kuten suora ja neliö ovat itsesimilaarisia. Fraktaalien täsmällinen määrittely vaihtelee kontekstista riippuen, mutta tämän tutkielman tarkoituksiin riittää ei-triviaali itsesimilaarisuus tai -affinisuus. Itseaffinisuus on itsesimilaarisuuden yleistys, jossa aliyksiköihin jaossa suurennos voi olla anisotroop-



**Kuva 2.1:** Neljä ensimmäistä iteraatiota Kochin käyrästä. Kochin käyrä voidaan muodostaa jakamalla jana kolmeen osaan, asettamalla tasasivuisen kolmion kanta jaon keskimmäiseen osaan, poistamalla kanta ja toistamalla tätä uusille muodostuneille janoille [2]. Kochin käyrä on itseänsimilaarinen, koska kaikille uusille janoille (esimerkiksi alkuperäisen janan 1. kolmannekselle) muodostuu täydellinen (pienennetty) kopio kokonaisuudesta.

pinen, eli suurennoksen kertoimet eri akseleille poikkeavat toisistaan.

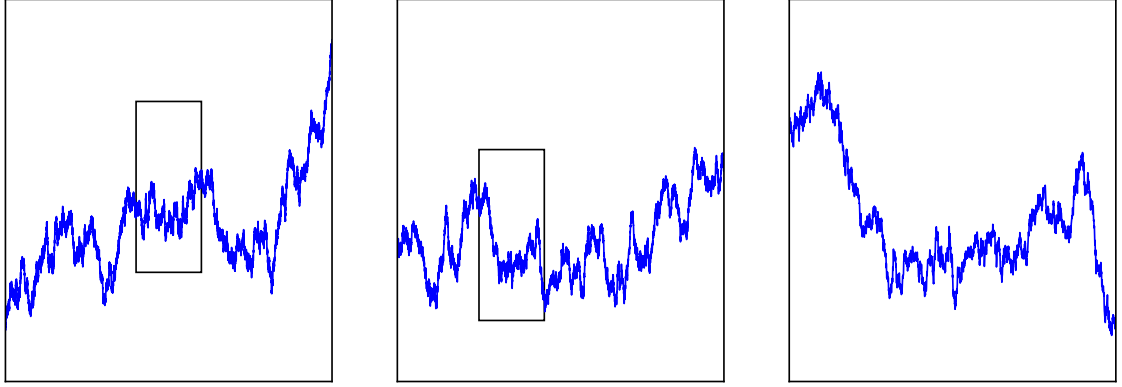
Aikasarjat eivät yleensä ole säännöllisiä, mutta niissä voi esiintyä tilastollista itseaffiinisyyttä. Itseaffiini jatkuva aikasarja  $x(t)$  säilyttää tilastolliset ominaisuutensa, kun se skaalataan relaation

$$x(t) \rightarrow a^H x(at), \quad 0 < H < 1 \quad (2.4)$$

mukaisesti, missä  $H$  on Hurstin eksponentti. Itseaffiineja aikasarjoja kutsutaan usein fraktaaliksi aikasarjoiksi. Fraktaalinen aikasarja ei ole relaation (2.4) mukaisen suurennoksen jälkeen täsmällinen kopio itsestään kuten vaikkapa Kochin käyrä, mutta se näyttää kvalitatiivisesti samanlaiselta. Tämä mittakaavainvarianssi on selvästi nähtävissä kuvassa 2.2<sup>2</sup>.

Esimerkki fraktaalista aikasarjasta on yksiulotteisen Brownin liikkeen (Wiener-prosessi) polku. Tavallisessa Brownin liikkeessä paikan lisäykset  $\Delta x(t,u) = x(t+u) - x(t)$  ovat riippumattomia ja ne ovat normaalijakautuneet jakaumalla  $\Delta x(t,u) \sim \mathcal{N}(0,u)$ . Brownin liikkeen polku on itseaffiini Hurstin eksponentilla  $H = 1/2$ . Voidaan osoittaa, että prosessi voidaan yleistää tapauksiin  $H \neq 1/2$ , jos luovutaan joko peräkkäisten lisäysten välisestä riippumattomuudesta (osittainen Brownin liike, *fractional Brownian motion*, lisäykset noudattavat kuitenkin normaalijakaumaa) tai lisäysten äärellisestä varianssista (*Lévy flight*, lisäykset eivät noudata normaalijakaumaa) [3]. Osittaista Brownin liikettä ja lisäysten korrelaatiota tarkastellaan seuraavassa kappaleessa. Osittaisen Brownin liikkeen ominaisuudet otetaan annettuina. Täsmällisiin matemaattisiin yksityiskohtiin voi tutustua esimerkiksi viitteessä [4].

<sup>2</sup>Kuvatekstissä käytetään suomen kielen yleisestä käytännöstä poikkeavaa desimaalierotinta merkinnässä  $a = 0.2$ . Tätä kansainvälisesti yleisempää merkintätapaa noudatetaan tässä tutkielmassa jatkossakin.



**Kuva 2.2:** Simuloidun yksiulotteisen Brownin liikkeen polku. Rajatut alueet on suurennettu oikeanpuoleiseen kuvaan. Rajauksien korkeus suhteessa leveyteen on relaan (2.4) mukainen, kun  $H = 1/2$  ja  $a = 0.2$ .

Reaalimaailman fraktaaleissa, niin fyysisissä kuin aikasarjoissakin, itseaffinisyyden aiheuttama mittakaavainvarianssi ei ole rajoittamatonta kuten matemaattisissa fraktaaleissa. Aikasarjat ovat aina lopulta diskreettejä, joko luonteeltaan tai rajallisen resoluution vuoksi, joten jatkuvasti suurentamalla ei voida nähdä aina vain lisää yksityiskohtia. Reaalimaailman aikasarjaa pidetään fraktaalina, jos sille voidaan mielekkäästi määrittää Hurstin eksponentti  $H$ , joka pätee useamman kertaluokan aikaskaaloilla.

### 2.3 Korrelaatiot aikasarjassa

Yleensä aikasarjojen peräkkäiset arvot eivät ole riippumattomia, vaan niiden välillä on korrelaatiota eli seuraava arvo riippuu edellisistä. Esimerkiksi päivittäinen lämpötila ei yleensä poikkea rajusti edellisen päivän lämpötilasta tai osakekurssien hinta edellisestä arvosta. Määritellään nyt täsmällisemmin, mitä korrelaatiolla tarkoitetaan.

Aikasarjan autokovarianssi  $C(s)$  kertoo etäisyydellä  $s$  toisistaan olevien havaintojen todennäköisyydestä saada samankaltaisia arvoja [5]:

$$C(s) = \langle x_i x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} x_i x_{i+s}. \quad (2.5)$$

Väljästi ilmaistuna, jos autokovarianssi on positiivinen, suurta arvoa seuraa todennäköisesti toinen suuri arvo ja pientä arvoa toinen pieni arvo, eli arvot ovat korreloituneita. Jos autokovarianssi on negatiivinen, niin suurta arvoa seuraa todennäköisesti pieni arvo ja päinvastoin, eli arvot ovat antikorreloituneita. Jos etäisyydellä  $s$  olevat arvot ovat riippumattomia, autokovarianssi lähestyy nollaa. Toisinaan autokovarianssin sijaan käytetään autokorrelaatiota, joka on autokovarianssi jaettuna varianssilla. Tämä normalisoi autokorrelaation arvot välille  $[-1,1]$ .

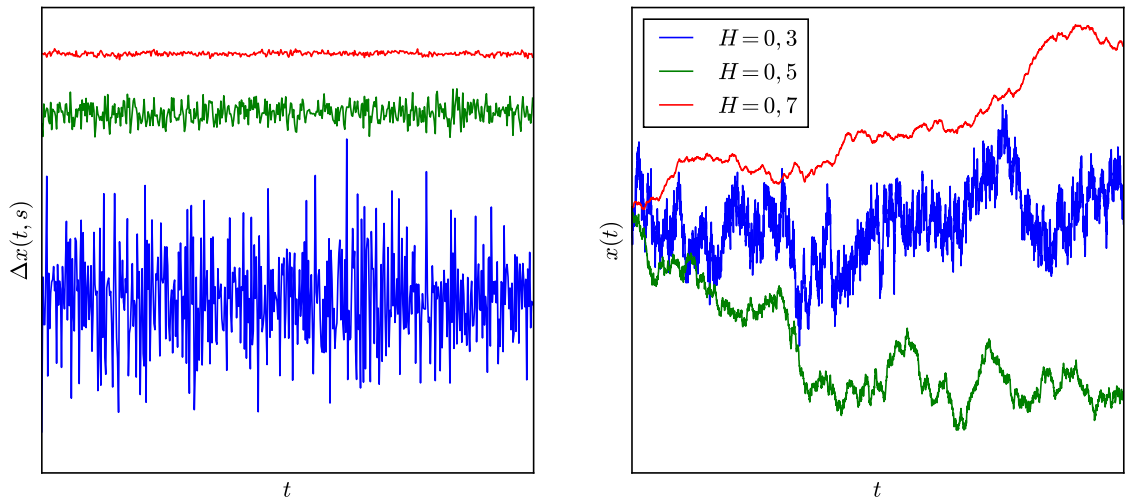
Tarkastellaan osittaista Brownin liikettä autokovarianssin avulla. Brownin liike tai osittainen Brownin liike eivät ole stationäärisiä<sup>3</sup>, mutta niiden lisäykset  $\Delta x_i = x_i - x_{i-1}$  ovat<sup>4</sup>. Tavallisen Brownin liikkeen,  $H = 1/2$ , lisäykset ovat riippumattomia eli lisäysten autokovarianssi on nolla. Osittaisen Brownin liikkeen lisäysten autokovarianssi puolestaan noudattaa potenssilakia

$$C(s) = \langle \Delta x_i \Delta x_{i+s} \rangle \propto s^{-\gamma}, \quad 0 < \gamma < 1 \quad (2.6)$$

kun  $1/2 < H < 1$ . Potenssilain eksponentin  $\gamma$  ja Hurstin eksponentin  $H$  välinen yhteys on

$$\gamma = 2 - 2H, \quad 1/2 < H < 1 \quad (2.7)$$

Osittainen Brownin liike lähestyy tavallista Brownin liikettä, kun lisäysten välinen korrelaatio heikkenee, eli kun  $\gamma \rightarrow 1$  ja ballistista liikettä ( $x(t) \approx at$ ), kun korrelaatio vahvistuu eli  $\gamma \rightarrow 0$ . Jos  $0 < H < 1/2$ , niin lisäykset ovat antikorreloituneita. Kuvassa 2.3 vertaillaan osittaista Brownin liikettä ja sen lisäyksiä eri Hurstin eksponenteilla. Polku on “sileämpi”, kun lisäysten korrelaatio on vahvempi.



**Kuva 2.3:** Osittaisen Brownin liikkeen lisäykset (vasemmalla) ja vastaavat polut (oikealla) eri Hurstin eksponenteilla. Polkujen visuaalisuuden vertailukelpoisuuden vuoksi ajautumisnopeus on vakioitu, minkä takia eri polkujen lisäysten varianssit vaihtelevat.

Potenssilakia (2.6) noudattavassa aikasarjassa sanotaan olevan pitkän kantaman korrelaatioita. Pitkän kantaman korrelaatioille autokovarianssi heikkenee niin hitaasti, että summa  $\sum_{s=1}^{\infty} C(s)$  hajaantuu. Vastaavasti niin sanotuille lyhyen kantaman korrelaatioille edellinen summa suppenee. Esimerkki lyhyen kantaman korre-

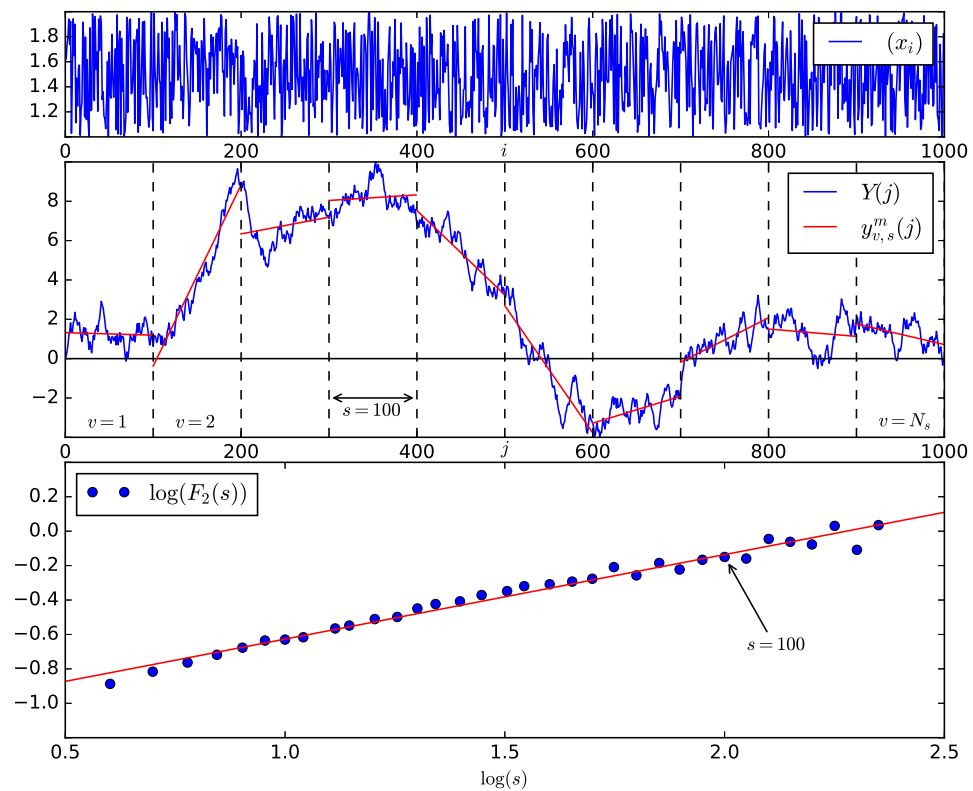
<sup>3</sup>Ajan kuluessa aikasarjan arvo on ajautunut todennäköisemmin kauemmaksi nolasta, joten realisoituneet arvot riippuvat aloitusajankohdasta.

<sup>4</sup>Nyt tarkastelemme diskreettiä aikasarjaa, vaikka aiemmin määrittelimme (osittaisen) Brownin liikkeen jatkuvaksi. Diskreetin (osittaisen) Brownin liikkeen voidaan ajatella olevan näytteistetty tasavälein jatkuvasta.

laatioista on aikasarja, jonka autokovarianssi vaimenee eksponentiaalisesti,  $C(s) \propto \exp(-s/t_x)$ . Autokovarianssien summien suppenemista tai hajaantumista intuitiivisempi ero potenssilain mukaisen ja eksponentiaalisen vaimenemisen välillä on suhde  $C(ns)/C(s)$ . Eksponentiaalisesti vaimenevalle autokovarianssille suhde lähestyy nol-  
laa  $s$ :n kasvaessa ja  $n$ :n pysyessä vakiona, kun taas potenssilain mukaan vaimenevalle suhde on vakio, kun  $n$  on vakio.

### 3. TRENDIT POISTAVA FLUKTUAATIOANALYYSI, DFA

Fraktaalisia ja pitkän kantaman korrelaatioita sisältäviä aikasarjoja on löydetty lukuisista eri järjestelmistä. Laajan suosion näiden analysoinnissa on saavuttanut trendit poistava fluktuatioanalyysi, DFA (*detrended fluctuation analysis*) [3, 5, 6]. DFA-algoritmilla määritetään skaalauseksponentti  $\alpha$ , joka voidaan joissain erityistapauksissa liittää suoraan Hurstin eksponenttiin  $H$ . DFA:ta on sovellettu sykevälivaihtelun lisäksi muun muassa DNA jaksoihin [6], osakekursseihin [7] sekä kvanttisysteemien energiatasoihin [8].



**Kuva 3.1:** DFA-algoritmi tasajakaumasta muodostetulle valkoiselle kohinalle, käyttäen lineaarista trendien poistoa eli  $m = 1$ . Skaalauseksponentiksi saadaan suoran kulmakertoimesta  $\alpha \approx 0.491$  teoreettisen arvon ollessa 0.5.

### 3.1 Algoritmin kulku

Esitetään DFA-algoritmin kulku aikasarjalle  $(x_i)$  viitettä [5] mukailleen. Aluksi aikasarjaa muokataan vähentämällä sarjan keskiarvo  $\langle x \rangle$  jokaisesta alkioista

$$\hat{x}_i = x_i - \langle x \rangle, \quad (3.1)$$

jolloin uuden sarjan  $(\hat{x}_i)$  keskiarvoksi saadaan 0<sup>1</sup>. Seuraavaksi saadusta sarjasta lasketaan kumulatiivinen summa<sup>2</sup>:

$$Y(j) = \sum_{i=1}^j \hat{x}_i, \quad j = 0, 1, 2, \dots, N. \quad (3.2)$$

Kumulatiivinen summa jaetaan  $N_s$ :ään ikkunaan, joiden koko on  $s$ . Kuhunkin ikkunaan sovitetaan  $m$ -asteinen polynomi pienimmän neliösumman menetelmällä, jonka jälkeen lasketaan jälleen uusi sarja kumulatiivisen summan ja polynomisovitteiden erotuksesta:

$$\tilde{Y}(j) = Y(j) - y_{v,s}^m(j). \quad (3.3)$$

Tästä sarjasta lasketaan neliöllinen keskiarvo fluktuaatioista kussakin ikkunassa:

$$F_{\text{DFAm}}^2(v, s) = \frac{1}{s} \sum_{j=1}^s \tilde{Y}^2(j). \quad (3.4)$$

Ikkunoiden neliöllisistä fluktuaatioista lasketaan edelleen keskiarvo, josta lasketaan neliöjuuri. Tätä kutsutaan ikkunakoosta  $s$  riippuvaksi keskimääräiseksi fluktuaatioksi:

$$F_2(s) = \left[ \frac{1}{N_s} \sum_{v=1}^{N_s} F_{\text{DFAm}}^2(v, s) \right]^{1/2}. \quad (3.5)$$

Algoritmi toistetaan usealle ikkunakoolle  $s$ , ja lopulta fluktuaatioista  $F_2$  sekä vastaavista ikkunoista  $s$  lasketaan 10-kantainen logaritmi. Saatuu pistejoukkoon, jota kutsuttakoon fluktuaatiokäyräksi, sovitetaan suora pienimmän neliösumman menetelmällä. Suoran kulmakertoimesta saadaan skaalauseksponentti  $\alpha$ , kun oletetaan, että  $F_2(s) \propto s^\alpha$ . Algoritmin eri vaiheita on havainnollistettu kuvassa 3.1.

<sup>1</sup>Tämä askel ei vaikuta algoritmin lopputulokseen, eikä siten ole välttämätön. Kirjallisuudessa algoritmi kuitenkin useimmiten esitetään näin.

<sup>2</sup>Usein kirjallisuudessa  $j = 1, 2, 3, \dots, N$  yhtälössä (3.2). Silloin algoritmi ei tuota identtistä tulosta käännetylle aikasarjalle  $(x_h)$  missä  $h = N, N-1, N-2, \dots, 1$ . Myös  $j = 1, 2, 3, \dots, N-1$  olisi perusteltu. Tulos olisi käännetylle aikasarjalle identtinen eikä alkuun lisittäisi ylimääräistä alkioita.



### 3.2 Skaalauseksponentti $\alpha$

DFA-algoritmissa alkuperäisen aikasarjan sijaan tutkitaan integroitua sarjaa (3.2). Integroitu sarja jaetaan ikkunoihin, joihin sovitetaan lokaali trendi ja tutkitaan ikkunakohtaisia keskimääräisiä neliöllistä poikkeamia lokaalista trendistä eli ns. neliöllisiä fluktuatiota (3.4). Näyttää intuitiivisesti selvältä, että neliöllinen fluktuatio kasvaa, kun ikkunakoko kasvaa. Näin myös keskimääräinen fluktuatio  $F_2$  (3.5) kasvaa ikkunakoon kasvaessa.

Empiirisin menetelmin (simulaatioiden avulla) on näytetty, että tunnetuille aikasarjoille, joilla on hyvin määritelty Hurstin eksponentti, keskimääräinen fluktuatio (3.5) noudattaa lakia  $F_2(s) \propto s^\alpha$ , ja että skaalauseksponentti  $\alpha$  voidaan liittää Hurstin eksponenttiin  $H$ . Simulaatioiden lisäksi osittaisen Brownin liikkeen lisäyksille on analyttisesti osoitettu, että asympotoottisesti ( $s \rightarrow \infty$ ) pätee

$$\alpha = H, \quad (3.6)$$

kun käytetään lineaarista trendien poistoa eli  $m = 1$ , ja  $H$  on lisäyksiä vastaavan Brownin liikkeen Hurstin eksponentti [9]<sup>3</sup>.

