



**TURUN  
YLIOPISTO**

SUOMEN SADEAIKASARJOJEN ANALYYSIA: TRENDEJÄ VAI EI?

Charlotta Sievänen

Pro gradu -tutkielma

Kesäkuu 2022

Tarkastajat:

VTT Henri Nyberg

FT Joni Virta

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä.

TURUN YLIOPISTO

Matematiikan ja tilastotieteen laitos

CHARLOTTA SIEVÄNEN: Suomen sadeaikasarjojen analyysia: trendejä vai ei?

Pro gradu -tutkielma, 26 s., 4 liites.

Tilastotiede

Kesäkuu 2022

---

Ennusteiden mukaan sademäärät ja rankkasateet tulevat lisääntymään Suomessa ja yleisesti pohjoisella pallonpuoliskolla ilmastonmuutoksen seurauksena. Muutokset sateisuudessa vaikuttavat moniin yhteiskunnan tärkeisiin toimintoihin, kuten viljelyyn ja rakentamiseen. Suomessa ei kuitenkaan ole juurikaan tutkittu sadeaikasarjoja yksikköjuuritestien avulla, vaikka tämä voisi parantaa ilmiön ymmärtämistä ja jatkoanalyysieja. Tässä työssä on tutkittu yksikköjuuritestien avulla sitä, ovatko Suomen sadeaikasarjat stationaarisia, trendistationaarisia vai yksikköjuuriprosesseja. Aikasarjat perustuvat Ilmatieteen laitoksen eri puolilta Suomea keräämiin kuukausittaisiin sadesummiin ja kattavat vuodet 1965–2021. Aikasarjoja tarkasteltiin kuukausikohtaisesti kaikkien havaintoasemien välistä keskiarvoa käyttäen.

Käytettyjen kolmen yksikköjuuritestin (ns. ADF-, PP- ja KPSS-testien) perusteella tammi-kuun sadesummiä kuvaava aikasarja vaikutti mahdollisesti trendistationaariselta, ja siinä voisi kuvankin perusteella olla loiva nouseva trendi. Tulos on yhteneväinen aiempien tutkimusten kanssa, joissa Suomen sademäärien on ennustettu kasvavan erityisesti talvella. Myös touko- ja joulukuun aikasarjat ovat mahdollisesti trendistationaarisia, mutta tulokset jäivät hieman epäselviksi. Lähes kaikkien muiden kuukausien sadeaikasarjat olivat testien mukaan tason suhteen stationaarisia, eli ne vaihtelevat vakiovarianssilla muuttumattoman odotusarvon ympärillä. Useimpien kuukausien osalta ei siis löytynyt näyttöä minkäänlaisesta trendistä sademäärissä. Lisäksi osa testituloksista viittasi siihen, että helmikuun aikasarja saattaisi olla yksikköjuuri-prosessi, mikä on tärkeää huomioida mahdollisissa jatkoanalyysieissa.

Asiasanat: aikasarja-analyysi, ilmastonmuutos, sade, yksikköjuuritestit, stationaarisuus.

# Sisällys

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Johdanto</b>                                       | <b>1</b>  |
| <b>2</b> | <b>Teoriaa</b>  | <b>3</b>  |
| 2.1      | Stokastinen prosessi ja stationaarisuus . . . . .     | 3         |
| 2.2      | Autoregressiivinen prosessi ja yksikköjuuri . . . . . | 4         |
| 2.3      | Trendistationaarinen prosessi . . . . .               | 6         |
| 2.4      | Yksikköjuuren käytännön merkitys . . . . .            | 6         |
| <b>3</b> | <b>Yksikköjuuritestit</b>                             | <b>8</b>  |
| 3.1      | Dickey–Fuller-testit . . . . .                        | 8         |
| 3.1.1    | DF-testi . . . . .                                    | 8         |
| 3.1.2    | ADF-testi . . . . .                                   | 10        |
| 3.2      | Phillips–Perron-testi . . . . .                       | 11        |
| 3.3      | KPSS-testi . . . . .                                  | 12        |
| <b>4</b> | <b>Ilmastoaikasarjat ja yksikköjuuriprosessit</b>     | <b>14</b> |
| <b>5</b> | <b>Aineisto ja menetelmät</b>                         | <b>17</b> |
| <b>6</b> | <b>Tulokset</b>                                       | <b>20</b> |
| <b>7</b> | <b>Pohdinta</b>                                       | <b>23</b> |
| <b>8</b> | <b>Viitteet</b>                                       | <b>25</b> |
| <b>9</b> | <b>Liitteet</b>                                       | <b>27</b> |

# 1 Johdanto

Ihmisen toiminnan seurauksena ilmasto on muuttunut ja muuttuu edelleen niin maailmanlaajuisesti kuin Suomessakin. Hallitustenvälisen ilmastomuutospaneeli IPCC:n (Intergovernmental Panel on Climate Change) vuonna 2014 julkaistun raportin mukaan maapallon keskilämpötila on noussut arviolta 0,85 °C ajanjaksolla 1880–2012, kun otetaan huomioon sekä maan että merten pintalämpötilat. Myös globaalissa veden kiertoalustassa on havaittu muutoksia, esimerkiksi pohjoisella pallonpuoliskolla sademäärät ovat kasvaneet. Myös merten muuttuneet suolapitoisuudet kertovat epäsuorasti sateiden ja veden kiertoalustan muutoksista: suolaiset alueet, joilla tapahtuu enemmän haihtumista, ovat tulleet entistä suolaisemmiksi. Toisaalta makeamman veden alueet, joilla sataa enemmän, ovat tulleet entistä makeammiksi. Lisäksi erityisesti pohjoinen napajäättikkö on sulanut kiihtyvällä vauhdilla viime vuosikymmeninä, ja merenpinta on noussut arviolta 0,19 m vuosina 1901–2010. [8]

Maapallon keskilämpötila tulee erittäin todennäköisesti nousemaan tulevaisuudessa, ja äärimmäisen lämpimät ajanjaksot yleistyvät. Merenpinnan on ennustettu edelleen nousevan ja rankkasateiden yleistyvän. Sateisuuden muutokset ovat kuitenkin erilaisia eri puolilla maapalloa. Ennusteiden mukaan korkeilla leveysasteilla sademäärät kasvavat edelleen, kun taas monilla subtrooppisilla alueilla sademäärät vähenevät. Päiväntasaajan alueella ennusteet vaihtelevat paikallisesti. [8]

Suomessa sekä sateiden intensiteetin että sadepäivien lukumäärän on ennustettu kasvavan, kuitenkin talvella enemmän kuin kesällä. Sade- ja poutapäivät saattavat myös tulevaisuudessa esiintyä entistä enemmän ryppäissä, sekä kuukausikohtaisen sademäärän vaihtelu kasvaa eri vuosien välillä. Siis hyvin kuivien ja hyvin vetisten kuukausien esiintyminen yleistyy. Tämä saattaa lisätä myös kuivia kesiä, jotka ovat entistä ongelmallisempia, kun lämpötilat samanaikaisesti nousevat. [15] Suomessakin ennustetut muutokset seuraavat siis jossain määrin globaalia sään äärevöitymisen trendiä.