Jos skaalauseksponentti  $\alpha$  voidaan mielekkäästi määrittää eli fluktuatiokäyrää voidaan perustellusti pitää suorana, niin  $\alpha$ :aa voidaan pitää aikasarjan fraktaalisuuden tai korrelaatioiden mittana. Taulukkoon 3.1 on koottu selitys skaalauseksponentin eri arvoille. Taulukosta puuttuvat arvot  $\alpha = 1$  ja  $\alpha = 2$  ovat rajatapauksia. Hurstin eksponentti on määritelmän mukaan välillä  $(0,1)$ , eli edelliset rajatapaukset eivät sisälly tähän. Kun  $H \rightarrow 1$ , niin osittaisen Brownin liikkeen polku lähenee ballistista liikettä. Toisaalta tehospektrin (ks. luku 5.1) kannalta  $\alpha = 1$  on hyvin määritelty ja vastaa niin sanottua vaaleanpunaista kohinaa eli  $1/f$ -kohinaa ja  $\alpha = 2$  integroitua vaaleanpunaista kohinaa.

**Taulukko 3.1:** Skaalauseksponentin  $\alpha$  merkitys osittaisen Brownin liikkeen viitekehysessä. Jos  $\alpha > 1$ , niin aikasarja vastaa osittaisen Brownin liikkeen polku. Jos taas  $\alpha < 1$ , niin aikasarja vastaa polun lisäyksiä. Rajatapaus  $\alpha = 1$  ei ole määritelty osittaiselle Brownin liikkeelle (ks. luku 5.1).

$\alpha$	$H$	selite
$0 < \alpha < 1/2$	–	antikorreloitunut
$\alpha = 1/2$	–	riippumaton (valkoinen kohina)
$1/2 < \alpha < 1$	–	korreloitunut
$\alpha = 1$	–	korreloitunut ( $1/f$ -kohina)
$1 < \alpha < 3/2$	$0 < H < 1/2$	lisäykset antikorreloituneet
$\alpha = 3/2$	$H = 1/2$	lisäykset riippumattomia
$3/2 < \alpha < 2$	$1/2 < H < 1$	lisäykset korreloituneet

<sup>3</sup>Samassa viitteessä [9] tuloksen todetaan pätevän myös FARIMA-mallin (*fractional autoregressive integrated moving average*) mukaisille aikasarjoille, mutta todistus on sivuutettu.

Huomautettakoon, että kumulatiivinen summa eli sarjan integrointi, lisää skaalauseksponentin arvoa yhdellä ja vastaavasti sarjan differentiointi pienentää skaalauseksponentin arvoa yhdellä. Skaalauseksponentin saama maksimiarvo on  $m + 1$ . Täten esimerkiksi kahteen kertaan integroidulle valkoiselle kohinalle ( $\alpha = 2.5$ ) tulisi käyttää vähintään toisen asteen polynomeja trendien poistoon tai sarja olisi differentioitava ensin.

### 3.3 Algoritmin ominaisuuksia

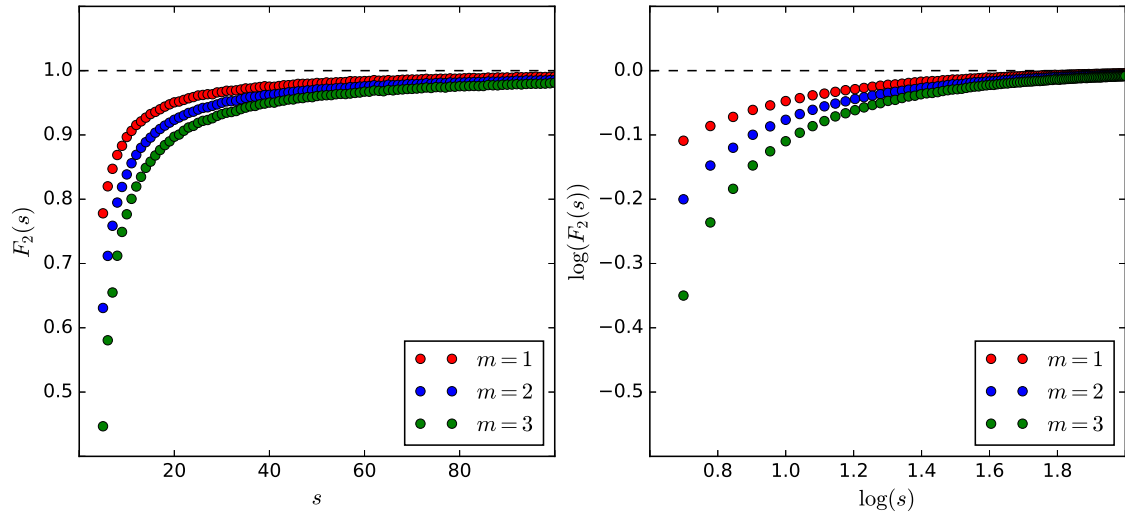
DFA-algoritmin on osoitettu toimivan hyvin ja luotettavasti tunnettuihin testitapauksiin, erityisesti osittaiseen Brownin liikkeeseen [9, 12]. Algoritmilla on kuitenkin joitakin yleisiä heikkouksia, jotka ilmenevät tunnetuissa testitapauksissakin. Lisäksi erilaiset poikkeavuudet aikasarjassa kuten ulkoiset trendit ja muut ei-stationäärisyydet vaikuttavat flukтуаatioon ja hankaloittavat tulosten tulkintaa. Tarkastellaan nyt algoritmin käyttäytymistä eri tilanteissa.

#### 3.3.1 Ikkunakoon vaikutus

Flukтуаatio aliarvioidaan, kun ikkunakoko on pieni. Pienillä  $s:n$  arvoilla sovitettava polynomitrendi ei mukaile vain todellista trendiä, vaan on myös erityisen altis satunnaisille muutoksille. Ilmiötä voidaan havainnollistaa soveltamalla DFA-algoritmia normaalijakaumaa noudattavan valkoisen kohinan lisäykseen (kuva 3.2), jolloin integroitu sarja noudattaa normaalijakaumaa. Integroidussa sarjassa ei ole todellista trendiä, joten trendien poistaminen pienentää ikkunakohtaista neliöllistä flukтуаatiota, koska sovitettava trendi ainoastaan mukailee satunnaisia muutoksia. Ilmiö aiheuttaa kaarevuutta logaritmisessa esityksessä pienillä  $s:n$  arvoilla ja on voimakkaampi mitä pienempi ikkunakoko on ja mitä suurempi on sovitettavan polynomien aste. Toisaalta todellinen lokaali trendi kuitenkin heikentää ilmiön vaikutusta eikä se siten tavallisesti ole niin voimakas kuin esimerkkikuvassa. Kun  $m = 1$  ja  $\alpha > 1$ , niin ilmiön vaikutus on jo varsin vähäinen.

Flukтуаation arvio on tilastollisesti epävarmempi, kun ikkunakoko on suuri. Suuremmilla ikkunakoilla  $s$  sovitettavaa trendiä hallitsee integroidun sarjan suuremman mittakaavan muoto ja yksittäisten arvojen satunnaisten heilahtelujen merkitys on vähäinen. Toisinaan sovitettava polynomi, esimerkiksi suora, soveltuu paremmin arvioimaan tietyn ikkunan trendiä, toisinaan taas huonommin. Eri ikkunoiden neliöllisten flukтуаatioiden vaihtelu saattaa olla suurta, kun taas toisaalta ikkunoiden määrä vähenee. Lasketun flukтуаation  $F_2$  ja laskettuihin flukтуаatioihin sovitettavan suoran erotuksien varianssi kasvaa ikkunakoon kasvaessa [9], eli flukтуаation tilastollinen epävarmuus kasvaa ikkunakoon mukana.

Edellisten seikkojen nojalla DFA-algoritmi soveltuu tässä mielessä parhaiten ns.



**Kuva 3.2:** DFA-algoritmin fluktuaatio tavallisella ja logaritmisella asteikolla, kun integroitu sarja noudattaa normaalijakaumaa  $\mathcal{N}(0,1)$  ja  $m$  on sovitettavan polynomin aste. Nyt fluktuaation ei pitäisi riippua ikkunakoosta, mutta trendien poistaminen vääristää tulosta.

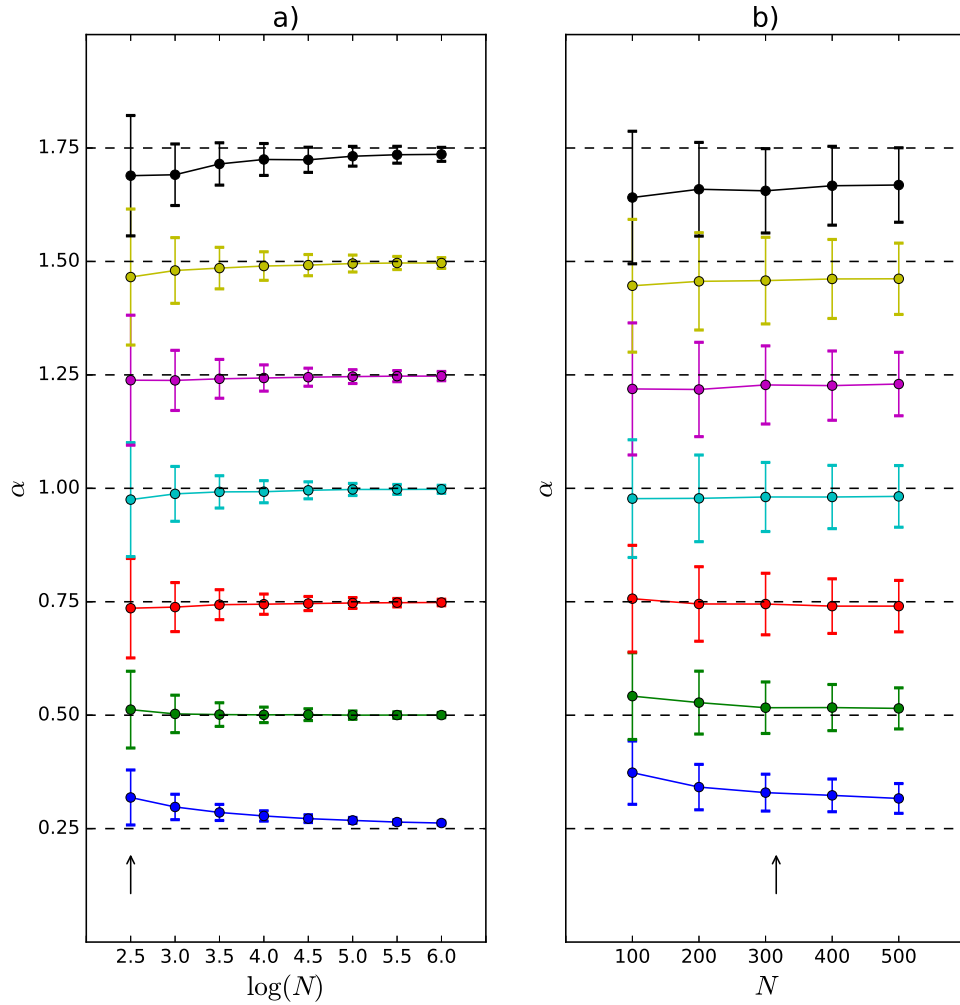
keskikokoisille ikkunoille. Pienten ikkunoiden vääristymän minimoimiseksi on myös syytä välttää korkean asteen polynomeja trendien poistossa, jos niiden käyttöön ei ole muuta erityistä syytä.

### 3.3.2 Aikasarjan pituus

DFA:n on teoreettisesti näytetty antavan oikean tuloksen, kun aikasarjan pituus lähestyy ääretöntä (ainakin osittaisen Brownin liikkeen tapauksessa) [9]. Tämä tulos ei vielä takaa, että DFA käyttäytyy hyvin äärellisten aikasarjojen kanssa. Tutkitaan nyt, miten aikasarjan pituus vaikuttaa arvioituun skaalauseksponenttiin  $\alpha$  ja sen vaihteluun generoimalla Fourier-suodatuksella [5] testiaikasarjoja, joiden  $\alpha$  on tunnettu<sup>4</sup>.

Kuvaan 3.3 (a) liittyvässä laskennassa on käytetty lineaarista trendien poistoa  $m = 1$ , ikkunakokoja  $s = 10, \dots, N/10$  ikkunakokojen välin ollessa logaritmisella asteikolla 0.1 Kuvaan 3.3 (b) liittyvä laskenta poikkeaa ikkunakokojen puolesta; siinä  $s = 4, \dots, N/4$  ja ikkunoita on tasavälein logaritmisella asteikolla 50 (vähemmän jos peräkkäiset logaritmiset ikkunakoot pyörivät samaan kokonaislukuun). Kuvissa (a) ja (b) on kullekin eri  $\alpha$ :lle ja aikasarjan pituudelle  $N$  käytetty otoskoko 1000. Laskennasta saadut numeeriset arvot löytyvät taulukosta 3.2.

<sup>4</sup>Aikasarjojen generointimenetelmää, Fourier-suodatusta, ei ole tutkittu tässä. On mahdollista, että testiaikasarjojen tavoiteltu  $\alpha$  ei vastaa täsmälleen lopputulosta, vaan siinä voi olla myös vaihtelua ja poikkeamaa. Tuloksia tarkastellessa on hyvä pitää mielessä, että nyt DFA-tuloksia verrataan nimenomaan Fourier-suodatuksella generoituihin testiaikasarjoihin. Fourier-suodatukselta käytetään yksinkertaisinta mallia eikä esimerkiksi Besselin funktioilla paranneltua mallia [5].



**Kuva 3.3:** Skaalauseksponttien  $\alpha$  keskiarvot ja standardipoikkeamat generoidulle aineistolle. Otokoko on 1000 jokaiselle  $\alpha$ :n ja  $N$ :n arvolle. Katkoviivat ovat teoreettisia arvoja. Kuvassa (a) on käytössä logaritminen asteikko ja aikasarjojen pituudet ovat noin kolmestasadasta miljoonaan ja käytetyt ikkunakoot ovat  $s = 10, \dots, N/10$ . Kuvassa (b) ikkunakoot ovat  $s = 4, \dots, N/4$ . Nuolet kuvissa osoittavat samaa aikasarjan pituutta  $N$ . Numeeriset arvot löytyvät taulukosta 3.2.

Tarkastellaan kuvaa 3.3 (a). Odotetusti  $\alpha$ :n arvion tarkkuus kasvaa ja vaihtelu pienenee aikasarjan pituuden kasvaessa. Antikorreloituneille ja riippumattomille aikasarjoille ( $\alpha = 0.25, 0.50$ )  $\alpha$  tyypillisesti yliarvioidaan, kun taas muille aliarvioidaan. Suurimmat  $\alpha$ :n keskiarvon poikkeamat löytyvät käytetyistä ääripäistä  $\alpha = 0.25, 1.75$ . Vaihtelu puolestaan kasvaa melko säännönmukaisesti  $\alpha$ :n kasvaessa lukuun ottamatta suurinta arvoa  $\alpha = 1.75$ , jonka vaihtelu on pienempää kuin edellisen arvon  $\alpha = 1.50$ . Vaihtelu on esitetty standardipoikkeamien avulla<sup>5</sup>. Lyhyillä aikasarjoilla (joitain satoja havaintoja) arvioidun  $\alpha$ :n tulee olla selvästi

<sup>5</sup>Vaihtelun oletetaan olevan normaalijakautunut. On hyvä pitää mielessä, että esimerkiksi kahden standardipoikkeaman verran keskiarvosta eroava havainto ei ole tavattoman harvainen.

yli tai alle 0.50, jos halutaan esimerkiksi päätellä onko aikasarja korreloitunut vai antikorreloitunut, koska virherajat ovat melko suuret.

**Taulukko 3.2:** Numeeriset arvot kuvasta 3.3.  $N$  on aikasarjan pituus,  $\alpha$  skaalauseksponentti, jonka mukaan aikasarjat on generoitu,  $\mu(\alpha)$  arvioidun  $\alpha$ :n keskiarvo ja  $\sigma(\alpha)$  arvioidun  $\alpha$ :n standardipoikkeama.

$\alpha$	$N$	$\log(N)$	$\mu(\alpha)$	$\sigma(\alpha)$
0.25	316	2.50	0.319	0.0606
0.25	1000	3.00	0.298	0.0280
0.25	3162	3.50	0.286	0.0177
0.25	10000	4.00	0.278	0.0114
0.25	31622	4.50	0.272	0.00856
0.25	100000	5.0	0.268	0.00644
0.25	316227	5.50	0.264	0.00520
0.25	1000000	6.00	0.262	0.00399
0.50	316	2.50	0.512	0.0846
0.50	1000	3.00	0.503	0.0413
0.50	3162	3.50	0.501	0.0260
0.50	10000	4.00	0.501	0.0171
0.50	31622	4.50	0.501	0.0130
0.50	100000	5.00	0.500	0.00920
0.50	316227	5.50	0.500	0.00723
0.50	1000000	6.00	0.500	0.00576
0.75	316	2.50	0.736	0.110
0.75	1000	3.00	0.738	0.0540
0.75	3162	3.50	0.744	0.0330
0.75	10000	4.00	0.745	0.0223
0.75	31622	4.50	0.746	0.0156
0.75	100000	5.00	0.747	0.0120
0.75	316227	5.50	0.748	0.00932
0.75	1000000	6.00	0.748	0.00722
1.00	316	2.50	0.975	0.126
1.00	1000	3.00	0.988	0.0605
1.00	3162	3.50	0.992	0.0356
1.00	10000	4.00	0.993	0.0244
1.00	31622	4.50	0.996	0.0187
1.00	100000	5.00	0.997	0.0133
1.00	316227	5.50	0.998	0.0109
1.00	1000000	6.00	0.998	0.00889
1.25	316	2.50	1.238	0.143
1.25	1000	3.00	1.238	0.0663
1.25	3162	3.50	1.241	0.0428
1.25	10000	4.00	1.243	0.0289
1.25	31622	4.50	1.245	0.0199
1.25	100000	5.00	1.246	0.0152
1.25	316227	5.50	1.247	0.0123
1.25	1000000	6.00	1.247	0.0102
1.50	316	2.50	1.466	0.150
1.50	1000	3.00	1.480	0.0724
1.50	3162	3.50	1.485	0.0458
1.50	10000	4.00	1.490	0.0315
1.50	31622	4.50	1.492	0.0234
1.50	100000	5.00	1.495	0.0187
1.50	316227	5.50	1.497	0.0145
1.50	1000000	6.00	1.497	0.0119
1.75	316	2.50	1.689	0.133
1.75	1000	3.00	1.691	0.0678
1.75	3162	3.50	1.715	0.0466
1.75	10000	4.00	1.725	0.0352
1.75	31622	4.50	1.724	0.0277
1.75	100000	5.00	1.732	0.0218
1.75	316227	5.50	1.735	0.0186
1.75	1000000	6.00	1.736	0.0154

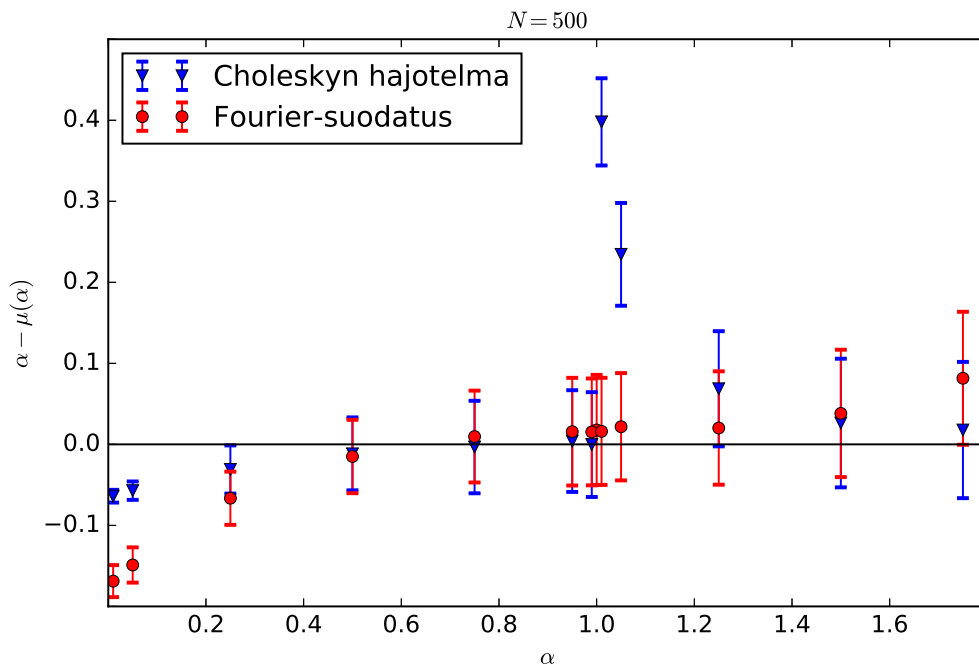
$\alpha$	$N$	$\log(N)$	$\mu(\alpha)$	$\sigma(\alpha)$
0.25	100	2.00	0.373	0.0698
0.25	200	2.30	0.342	0.0501
0.25	300	2.48	0.329	0.0406
0.25	400	2.60	0.323	0.0360
0.25	500	2.70	0.316	0.0328
0.5	100	2.00	0.542	0.0956
0.5	200	2.30	0.528	0.0693
0.5	300	2.48	0.516	0.0568
0.5	400	2.60	0.517	0.0509
0.5	500	2.70	0.515	0.0453
0.75	100	2.0	0.757	0.118
0.75	200	2.30	0.745	0.0821
0.75	300	2.48	0.745	0.0678
0.75	400	2.60	0.740	0.0602
0.75	500	2.70	0.740	0.0567
1.0	100	2.00	0.977	0.130
1.0	200	2.30	0.978	0.0955
1.0	300	2.48	0.981	0.0761
1.0	400	2.60	0.981	0.0697
1.0	500	2.70	0.982	0.0680
1.25	100	2.00	1.219	0.146
1.25	200	2.30	1.218	0.104
1.25	300	2.48	1.228	0.0861
1.25	400	2.60	1.226	0.0764
1.25	500	2.70	1.230	0.0700
1.5	100	2.00	1.446	0.146
1.5	200	2.30	1.456	0.107
1.5	300	2.48	1.458	0.0955
1.5	400	2.60	1.461	0.0871
1.5	500	2.70	1.462	0.0786
1.75	100	2.00	1.641	0.146
1.75	200	2.30	1.659	0.103
1.75	300	2.48	1.656	0.0929
1.75	400	2.60	1.667	0.0869
1.75	500	2.70	1.668	0.0821

Kun  $\alpha$ :aa arvioidaan, niin kuvassa 3.3 (a) sovitteessa käytettävien pisteiden (ikkunakokojen) määrä kasvaa, kun aikasarjan pituus kasvaa. Oletettavasti  $\alpha$ :n poikkeamat riippuvat jossain määrin pisteiden (ikkunakokojen) määrästä. Kuvassa 3.3 (b) käytetään suurta vakiomäärää ikkunakokoja lyhyille aikasarjoille. Tämä vähentää vaihtelua jonkin verran, mutta yleisesti tulokset ovat samankaltaisia kuin kuvassa (a). Skaalauseksponentin vaihtelu lyhyillä aikasarjoilla kuvassa (a) ei siis riipu merkittävästi pienestä ikkunakokojen määrästä.