Tulevaisuuden ilmaston ennustaminen on kuitenkin epävarmaa. Ilmastomuutoksen ilmeneminen riippuu siitä, miten kasvihuonekaasujen pitoisuus ilmakehässä muuttuu, ja miten ilmasto reagoi näihin muutoksiin. Siispä eri mallit antavat vaihtelevia tuloksia. Ilmasto vaihtelee myös luonnostaan: auringon aktiivisuuden muutokset, tulivuorenpurkaukset, sekä ilmakehän ja merten kiertoalustat ovat tärkeimpiä satunnaisvaihtelun lähteitä. Luonnollisesta satunnaisvaihtelusta johtuva epävarmuus on sitä suurempaa, mitä paikallisemmin tai lyhyemmällä aikavälillä ilmastoennusteita yritetään tehdä. [8, 15]

Viime vuosikymmeninä ilmastomuutoksen tutkimiseen on alettu soveltaa enemmän ekonometriassa käytettyjä menetelmiä. Koska sekä ilmasto että monet taloudelliset mittarit riippuvat ihmisen toiminnasta, ei ole kovin yllättävää, että esimerkiksi ekonometriassa käytetyt aikasarja-analyysimenetelmät soveltuvat hyvin myös tähän tarkoitukseen. [2]

Ennen monimutkaisempaa mallintamista on kuitenkin aina tärkeää tutkia käsiteltävien aikasarjojen ominaisuuksia yksiulotteisella tasolla. Yksi olennaisimmista ehdoista monien tilastollisten menetelmien käyttämiselle on aikasarjan stationaarisuus. Stationaarisen aikasarjan tilastolliset ominaisuudet eivät vaihtelee ajassa. Erityisesti sillä ei siis ole trendiä, eli käsiteltävässä muuttujassa ei ole tilastollisesti merkitsevää trendinomaista käyttäytymistä joko ylös- tai alaspäin [1, 9]. Mikäli aikasarja taas vaikuttaa epästationaariselta, on tärkeää tietää, onko kysessä niin sanottu trendistationaarinen aikasarja vai yksikköjuuri prosessi, sillä tämä vaikuttaa paitsi prosessin tulkintaan, myös siihen, miten aikasarjaa kannattaa käsitellä jatkoanalyysia varten. Tähän tarkoitukseen on kehitetty useita erilaisia testejä, joista kolmea käytetään ja esitellään tarkemmin tässä työssä.

Ilmastonmuutoksen kontekstissa varsinkin lämpötila-aikasarjoja on tutkittu stationaarisuuskysymyksen näkökulmasta melko paljon verrattuna sademääriin tai muihin muutujiin. Tämä saattaa johtua siitä, että lämpötilasta on saatavilla laadukkaampia aineistoja pidemmältä aikaväliltä kuin monista muista ilmastomuuttujista. Lämpötilaa pidetään myös kaikkein suorimpana ilmastonmuutoksen indikaattorina, ja se on noudattanut globaalisti selkeämpää trendiä kuin esimerkiksi sademäärät [8, 10]. Sademäärienkin stationaarisuutta on silti hyödyllistä tutkia, sillä tämä voi parantaa ennusteita ja auttaa ymmärtämään aikasarjaa generoivan prosessin luonnetta. Sateisuuden pitkän aikavälin muutokset vaikuttavat muun muassa viljelyyn, pohjavesiin, rakentamiseen, tulvariskeihin ja moniin eliölajeihin, joiden selviytymistä esimerkiksi pitkät kuivat jaksot usein vaikeuttavat.

Etenkään Suomessa sadeaikasarjojen stationaarisuutta ei ole juurikaan tutkittu. Tässä opinnäytetyössä keskitytäänkin Suomen sademäärien tutkimiseen aikasarja-analyysin näkökulmasta. Työssä käytetään eri puolilta Suomea 56 vuoden aikana kerättyjä kuukausittaisia sademääriä, joista lasketaan koko Suomen keskiarvot. Tavoitteena on selvittää, ovatko kuukausikohtaiset aikasarjat stationaarisia, trendistationaarisia vai yksikköjuuri prosesseja. Tarkoitukseen käytetään kolmea yksikköjuuritestistä (ns. ADF-, PP- ja KPSS-testiä), joiden tuloksia vertaillaan. Opinnäytetyö koostuu teoriaosuudesta, lyhyestä kirjallisuuskatsauksesta ja lopun empiirisestä osuudesta.

## 2 Teoriaa

Käydään ensin läpi muutamia tämän tutkielman kannalta oleellisia käsitteitä. Tämä luku perustuu Boxin, Jenkinsin ja Reinselien kirjaan *Time Series Analysis. Forecast and Control* [1], ellei toisin mainita.

### 2.1 Stokastinen prosessi ja stationaarisuus

Olkoon

$$z = (z_1, z_2, \dots, z_N)$$

$N$ :stä peräkkäisestä havainnosta koostuva aikasarja, jossa indeksi  $t = 1, \dots, N$  kuvaa havaintojen ajankohtaa. Tässä tutkielmassa oletetaan, että  $t$  on diskreetti ja havainnot ovat ajallisesti yhtä kaukana toisistaan. Aikasarjan  $z$  ajatellaan olevan jonkin stokastisen prosessin havaittu realisaatio, siis yksi kaikista (äärettömän monista) aikasarjoista, jotka tämä prosessi olisi voinut tuottaa. Stokastiseksi kutsutaan sellaisia matemaattisia malleja, joissa epävarmuutta mallinnetaan satunnaismuuttujien ja näiden jakaumien avulla. Stokastista prosessia ei siis voi koskaan ennustaa täysin varmasti, toisin kuin täysin determinististä prosessia, jonka tulevaisuuden arvot tiedetään. Käytännössä mikään luonnollinen prosessi ei näyttäydy meille täysin deterministisenä, koska prosessin havainnoissa on aina mukana ulkoisten tekijöiden aiheuttamaa epävarmuutta.

Stokastiset prosessit voidaan edelleen jakaa stationaarisiiin ja epästationaarisiiin prosesseihin. Stationaarinen prosessi tai malli vaihtelee vakiovarianssilla tietyn, muuttumattoman odotusarvon ympärillä. Visuaalisesti katsottuna tällaisessa aikasarjassa ei siis ole esimerkiksi nousevaa tai laskevaa trendiä, eikä se keskimäärin vaihtelee enemmän jollakin toisella aikavälillä kuin toisella. Ajatellaan siis, että aikasarjaprosessi  $z$  tuottaa  $N$ -ulotteisen satunnaisvektorin yhteistodennäköisyysjakaumalla  $p(z_1, z_2, \dots, z_N)$ . Prosessia  $z$  kutsutaan vahvasti stationaariseksi, jos tämä  $N$ -ulotteinen yhteisjakauma pysyy samana, vaikka kaikkien havaintojen aikaindeksiä  $t$  siirrettäisiin jollakin kokonaisluvulla  $k$ . Toisin sanoen

$$p(z_1, z_2, \dots, z_N) = p(z_{1+k}, z_{2+k}, \dots, z_{N+k}),$$

kun  $k \in \mathbb{Z}$ , tai yleisemmin

$$p(z_{t_1}, z_{t_2}, \dots, z_{t_m}) = p(z_{t_1+k}, z_{t_2+k}, \dots, z_{t_m+k}),$$

jossa  $m$  on havaintojen lukumäärä. Tällöin kunkin havainnon odotusarvo  $E(z_t) = \mu$  ja varianssi  $Var(z_t) = \sigma^2$  jokaisella  $t$ . Vahvan stationaarisuuden määritelmästä seuraa myös

se, että havaintojen  $z_t$  ja  $z_{t+k}$  välinen kovarianssi

$$\text{cov}(z_t, z_{t+k}) = E[(z_t - \mu)(z_{t+k} - \mu)] = \gamma_k$$

on sama jokaisella  $t$ , jolloin kovarianssi riippuu ainoastaan havaintojen välisestä ajallisesta etäisyydestä  $k$ . Tätä kutsutaan prosessin  $z$  autokovarianssiksi viiveellä  $k$ . Huomataan, että  $\gamma_0 = \text{Var}(z_t)$ .

Prosessia kutsutaan heikosti stationaariseksi kertaluvulla  $f$ , jos sen momentit  $f$ :nteen momenttiin asti eivät riipu ajankohdasta  $t$ . Tällöin esimerkiksi toisen kertaluvun mielessä heikosti stationaarisella prosessilla on vakio-odotusarvo

$$E(z_t) = \mu, \quad \forall t$$

ja havaintojen välinen autokovarianssi  $\gamma_k$  riippuu vain viiveestä  $k$ . Toisen kertaluvun heikko stationaarisuus ja normaalijakaumaoletus yhdessä riittävät vahvan stationaarisuuden toteamiseksi. Käytännön ongelmissa vahvan stationaarisuuden toteaminen voi kuitenkin olla hyvin vaikeaa, joten useimmiten tyydytään toisen kertaluvun heikkoon stationaarisuuteen, jota kutsutaan myös kovarianssistationaarisuudeksi [5]. Koska heikko stationaarisuus määritellään vain kahden ensimmäisen momentin kautta, niin sitä voidaan arvoida myös visuaalisesti.

## 2.2 Autoregressiivinen prosessi ja yksikköjuuri

Monet stokastiset prosessit, joissa peräkkäisten havaintojen välillä on riippuvuutta, voidaan esittää muodossa

$$z_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \psi(B)\varepsilon_t,$$

jossa  $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$  on ns. valkoista kohinaa, eli sarja satunnaisia ja riippumattomia muuttujia jostakin jakaumasta. Usein oletetaan vielä vahvemmin, että muuttujat ovat lisäksi normaalijakautuneita, eli  $\varepsilon_t \sim n.i.d(0, \sigma^2) \forall t$ . Näihin satunnaisshokkeihin on sovitettu eräänlainen lineaarinen “suodatinfunktio” tai operaattori

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

Tässä  $B$  on niin kutsuttu viivästysoperaattori:  $B\varepsilon_t = \varepsilon_{t-1}$ ,  $B^2\varepsilon_t = \varepsilon_{t-2}$  ja niin edelleen. Funktio  $\psi(B)$  siis muuttaa valkoisen kohinan aikasarjaksi  $z$ . Jos kertoimien sarja  $\psi_1, \psi_2, \dots$  on äärellinen, tai  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ , niin aikasarja  $z$  on stationaarinen ja vaihtelee tason  $\mu$  (eli odotusarvon) ympärillä. Muussa tapauksessa  $z$  on epästationaarinen eikä aikasarjan



taso ole pysyvä.

Autoregressiivinen prosessi astetta  $p$  eli  $AR(p)$ -prosessi on erikoistapaus ylläolevasta mallista, jossa aikasarjan jokainen havainto riippuu  $p$ :stä edellisestä havainnosta sekä virhetermistä  $\varepsilon_t$ . Olkoon  $\tilde{z}_t = z_t - \mu$  havaintojen poikkeama odotusarvosta. Tällöin prosessia, jossa

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + \varepsilon_t$$

kutsutaan  $AR(p)$ -prosessiksi. Jos merkitään edelleen operaattoria

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p,$$

niin  $AR(p)$ -prosessi voidaan kirjoittaa myös muodossa

$$\phi(B)\tilde{z}_t = \varepsilon_t,$$

josta nähdään yhtäpitävästi, että

$$\tilde{z}_t = \phi^{-1}(B)\varepsilon_t = \psi(B)\varepsilon_t$$

kun

$$\phi^{-1}(B) = \psi(B) = \sum_{j=0}^{\infty} \psi_j B^j,$$

jossa kertoimet  $\psi_j$  riippuvat operaattorin  $\phi(B)$  kertoimista  $\phi_j$ . Edellä mainitut stationaarisuuden ehdot kertoimien  $\psi_1, \psi_2, \dots$  suhteen ovat edelleen voimassa.  $AR(p)$ -prosessin stationaarisuus voidaan todeta myös käsittelemällä operaattoria  $\phi(B)$   $p$ :nnen asteen polynomina, jonka juuret toteuttavat yhtälön

$$\phi(B) = 0.$$

Jos kaikki  $\phi(B)$ :n juuret ovat itseisarvoltaan suurempia kuin yksi (ts. kompleksitasossa yksikköympyrän ulkopuolella), niin prosessi on stationaarinen. Kun  $p = 1$ , niin ehdon täyttymiseksi riittää, että  $|\phi_1| < 1$ . Mikäli kuitenkin  $\phi_1 = 1$ , niin  $AR(1)$ -prosessilla sanotaan olevan yksikköjuuri. Tällöin prosessi on epästationaarinen. Korkeamman asteen  $AR(p)$ -prosessit ovat yksikköjuuriprosesseja silloin, kun yksi tai useampi juurista on yksi. Käytännön aikasarjojen analyysin kannalta mitä lähempänä ykköstä jokin juurista on, sitä enemmän aikasarja alkaa käyttäytyä yksikköjuuriprosessille ominaisella tavalla (ks. luku 2.4). Juuret voivat olla myös kompleksitasossa yksikköympyrän sisäpuolella, jolloin aikasarjassa tyypillisesti esiintyy eksponentiaalista käyttäytymistä. Tässä tutkielmassa ei kuitenkaan olla kiinnostuneita tällaisista tapauksista.

Yksinkertaisin yksikköjuuriprosessi on niin kutsuttu satunnaiskulku (*random walk*), jossa

$$z_t = z_{t-1} + \varepsilon_t$$

ja virhetermiä  $\varepsilon_t$  koskevat samat oletukset kuin edellä, eli se on valkoista kohinaa [11].

Aikasarjan kahden peräkkäisen havainnon erotusta merkitään tavallisesti differenssioperaattorilla

$$\Delta z_t = z_t - z_{t-1}.$$

Yksikköjuuriprosessista saadaan stationaarinen differensoimalla sitä  $d$  kertaa, kun  $d$  on yksikköjuurten lukumäärä. Prosessin sanotaan siis olevan integroitunut asteella  $d$  eli olevan  $I(d)$ -prosessi. Vastaavasti stationaariseen prosessiin voidaan viitata  $I(0)$ -prosessina. Monesti aikasarjan mallintamiseen on syytä lisätä autoregressiivisen termin lisäksi niin sanottu liukuvan keskiarvon termi eli  $MA(q)$ -termi. Mallia, jossa on kummankin tyyppinen termi, kutsutaan  $ARMA(p, q)$ -prosessiksi, ja vielä yleisemmin  $ARIMA(p, d, q)$ -prosessiksi, jossa  $d$  on integroituneisuusaste.