Jos aikasarjan pituuden vaikutusta DFA:n luotettavuuteen haluaa tutkia kattavammin, niin huomioitavia tekijöitä ovat mm. ikkunakoot (määrä, välit), sovitusalue, trendien poiston aste, otoksen koko, sekä aikasarjojen generoinnissa käytettävä algoritmi. Huomautettakoon kuitenkin, että taulukon tulokset ovat varsin yhteneviä viitteen [9] kanssa, jossa on tutkittu  $\alpha$ :n ( $H$ :n) arvion tarkkuutta 10000:n alkion aikasarjoille eri menetelmillä.

### 3.3.3 Aikasarjan generointimenetelmän vaikutus

Aiemmin näimme, että DFA:lla arvioidut skaalauseksponentit saattavat poiketa generoitujen aikasarjojen tavoitelluista arvoista. Tämä herättää kysymyksen poikkeaman alkuperästä: johtuuko se DFA:sta vai aikasarjan generointimenetelmästä? Kysymystä voi olla vaikeaa selvittää aukottomasti, mutta tarkastellaan kuitenkin Fourier-suodatuksen lisäksi lyhyesti myös toista generointimenetelmää.



**Kuva 3.4:** Vaaka-akselilla on aikasarjan generointimenetelmällä tavoiteltu skaalauseksponentti ja pystyakselilla tavoitellun skaalauseksponentin ja DFA:lla arvioidun skaalauseksponentin keskiarvon erotus. Otoskoko on 1000 ja standardipoikkeamat on merkitty palkeilla.

**Taulukko 3.3:** Numeeriset arvot kuvasta 3.4 (ilman erotusta). Skaalauseksponenttia, jonka mukaan aikasarjat on generoitu, on merkitty  $\alpha$ :lla,  $\mu(\alpha)$  on arvioidun  $\alpha$ :n keskiarvo ja  $\sigma(\alpha)$  arvioidun  $\alpha$ :n standardipoikkeama. Aikasarjojen pituudet olivat 500 alkia ja otoskoko on 1000. Osa taulukon aineistosta löytyy myös taulukosta 3.2.

$\alpha$	Fourier		Cholesky	
	$\mu(\alpha)$	$\sigma(\alpha)$	$\mu(\alpha)$	$\sigma(\alpha)$
0.01	0.179	0.020	0.074	0.008
0.05	0.199	0.022	0.107	0.011
0.25	0.317	0.033	0.281	0.030
0.50	0.515	0.045	0.512	0.045
0.75	0.740	0.057	0.753	0.057
0.95	0.934	0.066	0.946	0.063
0.99	0.975	0.066	0.990	0.065
1.00	0.982	0.068	–	–
1.01	0.994	0.066	0.612	0.054
1.05	1.028	0.066	0.816	0.063
1.25	1.230	0.070	1.181	0.071
1.50	1.462	0.079	1.474	0.079
1.75	1.668	0.082	1.732	0.084

Osittaista Brownin liikettä voidaan simuloida eksaktisti Choleskyn hajotelmaan perustuvalla menetelmällä [10], jota kutsuttakoon nyt Choleskyn menetelmäksi. Menetelmän käyttöä rajoittaa sen laskennallinen raskaus [kompleksisuus  $\mathcal{O}(n^3)$ ], joten tarkastellaan vain lyhyitä 500 alkion aikasarjoja. Kuvassa 3.4 vertaillaan Choleskyn menetelmää ja Fourier-suodatusta suhteessa  $\alpha$ :n arvioon. Laskennassa on käytetty ikkunakokoja  $s = 4, \dots, N/4$  ja ikkunoita on tasavälein logaritmisella asteikolla 50 kuten kuvassa 3.3 (b). Otoskoko oli jälleen 1000. Laskennasta saadut numeeriset arvot löytyvät taulukosta 3.3.

Kuvasta havaitaan välittömästi, että tulokset riippuvat aikasarjan generointimenetelmästä. Choleskyn menetelmän tulokset poikkeavat eniten Fourier-suodatuksen tuloksista, kun tavoiteltu  $\alpha$  on lähellä nollaa tai hieman yli yksi. Lähellä nollaa Choleskyn menetelmällä generoitujen aikasarjojen  $\alpha$  yliarvioidaan vähemmän ja standardipoikkeama kutistuu enemmän. Toisaalta, kun tavoiteltu  $\alpha$  on hieman yli yksi, niin Choleskyn menetelmällä generoitujen aikasarjojen  $\alpha$  aliarvioidaan rajusti. Tämä on mielenkiintoinen tulos, koska esimerkiksi Choleskyn menetelmällä generoitu  $\alpha = 0.01$  vastaa integroinnin jälkeen Choleskyn menetelmällä generoitua arvoa  $\alpha = 1.01$ . Kuitenkin ennen integrointia  $\alpha$ :n arvio on merkittävästi parempi<sup>6</sup>. Huomautettakoon kuitenkin, että kun  $\alpha$  on jotain muuta kuin hieman yli yksi tai lähellä nollaa, niin standardipoikkeamat eivät juurikaan riipu generointimenetelmästä.

Tällä melko suppealla tarkastelulla olemme osoittaneet, että esimerkiksi arvioi-

<sup>6</sup>Aikasarjan integroinnin pitäisi lisätä  $\alpha$ :n arvoa yhdellä. Tämä osoittaa, että ainakin lyhyestä aikasarjasta arvioitu  $\alpha$  saattaa poiketa tästä säännöstä merkittävästi.

taessa DFA:n luotettavuutta, valittu aikasarjojen generointimenetelmä vaikuttaa tuloksiin. Valitettavasti tästä ei voida päätellä mikä generointimenetelmä olisi parempi luotettavuustarkasteluihin ja missä mielessä. Aikaisemmin luvussa 3.3.2 käytimme Fourier-suodatusta, jotta pystyimme generoimaan pitkiäkin aikasarjoja nopeasti.

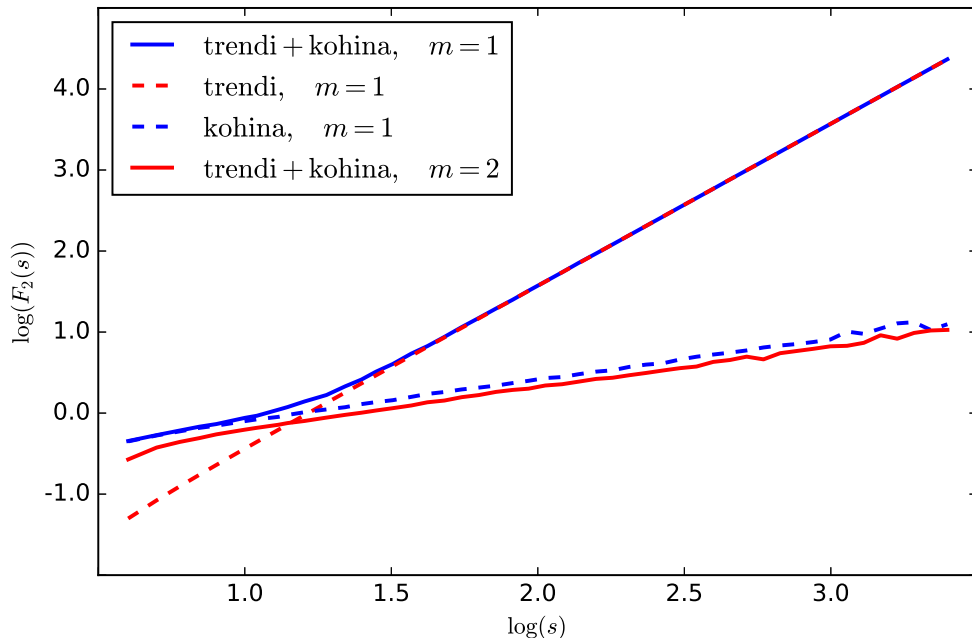
### 3.3.4 Ulkoiset trendit

Ulkoisilla trendeillä tarkoitetaan prosessin sisäisestä dynamiikasta riippumatonta trendiä, joka voi olla esimerkiksi lineaarinen tai periodinen. DFA-algoritmin trendien poisto eliminoi ulkoisten trendien vaikutuksen vain joissakin erityistapauksissa. Tarkastellaan milloin näin käy ja mikä on ulkoisten trendien vaikutus muulloin.

Jos muodostetaan superpositio kahdesta riippumattomasta aikasarjasta siten, että  $x_i = x_i^f + x_i^g$ , niin voidaan osoittaa että flukтуаation neliö on likimain superpositio erillisten aikasarjojen flukтуаatioiden neliöistä [13, 15]:

$$[F_2]^2 = \frac{1}{s} \left[ \sum_{j=1}^s [\tilde{Y}_f(j)]^2 + \sum_{j=1}^s [\tilde{Y}_g(j)]^2 + \sum_{j=1}^s \tilde{Y}_f(j)\tilde{Y}_g(j) \right] \approx [F_2^f]^2 + [F_2^g]^2. \quad (3.7)$$

Likiarvossa yhtäsuuruuden oikean puolen viimeinen termi voidaan jättää pois, koska se on lähellä nollaa  $\tilde{Y}_f$ :n ja  $\tilde{Y}_g$ :n riippumattomuuden nojalla.



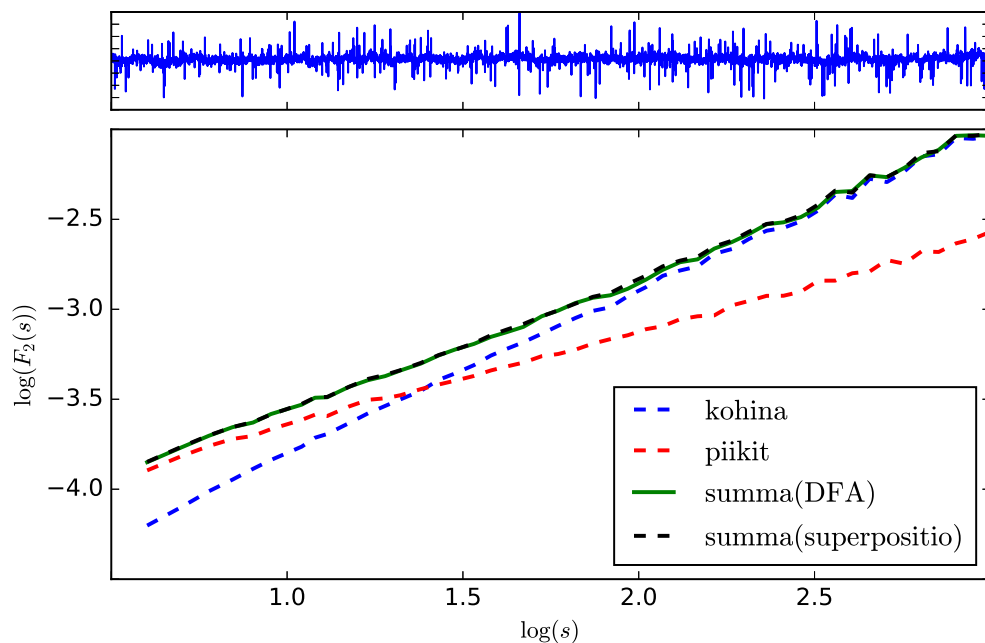
**Kuva 3.5:** Lineaarisen trendin vaikutus DFA-algoritmiin. Trendien poisto ei kumoa lineaarista trendiä valkoisessa kohinassa, kun  $m = 1$ . Jos  $m = 2$ , niin yhtälössä (3.7) trendin flukтуаatio on nolla, joten sen vaikutus kumoutuu.



Olkoon  $F_2^f$  varsinaisen aikasarjan flukтуаatio ja  $F_2^g$  siihen liittyvän trendin flukтуаatio. Yhtälöstä (3.7) nähdään, että jos trendin flukтуаatio on nolla, sillä ei ole vaikutusta kokonaisflukтуаatioon. Näin käy, jos integroidun sarjan trendi on samaa muotoa kuin trendien poistossa käytettävä polynomi. Integrointi lisää polynomitrendin astetta yhdellä, joten trendien poistossa käytettävä  $m$ -asteinen polynomi kumoaa korkeintaan  $m - 1$ -asteisten polynomitrendien vaikutuksen. Tätä on havainnollistettu kuvassa 3.5. Jos trendin vaikutus ei kumoudu, se tyypillisesti aiheuttaa siirtymiä (*crossovers*) logaritmiiseen flukтуаatiokäyrään. Siirtymällä tarkoitetaan flukтуаatiokäyrän poikkeamaa suorasta siten, että jonkin aikaskaalan jälkeen käyrä siirtyy noudattamaan toista suoraa. Viitteessä [15] on tutkittu erilaisten trendien vaikutusta kattavammin.

### 3.3.5 Muut ei-stationäärisyydet

Trendien poiston ansiosta DFA-algoritmi hallitsee Brownin liikkeen ajautumisen kaltaisen ei-stationäärisyyden, joka aiheutuu prosessin sisäisestä dynamiikasta. Ulkoiset trendit käsiteltiin yllä, mutta tarkastellaan nyt muunkaltaisia ei-stationäärisyyksiä. Viitteessä [13] on tutkittu kolmea yleistä ei-stationäärisyyden muotoa: aikasarjasta puuttuvia lohkoja, satunnaisia piikkejä sekä aikasarjan vaihtelevaa lokaalia käyttäytymistä.



**Kuva 3.6:** Aikasarja koostuu osittaisen Brownin liikkeen lisäyksistä (kohina,  $H = 0.9$ ) ja riippumattomista normaalijakautuneista piikeistä (5%,  $H = 0.5$ ). Piikit aiheuttavat siirtymän logaritmiiseen flukтуаatiokäyrään. Superpositiosäännöllä erillisistä aikasarjoista saadaan käytännössä sama tulos, kuin DFA-algoritmillä summasarjasta.

Jos satunnaiset piikit ovat aikasarjasta muutoin riippumattomia, voidaan niihin soveltaa superpositiosääntöä (3.7). Piikkien vaikutusta on havainnollistettu kuvassa 3.6. Vaihtelevalla lokaalilla käytöksellä tarkoitetaan, että esimerkiksi osalle sarjasta  $\alpha = 0.6$  ja osalle  $\alpha = 0.9$ . Täten fluktuaatio on summa näiden osien fluktuaatioista, josta poikkeamaa aiheuttavat ainoastaan ikkunat, joihin sisältyy molempia osia. Puuttuvilla lohkoilla puolestaan tarkoitetaan, että osia aikasarjasta on poistettu ja jäljellä olevat osat on ns. liimattu yhteen. On havaittu, että puuttuvat lohkot eivät juuri vaikuta lopputulokseen, jos aikasarja on korreloitunut ( $\alpha > 0.5$ ), kun taas antikorreloituneille sarjoille ( $\alpha < 0.5$ ) vaikutus on merkittävä.

### 3.3.6 Lyhyen kantaman korrelaatiot

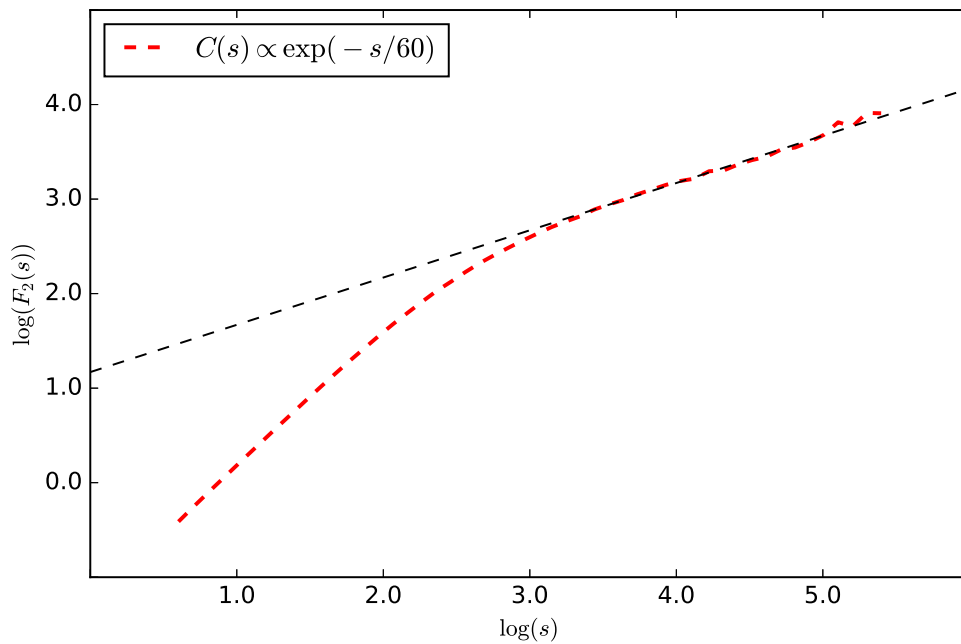
Lyhyen kantaman korrelaatiot ovat tavallinen siirtymän aiheuttaja stationäärisen aikasarjan fluktuaatiokäyrässä. Autokovarianssin käyttäytyminen poikkeaa lyhyellä aikaskaalalla asympotoottisesta käyttäytymisestä. Kenties yleisin esimerkki lyhyen kantaman korrelaatioista on autoregressiivinen prosessi. Autoregressiiviselle prosessille pätee

$$\Delta x_i = c\Delta x_{i-1} + \epsilon_i. \quad (3.8)$$

Autoregressiivisen prosessin autokovarianssi vaimenee eksponentiaalisesti  $C(s) \propto \exp(-s/t_x) = c$ , missä  $t_x$  on karakteristinen vaimenemisaika ( $C(s)$  vaimenee aina  $1/e \approx 0.37$  osaan itsestään ajassa  $t_x$ ).