### 2.3 Trendistationaarinen prosessi

Stokastinen prosessi voi koostua myös deterministisestä trendistä ja stationaarisesta virhetermistä. Tällainen niin sanottu trendistationaarinen prosessi on yksikköjuuriprosessin tapaan epästationaarinen (koska sen odotusarvo ei pysy samana ajan kuluessa), mutta sillä ei ole yksikköjuuria. Trendistationaarisen prosessin voi esittää muodossa

$$z_t = \mu_t + \varepsilon_t,$$

jossa  $\mu_t$  on deterministinen osa, kuten  $\mu_t = \beta t$  ( $\beta \neq 0$ ), ja stationaariset virhetermit  $\varepsilon_t$  ovat samoin jakautuneita ja riippumattomia, ts.  $\varepsilon_t \sim i.i.d(0, \sigma^2) \forall t$ .

Sekä trendistationaarisen prosessin että yksikköjuuriprosessin tyypilliset realisaatiot ovat trendimäisiä. Erona on se, että yksikköjuuriprosessin trendi on stokastinen, eli se voi muuttua satunnaisesti ajan kuluessa, kun taas trendistationaarisen prosessin trendi on deterministinen. [5] Trendistationaarista aikasarjaa ei tyypillisesti muuteta stationaariseksi differensoimalla, kuten yksikköjuuriprosessia, vaan poistamalla siitä trendi [14].

### 2.4 Yksikköjuuren käytännön merkitys

Yksikköjuuriprosessin ominaisuuksiin kuuluu niin sanottujen shokkien pysyvyys. Siis jos jokin ulkoinen tapahtuma tai tekijä aiheuttaa muutoksen esimerkiksi aikasarjan tasossa, niin tämä ”hyppäys” on pysyvä, eikä aikasarja palaudu takaisin entiselle tasolleen shokin

vaikutuksen lakattua. Toinen tyypillinen ominaisuus yksikköjuuriprosesseilla on varianssin kasvaminen ajan kuluessa. [5] Visuaalisella tarkastelullakin voi siis jo saada viitteitä aikasarjan luonteesta: mikäli aikasarja näyttää ”vaeltelevalta”, sen taso muuttuu porrasmaisesti tai vaihtelu on suurempaa jollakin aikavälillä, kannattaa yksikköjuurta testata. Myös trendimäisesti käyttäytyvän aikasarjan tapauksessa yksikköjuuritesteillä voidaan erottaa, onko kyseessä deterministinen vai stokastinen trendi eli yksikköjuuri. Tämä tieto on itsessäänkin mielenkiintoinen ja auttaa ymmärtämään taustalla vaikuttavaa prosessia, sillä deterministisen trendin voi olettaa jatkuvan samanlaisena, kun taas stokastisen trendin voi odottaa muuttuvan ajan kuluessa.

Yksikköjuuritestaaminen on kuitenkin tärkeää myös jatkoanalyysien kannalta, sillä monet mallinnusmenetelmät olettavat aikasarjan olevan stationaarinen. Jos yksikköjuuri-prosessiin sovelletaan esimerkiksi regressioanalyysia ikään kuin se olisi stationaarinen, ovat tulokset usein harhaanjohtavia [3, 9, 14]. Myös tutkittaessa kahden aikasarjan suhdetta toisiinsa tuloksena voi olla näennäinen riippuvuus, jos toinen tai molemmat näistä aikasarjoista ovat yksikköjuuriprosesseja [10, 14], tai jos aikasarjojen integroituneisuusasetta ei ole määritelty oikein [5]. Toisaalta yksikköjuuren olemassaolosta voi olla myös hyötyä: Rahman ym. (2013) muotoilevat stokastisen trendin olevan kuin prosessin yksilöllinen sormenjälki. Jos jostakin toisesta aikasarjasta löytyy samanlainen stokastinen trendi, voidaan näiden kahden aikasarjan välillä nähdä olevan vahva riippuvuus. Tällöin aikasarjojen sanotaan olevan yhteisintegroituneita. [14]

## 3 Yksikköjuuritestit

Tässä opinnäytetyössä käytetään kolmea yksikköjuuritestää: ADF-, PP- ja KPSS-testiä. Seuraavissa luvuissa esitellään näiden testien periaatteet kuten Kočendan kirjassa *Elements of Time Series Econometrics: An Applied Approach* [11].

### 3.1 Dickey–Fuller-testit

Kehittäjiensä mukaan nimettyihin Dickey–Fuller-testeihin kuuluvat yksinkertainen DF-testi ja laajennettu DF-testi eli ADF-testi (*augmented Dickey–Fuller test*). ADF-testi on näistä yleisemmin käytetty, sillä se soveltuu paremmin useimpiin käytännön tilanteisiin. ADF-testin idea selitetään seuraavassa DF-testin kautta.

#### 3.1.1 DF-testi

Aloitetaan tarkastelemalla yksinkertaista AR(1)-mallia

$$z_t = \phi z_{t-1} + \varepsilon_t, \quad (1)$$

jossa  $\varepsilon_t \sim i.i.d(0, \sigma^2) \forall t$ . Kuten ylempänä todettiin, niin AR(1)-prosessi on stationaarinen, kun  $|\phi| < 1$ , ja sillä on yksikköjuuri, kun  $\phi = 1$ . Jos  $|\phi| > 1$ , niin prosessi on epästationaarinen, ja sen varianssi kasvaa eksponentiaalisesti ajan kuluessa. Keskitytään kuitenkin tapauksiin, joissa  $|\phi| \leq 1$  ja  $\phi \neq -1$ . Ajatellaan AR(1)-mallia lineaarisena regressiona ja asetetaan hypoteesit

$$H_0 : \phi = 1$$

ja

$$H_1 : \phi < 1.$$

Regressiokertoimen  $\phi$  suurimman uskottavuuden estimaatti  $\hat{\phi}$  saadaan pienimmän neliösumman (PNS) menetelmällä

$$\hat{\phi} = \frac{\sum_{t=1}^N z_{t-1} z_t}{\sum_{t=1}^N z_{t-1}^2},$$

jossa  $z_0 = 0$  ja  $N$  on havaintojen lukumäärä. Nyt tuntuisi luontevalta käyttää nollahypoteesin testaamiseen tavallista yksisuuntaista t-testiä. Ongelmana on kuitenkin se, että t-testisuure ei noudata tavanomaista t-jakaumaa nollahypoteesin ollessa voimassa, sillä PNS-estimaatti on (alaspäin) harhainen, eikä se noudata asympotoottisesti normaalijakaumaa. Dickey ja Fuller ovatkin johtaneet yksikköjuuren testaamiseen tarkoitettulle t-testisuurelle uudet kriittiset arvot [4], jotka ovat t-jakauman kriittisiä arvoja matalammat. Näin nollahypoteesia (yksikköjuurta) voidaan testata t-testillä, kun vain verrataan saatua

t-testisuureen arvoa Dickeyn ja Fullerin kriittisiin arvoihin.

Vähentämällä yhtälön (1) molemmilta puolilta  $z_{t-1}$  saadaan AR(1)-prosessille, ilman vakiota tai trendikomponenttia, esitysmuoto

$$\begin{aligned}\Delta z_t &= (\phi - 1)z_{t-1} + \varepsilon_t \\ &= \gamma z_{t-1} + \varepsilon_t.\end{aligned}\tag{2}$$

Alkuperäisen nollahypoteesin voi nyt esittää muodossa

$$H_0 : \gamma = 0$$

ja vaihtoehdoisen hypoteesin muodossa

$$H_1 : \gamma < 0.$$

DF-testiä voi laajentaa myös tapauksiin, joissa AR(1)-prosessin lausekkeessa on vakio (3) tai vakio ja trendikomponentti (4). Tällöin saadaan esitykset

$$\Delta z_t = c + \gamma z_{t-1} + \varepsilon_t\tag{3}$$

ja

$$\Delta z_t = c + \beta t + \gamma z_{t-1} + \varepsilon_t,\tag{4}$$

joissa siis vakio on  $c$  ja trendikomponentti  $\beta t$ . Testattavaksi malliksi valitaan (2), (3) tai (4) yleensä aikasarjan visuaalisen tarkastelun perusteella. Jos aikasarja näyttää keskittyneen nollatason ympärille, valitaan (2) ja muuten, ellei havaittavissa ole trendiä, valitaan (3). Jos aikasarjassa näyttää olevan trendi, valitaan (4), ja tällöin nollahypoteesin hylkääminen tarkoittaa, että aikasarja on trendistationaarinen. Kaikille malleille t-testisuureen lauseke on

$$t_{DF} = \frac{\hat{\gamma}}{SE(\hat{\gamma})},$$

joissa  $\hat{\gamma}$  on kertoimen  $\gamma$  suurimman uskottavuuden estimaatti ja  $SE(\hat{\gamma})$  tämän keskivirhe. Dickeyn ja Fullerin kriittiset arvot t-testisuurelle ovat erilaiset riippuen valitusta mallista ja havaintojen lukumäärästä  $N$ . Jos  $t_{DF}$  on pienempi kuin kriittinen arvo, niin nollahypoteesi hylätään.

Käytännössä DF-testi on kuitenkin usein liian rajoittava, sillä se soveltuu vain AR(1)-malleihin. Yleensä käytetäänkin laajennettua DF-testiä eli ADF-testiä, jolloin mallin asteen osalta ei rajoitauduta AR(1)-tapaukseen. ADF-testi soveltuu myös malleille, joissa on MA( $q$ )-osa.

### 3.1.2 ADF-testi

Esitetään AR( $p$ )-prosessi muodossa

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_{p-2} z_{t-p+2} + \phi_{p-1} z_{t-p+1} + \phi_p z_{t-p} + \varepsilon_t.$$

Lisäämällä ja vähentämällä  $\phi_p z_{t-p+1}$  yhtälön molemmilta puolilta saadaan

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_{p-2} z_{t-p+2} + (\phi_{p-1} + \phi_p) z_{t-p+1} - \phi_p \Delta z_{t-p+1} + \varepsilon_t.$$

Lisäämällä ja vähentämällä jälleen  $(\phi_{p-1} + \phi_p) z_{t-p+2}$  saadaan

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + (\phi_{p-2} + \phi_{p-1} + \phi_p) z_{t-p+2} - (\phi_{p-1} + \phi_p) \Delta z_{t-p+2} - \phi_p \Delta z_{t-p+1} + \varepsilon_t$$

ja jatkamalla tällä tavalla saadaan yhtälölle lopulta muoto

$$z_t = \left( \sum_{i=1}^p \phi_i \right) z_{t-1} - \left( \sum_{i=2}^p \phi_i \right) \Delta z_{t-1} - \cdots - \left( \sum_{i=p-1}^p \phi_i \right) \Delta z_{t-p+2} - \phi_p \Delta z_{t-p+1} + \varepsilon_t.$$

Kun vielä vähennetään  $z_{t-1}$  yhtälön molemmilta puolilta, niin tuloksena on

$$\Delta z_t = \gamma z_{t-1} + \sum_{i=2}^p \delta_i \Delta z_{t-i+1} + \varepsilon_t, \quad (5)$$

jossa

$$\gamma = \sum_{i=1}^p \phi_i - 1$$

ja

$$\delta_i = - \sum_{j=i}^p \phi_j.$$

Nähdään, että saatu yhtälö (5) on sama kuin yksinkertaisen DF-testin yhtälö (2), mutta tähän on lisätty summalauseke  $\sum_{i=2}^p \delta_i \Delta z_{t-i+1}$ , jonka tarkoitus on korjata mahdollista virhetermien autokorrelaatiota. Kiinnostava parametri yksikköjuuritestissä on jälleen  $\gamma$ , ja hypoteesit voi muotoilla samoin kuin DF-testissäkin. Jos  $\gamma = 0$ , niin  $\sum_{i=1}^p \phi_i = 1$ , ja prosessissa on yksikköjuuri, ja jos  $\gamma < 0$ , niin  $\sum_{i=1}^p \phi_i < 1$ , ja prosessi on stationaarinen. Yksikköjuurihypoteesin testaaminen onnistuu täysin samalla tavalla kuin yksinkertaisemmassa DF-testissä: t-testisuure ja kriittiset arvot ovat samat. Myös ADF-testissä malliin voidaan lisätä vakio ( $c$ ) tai vakio ja trendikomponentti ( $c + \beta t$ ). [11]

Yhtenä käytännön ongelmana ADF-testissä on viiveparametrin  $p$  valinta, sillä se ei yleensä ole tiedossa etukäteen. Käytetään valitusta  $p$ :n arvosta selvyuden vuoksi merkintää

$p_v$ . Valitun  $p_v$ :n tulisi olla suurempi tai yhtäsuuri kuin mallin todellinen viiveparametri  $p$ , jotta testi toimisi hyvin. Testin voima kuitenkin heikkenee, mitä suurempi arvo  $p_v$ :lle asetetaan. Voiman heiketessä testi ei kykene hylkäämään virheellistä (paikkansapitämätöntä) nollahypoteesia, eli antaa tulokseksi epästationaarisuuden, vaikka aikasarja olisi oikeasti stationaarinen. Täten  $p_v$ :n valinta vaikuttaa olennaisesti testin tulokseen ja sen perusteella tehtäviin päätelmiin, mikä on yksi ADF-testin suurimmista heikkouksista. Erilaisia menetelmiä  $p_v$ :n valintaan on listattu esimerkiksi Maddalan ja Kimin kirjassa *Unit Roots, Cointegration, and Structural Change* [13]. Viiveparametrin valintaan voidaan käyttää esimerkiksi informaatiokriteerejä tai Hallin (1994) menetelmää, jossa  $p_v$ :tä kasvatetaan tai pienennetään asteittain jostakin lähtöarvosta, kunnes löydetään suurin mahdollinen tilastollisesti merkitsevä arvo. Viiveparametri suositellaan valittavan jollakin aineistoon perustuvalla menetelmällä, eikä sitä yleensä kannata kiinnittää sattumanvaraiseksi luvuksi. [7, 13]

### 3.2 Phillips–Perron-testi

Phillips–Perron-testi eli PP-testi on yksi monista ADF-testin pohjalta kehitetyistä yksiköjuuritesteistä. Sen idea on sama kuin ADF-testissä, mutta t-testisuure on

$$t_{PP} = \frac{\hat{\sigma}_\delta}{\hat{\sigma}} t_{DF} - 0.5(\hat{\sigma}^2 - \hat{\sigma}_\delta^2) \left[ \hat{\sigma} \left( N^{-2} \sum_{t=2}^N (z_{t-1} - \bar{z}_{-1})^2 \right)^{1/2} \right]^{-1},$$

jossa  $N$  on jälleen havaintojen lukumäärä ja

$$\bar{z}_{-1} = (N - 1)^{-1} \sum_{t=1}^{N-1} z_t.$$

Lisäksi

$$\hat{\sigma}_\delta^2 = N^{-1} \sum_{t=1}^N \hat{\varepsilon}_t^2$$

on estimaatti parametrille

$$\sigma_\delta^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{t=1}^N E(\varepsilon_t^2).$$