Kuvassa 3.7 esitetään autoregressiivisen prosessin fluktuaatiokäyrä. Käyrä näyttää aikasarjan olevan korreloitunut lyhyemmillä aikaskaaloilla ( $\alpha > 0.5$ ), mutta suuremmilla aikaskaaloilla tapahtuu siirtymä, jonka jälkeen fluktuaatiokäyrä noudattaa kulmakertoimen 0.5 suoraa, mikä viittaa korreloitumattomaan käytökseen. Siirtymän paikka fluktuaatiokäyrässä ei ole triviaalisti pääteltävissä autokovarianssista; kuvasta katsoen siirtymän voidaan katsoa tapahtuneen noin ikkunakoon  $s = 1000$  jälkeen. Autokovarianssi on jo lähellä nollaa huomattavasti pienemmillä etäisyyksillä ajassa. Siirtymän paikkaa, kun autokovarianssin käytös muuttuu, on tutkittu viitteessä [14] ja siihen vaikuttaa myös trendien poistossa käytetyn polynomien aste.

Lyhyen kantaman korrelaatioiden mahdollisuus on erityisen kiusallinen, jos tutkitaan lyhyitä aikasarjoja. Vaikka pitkän kantaman korrelaation skaalauseksponentti voidaankin arvioida suhteellisen lyhyestä aikasarjasta, niin jos korrelaation ei etukäteen tiedetä olevan pitkän kantaman, sitä ei välttämättä voi erottaa lyhyen kantaman korrelaatiosta. Huomautettakoon lopuksi, että siirtymiä voi aiheuttaa myös muunlainen vaihtelu autokovarianssin käytöksessä kuin lyhyen kantaman korrelaatiot, esimerkiksi siirtyminen pitkän kantaman korrelaatioista toiseen pitkän kantaman korrelaatioon (eksponentin muutos potenssilaisissa tietyn etäisyyden jälkeen ajassa).



**Kuva 3.7:** Autoregressiivisen prosessin fluktuaatiokäyrä (punainen katkoviiva), kun  $t_x = 60$  ja käytetään lineaarista trendien poistoa  $m = 1$ . Mustalla katkoviivalla on merkitty suoraa, jonka kulmakerroin on 0.5.

### 3.3.7 Mahdolliset valesiirtymät

Tähän mennessä olemme nähneet, että yleensä ei-stationäärisyydet ja poikkeamat pitkän kantaman korrelaatiosta aiheuttavat siirtymiä fluktuaatiokäyrään. Entä onko mahdollista, että fluktuaatiokäyrän siirtymä johtuisi satunnaisesta vaihtelusta, eikä todellisesta aikasarjan piirteestä? Tutkitaan tätä jälleen generoimalla testiaikasarjoja Fourier-suodatuksella. Yleisimmät sovituksen laadun mitat eivät sovellu siirtymien arviointiin, joten käytetään kahta sovitusaluetta ja verrataan saatujen skaalauseksponenttien erotuksia.

Taulukossa 3.4 esitetään tuhannen aikasarjan otoksesta lasketut lyhyen ja pitkän skaalan skaalauseksponenttien erotuksien keskiarvot ja standardipoikkeamat. Kullekin eri  $\alpha$ :lle käytetään aikasarjan pituutta  $N = 1000$ . Lyhyen kantaman skaalauseksponentille  $\alpha_s$  käytetään ikkunakokoja  $s = 10, \dots, 1000$  ja pitkän kantaman skaalauseksponentille ikkunakokoja  $s = 1000, \dots, 100000 = N/10$ . Peräkkäisten ikkunakokojen väli logaritmisella asteikolla on 0.1. Erotukset ja standardipoikkeamat ovat pieniä, joten tulosten perusteella ei ole syytä olettaa, että fluktuaatiokäyrään syntyisi siirtymiä sattumalta.

**Taulukko 3.4:** Arvioitujen lyhyen ja pitkän skaalan skaalauseksponenttien erotuksien keskiarvot  $\mu(\alpha_s - \alpha_l)$  ja standardipoikkeamat  $\sigma(\alpha_s - \alpha_l)$  Fourier-suodatuksella generoiduille aikasarjoille. Otokoko on 1000.

$\alpha$	$\mu(\alpha_s - \alpha_l)$	$\sigma(\alpha_s - \alpha_l)$
0.25	0.027	0.011
0.50	0.0042	0.018
0.75	-0.00020	0.022
1.00	0.0024	0.025
1.25	0.0046	0.030
1.50	0.010	0.032
1.75	0.034	0.034

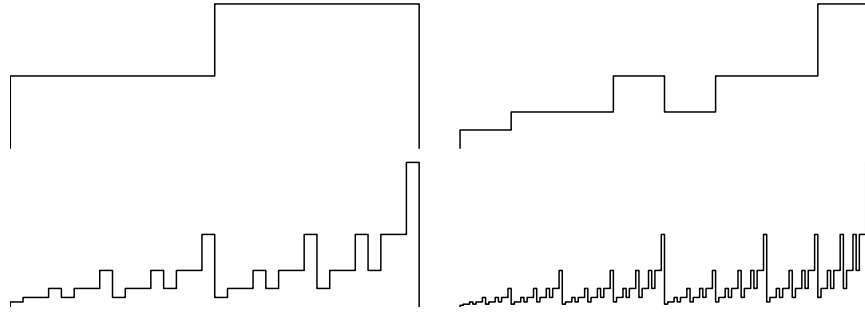
### 3.3.8 Tuntemattomat aikasarjat

Monesti todellisten sovelluskohteiden kanssa aikasarja saattaa olla täysin tuntematon ja osittain epätäydellinen. Mahdollisten trendien luonne ei ole aina selvä, ei ole selvää muuttuuko systeemin sisäinen dynamiikka ajan kuluessa, ja lisäksi osa mittauksista saattaa olla huonolaatuisia tai puuttuvia. Vaikka skaalauseksponentin  $\alpha$  voi aina laskea logaritmiin fluktuatioihin sovitetusta suorasta, on tuntemattoman aikasarjan kanssa yritettävä huolellisesti varmistua tämän luvun mielekkyydestä. Onko suora todella suora ja kertooko se jotain koko systeemin dynamiikasta, vai onko se pikemminkin likimain suora ja tulosta useamman ilmiön yhteisvaikutuksesta? Mielekkyydestä varmistuminen täysin tuntemattoman aikasarjan suhteen on haastavaa, koska algoritmin ominaisuuksia on yleensä tutkittu hyvin käyttäytyville testitapauksille. Ei ole varmaa, onko vaikkapa ei-stationäärisyyksien vaikutus riippumaton ns. isäntäsarjasta.

Lisäksi tulkinnan kanssa on edelleen noudatettava varovaisuutta, vaikka näyttäisi että skaalauseksponentti todella voidaan määrittää mielekkäästi. Jos  $\alpha > 1$ , niin voidaanko aikasarjan sanoa olevan fraktaalinen sillä perusteella, että osittainen Brownin liike on fraktaalinen ja sille  $\alpha > 1$ ? Näihin kysymyksiin palataan edellä, kun DFA-algoritmia sovelletaan sykevälivaihteluun.

## 3.4 Multifraktaalisuus ja MF-DFA

Aikaisemmin on tarkasteltu niin sanottua monofraktaalisuutta ja monofraktaalisia aikasarjoja. Monofraktaalien skaalautumisominaisuudet voidaan määrittää yhdellä skaalauseksponentilla. Multifraktaalien tai multifraktaalisien aikasarjan skaalautumisominaisuuksien määrittämiseen tarvitaan useampia, mahdollisesti ääretön määrä skaalauseksponentteja. Tässä tutkielmassa keskitytään enimmäkseen monofraktaalisuuteen, mutta esitellään seuraavaksi lyhyesti myös DFA:n laajennus multifraktaaleihin.



**Kuva 3.8:** Binomisen vesiputouksen neljä ensimmäistä iteraatiota.

Intuitiivisesti voidaan ajatella, että multifraktaali tai multifraktaalinen aikasarja sisältää ikään kuin sisäkkäin useita monofraktaaleita. Kun kuviota tai sarjaa suurennetaan, niin riippuen suurennoksessa käytetystä skaalauseksponentista eri piirteet pysyvät samaan tapaan muuttumattomina, kuin kokonaisuus monofraktaalissa. Kenties yksinkertaisin multifraktaali on binominen vesiputous (binomial cascade) [5]. Binominen vesiputous muodostetaan jakamalla aluksi janalle tasaisesti jakautunut yksikkömassa uudelleen,  $0 < a < 1$  massasta ensimmäiselle puolikkaalle ja  $1 - a$  toiselle puolikkaalle. Prosessia toistetaan rekursiivisesti muodostuneille osille, joille massa on nyt jakautunut tasaisesti. Prosessin neljä ensimmäistä iteraatiota on esitetty kuvassa 3.8.

DFA:n laajennus multifraktaaleihin on yksinkertainen. Yhtälö (3.5) muuttuu muotoon

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [F_{\text{DFAm}}^2(v,s)]^{q/2} \right\}^{1/q}, \quad q \neq 0, \quad (3.9)$$

missä  $F_q$  on  $q$ :sta riippuva flukтуаatio. Tavallisessa DFA:ssa  $q = 2$ . Kun  $q = 0$ , niin

$$F_0(s) = \exp \left\{ \frac{1}{2N_s} \sum_{v=1}^{N_s} \ln [F_{\text{DFAm}}^2(v,s)] \right\}, \quad q = 0. \quad (3.10)$$

Vastaavasti kuin tavallisessa DFA:ssa, flukтуаatiokäyrään sovitetaan suora, josta saadaan  $q$ :sta riippuva skaalauseksponentti  $h(q)$ .

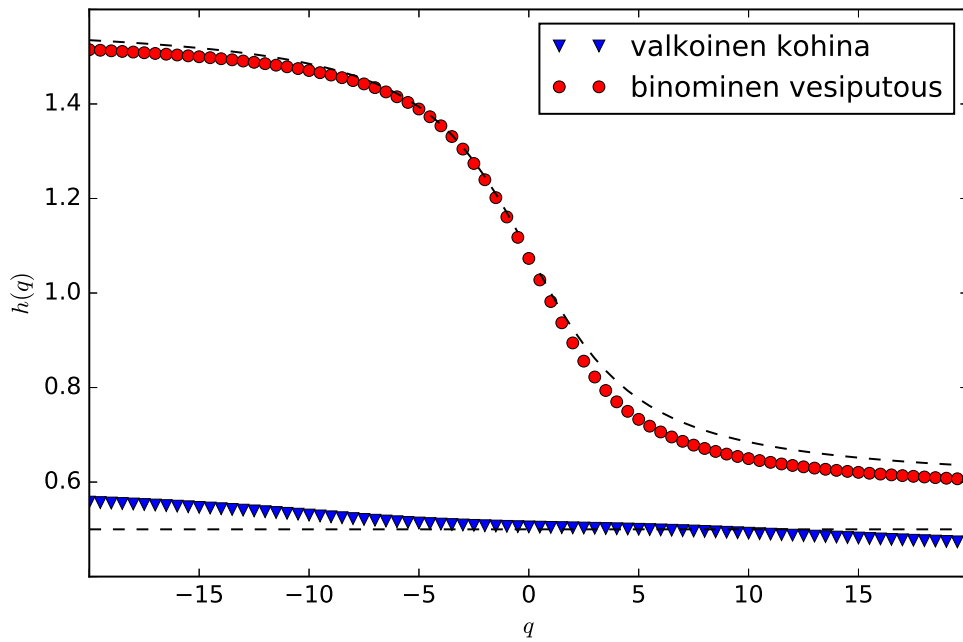
Miksi skaalauseksponentti riippuisi  $q$ :sta? Tarkastellaan seuraavaksi ikkunakohtaisia neliöllisiä flukтуаatioita  $F_{\text{DFAm}}^2$ . Jos  $q$  on positiivinen, ikkunat joiden neliöllinen flukтуаatio on suuri, dominoivat keskiarvoistamisessa. Toisaalta taas jos  $q$  on negatiivinen, niin ikkunat, joiden neliöllinen flukтуаatio on pieni, dominoivat keskiarvoistamisessa. Kun  $q < 0$ ,  $h(q)$  kuvaa pienten flukтуаatioiden skaalautumista ja kun  $q > 0$ ,  $h(q)$  kuvaa suurten flukтуаatioiden skaalautumista. Binomiselle vesipu-

toukselle pätee [5]

$$h(q) = \frac{1}{q} - \frac{\ln(a^q + b^q)}{q \ln 2} + \frac{\ln(a + b)}{\ln 2}, \quad 0 < a < 1, \quad b = 1 - a. \quad (3.11)$$

Monofraktaalille puolestaan ikkunakohtaiset neliölliset flukтуаatiot ovat samaa suuruusluokkaa (samalle ikkunakoolle), joten  $h(q)$  on vakio.

Kuvassa 3.9 esitetään  $h(q)$  valkoiselle kohinalle ja binomiselle vesiputoukselle. Binomisen vesiputouksen multifraktaalisuus havaitaan selvästi, mutta monofraktaalinkaan  $h(q)$  ei pysy aivan vakiona pitkällekin aikasarjalle (pituus  $N = 2^{20}$ ), joten multifraktaalisuuden toteamisessa on oltava varovainen. Myöskään ei-stationäärisyyksien vaikutusta ei ole selvitetty, joten on oletettava, että ne saattavat aiheuttaa näennäistä multifraktaalisuutta. Lisäksi mainittakoon, että aikasarjan leveä arvojakouma (ei äärellistä varianssia) aiheuttaa multifraktaalisuutta. Tämä voidaan erottaa suurten ja pienten flukтуаatioiden erilaisesta skaalautumisesta sekoittamalla aikasarjan arvot. Jos sekoitus ei poista multifraktaalisuutta, se on peräisin leveästä jakaumasta.



**Kuva 3.9:** MF-DFA:lla laskettu  $h(q)$  valkoiselle kohinalle ja binomiselle vesiputoukselle. Teoreettiset arvot on merkitty katkoviivoin.

## 4. DFA JA SYKEVÄLIANALYYSI

Sykeväleillä tarkoitetaan peräkkäisten sydämenlyöntien välistä etäisyyttä ajassa. Sykevälit voidaan mitata niin sanottujen R-piikkien välisenä etäisyytenä [17]. Täten RR-väliä käytetään synonyyminä sykevälille. Sykevälien vaihtelu ei ole korreloitumaton kohinaa, vaan sen on havaittu sisältävän pitkän kantaman korrelaatioita. Sykevälivaihtelua ohjaa autonominen hermosto, ja sykevälianalyysin avulla saadaan epäsuorasti tietoa autonomisen hermoston tilasta [18]. Tässä tutkielmassa ei kuitenkaan paneuduta sykevälivaihtelun fysiologiaan, vaan tutkitaan mitä sykevälivaihteluaikasarjoista (RR-sarjoista)<sup>1</sup> voidaan saada irti DFA:n avulla. Fysiologiseen puoleen voi perehtyä esimerkiksi oppikirjojen kuten viitteen [17], kautta. Viite [19] on suomenkielinen johdatus sykevälianalyysin ja sen menetelmiin, jossa avataan myös fysiologista puolta jonkin verran.

Tässä luvussa tehdään aluksi lyhyt katsaus sykevälianalyysin taustaan DFA:n kontekstissa. Tämän jälkeen tarkastellaan yksittäistä RR-sarjaa DFA:n avulla. Lopuksi palataan vielä erityisesti aiempien tutkimusten tulkintaan RR-sarjasta havaittujen seikkojen valossa.

### 4.1 Taustaa

Alunperin DFA esiteltiin artikkelissa, jossa tutkittiin pitkän kantaman korrelaatioita nukleotidien järjestymisessä DNA-jaksoissa [6]. Pian tämän jälkeen menetelmää sovellettiin myös sykevälianalyysiin [20]. Tässä tutkimuksessa (johon on lähes 3000 viittausta) havaittiin, että sykevälivaihtelu on lähellä  $1/f$ -kohinaa (mitä oli raportoitu aiemmin muiden menetelmien avulla), ja että se sisältää usein selvän siirtymän eli flukтуаatiokäyrän kulmakertoimen muutoksen lyhyehköllä aikaskaalalla. Siirtymän tulkittiin tarkoittavan korrelaatioiden erilaista käyttäytymistä eri aikaskaaloilla (mikä onkin mahdollinen tulkinta, ks. luku 3.3.6). Tämän vuoksi ehdotettiin kahden skaalausekspONENTIN käyttöä. Toinen skaalausekspONENTTI lyhyille kantamille ja toinen hieman pidemmille. Tutkimuksessa käytettiin ikkunakokoja  $s = 4, \dots, 16$  ja  $s = 16, \dots, 64$  ja RR-sarjat olivat 24:n tunnin pituisia. Lyhyehkön kantaman

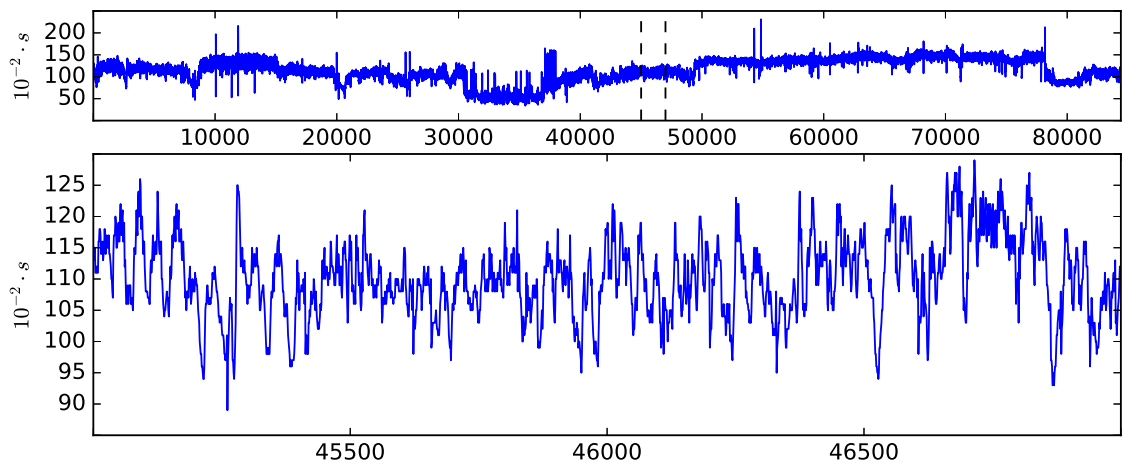
<sup>1</sup>Tarkalleen ottaen RR-sarjat eivät ole aikasarjoja. Ajan sijasta muuttujana on sydämenlyöntien juokseva lukumäärä ja sarjan arvoina on sydämenlyöntien väliset ajat. Aikasarja-analyysin työkalujen kannalta tällä ei kuitenkaan ole merkitystä; kyseessä on diskreetti sarja, jonka havainnot ovat tasavälein.

eksponentit ovat houkuttelevia, koska niitä varten ei tarvita välttämättä pitkiä mittauksia. Tutkimuksessa löydettiin jonkinlaista tilastollista eroavaisuutta terveiden ja sairaiden koehenkilöiden skaalauseksponenttien välillä.

Sittemmin DFA:ta on sovellettu vastaavalla tavalla monenlaisiin tapauksiin, joissa vertaillaan jollain tavalla luokiteltujen ryhmien tuloksia tilastollisen päättelyn avulla. Muutamia esimerkkejä, joissa tutkitaan iän, sydämenpysähdyksestä selviämisen, uniapnean ja raskauden vaikutuksia sykevälivaihteluun löytyy viitteistä [21, 22, 23, 24]. Usein näissä tutkimuksissa löydetäänkin tilastollisia eroavaisuuksia eri ryhmien välillä. Yksityiskohdat analyysissa vaihtelevat jonkin verran. Toisinaan mittaukset ovat lyhyitä ja kontrolloituja, toisinaan pidempiä ja kontrolloimattomampia. Kahden skaalauseksponentin käyttö on vakiintunut, mutta niiden sovitusväleille ei ole standardeja.

Näiden vakiintuneella tavalla DFA:ta sykevälianalyysiin soveltavien tutkimuksien lisäksi löytyy joukko, joissa käytetään jotain DFA:n laajennusta (esimerkiksi MF-DFA:ta) tai analysoidaan flukтуаatiokäyriä toisin. Esimerkiksi viitteessä [25] RR-sarjoihin käytetään MF-DFA:ta ja liukuvaa sovitusväliä flukтуаatiokäyriin ja näin lasketaan ns. flukтуаatiopinta. Viitteessä [26] puolestaan arvioidaan flukтуаatiokäyrän derivaattaa ja käytetään koneoppimista tulosten luokitteluun. Näillä menetelmillä toivotaan saavan jotain lisäinformaatiota tavalliseen DFA:han verrattuna.