Vastaavasti

$$\hat{\sigma}^2 = N^{-1} \sum_{t=1}^N \hat{\varepsilon}_t^2 + 2N^{-1} \sum_{s=1}^N w(s, l) \sum_{t=s+1}^N \hat{\varepsilon}_t \hat{\varepsilon}_{t-s}$$

on estimaatti parametrille

$$\sigma^2 = \lim_{N \rightarrow \infty} E \left[ N^{-1} \left\{ \sum_{t=1}^N \varepsilon_t \right\}^2 \right].$$

Painot  $w(s, l) = 1 - s/(l + 1)$  takaavat estimaatin  $\hat{\sigma}^2$  positiivisuuden. PP-testin käyttäjän ei tarvitse valita viiveparametrin  $p$  arvoa korjataksean mahdollista virhetermien autokorrelaatiota kuten ADF-testissä, mutta PP-testissä on valittava painoissa  $w(s, l)$  käytettävä kaistanleveysparametrin  $l$  arvo. [11] Tässä opinnäytetyössä käytetyssä R-ohjelmiston tseries-kirjaston pp.test-funktiossa vaihtoehtoina ovat Schwertin (1989) ehdottamat  $l = \text{int}\{4(N/100)^{0.25}\}$  ja  $l = \text{int}\{12(N/100)^{0.25}\}$ , jossa  $\text{int}\{\}$  tarkoittaa sitä, että  $l$  pyöristetään alaspäin lähimpään kokonaislukuun [17]. Testisuureiden eroista huolimatta PP-testi tuottaa kuitenkin yleensä samankaltaisia tuloksia kuin ADF-testi [11].

### 3.3 KPSS-testi

KPSS-testi on nimetty kehittäjiensä Kwiatkowskin, Phillipsin, Schmidtin ja Shinin mukaan, ja se on kehitelty vastauksena ADF-testin heikkoon voimaan eli taipumukseen jättää nollahypoteesi voimaan. Sen tärkein ero verrattuna ADF- ja PP-testeihin on hypoteesien kääntäminen toisin päin, eli KPSS-testin nollahypoteesi on stationaarisuus ja vaihtoehtoinen hypoteesi yksikköjuuren olemassaolo. Prosessi, jonka juuret ovat lähellä ykköstä, mutta eivät kuitenkaan yksikköjuuria, tulee KPSS-testin avulla useammin identifioitua oikein stationaariseksi.

KPSS-testi on myös teknisesti erilainen kuin ADF- ja PP-testit. Siinä aikasarjaa  $z$  mallinnetaan seuraavasti:

$$z_t = \beta t + r_t + \varepsilon_t,$$

jossa  $\beta t$  on deterministinen trendi,  $\varepsilon_t$  on stationaarinen virhetermi ja  $r_t = r_{t-1} + u_t$ , jossa  $u_t \sim N(0, \sigma_u^2)$  eli  $r_t$  on satunnaiskulku. Trendiosa voidaan edellisten testien tapaan myös jättää pois, jollei sille ole perusteita. Hypoteesit ovat

$$H_0 : \sigma_u^2 = 0$$

ja

$$H_1 : \sigma_u^2 \neq 0.$$

Nollahypoteesin tapauksessa aikasarja on (trendi)stationaarinen, sillä silloin

$$r_t = r_{t-1} = r_0, \quad \forall t,$$



eli  $r_t$  supistuu vakiotermitiksi. Jos taas  $\sigma_u^2 \neq 0$  eli vastahypoteesi on voimassa, niin aikasarja on yksikköjuuriprosessi.

Testausta varten estimoidaan aikasarjalle  $z$  ensin PNS-menetelmällä regressiosuora

$$z_t = a_0 + \beta t + e_t$$

tai

$$z_t = a_0 + e_t$$

sen mukaan, sisällytetäänkö malliin trendikomponentti vai ei. Residuaalien  $e_t$  avulla lasketaan niin sanottu LM-testisuure

$$\eta_{\mu/\tau} = N^{-2} \frac{1}{s^2(l)} \sum_{t=1}^N S_t^2,$$

jossa  $S_t$  on residuaalisarjan  $t$ :s osasumma eli

$$S_t = \sum_{i=1}^t e_i$$

ja  $s^2(l)$  on kaistanleveysparametrilla  $l$  riippuva residuaalien varianssin estimaattori

$$s^2(l) = N^{-1} \sum_{t=1}^N e_t^2 + 2N^{-1} \sum_{s=1}^l w(s, l) \sum_{t=s+1}^N e_t e_{t-s},$$

jossa  $w(s, l) = 1 - s/(l + 1)$  on niin sanottu Bartlett-ikkuna. Residuaalien mahdollista autokorrelaatiota korjataan kasvattamalla  $l$ :ää suuremmaksi kuin 0. LM-testisuureesta käytetään merkintää  $\eta_{\mu}$ , jos mallissa ei ole trendikomponenttia, ja  $\eta_{\tau}$ , jos mallissa on trendikomponentti. Testisuureen arvoa verrataan kriittisiin arvoihin [12]. Jos testisuureen arvo on suurempi kuin kriittinen arvo, niin nollahypoteesi stationaarisuudesta hylätään.

KPSS-testin käyttäjä törmää samankaltaiseen ongelmaan kuin ADF- ja PP-testienkin tapauksessa: testin tulos riippuu jälleen omavalintaisen kaistanleveysparametrin  $l$  arvosta. Tavallisesti  $l$  asetetaan nollan ja kahdeksan välille, ja mitä suurempi sen arvo on, sitä helpommin nollahypoteesi jää voimaan. Kuitenkin jos nollahypoteesi jää voimaan pienilläkin  $l$ :n arvoilla (0, 1 tai 2), niin voidaan melko varmasti sanoa aikasarjan olevan stationaarinen. [11] Tässä opinnäytetyössä käytetään R-ohjelmiston tseries-kirjaston kpss.test-funktiota, jossa valmiina vaihtoehtoina ovat PP-testin tapaan  $l = \text{int}\{4(N/100)^{0.25}\}$  ja  $l = \text{int}\{12(N/100)^{0.25}\}$ .

## 4 Ilmastoaikasarjat ja yksikköjuuriprosessit

Ilmastonmuutokseen liittyvistä muuttujista erityisesti maapallon lämpötiloja on tutkittu paljon: onko globaali lämpeneminen todella seurausta ihmisen toiminnasta, vai osuuko syypäänä pidetty kasvihuonekaasujen nousu vain sattumalta samaan ajankohtaan, jolloin lämpötilat nousisivat muutenkin? Tiedetään, että maapallon keskilämpötila on aina vaihdellut luonnollisesti, kuten esimerkiksi jääkaudet osoittavat. Siispä olisi teoriassa mahdollista, että maapallolla olisi nytkin käynnissä vain tällaisen luonnollisen syklin vaihtelu. Tilastotieteen näkökulmasta on ajateltu, että luonnollisen vaihtelun hypoteesi saa tukea silloin, kun lämpötila-aikasarjoissa on yksi tai useampia yksikköjuuria. Tällöin lämpötila-aikasarjaa generoivan prosessin trendikomponentti olisi stokastinen eli vaihtelisi satunnaisesti ajan kuluessa. [5, 10] Samaa logiikkaa voisi luontevasti soveltaa myös muihin ilmastomuuttujiin: esimerkiksi Suomessa ennustettu sademäärien kasvu tulkittaisiin osaksi sateisuuden luonnollista vaihtelua, jos kyseessä olisi yksikköjuuriprosessi.