Usein tutkimuksissa myös yritetään tulkita ja mahdollisesti jopa ennustaa, tuloksia biologisin argumentein. Tulkinnoissa ja ennustuksissa on kuitenkin keskeisiä heikkouksia, jotka kumpuavat pääasiassa aineiston laadusta suhteessa DFA:n ominaisuuksiin. Tarkastellaan nyt yksittäistä RR-sarjaa ja tehdään sille muutamia testejä, jonka jälkeen palataan vielä aiempiin tutkimuksiin ja niiden tulkintaan.



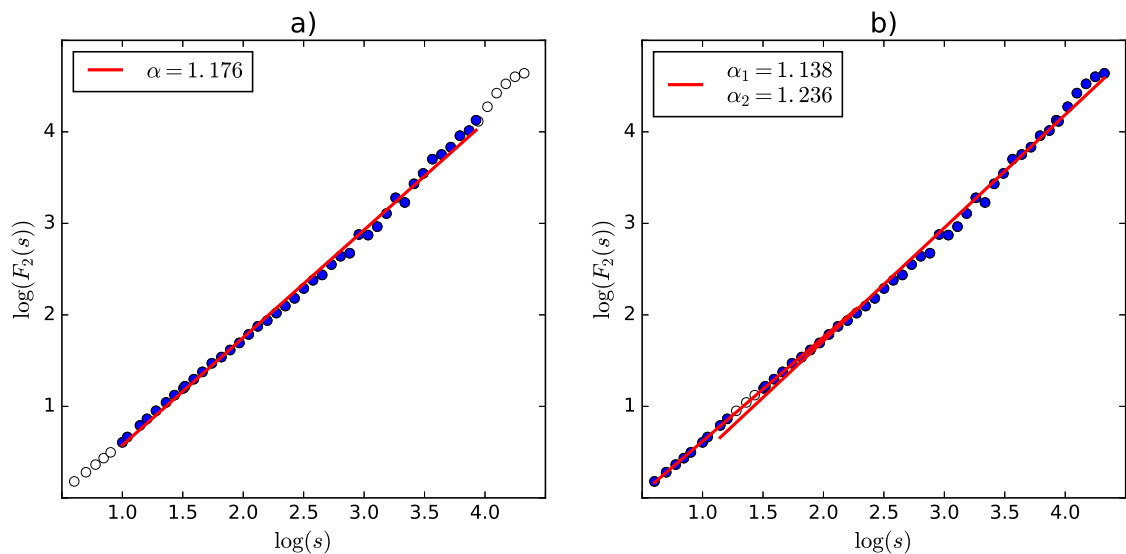
**Kuva 4.1:** Noin vuorokauden pituinen terveen henkilön RR-sarja (viitteestä [27]). Alemmassa kuvassa on suurennos katkoviivoin merkitystä kohdasta.



## 4.2 Pitkän ja lyhyen kantaman skaalausekspONENTIT

Kuvassa 4.1 on noin vuorokauden pituinen terveen henkilön RR-sarja. Pelkällä silmämääräisellä tarkastelulla näyttää välittömästi selvältä, että sarja ei voi olla stationäärinen. Ennen kaikkea keskisyke vaihtelee, mihin vaikuttavat erilaiset aktiviteetit. Esimerkiksi levossa syke on matala, kun taas urheillessa korkea<sup>2</sup>. Keskisyke määräytyy aktiviteetin ja yksilöllisten ominaisuuksien kautta. Keskisykkeen vaihtelun lisäksi RR-sarjassa on jaksoja, joissa vaihtelun absoluuttinen suuruus poikkeaa selvästi toisistaan.

On periaatteessa mahdollista, että keskisykkeen vaihtelut ja vaihtelun absoluuttinen suuruus eivät juuri vaikuta sykeväli vaihtelun korrelaatio-ominaisuuksiin. Missään nimessä tätä ei kuitenkaan voida olettaa, vaan pikemminkin päinvastoin. On oletettava, että korrelaatio-ominaisuudet saattavat muuttua muiden ominaisuuksien mukana tai jopa niistä riippumatta. Lisäksi toisistaan poikkeavien osuuksien ei-stationäärisyydet eivät välttämättä ole riippumattomia muista ominaisuuksista [vrt. yhtälö (3.7)]. Tämän pohdinnan perusteella on odotettavissa, että DFA:n fluktuatiokäyrä sisältää siirtymiä, joiden alkuperä on epäselvä. Näistä seikoista huolimatta tarkastellaan DFA:n tuloksia, kun sovelletaan sitä yksinkertaisesti koko sarjaan.



**Kuva 4.2:** Kuvan 4.1 RR-sarjan fluktuatiokäyrä. Kuvassa (a) on arvioitu sovitteesta yksi skaalausekspONENTTI ja kuvassa (b) kaksi, pitkän ja lyhyen kantaman skaalausekspONENTTI.

Kuvassa 4.2 (a) on määritetty skaalausekspONENTTI  $\alpha$  koko kuvan 4.1 aikasarjaa käyttäen. Silmämääräisesti sovite näyttää pätevältä. Lisätarkastelu kuitenkin paljastaa, että yksi skaalausekspONENTTI on vain approksimaatio. Kuten totesimme, sykeväli aikasarjojen on havaittu usein sisältävän siirtymän pienehköillä ikkunakoilla,

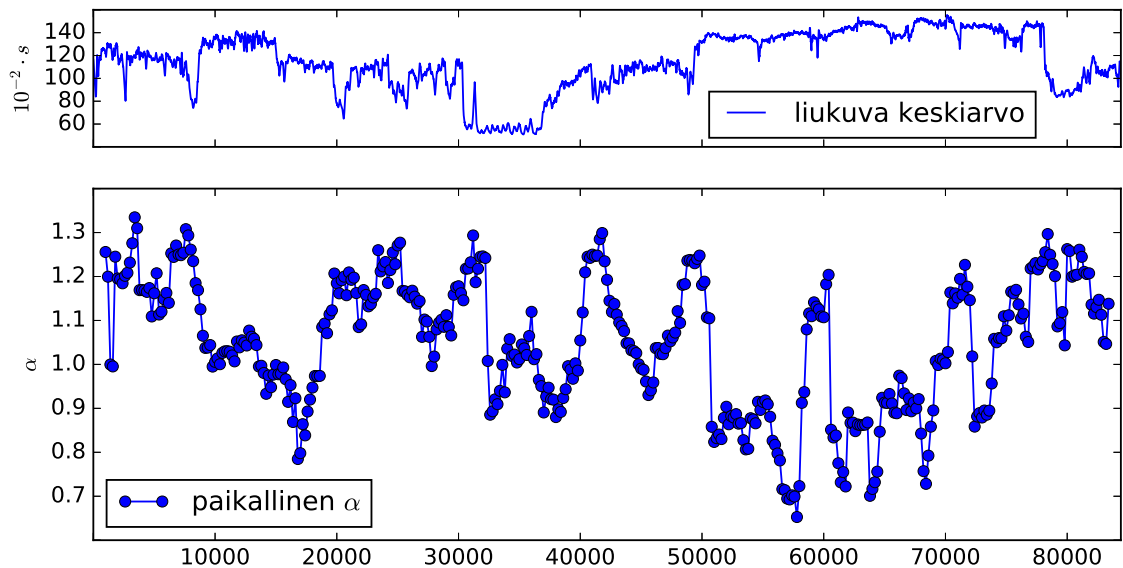
<sup>2</sup>Tavallisesti sykettä mitataan lyönteinä minuutissa, joten matala syke tarkoittaa pitkiä sykevälejä ja korkea lyhyitä.

joten arvioidaan kaksi skaalauseksponenttia, lyhyen kantaman  $\alpha_1$  ja pitkän kantaman  $\alpha_2$ , jotka on määritetty kuvassa 4.2. Ikkunakokojen täsmällinen valinta vaihtelee tapauskohtaisesti, ja nyt ilman ennakkotarkastelua on valittu  $s_{min} = 4$  ja  $s_{max} = 16$   $\alpha_1$ :lle ja  $s_{min} = 32$  ja  $s_{max} = N/4$   $\alpha_2$ :lle. Skaalauseksponenttien erotus on noin 0.1, mikä taulukon 3.4 perusteella viittaa todelliseen siirtymään. Lisäksi kuvasta (b) nähdään, että erotus kasvaisi suuremmaksi, jos ikkunakoot valittaisiin toisin.

Analysoidun RR-sarjan flukтуаatiokäyrä kaikkiaan vaikuttaa siltä, mitä monesti on raportoitu: pitkällä aikaskaalalla lähellä  $1/f$ -kohinaa ( $\alpha = 1$ ) ja lisäksi löydetään jonkinlainen siirtymä lyhyehköllä aikaskaalalla. Siirtymä kuitenkin voi johtua lukuisista muistakin syistä kuin eri aikaskaaloilla toisistaan poikkeavista korrelaatioominaisuuksista. Tarkastellaan seuraavaksi RR-sarjaa jaksoittain.

### 4.3 Paikalliset skaalauseksponentit

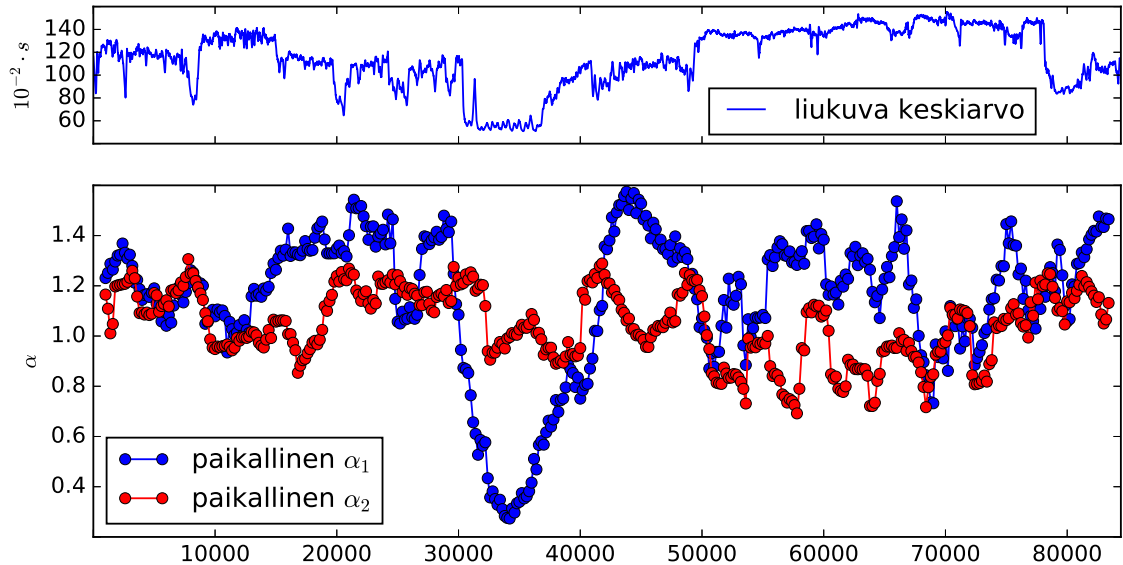
Kuvan 4.1 RR-sarjasta voisi yrittää erottaa stationäärisiä osia ja tarkastella näiden osuuksien flukтуаatiokäyriä. Tämä olisi kuitenkin mutkikasta ja stationäärisyydestä olisi vaikea varmistua. Jaetaan sarja siksi yksinkertaisesti yhtäsuuriin osiin, tarkastellaan niitä ja katsotaan, mitä tulokset ovat.



**Kuva 4.3:** Paikallinen  $\alpha$  jaksoille, joiden pituus on 2000 RR-väliä. Peräkkäisillä jaksoilla on 90% päällekkäisyyttä. Laskennassa käytettiin ikkunakokoja  $s = 10, \dots, N/10 = 200$ . Alemman kuvan vaaka-akseli on välin keskipiste. Ylemmässä kuvassa on RR-sarjan 100:n RR-välin liukuva keskiarvo.

Kuvassa 4.3 RR-sarja on jaettu 2000:n RR-välin jaksoihin ja kullekin jaksolle on laskettu paikallinen  $\alpha$ . Odotetusti  $\alpha$  ei pysy vakiona, mutta vaihtelu on yllättävän suurta:  $\alpha$  minimin ja maksimin erotus on jopa yli 0.5. Taulukon 3.2 arvoista interpoloimalla paikallisen  $\alpha$ :n standardipoikkeaman voisi ennustaa olevan luokkaa

0.05, jos kaikki jaksot olisivat samasta stationäärisestä aikasarjasta. Nyt standardipoikkeama on lähes kolminkertainen. Lisäksi  $\alpha$ :n jakauma ei myöskään noudata normaalijakaumaa, vaan yleisimmät arvot näyttävät olevan arvovälin yläpäässä.



**Kuva 4.4:** Lyhyen ( $\alpha_1$ ) ja pitkän ( $\alpha_2$ ) kantaman skaalauseksponentit 2000:n RR-välin jaksoille. Peräkkäisillä jaksoilla on 90% päällekkäisyyttä. Laskennassa käytettiin  $\alpha_1$ :lle ikkunakokoja  $s = 4, \dots, 16$  ja  $\alpha_2$ :lle ikkunakokoja  $s = 32, \dots, N/4 = 500$ . Alemman kuvan vaaka-akseli on välin keskipiste. Ylemmässä kuvassa on jälleen RR-sarjan 100:n RR-välin liukuva keskiarvo.

Kuvassa 4.4 on puolestaan tarkasteltu pitkän ja lyhyen kantaman skaalauseksponentteja vastaavasti kuin yhtä skaalauseksponenttia kuvassa 4.3. Nämä skaalauseksponentit toisinaan eroavat samassa jaksossa merkittävästi toisistaan, eli fluktuaatiokäyrässä on joissain jaksoissa huomattava siirtymä. Kumpikaan skaalauseksponentti ei ole systemaattisesti toista suurempi, eli siirtymän “muoto” vaihtelee. Lisäksi erotuksien itseisarvo on suurimmillaan jopa yli 0.7, kun  $\alpha_1$  on noin 0.3 ja  $\alpha_2$  noin 1.0. Naiivin tulkinnan mukaan tämä tarkoittaisi, että tällöin lyhyellä aikaskaalalla arvot olisivat antikorreloituneita ja pitkällä aikaskaalalla korreloituneita.

Seuraavaksi tehdään muutamia huomioita näistä yksinkertaisista testeistä. Nyt on viimeistään selvää, että tarkasteltava RR-sarja ei ole stationäärinen. Fluktuaatiokäyrän muoto vaihtelee huomattavasti, kun tarkastellaan lyhyempiä jaksoja. Toisaalta lyhyemmätkään jaksot eivät varmasti ole ainakaan kaikki stationäärisiä; osa on ehkä enemmän kuin toiset<sup>3</sup>, mutta tämän perusteella ei voida sanoa, mitä jaksoja voitaisiin pitää stationäärisinä, saati miten voitaisiin tuottaa tai löytää toinen vertailukelpoinen jakso. Tämä on keskeinen ongelma. Palataan nyt muihin tutkimuksiin

<sup>3</sup>Sykevälien tapauksessa muuttumattoman ja tasaisen toiminnan (esimerkiksi levon tai tasaisen kävelyn) voisi ajatella tuottavan siinä mielessä stationäärisen sarjan, että jos prosessi “jäädyytetäisiin” tähän tilaan, niin sarja todella olisi stationäärinen sen varsinaisessa merkityksessä. Muuttuvan toiminnan (esimerkiksi kiihtyvä fyysinen suoritus tai unen eri vaiheet) voisi puolestaan ajatella tuottavan ei-stationäärisen, tai “vähemmän stationäärisen” sarjan.

ja tulosten tulkintaan.

#### 4.4 Tuloksista ja niiden tulkinnasta

Tähän mennessä olemme nähneet, että ei-stationäärisyydet vaikuttavat yleisesti DFA-algoritmin lopputulokseen. Toisaalta RR-sarjat ovat usein tuntemattomalla tavalla ei-stationäärisiä, ja skaalausekspONENTIT näyttävät vaihtelevan runsaasti ns. tavallisissakin olosuhteissa. Näiden seikkojen vuoksi tulosten tulkinta saattaa olla ongelmallista: melkein mikä tahansa eroavaisuus skaalausekspONENTEISSA eriteltyjen ryhmien välillä voidaan hypoteettisesti selittää sopiviin oletuksiin vetoamalla. Näitä hypoteeseja on kuitenkin vaikeahko todentaa, ja toisinaan tulkinnat vaikuttavat jopa jälkikäteen esitetyltä narraatiolta, joka on sovitettu tuloksiin. Tästä syystä tulkinnan sijaan on mieluummin keskityttävä empiirisiin seikkoihin kuten tilastollisella päättelyllä todennettuihin havaintoihin.

Vaikuttaa siltä, että metodologisesti tämä tutkimusala ei ole edistynyt merkittävästi reilun kahdenkymmenen vuoden kuluessa. Erilaisten ja aktiviteettien ja fysiologisten tilojen on kyllä näytetty vaikuttavan sykevälivaihteluun ja sitä kautta flukтуаatiokäyriin, mutta kuitenkin täsmällistä vaikutusta ei osata ennustaa tai usein edes selittää jälkeenpäin. Hienostuneempien menetelmien (MF-DFA jne.) merkitys jää täten melko vähäiseksi. Saadulla lisäinformaatiolla voidaan mahdollisesti tuoda lisäarvoa esimerkiksi tilastolliseen päättelyyn täsmällisesti määritellyssä koetapauksessa, mutta tästä ei kuitenkaan ole apua keskeisempiin ongelmiin.

Yhteenvetona voitaisiin sanoa, että keskeiset ongelmat kumpuavat aineiston luonteesta, joka rajoittaa analyysin lähinnä empiirisiin havaintoihin. Toisaalta tämän lähestymistavan haasteena on aineistojen vertailukelpoisuus. SkaalausekspONENTIT vaihtelevat tuntemattomalla tavalla olosuhteiden mukaan. Tästä syystä kahden tyyppisen aineiston voisi ajatella olevan vertailukelpoista: lyhyiden ja tiukasti kontrolloitujen mittausten tai kontrolloimattomien ja hyvin pitkien mittausten (mieluiten päiviä). Lyhyet ja kontrolloidut mittaukset olisivat vertailukelpoisia siitä ilmeisestä syystä, että olosuhteet olisivat mahdollisimman vakioituja. Pitkät vapaamuotoiset taas siksi, että vuorokauden voi ajatella olevan yksi jakso, joka sisältää keskimäärin tietyt määrät eri aktiviteetteja. Mitä useampia jaksoja (pidempi mittausta), sitä vähemmän flukтуаatiokäyrän voisi ajatella muuttuvan mittausta pidentämällä. Näissäkin vaihtoehdoissa on omat haasteensa. Kummankaan luokan mittausten hankkiminen ei välttämättä ole yksinkertaista. Lyhyet ja kontrolloidut sarjat ovat potentiaalisesti virhealttiita. Pitkien sarjojen vertailukelpoisuus taas on jossain määrin kyseenalainen.

Luvussa 6 tutkimme pidempiä ja vähäisesti kontrolloituja mittauksia ja selvitämme jälleen, pystymmekö DFA:n ja tilastollisen päättelyn avulla erottamaan eri ryhmiä toisistaan. Lisäksi vahvistamme tässä luvussa tekemiämme hypoteeseja skaa-

lauseksponenttien vaihtelusta, koska tarkastelimme toistaiseksi vain yksittäistä RR-sarjaa.

## 5. MUITA MENETELMIÄ SYKEVÄLIANALYYSISSÄ

Toistaiseksi olemme keskittyneet vain yhteen aikasarja-analyysin menetelmään, trendit poistavaan fluktuatioanalyysiin (DFA). Tämä antaa varsin suppean kuvan kokonaisuudesta, joten tehdään nyt lyhyt katsaus muihin sykeväliallyysissä ja pitkän kantaman korrelaatioiden kanssa käytettyihin aikasarja-analyysin työkaluihin. Katsaus ei ole kattava, vaan sen tarkoituksena on tuoda lukijan tietoisuuteen valikoituja muita näkökulmia ja menetelmiä.