Asia ei ole kuitenkaan välttämättä näin yksinkertainen. Esimerkiksi Kelly (2000) argumentoi, että globaali lämpeneminen vaikuttaisi olevan yksikköjuuriprosessi, mutta tämä ei silti tarkoita sitä, etteivätkö kasvihuonekaasut vaikuttaisi siihen. Itse asiassa kasvihuonekaasupäästöjen nousu on saattanut toimia yksikköjuuriprosessille ominaisena pysyvänä shokkina, joka voi saada lämpötilan nousun jatkumaan hyvin pitkän aikaa. [10] Ehkä yksikköjuuritestien tuloksia ei siis pitäisikään käyttää suoraviivaisesti ilmastollisten trendien syntymekanismien tulkintaan, vaan enemmänkin apuvälineenä niiden tulevan käyttäytymisen ennustamiseen ja mallien sovittamiseen.

Yksikköjuuriprosessin identifioiminenkaan ei aina ole helppoa. Luvussa 3 mainittujen testien sisäisten ongelmien (kuten viiveparametrin valinnan) lisäksi eri testit voivat antaa erilaisia tuloksia samalle aikasarjalle. Esimerkiksi Estradan ja Perronin (2019) artikkelissa todetaan, että perinteinen ADF-testi puhuu vahvasti yksikköjuuren puolesta globaaleissa lämpötila-aikasarjoissa, mutta yksikköjuuritesti, joka sallii muutoksen aikasarjan trendissä, antaa erilaisen tuloksen. Globaaleissa lämpötila-aikasarjoissa näkyy nimittäin trendin kulmakertoimen muutos viime vuosisadan puolivälin jälkeen, jolloin sekä lämpeneminen että päästöjen kasvu kiihtyivät. Kun käytetään yksikköjuuritestejä, jotka ottavat tällaisen rakenteellisen muutoksen huomioon, niin testit hylkäävät yksikköjuurihypoteesin ja tukevat aikasarjojen trendistationaarisuutta. [6] Yleisesti ottaen ADF-testi luokittelee aikasarjan usein yksikköjuuriprosessiksi, jos siinä esiintyy pysyviä muutoksia vakiotermin tai trendin suhteen, tulkiten ne yksikköjuurelle tyypillisiksi pysyviksi shokeiksi. Tämän ongelman ratkaisemiseksi on kehitetty omat yksikköjuuritestinsä. [11] Tässä opinnäytetyössä analysoidussa sadeaineistossa ei kuitenkaan näytä olevan selviä rakenteellisia muutoksia (kuva 3), jotka aiheuttaisivat ongelmia niin sanottujen perinteisten yksikköjuuritestien käytössä. Joka tapauksessa on aina turvallisempaa käyttää useita yksikköjuuritestejä ja

verrata tulosten yhteneväisyyttä. Esimerkiksi Kočenda suosittelee käyttämään sekä ADF-että KPSS-testiä niiden erilaisen hypoteesinasettelu takia (ks. luku 3). Jos molemmat testit antavat saman tuloksen, sitä voi pitää luotettavana. [11] Kirjallisuudessa näkeekin usein käytettävän molempia testejä.

Sateisuuteen liittyvien muuttujien aikasarjaominaisuuksia on tutkittu jonkin verran eri puolilla maailmaa, mutta yksikköjuuritestauksen näkökulmasta melko vähän. Kuitenkin esimerkiksi Rahman ym. (2013) ovat käyttäneet yksikköjuuritestejä Bangladeshista Dinajpurin alueelta vuosina 1948–2002 kerättyyn ilmankosteus-, lämpötila- ja sadeaineistoon. Sademäärät Dinajpurissa noudattavat hyvin erityyppistä kaavaa kuin Suomessa: kesäkuukausina sataa rankasti, ja talvikuukaudet ovat hyvin kuivia. Rahman ym. eivät tarkastele kuukausia tai vuodenaikoja erikseen yksikköjuuritesteissä, vaan testattavana on vuoden keskiarvo, minimi, maksimi, sekä minimin ja maksimin erotus. Käytetyt testit ovat ADF, PP, KPSS sekä eräänlainen saapasremmimenetelmä. Kaikki käytetyt testit viittaavat sadeaikasarjojen trendistationaarisuuteen, kun nollahypoteesina on satunnaiskulku stokastisella trendillä ml. vakiotermin (*drift*). Kuitenkin jos nollahypoteesiksi asetetaan satunnaiskulku ilman trendikomponenttia, niin testit tukevat aikasarjojen epästationaarisuutta lukuunottamatta minimisademäärää, joka on stationaarinen. Kuvan perusteella tämän alueen sademäärissä ei olekaan kovin selvää trendiä. Kirjoittajat tulkitsevat kaikkien mitattujen muuttujien olevan minimisademäärää lukuun ottamatta  $I(1)$ -prosesseja. [14]

Jarvis ym. (2013) ovat niin ikään käyttäneet ADF-, PP- ja KPSS-testejä sadeaikasarjoihin tutkiessaan sateen vaikutusta joen virtaamaan Australiassa. ADF- ja PP-testit tukevat suurimmaksi osaksi sademuuttujien stationaarisuutta, mutta KPSS-testin tulokset ovat välillä erisuuntaisia. Kirjoittajat nostavat kuitenkin esiin erityisesti KPSS-testin heikkouksia, kuten taipumuksen hylätä nollahypoteesi stationaarisuudesta liian helposti ja herkkyyden edellämaintun kaistanleveysparametrin valinnan osalta. Tämän takia tutkimuksessa on annettu enemmän painoarvoa ADF- ja PP-testien tuloksille, vaikkakin näillä testeillä heikkoudet ovat melko samankaltaisia (ks. luku 3). Tulkinallinen johtopäätös ADF- ja PP-testien perusteella on se, että ilmastonmuutos ei näy tämän alueen sademäärissä. [9] Myös Saidi ym. (2014) ovat käyttäneet KPSS- ja PP-testejä sadeaineistoon. Tutkimuksessa on käytetty Luoteis-Italiasta jopa 5 minuutin välein kerättyjä havaintoja ja keskitytty rankkasateiden esiintymisiin. Toisin kuin Jarvisin ym. (2013) tutkimuksessa, niin testien perusteella suurimmassa osassa aikasarjoja ei voida hylätä deterministisen trendin mahdollisuutta tässä aineistossa. [16]