### 5.1 Tehospektri

Eräs perinteinen tapa pitkän kantaman korrelaatioiden analysointiin on tehospektrin muoto [5]. Aikasarjan  $(x_i)$  tehospektri  $S(f)$ , missä  $f$  on taajuus, saadaan aikasarjan  $(\hat{x}_i)$  (3.1) Fourier-muunnoksesta. Potenssilain mukaan pitkän kantaman korreloituneelle aikasarjalle tehospektri noudattaa verrannollisuutta

$$S(f) \propto f^{-\beta}, \quad \beta = 1 - \gamma, \quad (5.1)$$

missä  $\gamma$  saadaan yhtälöstä (2.5) [5]. Tämä relaatio pätee vain stationääriselle aikasarjalle, ei esimerkiksi Brownin liikkeelle (tai jos aikasarja sisältää muita ei-stationäärisyyksiä). Toisaalta Brownin liikkeen lisäykset ovat stationäärisiä, joten niitä voidaan analysoida tehospektrin avulla. Huomautettakoon, että autokovarianssin yhteys spektriexponenttiin  $\beta$  yhdistää myös spektriexponentin DFA:n skaalauseksponenttiin,

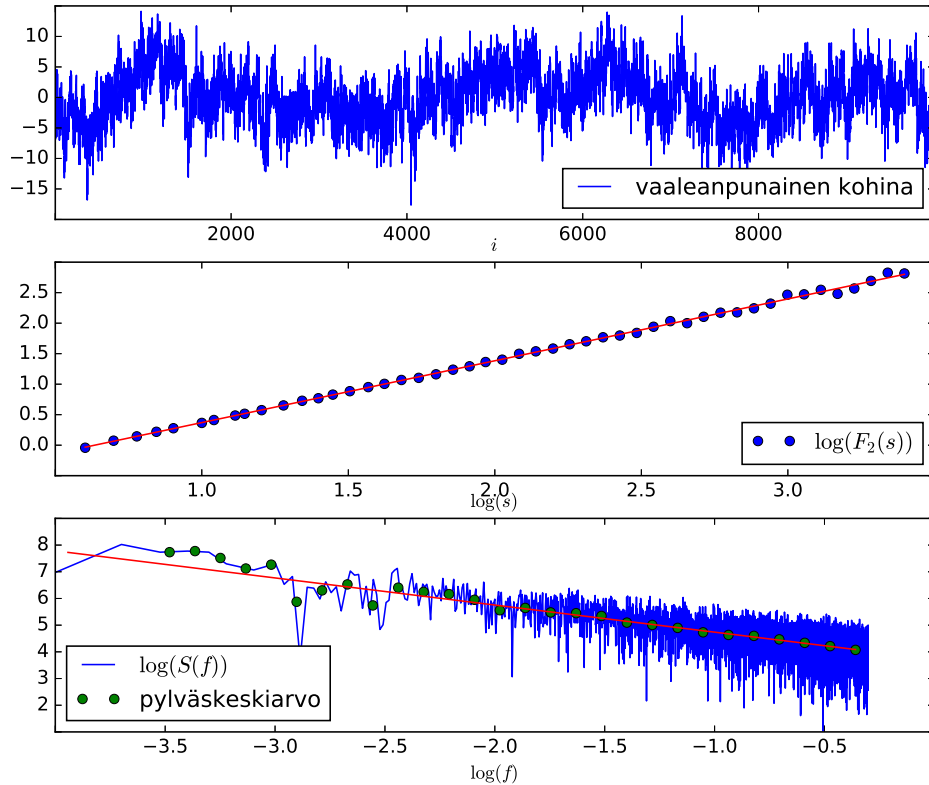
$$\beta = 2\alpha - 1, \quad (5.2)$$

stationääristen aikasarjojen viitekehysessä.

Spektriexponentti  $\beta$  voidaan määrittää sovittamalla suora tehospektriin logaritmisella asteikolla. Pylvässtatistiikan (*binned stastics*) käyttö parantaa  $\beta$ :n arviota; tehospektristä logaritmisella asteikolla lasketaan keskiarvo vakiolevyisistä kaistoista (pylväistä), joihin suoran sovitus tehdään [5]. Kuvassa 5.1 on analysoitu vaaleanpunaista kohinaa DFA:lla ja tehospektrin avulla. Molemmat menetelmät paljastavat pitkän kantaman korrelaation. Spektriexponentin arvosta  $\beta = -1.0$  selviää myös

syy vaaleanpunaisen kohina toiselle nimelle,  $1/f$ -kohinalle<sup>1</sup>.

Tehospektrin käyttöä rajoittaa se, että sitä ei pidetä erityisen soveltuvana ei-stationäärisille aikasarjoille [18], vaan esimerkiksi DFA:ta pidetään usein parempana vaihtoehtona – joskin olemme nähneet, että sekään ei ole läheskään ongelmaton.



**Kuva 5.1:** DFA:lla vaaleanpunaiselle kohinalle arvioidaan  $\alpha = 1.013$  ja tehospektristä puolestaan  $\beta = -1.016$ . Teoreettiset arvot ovat 1.0 ja  $-1.0$ .

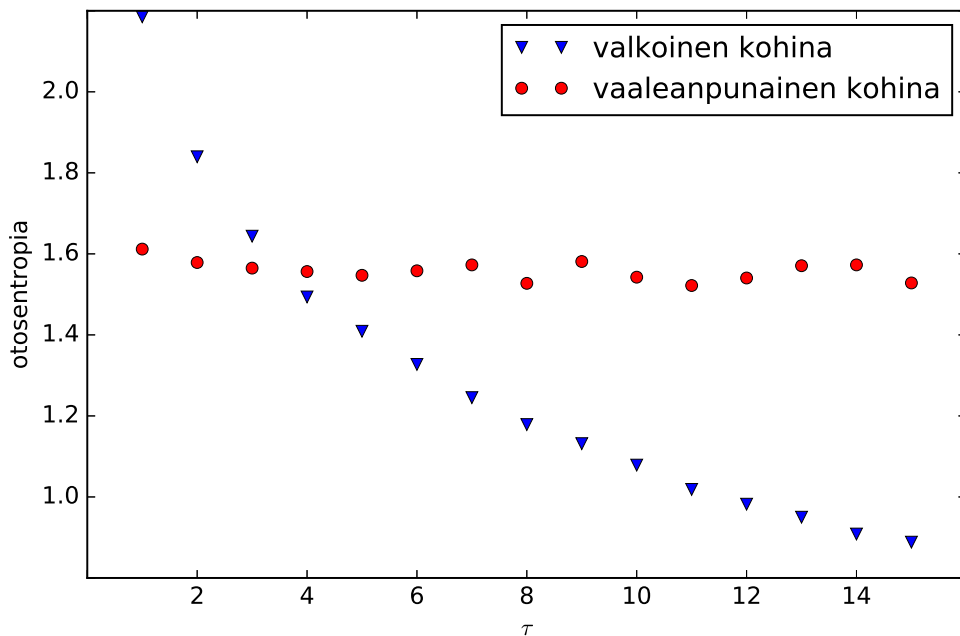
## 5.2 Entropiat

Termodynamiikassa entropia on systeemin epäjärjestyksen määrää kuvaava suure. Konsepti voidaan yleistää koskemaan informaation määrää yleisemmissä todennäköisyysjakaumissa [3, s. 217] ja siten soveltaa aikasarjoihin. Esimerkiksi korreloitunut aikasarja sisältää enemmän informaatiota kuin riippumaton valkoinen kohina. Korreloituneen aikasarjan menneisyyttä tarkastelemalla voidaan saada enemmän informaatiota tulevasta kuin valkoisen kohinan menneisyyttä tarkastelemalla.

Erilaisia entropiasuureita on useita. Sykevälianalyysiin ehdotettiin likimääräisen

<sup>1</sup>Skaalauseksponentin arvolle  $\alpha = 1.0$  ei ollut hyvin määriteltyä merkitystä osittaisen Brownin liikkeen viitekehysessä, mutta nyt tehospektrin viitekehysessä on. Se vastaa  $1/f$ -kohinaa.

entropian (*approximate entropy*) käyttöä vuonna 1994 [28]. Sen etuna muihin entropiasuureisiin on melko vaatimaton tarvittavan datan määrä ja parempi kohinan sieto. Likimääräinen entropia on suurempi mitä vähemmän ennustettava tai säännöllinen aikasarja on. Likimääräisen entropian absoluuttinen arvo ei ole erityisen merkityksellinen, koska se riippuu joistakin algoritmin säädettävistä parametreista. Merkityksellisempää on likimääräisten entropioiden suhteelliset arvot toisiinsa nähden. Sykeväliallyysissä on näytetty, että esimerkiksi joillekin patologisille tiloille RR-sarjojen suhteellinen entropia on tilastollisesti merkittävästi pienempi kuin normaaleille RR-sarjoille. Myöhemmin yleisempään käyttöön on tullut likimääräisen entropian johdannainen, otosentropia (*sample entropy*) [29], joka soveltuu likimääräistä entropiaa paremmin lyhyille ja paljon kohinaa sisältäville aikasarjoille.



**Kuva 5.2:** Multiskaalaentropiassa valkoisen kohinan otosentropia pienenee karkearakeistamisen seurauksena ( $\tau$  kasvaa), kun taas vaaleanpunaisen kohinan otosentropia pysyy likimain vakiona.

Eräs toinen sykeväliallyysiin tuotu entropiamenetelmien johdannainen on niin sanottu moniskaalaentropia (*multiscale entropy*) [30]. Moniskaalaentropiassa jokin entropiasuure (yleensä otosentropia) lasketaan aikasarjan lisäksi karkearakeistetuille (*coarse-grained*) aikasarjoille. Karkearakeistaminen tekijällä  $\tau = 2$  tarkoittaa, että aikasarjan 1. arvo on alkuperäisen 1. ja 2. arvon keskiarvo, 2. arvo alkuperäisen 3. ja 4. arvon keskiarvo ja niin edelleen. Karkearakeistaminen lyhentää aikasarjan  $1/\tau$ -osaan alkuperäisestä. Tavallisesti entropiasuureet ottavat huomioon vain aikasarjan peräkkäiset arvot eli lyhimmän mahdollisen aikaskaalan. Multiskaalaentropian ajatuksena on erottaa toisistaan aikasarjat, joissa on merkittävää pidemmän

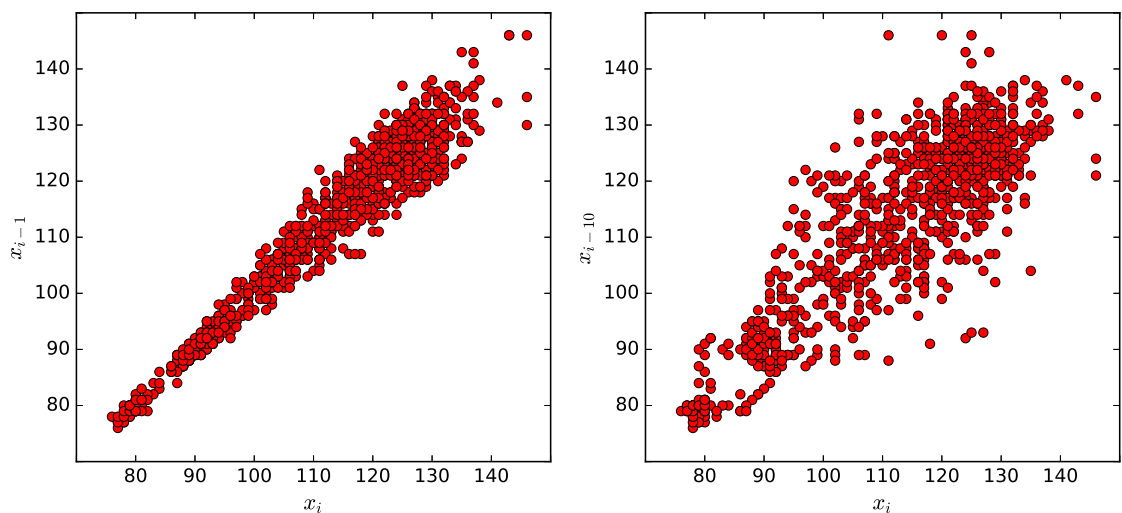


aikaskaalan rakennetta niistä, joissa sitä ei ole. Perusajatus selviää kuvasta 5.2, jossa vertaillaan valkoista ja vaaleanpunaista kohinaa. Pitkän kantaman korrelaatiot tekevät vaaleanpunaisesta kohinasta ns. yhtä kompleksisen kaikilla aikaskaaloilla.

### 5.3 Paluukartat

Paluukartat eli Poincarén kartat [31, 18, 19], ovat pääasiassa visuaalinen työkalu, joiden avulla voi helposti havaita aikasarjan korrelaation tai antikorrelaation tietyllä aikaskaalalla. Paluukartassa aikasarjan alkio  $i - n$  piirretään alkion  $i$  funktiona ja useimmiten  $n = 1$ . Korrelaation tai antikorrelaation voi päätellä paluukartan pisteryppään muodosta. Jos rypäs on sijoittunut nousevalle suoralle, niin aikasarjan arvot viiveellä  $n$  ovat korreloituneita ja vastaavasti antikorreloituneita, jos rypäs on sijoittunut laskevalle suoralle. Pallomainen rypäs viittaa arvojen riippumattomuuteen viiveellä  $n$ .

Kuvassa 5.3 on esitetty edellisessä luvussa käytetyn aikasarjan 1000:n ensimmäisen alkion paluukartat kahdella eri viiveellä. Viiveellä  $n = 1$  aikasarja näyttää voimakkaasti korreloituneelta. Viiveellä  $n = 10$  korrelaatio näyttää heikentyneen, mutta olevan vielä selvästi läsnä. Tämä on odotettava tulos aikasarjalle, jonka voidaan olettaa sisältävän pitkän kantaman korrelaatioita. Huomautettakoon kuitenkin, että paluukartat eivät siedä esimerkiksi trendejä tai vaihtelevaa keskiarvoa hyvin. Nouseva trendi venyttää paluukartan pisteryppään nousevalle suoralle, vaikka satunnainen vaihtelu trendin ympärillä olisi täysin riippumaton. Tämän vuoksi erityisesti sykeväliallyksissä ei kannata tehdä liian pitkälle meneviä päätelmiä pelkän paluukartan perusteella; pelkkä vaihteleva keskisyke vääristää pisteryppään muotoa.



**Kuva 5.3:** Edellisen luvun RR-sarjan 1000:n ensimmäisen alkion paluukartat viiveellä  $n = 1$  ja  $n = 10$ . Kuvan asteikoissa yksikkö on  $10^{-2}s$ .

Paluukarttoihin liitetään sykevälianalyysissä tavallisesti myös joitakin kvantitatiivisia suureita kuten standardipoikkeamat pisteryppääseen sovitetun ellipsin puoliakseleista tai Pearsonin korrelaatiokerroin. Voidaan kuitenkin osoittaa, että nämä ovat suoraan yhteydessä tavanomaisempiin tilastollisiin suureisiin kuten RR-välien standardipoikkeamaan, peräkkäisten RR-välien erotusten standardipoikkeamaan ja autokovarianssiin [tietyllä  $s$ :n arvolla laskettuun autokovarianssin arvoon, ks. (2.5)] [32]. Siten nämä paluukarttoihin liitetyt suureet eivät edusta aikasarjan epälineaarisia ominaisuuksia, vaikka näin usein väitetään [32].

Sydämen sykevälivaihtelun lisäksi paluukarttoja on sovellettu myös kantasoluista kasvatettujen sydänsolujen ryppäiden sykevälivaihteluun. Sydämen sykevälivaihtelun tapaan [18, 19] myös sydänsolujen sykevälivaihtelusta muodostetut paluukartat tuottavat nousevalle suoralle sijoittuvan pisteryppään [33], mikä viittaa korrelaatioihin. Antikorrelaatioita paluukarttojen avulla on puolestaan havaittu esimerkiksi rumpujen soitosta [34], jolloin pisteryppäs sijoittuu päinvastoin laskevalle suoralle.

## 6. TUTKIMUSESIMERKKI: SYKEVÄLIVAIHTELU AVANTOUIMAREILLA

Avantouinti ja saunominen ovat Suomessa suosittuja vapaa-ajan aktiviteetteja [35]. Erityisesti avantouinnin harrastajat pitävät toimintaa rentouttavana kokemuksena, minkä vuoksi mahdollisista terveysvaikutuksista ollaan kiinnostuttu ja asiaa onkin tutkittu [36, 37]. Tässä tutkielmassa ei kuitenkaan varsinaisesti paneuduta mahdollisiin terveysvaikutuksiin, vaan sovellamme esimerkinomaisesti avantouimariaineistoa DFA:n tutkimiseen menetelmänä. Toissijaisesti selvitämme avantouinnin ja saunan mahdollisia vaikutuksia sykevälivaihteluun DFA:n avulla, mutta tuloksista ei tässä tehdä fysiologisia tulkintoja.

### 6.1 Aineisto

Tässä työssä käytetään aineistona RR-sarjoja, jotka ovat peräisin 2008 alkaneista tutkimuksista, joissa selvitettiin avantouinnin ja saunan terveysvaikutuksia [36, 37]. Alkuperäinen aineisto on kerätty 27 vapaaehtoiselta (11 miestä ja 16 naista). Vapaaehtoisille oli seuraavat kriteerit:

- Koehenkilön on oltava kokenut avantouimari, joka käy säännöllisesti avannossa saunasta.
- Koehenkilön on oltava 35–55-vuotias.
- Koehenkilön on käytävä päivätöissä.
- Koehenkilön on oltava terve eikä hän saa käyttää säännöllistä lääkitystä.
- Koehenkilöllä ei saa olla unihäiriöitä, tai vakavaa kuorsaamista.
- Koehenkilöllä on oltava normaali painoindeksi (BMI) ja säännölliset elintavat.

Jokaiselta koehenkilöltä pyrittiin keräämään kolme mittausta, joissa mitataan koehenkilöiden sydänsähkökäyrää (ECG): referenssimittaus (R), saunamittaus (S), sauna- ja avantouintimittaus (A). Kukin mittaus aloitettiin ensimmäisen mittauspäivän aamuna ja lopetettiin neljännen mittauspäivän aamuna, jolloin yksittäisten mittaus-ten kokonaispituus on noin 72 tuntia. S- ja A-mittauksissa 1. ja 2. päivä sisälsivät saunomista (S) tai saunomista ja avantouintia (A). Referenssimittauksessa ei

saanut olla kumpaakaan aktiviteettia. Mittauslaitteena oli Suunnon [38] mittausvyö (Suunto Memory Belt), jota koehenkilöt pitivät koko mittauksen ajan (myös saunassa ja avannossa). RR-sarjat määritettiin sydänsähkökäyristä Firstbeat Technologies -yrityksen [39] ohjelmistolla.

Tämän työn tarpeisiin tämä lyhyt kuvaus aineiston keruusta on riittävä. Tarkempia yksityiskohtia mittauksen ohjeistuksesta, koehenkilöiden tiedoista ja raaka-aineiston prosessoinnista on saatavilla viitteistä [36, 37].

## 6.2 Tutkimuskysymykset

Tutkimus rajoitetaan kolmeen pääkysymykseen:

1. Vaihtelevatko saman henkilön skaalausekspONENTIT selvästi mittausajankohdasta riippuen kuten yksittäinen RR-sarja edellisessä luvussa antaa olettaa?
2. Onko saman henkilön skaalausekspONENTEILLA tilastollisesti merkittävää eroavaisuutta mittauksesta toiseen?
3. Miten edellisen kysymyksen tulokset vertautuvat aiempaan tutkimukseen samalla aineistolla?

Kappaleessa 6.4 käsitellään ensimmäistä kysymystä ja kappaleessa 6.5 toista ja kolmatta. Kappaleessa 6.6 pohditaan tutkimuksen mahdollisia laajennuksia, sekä rajoitteita.

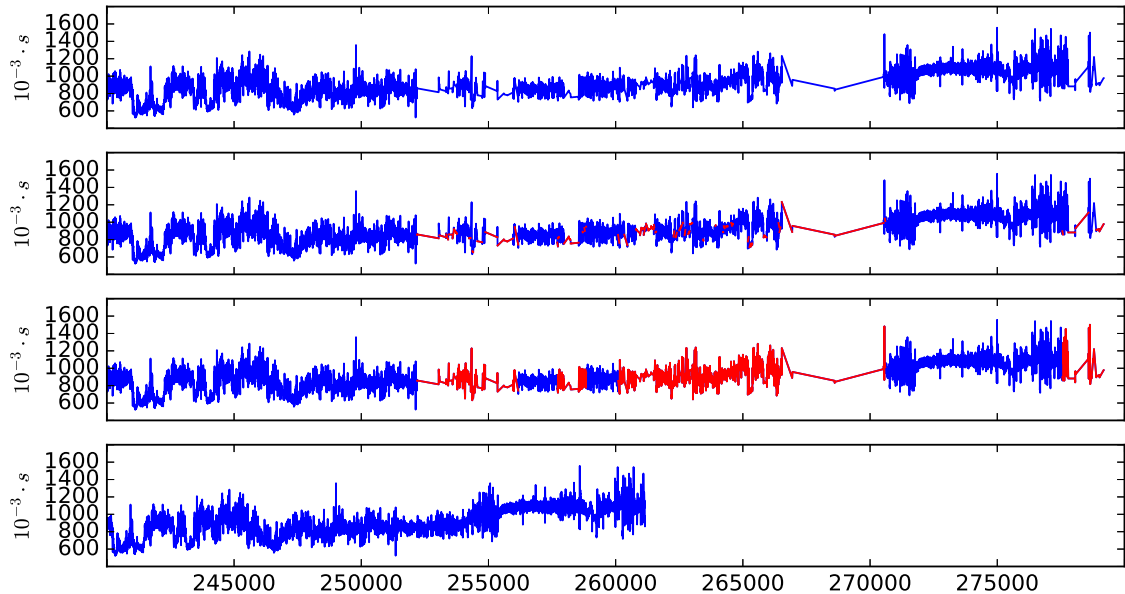
## 6.3 RR-sarjojen esiprosessointi

Aineiston RR-sarjojen lähempi silmämääräinen tarkastelu paljastaa, että useat sarjat sisältävät selvästi puuttuvia tai virheellisiä osioita, jotka saattavat johtua esimerkiksi mittausvyön irtoamisesta paikaltaan. Nämä osiot näkyvät aineistossa tasaisina monotonisina (nousevina, laskevina, vakiomuotoisina) jaksoina.

Luvussa 3.3 totesimme, että korreloituneelle ja stationääriselle sarjalle (tai sarjalle, jonka lisäykset ovat stationäärisiä) pois leikatut lohkot eivät juurikaan vaikuta skaalausekspONENTIN arviointiin. Virheellisten osioiden pois leikkaaminen vaikuttaa parhaalta toimintatavalta aineiston kanssa; RR-sarjat ovat korreloituneita, joskaan eivät stationäärisiä. Lisäksi viitteessä [13] neuvotaan leikkaamaan määrällisesti vähemmän lohkoja aineistosta jos mahdollista eli yhdistämään virheellisiä osioita, vaikka leikattujen lohkojen summan kokonaispituus näin kasvaakin. Näiden tietojen, koikeilun ja silmämääräisen tarkastelun perusteella päädyttiin seuraavaan algoritmiin virheellisten osioiden poistamiseksi RR-sarjoista:

1. Etsitään ja merkataan monotoniset jaksot, joiden pituus on vähintään 10 ja joille  $\Delta x_i \leq |2.0 \cdot 10^{-3}s|$  kaikille arvoille  $i$ <sup>1</sup>.
2. Yhdistetään peräkkäiset merkityt jaksot, jos väliin jäävän osuuden pituus on vähemmän kuin merkittyjen jaksojen pituuksien summa ja alle 750.
3. Toistetaan kohta 2 kunnes kaikki kriteerit täyttävät jaksot on yhdistetty.

Kuvassa 6.1 esitetään esimerkki algoritmista käytännössä.



**Kuva 6.1:** Virheellisen aineiston poistamista koehenkilön 6 referenssisarjasta. Ylimmässä kuvassa on loppuosa kyseisestä RR-sarjasta. Seuraavassa kuvassa algoritmisesti löydetty virheelliset osuudet on väritetty punaisella. Toiseksi alimmassa virheellisiä osuuksia on yhdistelty ja jälleen väritetty punaisella. Alimmassa kuvassa virheelliset osuudet on lopulta leikattu pois.

Taulukossa 6.1 on esitetty aineiston RR-sarjoista esiprosessoinnissa pois leikat-  
tujen osuuksien suhteelliset pituudet. Useat sarjat ovat lähes virheettömiä tässä  
suhteessa, mutta joistain on jouduttu poistamaan jopa yli puolet. Taulukosta puut-  
tuvia mittauksia ei ole alunperin ollutkaan tai ne on jouduttu hylkäämään, koska  
saadussa aineistossa oli sama RR-sarja eri mittauksille.

## 6.4 Paikallisten skaalauseksponttien vaihtelu

Luvussa 4.3 totesimme, että RR-sarjasta DFA:lla määritetyt paikalliset skaalauseks-  
ponentit vaihtelevat huomattavan paljon eikä vaihtelua voida selittää menetelmän  
virheellä, vaan kyseessä on aineiston ei-stationäärisyys. Päätelimme tämän kos-  
kevan RR-sarjoja yleisesti ei-kontrolloiduissa oloissa. Teimme päätelmät kuitenkin

<sup>1</sup>RR-sarjojen resoluutio oli  $10^{-3}s$  eli peräkkäisten arvojen pitää olla korkeintaan kahden yksikön  
päässä toisistaan.

**Taulukko 6.1:** Esiprosessoinnissa pois leikattujen osuuksien suhteelliset pituudet  $M/N$ , missä  $M$  on pois leikatun osuuden pituus ja  $N$  sarjan alkuperäinen pituus.

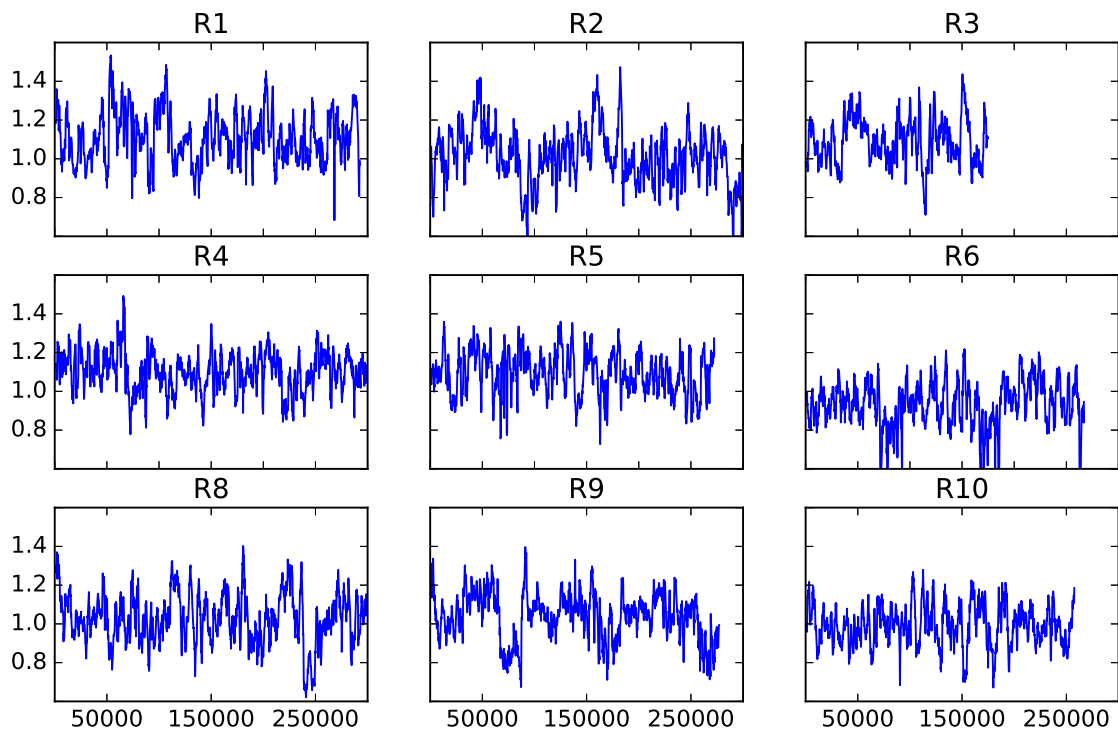
	Mittaus		
	R	S	A
1	0.0011	0.0747	0.0204
2	0.0364	0.0005	0.0057
3	0.0751	0.0304	-
4	0.0000	0.0003	0.0003
5	0.0034	0.0004	0.0153
6	0.0405	0.0221	0.0564
7	0.8804	0.4745	0.4300
8	0.0001	0.0013	0.0002
9	0.0000	0.0023	0.0042
10	0.1307	0.0955	0.0801
11	0.2478	0.0398	-
12	0.0489	0.2342	0.0458
13	0.0267	0.0036	-
14	0.2466	0.4560	0.2420
15	0.0894	0.0442	0.1038
16	0.2211	0.4535	0.2755
17	0.1076	0.0781	0.0280
18	0.0018	0.0047	0.2000
19	0.0721	0.2313	0.1863
20	0.4154	0.5196	0.5861
21	0.1330	-	0.0654
22	0.0085	-	0.0105
23	0.0002	0.0003	0.0067
24	0.1229	-	0.0254
25	0.0002	0.0152	0.0139
26	0.0002	0.0120	0.0054
27	0.0145	0.0003	0.0132

**Taulukko 6.2:** Yksittäisten mittausten paikallisten skaalauseksponenttien keskiarvojen ja standardipoikkeamien jakaumien keskiarvot 95% luottamusvälillä ( $s$  ilmaisee käytetyt ikkunakoot). Paikallisten skaalauseksponenttien laskennassa käytettiin 2000:n RR-välin jaksoja ja 95% päällekkäisyyttä peräkkäisille jaksolle.

	$\alpha$	$\alpha_1$	$\alpha_2$
s	10, ..., $N/10 = 100$	4, ..., 16	32, ..., $N/4 = 500$
$\mu(\mu)$	$1.048 \pm 0.014$	$1.348 \pm 0.030$	$1.071 \pm 0.012$
$\mu(\sigma)$	$0.129 \pm 0.004$	$0.156 \pm 0.011$	$0.126 \pm 0.004$

vain yhden RR-sarjan perusteella. Tutkitaan nyt vastaavalla tavalla avantouimariaineistoa ja katsotaan, vahvistaako tämä aiempaa hypoteesia.

Kaikille aineiston RR-sarjoille laskettiin vastaavat kolme paikallista skaalauseksponenttia kuin luvussa 4.3 käyttäen jälleen 2000:n RR-välin jaksoja ja 95% päällekkäisyyttä peräkkäisille jaksoille. Kullekin mittaukselle laskettiin kunkin paikallisen skaalauseksponentin keskiarvo ja keskimääräinen standardipoikkeama<sup>2</sup>. Näiden suureiden jakaumien keskiarvoa on edelleen arvioitu taulukossa 6.2. Mittaukset, joista pois leikattu osuus oli alle 25% otettiin huomioon analyysissä. Kuvassa 6.2 on esitetty muutamille mittauksille paikalliset skaalauseksponentit samaan tapaan kuin aiemmin kuvassa 4.3.



**Kuva 6.2:** Paikallinen  $\alpha$  2000:n RR-välin jaksoille mittauksille R1-R6 ja R8-R10 (R7 jätettiin pois, koska pois leikattu osuus oli suuri, ks. taulukko 6.1). Laskennassa käytettiin ikkunakokoja  $s = 10, \dots, N/10 = 100$ . Vaaka-akseli on jakson välin keskipiste ja pystyakseli paikallinen  $\alpha$ . Peräkkäisten jaksosten päällekkäisyys on 95%.

Standardipoikkeamien keskiarvoa voidaan verrata taulukon 3.2 arvoihin, vaikka kyseessä ei olekaan täysin samat suureet. 3162:n ja 1000:n pituisille testiaikasarjoille, jotka on generoitu skaalauseksponentilla 1.0, taulukko antaa arvioidulle skaalauseksponentille standardipoikkeamat 0.036 ja 0.061. Jos avantouimariaineisto olisi DFA:n kannalta samankaltaista kuin testiaikasarjat, standardipoikkeamien jakouman keskiarvon voisi olettaa osuvan näiden välille. Se on kuitenkin yli kaksinker-

<sup>2</sup>Yksittäisten mittauksien paikallisten skaalauseksponenttien jakaumien ei voida olettaa noudattavan likimain normaalijakaumaa. Näiden jakaumien keskiarvojen ja standardipoikkeamien kuitenkin voidaan olettaa noudattavan normaalijakaumaa.

tainen edellä arvioituun ylärajaan nähden:  $0.129 \pm 0.004$  ( $\alpha$ ). Näin ollen paikallinen skaalauseksponentti vaihtelee selvästi enemmän, mitä voidaan selittää menetelmän epätarkkuudella.

Toisaalta voidaan tarkastella paikallisten lyhyen ja pitkän kantaman skaalauseksponenttien ( $\alpha_1$ ,  $\alpha_2$ ) keskiarvojen jakauman keskiarvoa. Merkille pantavaa on, että nämä eroavat toistaan selvästi:  $\alpha_1$ :lle  $1.348 \pm 0.030$  ja  $\alpha_2$ :lle  $1.071 \pm 0.012$ . Näitä arvoja ei voida suoraan verrata testiaikasarjoille tehtyihin simulaatioihin (taulukko 3.4), koska sovitukset pitkän ja lyhyen kantaman eksponenteille poikkeavat tässä. Sovitukset valittiin osin kirjallisuuden perusteella. Huomautettakoon, että vaikka  $\alpha_1$ :n sovituskäyriä on erittäin lyhyt ( $s = 4, \dots, 16$ ) ja DFA hieman aliarvioi fluktuaatiota pienillä ikkunakoilla (luku 3.3), niin tämä tuskin selittää koko eroa keskiarvoissa. Lisätarkastelu myös paljastaa, että siirtymä ei ole systemaattisesti samansuuntainen, siten että  $\alpha_1 < \alpha_2$ .

Hypoteesi kontrolloimattomien RR-sarjojen ei-stationäärisyydestä (sen laatuudesta, joka vaikuttaa DFA:han) voidaan edellisen perusteella katsoa pitäväksi.

## 6.5 Skaalauseksponenttien tilastollinen analyysi

Aineiston RR-sarjoille ei voida määrittää täsmällistä skaalauseksponenttia, jolla olisi selvä tulkinta. Erilaiset aktiviteetit vaikuttavat tuntemattomalla tavalla arvioituun skaalauseksponenttiin. Siirrytään nyt selvittämään, voidaanko eri mittausten (R, S, A) välillä havaita systemaattisia eroja DFA:n ja tilastollisen päättelyn avulla.

Lähestytään ongelmaa yksinkertaisimmalla mahdollisella tavalla DFA:n näkökulmasta. Arvioidaan koko RR-sarjoista yksittäiset skaalauseksponentit (lyhyen ja pitkän kantaman sekä koko skaalan) ja vertaillaan näitä eri mittausten ja yksilöiden kesken. Koska sarjat ovat nyt pitkiä (verrattuna 2000:n RR-välin lohkoihin, joita analysoitiin edellisessä kappaleessa), niin valitaan sovituskäyriä hieman toisin. Käytetään konservatiivisempia ala- ja ylärajoja 10 ja  $N/10$ . Lisäksi lyhyen kantaman skaalauseksponentille  $\alpha_1$  valitaan yläraja 32 ja pitkän kantaman skaalauseksponentille alaraja 128. Lasketut arvot löytyvät taulukosta 6.3.

Arvioidaan nyt yksittäisten henkilöiden skaalauseksponenttien mahdollista vaihtelua mittauskerrasta toiseen parittaisten otosten  $t$ -testillä. Nollahypoteesin mukaan saman henkilön eri mittausten skaalauseksponenttien erotuksien odotusarvo on 0. Suuri  $p$ -arvo tukee nollahypoteesia ja pieni  $p$ -arvo nollahypoteesin hylkäämistä. Taulukossa 6.4 on esitetty  $p$ -arvot kaikkien mittausparien (R-S, R-A, A-S) parittaisten otosten  $t$ -testille. Parittaisten otosten  $t$ -testi olettaa aineiston (eri henkilöiden skaalauseksponenttien arvojen samasta mittausluokasta) olevan normaalijakautunutta, mikä on perusteltu oletus.

Taulukosta voidaan välittömästi tehdä muutama merkittävä havainto ja päätelmä. Lasketut  $p$ -arvot tukevat nollahypoteesin hylkäämistä R-S mittausparille  $\alpha$ :n



**Taulukko 6.3:** Koko RR-sarjoista, joista leikattu pois alle 25%, lasketut skaalausekspONENTIT. Sovitusvälit:  $\alpha$ :lle 10,  $\dots$ ,  $N/10$ ,  $\alpha_1$ :lle 10,  $\dots$ , 32 ja  $\alpha_2$ :lle 128,  $\dots$ ,  $N/10$ .

henkilö	mittaus	$\alpha$	$\alpha_1$	$\alpha_2$	mittaus	henkilö	$\alpha$	$\alpha_1$	$\alpha_2$
1	R	1.0824	1.2233	1.0674	14	R	1.1026	1.0361	1.1556
	A	1.1009	1.2329	1.0702		A	1.1000	1.0397	1.1161
	S	1.1066	1.1995	1.0875		S	1.0819	0.9946	1.0970
2	R	1.0831	1.0137	1.1043	15	A	1.1072	1.0412	1.1281
	A	1.0841	1.0115	1.1310		S	1.0952	1.0209	1.0636
	S	1.0751	1.0002	1.1214		16	R	1.0597	1.0278
3	R	1.1199	1.1669	1.1327	17	R	1.1207	1.1013	1.1726
	S	1.1362	1.1138	1.1868		A	1.0923	1.0677	1.1022
4	R	1.0859	1.2166	1.0815		S	1.1227	1.0749	1.1732
	A	1.0806	1.2236	1.0648	18	R	1.0694	0.9758	1.1358
	S	1.1241	1.2101	1.1427		A	1.0706	0.9451	1.0778
5	R	1.1189	1.1758	1.1389		S	1.0629	0.9629	1.1095
	A	1.1310	1.1639	1.1685	19	R	1.0834	1.2427	1.0734
	S	1.1213	1.0979	1.1636		A	1.1113	1.2479	1.0999
6	R	1.0798	0.9968	1.1436		S	1.1314	1.2559	1.1301
	A	1.0752	0.9765	1.1315	21	R	1.0699	1.0934	1.1183
	S	1.0733	0.9759	1.1429		A	1.0639	1.0746	1.1005
8	R	1.0936	1.2127	1.1234	22	R	1.1421	1.0409	1.2258
	A	1.0949	1.2217	1.0836		A	1.1380	1.0379	1.1871
	S	1.0913	1.2230	1.1026	23	R	1.1230	1.0441	1.2008
9	R	1.1143	1.1858	1.1623		A	1.1247	1.0294	1.1726
	A	1.1094	1.1666	1.1172		S	1.1122	1.0741	1.1489
	S	1.1321	1.1461	1.1724	24	R	1.1611	1.0629	1.1887
10	R	1.0981	1.0128	1.1273		A	1.1897	1.1199	1.2384
	A	1.1311	0.9515	1.2205	25	R	1.0308	1.1369	1.0549
	S	1.1297	0.9799	1.1809		A	1.0986	1.0950	1.1507
11	R	1.0846	1.1734	1.0587		S	1.0414	1.0887	1.0598
	S	1.0786	1.2078	1.1104	26	R	1.0950	1.1028	1.0968
12	R	1.0641	1.1825	1.0744		A	1.0801	1.0945	1.0947
	A	1.1041	1.1878	1.0967		S	1.0864	1.0122	1.1153
	S	1.1061	1.2014	1.0823	27	R	1.1251	1.1902	1.1260
13	R	1.0782	1.1000	1.0991		A	1.1223	1.1313	1.1236
	S	1.1338	1.1075	1.1978		S	1.1368	1.1763	1.1413

ja  $\alpha_2$ :n suhteen. Toisaalta myös  $\alpha_1$ :n  $p$ -arvo on lähes tilastollisesti merkittävä R-S mittausparille. Muille mittauspareille ja eri skaalausekspONENTEILLE  $p$ -arvot eivät tue nollahypoteesin hylkäämistä. On kuitenkin huomioitava, että mittausparin R-A ja skaalausekspONENTIN  $\alpha_2$   $p$ -arvo on lähellä yhtä (0.9729), mikä puolestaan tukee nollahypoteesia. Tämä tarkoittaisi sitä, että eri A-mittauksen aktiviteetit eivät vaikuta  $\alpha_2$ :n arvoon suhteessa R-mittaukseen. Tämä on erikoiselta vaikuttava tulos kahdesta syystä. Ensinnäkin R-S mittausparin tulos oli päinvastainen, vaikka sekä S- että A-mittaus sisälsi saunomista. Erona oli se, että A-mittaus sisälsi myös avantouintia. Toiseksi, R-A mittausparin  $\alpha$ :n  $p$ -arvo on melko pieni (0.0800), vaikka sekä  $\alpha$  että  $\alpha_2$  painottuvat enemmän suurille ikkunakoille (alarajat eroavat), joten  $p$ -arvojen voisi olettaa olevan saman suuntaisia.

Yhteenvedona voidaan todeta, että skaalausekspONENTEISSA tapahtuu tilastollisesti merkittävä muutos R- ja S-mittauksen välillä, kun taas A- ja S-mittauksen välillä pitkän kantaman skaalausekspONENTTI  $\alpha_2$  ei juurikaan muutu. Katsotaan nyt, miten tämä vertautuu aiempiin tutkimuksiin samalla aineistolla.

Viitteissä [36] ja [37] on analysoitu aineistoa perinteisempien sykeväliparametrien

**Taulukko 6.4:** Parittaisten otosten  $t$ -testien  $p$ -arvot ja testisuureiden arvot kaikille mittauspareille ja kullekin lasketulle skaalauseksponentille. Tähdellä (\*) on merkitty tilastollisesti merkittävät  $p$ -arvot (nollahypoteesin hylkäämiseksi), jos rajana pidetään  $p$ -arvoa 0.05 ja kahdella tähdellä (\*\*), jos rajana pidetään  $p$ -arvoa 0.01.

	$p$ -arvo			testisuure		
	$\alpha$	$\alpha_1$	$\alpha_2$	$\alpha$	$\alpha_1$	$\alpha_2$
R-S	0.0098**	0.0525	0.0400*	-2.8702	2.0681	-2.2048
R-A	0.0800	0.1869	0.9729	-1.8444	1.3666	-0.0344
A-S	0.0970	0.0675	0.7848	-1.7627	1.9611	-0.2777

avulla. Viitteessä [36] on vertailtu erityisesti unen ensimmäistä neljää tuntia altistuksien (sauna, avantouinti ja sauna) jälkeen verrattuna referenssiviikkoon. Tutkimuksessa havaittiin tilastollisesti merkittävä vaikutus joihinkin sykeväliparametreihin. Toisaalta myöhempi tutkimus [37] näyttää ristiriitaiselta tämän kanssa, jossa tarkasteltiin myös unen ensimmäistä neljää tuntia. Tilastollisesti merkittäviä eroja ei löytynyt. Tässä hieman kattavammassa tutkimuksessa päähavainto oli se, että altistuksien vaikutus sykeväliparametreihin on välitön eli akuutti. Saunan ja saunan ja avantouinnin akuutit vaikutukset erosivat toisistaan.

Edellisiä tuloksia ei ole suoraviivaista verrata saatuihin DFA-tuloksiin. DFA:lla tutkittiin kokonaisia usean päivän RR-sarjoja eikä joitain erityisiä jaksoja. Toisaalta DFA-tulokset eivät ainakaan ole ilmeisessä ristiriidassa tutkimuksen [37] perusteella. Vaikka eri altistuksilla olisi vain akuutti vaikutus, niin näiden osuuksien merkitys flukтуаatiokäyrään saattaa olla merkittävä. Myöskään S- ja A-mittaukset eivät olleet samanlaisia suhteessa R-mittaukseen, mikä voidaan mahdollisesti selittää akuuttien altistusten vaikutusten eroavaisuudella.

Helpompi ja kenties mielenkiintoisempi vertailukohde on oma tutkimuksemme [40], jossa avantouimariaineistoa on DFA:n lisäksi analysoitu multiskaalaentropialla (ks. luku 5.2). Multiskaalaentropianalyysissä käytettiin vastaavasti kokonaisia RR-sarjoja, joista virheelliset osuudet olis poistettu samalla algoritmilla kuin tässä tutkielmassa. Parillisten otosten  $t$ -testissä vertailtavana suurena käytettiin ns. kompleksisuusindeksiä, joka oli otosentropioiden summa 10:llä ensimmäisellä  $\tau$ :n arvolla. Tämä analyysi tuotti vastaavan tuloksen, kuin DFA: R-S mittausparin kompleksisuusindeksit erosivat tilastollisesti merkittävästi toisistaan mutta muiden mittausparien eivät.

## 6.6 Analyysin rajoituksia

Tilastollisen analyysin selvin heikkous on tulkinnan vaikeus. Skaalauseksponenttien arviointiin käytettiin kokonaisia, usean päivän mittaisia RR-sarjoja. Tämän takia on erittäin vaikeaa sanoa mistä tulosten erot johtuvat: onko kyseessä pelkästään sauna- tai avantoaltistuksien akuutit vaikutukset vai myös mahdolliset pidempiai-

kaiset vaikutukset? Aiemmat tutkimukset viittaa lähinnä akuutteihin vaikutuksiin [37]. Toisaalta tällöin avoimeksi kysymykseksi jää se, miksi avannon ja saunan akuutti yhteisvaikutus ei näy DFA- tai multiskaalaentropiatuloksia [40] hyödyntävässä tilastollisessa analyysissä, kun taas pelkän saunan akuutti vaikutus näkyy.

Toinen heikkous on aineiston yksinkertainen esiprosessointi. Vaikka esiprosessointi löytää luotettavasti selvästi virheelliset mittaukset, niin siinä ei kuitenkaan ole tarkasteltu, mitä aineistosta on poistettu. Jos hypoteesina on skaalauseksponenttien tilastollisten erojen johtuminen akuuteista altistuksista, niin esiprosessointi todennäköisesti vääristäisi tuloksia. Nyt analyysissä kuitenkin käytettiin kaikkia sarjoja, joista löydettiin riittävän vähän (alle 25%) virheellisiä mittauksia.

Edellisiä heikkouksia ja näistä seuraavia kysymyksiä voitaisiin selvittää laajentamalla analyysiä. Kokonaisten sarjojen lisäksi DFA:lla voisi tutkia pelkkiä altistusjaksoja, suunnilleen altistusjaksojen pituisia muita jaksoja, sekä S- ja A-mittauksia, joista on leikattu altistusjaksot pois. Näin voitaisiin mahdollisesti selvittää tilastollisten eroavaisuuksien alkuperää. Painotettakoon, että vaikka tilastollisten eroavaisuuksien syy löydettäisiin pelkistä altistusjaksoista, niin se ei edelleenkään antaisi skaalauseksponenteille selvää tulkintaa.

## 7. YHTEENVETO

Tässä työssä tutkittiin aikasarjojen pitkän kantaman korrelaatioita ja fraktaalisia ominaisuuksia. Työkaluna käytettiin erityisesti trendit poistavaa flukтуаatioanalyysiä, DFA:ta, jonka ominaisuuksia eri tilanteissa tarkasteltiin huolellisesti menetelmäosiossa. Totesimme, että DFA soveltuu hyvin pitkän kantaman korrelaatioiden ja fraktaalisuuden kvantitatiiviseen määrittämiseen tietyn edellytyksin. Jos aikasarjassa on pitkän kantaman korrelaatioita tai se on fraktaalinen Hurstin eksponentilla  $H$ , niin flukтуаatio noudattaa potenssilakia  $F_2(s) \propto s^\alpha$  ja skaalauseksponentti  $\alpha$  voidaan määrittää suorasta logaritmisella asteikolla. Näimme kuitenkin, että mahdolliset poikkeavuudet pitkän kantaman korrelaatioista aikasarjassa tai erilaiset ei-stationäärisyydet saavat flukтуаation usein poikkeamaan potenssilaista. Tämä näkyy yleensä siirtymänä flukтуаatiokäyrässä eli flukтуаatiokäyrä ei ole täsmälleen suora.

Tiivistettynä voidaan sanoa, että DFA:lla voidaan päätellä ja määrittää pitkän kantaman korrelaatioita ja fraktaalisuutta seuraavien ehtojen mukaisesti. Pitkän kantaman korrelaatio (skaalauseksponentti  $\alpha$ ) voidaan todeta ja määrittää, jos flukтуаatio  $F_2(s)$  noudattaa tarkasti potenssilakia useamman kertaluokan aikaskaaloilla. Jos aikasarja on lyhyehkö ja flukтуаatio noudattaa potenssilakia, niin skaalauseksponentti voidaan määrittää kohtalaisella tarkkuudella (riippuen aikasarjan pituudesta, tätä tarkasteltiin osiossa 3.3.2), mutta sen ei voida varmasti sanoa johtuvan pitkän kantaman korrelaatioista. Mikäli flukтуаatio taas poikkeaa potenssilaista, niin silloin on huolellisesti otettava eri mahdollisuudet huomioon ja pohdittava, voidaanko skaalauseksponentti määrittää ja tulkita mielekkäästi.

Menetelmäosion jälkeen siirryttiin sovelluskohteeseen eli sydämen sykevälivaihteluun. Sykevälivaihtelun on havaittu sisältävän pitkän kantaman korrelaatioita ja olevan  $1/f$ -kohinan kaltaista ( $\alpha \approx 1$ ). Toisaalta sykevälivaihteluun voi ainakin potentiaalisesti vaikuttaa melkein mikä tahansa fysiologinen toiminta, joten sykeväli- eli RR-sarjojen ei voida olettaa olevan stationäärisiä. Tämä näkyy DFA-tarkastelussa: pitkät (esimerkiksi vuorokauden) RR-sarjat eivät yleensä noudata potenssilakia täsmällisesti, vaan tyypillisesti flukтуаatiokäyrä sisältää siirtymän lyhyehköillä aikaskaaloilla. Jos puolestaan tarkastellaan lyhyempiä RR-sarjoja, niin flukтуаatiokäyrä (ja  $\alpha$ , jos sellainen jollain sovitteella arvioidaan) ei pysy likimainkaan vakiona, vaan vaihtelee merkittävästi. RR-sarjat eivät siis noudata mitään tunnettua teoreettista

mallia ja tämän takia niille arvioitujen skaalauseksponenttien ei voida tulkita olevan täsmällisesti sidoksissa teoreettisiin suureisiin kuten Hurstin eksponenttiin.

Nämä aineiston ominaisuudet rajoittavat DFA:n käyttöä sykevälianalyysissa. Empiiristen havaintojen kuten flukтуаatiokäyrän siirtymien perusteella tavaksi on muodostunut, että RR-sarjoille arvioidaan ns. pitkän ja lyhyen kantaman skaalauseksponentit. Näille skaalauseksponenteille ei ole selvää teoreettista tulkintaa, ja niiden vertailukelpoisuus edellyttää jonkinlaista mittausten vakiointia. Tyypillisesti nämä tunnusluvut soveltuvat lähinnä tilastolliseen analyysiin, jonka perusteella voidaan mahdollisesti luoda heuristisia sääntöjä. Tämä johtopäätös ei ole erityisen rohkaiseva, mutta todettakoon että heuristiset säännötkin voivat olla käytännössä erittäin hyödyllisiä, jos niitä pystytään soveltamaan esimerkiksi suurten aineistojen automaattisessa analysoinnissa.

Tutkielman viimeisessä osiossa tarkasteltiin avantouimarien RR-sarjoja DFA:n avulla. Koehenkilöiltä oli mitattu kolme usean päivän mittaista RR-sarjaa, joista yksi sisälsi saunomista (S), toinen saunomista ja avantouintia (A) ja kolmas referenssarja ei kumpaakaan näistä aktiviteeteista (R). Aineistolla pystyttiin todentamaan RR-sarjojen korrelaatio-ominaisuuksien vaihtelu tavallisissa oloissa. Määritetyt skaalauseksponentit vaihtelivat merkittävästi mittausten sisällä riippuen, siitä mitä osaa mittauksesta tarkasteltiin. Lisäksi saman henkilön skaalauseksponenttien vaihtelua mittauksesta toiseen tutkittiin tilastollisen analyysin avulla. R- ja S-mittausten skaalauseksponenttien välillä havaittiin tilastollisesti merkittävä eroavaisuus, mutta muiden mittausparien välillä ei. Eroavaisuuden alkuperää ei selvitetty, mutta aiempien tutkimuksien perusteella voidaan tehdä hypoteesi, että eroavaisuus johtuu sauna-altistuksen akuutista vaikutuksesta sykevälivaihteluun pidempiaikaisten vaikutusten sijaan. Vastaavasti avantoaltistuksen akuutin vaikutuksen voidaan olettaa kumoavan saunan-altistuksen vaikutuksen skaalauseksponenteissa, koska R-A-mittausparin väliltä ei löydetty tilastollisia eroavaisuuksia.

Kaikkiaan tutkielmassa paneuduttiin DFA:n ominaisuuksiin menetelmänä sekä sen haasteisiin erityisen sovelluskohteen eli sykevälivaihtelun parissa. Perimmäinen syy mainittuihin haasteisiin oli se, että aineiston ei voitu olettaa noudattavan tiettyjä teoreettisia taustaoletuksia (stationäärisyys). Tämä ei ole epätavallista monimutkaisista systeemeistä mitatuissa aikasarjoissa. Niinpä analyysistä saatuja tuloksia voi mahdollisesti hyödyntää muissakin sovelluskohteissa.

Jatkotutkimukselle olisi tarvetta niin menetelmäpuolella kuin tutkitussa aineistossakin. DFA:ta on tutkittu paljon, mutta esimerkiksi  $\alpha$ :n arvio suhteessa eri aikasarjojen generointialgoritmeihin on jäänyt vähemmälle huomiolle. Avantouimariaineiston tutkimus oli melko suppea ja siitä jäi selviä avoimia kysymyksiä kuten se, miten altistukset akuutisti vaikuttavat  $\alpha$ :aan. Lisäksi sykevälivaihtelun tutkimuksessa yleisesti on tietysti pidemmälle meneviä tavoitteita, esimerkiksi patologisten

tilojen automaattinen seulonta suurista aineistoista tilastollisen päättelyn avulla. DFA on yksi mahdollinen työkalu, jolla voidaan saada sopivaa aineistoa tilastollista päättelyä varten.

## VIITTEET

- [1] Chatfield C. 1991. *The Analysis of Time Series: An Introduction*. 4th ed. Chapman and Hall.
- [2] Zhu Z. W., Zhou Z. L., Jia B. G. 2003. On the Lower Bound of the Hausdorff Measure of the Koch Curve. *Acta Mathematica Sinica, English Series*, vol. 19 No. 4.
- [3] Kantz H., Schreiber T. 2004. *Nonlinear Time Series Analysis*. 2nd ed. Cambridge University Press.
- [4] Mandelbrot B, van Ness J. W. 1968. Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10.
- [5] Kantelhardt J. W. 2008. Fractal and Multifractal Time Series. arXiv:0804.0747 [physics.data-an].
- [6] Peng C.-K., Buldyrev S. V., Havlin S., Simons M., Stanley H. E., Goldberger A. L. 1994. Mosaic organization of DNA nucleotides. *Physical Review E* 49.
- [7] Liu Y., Cizeau P., Peng C.-K., Stanley H. E. 1997. Correlations in Economic Time Series. *Physica A* 245.
- [8] Santhanam M. S., Jayendra N. B., Angom D. 2006. Quantum spectrum as time series: Fluctuation measures. *Physical Review E* 73.
- [9] Taqqu M. S., Teverovsky W., Willinger W. 1995. Estimators for long-range dependence: an empirical study. *Fractals* 3.
- [10] Dieker T. 2004. Simulation of fractional Brownian motion. Master's thesis, University of Twente, The Netherlands.
- [11] Bryce R. M., Sprague K. B. 2012. Revisiting detrended fluctuation analysis. *Scientific Reports* 2.
- [12] Pilgram B, Kaplan D. T. 1997. A comparison of estimators for  $1/f$  noise. *Physica D* 114.
- [13] Chen Z., Ivanov P. C., Hu K., Stanley H. E. 2002. Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E* 65.
- [14] Kantelhardt J. W., Koscielny-Bunde E., Rego H. H. A., Havlin S., Bunde A. 2001. Detecting long-range correlations with detrended fluctuation analysis. *Physica A* 295.

- [15] Hu K. Ivanov P. C., Chen Z., Carpena P., Stanley H. E. 2001. Effect of trends on detrended fluctuation analysis. *Physical Review E* 64.
- [16] Peng C.-K., Hausdorff J. M., Goldberger A. L. 2000. *Self-Organized Biological Dynamics and Nonlinear Control*, luku 3: Fractal mechanisms in neuronal control: human heartbeat and gait dynamics in health and disease. Cambridge University Press.
- [17] Guyton A. C., Hall J. E. 2006. *Textbook of Medical Physiology*. 11th ed. Elsevier Inc.
- [18] Acharya U. R., Joseph K. P., Kannathal N., Lim C. H., Suri J. S. 2006. Heart rate variability: a review. *Medical & Biological Engineering & Computing*. vol. 44:1031-1051.
- [19] Laitio T., Scheinin H., Kuusela T., Mäenpää M., Jalonen J. 2001. Mitä sydämen sykeväilvaihtelu kertoo? *FINNANEST* Vol. 34, No. 3.
- [20] Peng, C.-K., Havlin S., Stanley H. E., Goldberger A. L. 1995. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *CHAOS*, vol. 5, No. 1.
- [21] Iyengar N., Peng C.-K., Morin R., Goldberger A. L., Lipsitz L. A. 1996. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology*, vol. 271, No. 4.
- [22] Ho K. K., Moody G. B., Peng C.-K., Mietus J. E., Larson M. G., Levy D., Goldberger A. L. 1997. Predicting Survival in Heart Failure Case and Control Subjects by Use of Fully Automated Methods for Deriving Nonlinear and Conventional Indices of Heart Rate Dynamics. *Circulation* vol. 96 No. 3.
- [23] Penzel N., Kantelhardt J. W., Grote L., Peter J.-H., Bunde A. 2003. Comparison of Detrended Fluctuation Analysis and Spectral Analysis for Heart Rate Variability in Sleep and Sleep Apnea. *Transactions on Biomedical Engineering*, vol. 50, No. 10.
- [24] Baumert M., Javorka M., Seeck A., Faber R., Sanders P., Voss A. 2012. Multiscale entropy and detrended fluctuation analysis of QT interval and heart rate variability during normal pregnancy. *Computers in Biology and Medicin* vol. 42.
- [25] Gieraltowski J., Zebrowski J. J., Baranowski R. 2012. Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia. *Physical Review E* 85.



- [26] Echeverria J. C., Woolfson M. S., Crowe J. A., Hayes-Gill B. R., Croaker G. D. H., Vyas H. 2003. Interpretation of heart rate variability via detrended fluctuation analysis and  $\alpha\beta$  filter. *CHAOS*, vol. 13, No. 2.
- [27] Kuusela J, Kim J., Räsänen E., Aalto-Setälä K. 2016. The Effects of Pharmacological Compounds on Beat Rate Variations in Human Long QT-syndrome Cardiomyocytes. *Stem Cell Review*, 12(6): 698–707.
- [28] Pincus S. M., Golberger A. L. 1994. Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology*, vol. 266.
- [29] Richman J. S., Moorman J. R. 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology*, vol. 278.
- [30] Costa M., Goldberger A. L., Peng C-K. 2002. Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters* vol. 89 No. 6.
- [31] Kamen P., Krum K., Tonkin A. 1996. Poincaré plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans. *Clin Sci (Lond)*. 1996;91: 201–208.
- [32] Brennan M., Palaniswami M., Kamen P. 2001. Do Existing Measures of Poincaré Plot Geometry Reflect Nonlinear Features of Heart Rate Variability? *Transactions on Biomedical Engineering*, vol. 48, No. 11.
- [33] Kim J., Kuusela J., Aalto-Setälä K., Räsänen E. 2017. Short- and long-term correlations in beat rate variability of human induced pluripotent stem cell-derived cardiomyocytes. *Computing in Cardiology* (in print).
- [34] Räsänen E., Pulkkinen O., Virtanen T., Zollner M., Hennig H. 2015. Fluctuations of Hi-Hat Timing and Dynamics in a Virtuoso Drum Track of a Popular Music Recording. *PLoS ONE* 10(6): e0127902.
- [35] Heikura P., Huttunen P., Kinnunen T. 2000. *Hyinen hurmio: avantouimarin käsikirja*. Edita.
- [36] Riikonen R. 2008. The effects of sauna bath and winter swimming on night time recovery: Heart rate variability and actigraphy analysis. Master of Science thesis, Faculty of Computing and Electrical Engineering, Tampere University of Technology.
- [37] Rajala S. I. 2011. The effects of sauna bath and winter swimming on sympathovagal balance of the heart. Masters's Thesis in Biomechanics, University of Jyväskylä, Finland.

- [38] Suunto Oy. [WWW]: *www.suunto.com*
- [39] Firstbeat Technologies Oy. [WWW]: *www.firstbeat.com*
- [40] Potapov I., Haverinen S., Smolander J., Viik J., Räsänen E. 2017. Nonlinear Effects of Winter Swimming and Sauna Recreational Activities on the Heart Rate Variability. *Computing in Cardiology* (in print).