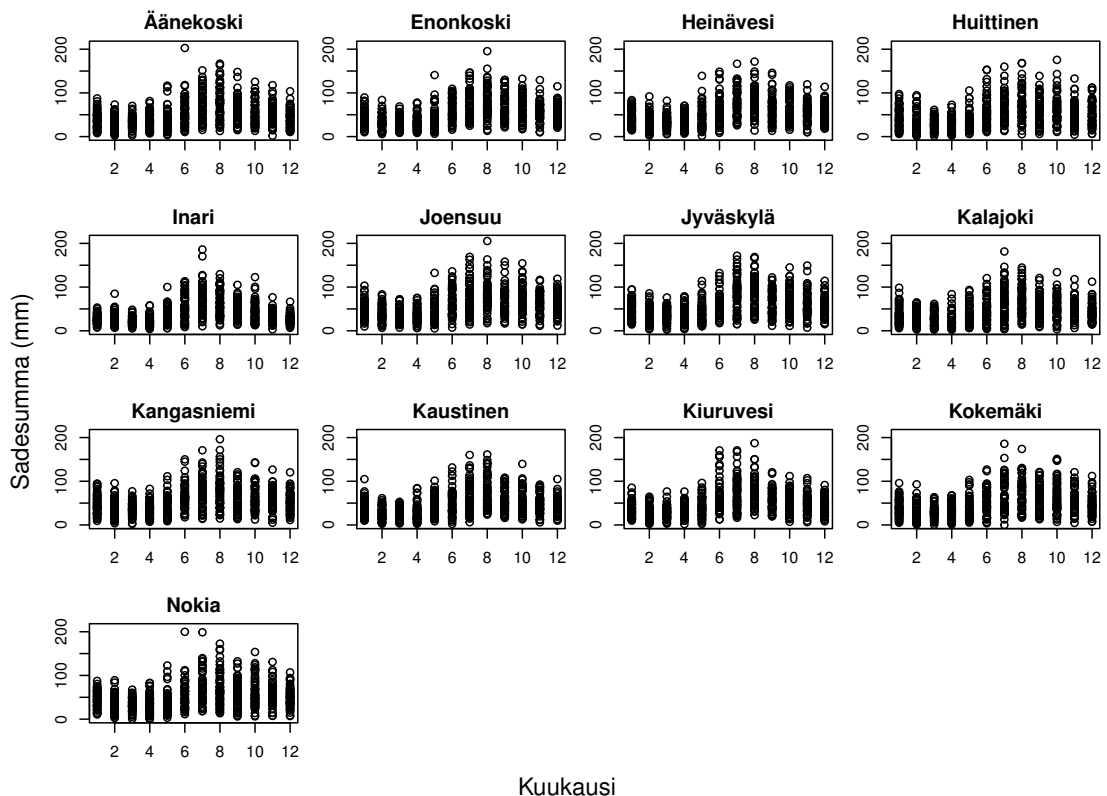
Clarkin ym. (2000) tutkimuksessa lämpötila- ja sadeaineistoa on kerätty Kanadasta, joka on ilmastollisesti lähempänä Suomea ja siten ehkä edellisiä vertailukelpoisempi tämän tutkielman kannalta. Tutkimuksessa analysoidaan koko vuoden sadesummaa yhdeksässä kanadalaisessa kaupungissa, eivätkä tulokset anna tukea yleiselle deterministiselle tren-

dille Kanadan sademäärissä, vaikka kahden kaupungin kohdalta tällainen ehkä löytyykin. Artikkelin johtopäätöksissä todetaan ilmastonmuutoksen vaikutuksilla olevan alueellisia eroja, ja ettei nouseva lämpötila aina korreloi kasvavien sademäärien kanssa. Tutkimus on tosin tehty yli 20 vuotta sitten, joten tuoreemmalla aineistolla tulokset saattaisivat olla erilaisia. Artikkelissa kritisoidaan myös Dickey–Fuller-yksikköjuuritestejä muun muassa siitä, että näiden testien käyttäjän tulee tehdä valinta deterministisen trendin sisällyttämisestä malliin ennen testaamista. Samaan aikaan testitulosten tarkoituksena on kuitenkin kertoa, onko aikasarjassa determinististä trendiä vai ei. Dickey–Fuller-testeihin liittyy siis kirjoittajien mukaan tietynlaista kehäpäättelyä, joten he käyttävätkin muita menetelmiä deterministisen trendin identifioimiseen. Artikkelin lopussa kuitenkin mainitaan, että myös Dickey–Fuller-testien tulokset olivat linjassa heidän käyttämiensä muiden testien kanssa. [3]

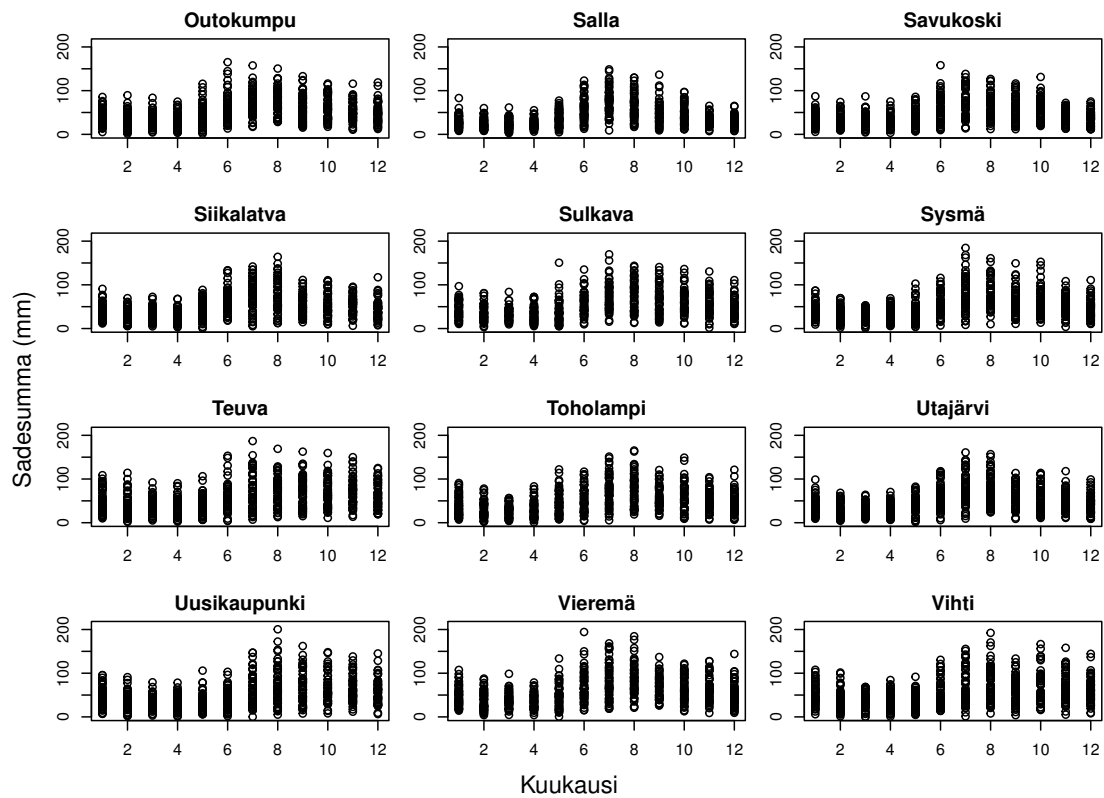
Näiden eri maissa tehtyjen tutkimusten vaihtelevista tuloksista huomataan, että toisin kuin lämpötiloista, niin sadeaikasarjojen käyttäytymisestä on vaikea tehdä minkäänlaisia globaaleja päätelmiä. Tämä tuntuukin aivan loogiselta, kun ottaa huomioon sateisuuden alueellisen vaihtelun ja esimerkiksi IPCC:n erilaiset ennusteet eri leveyspiireille [8]. Tutkimusten perusteella näyttäisi siltä, että joillakin alueilla yksikköjuuritestit tukevat determinististä trendiä sateisuudessa, kun taas toisaalla sadeaikasarjat vaikuttavat stationaarisilta eli odotusarvomielessä muuttumattomilta. Tämän takia onkin tärkeää tutkia sadeaikasarjojen stationaarisuutta sellaisen alueen sisällä, jossa suuria eroja keskimääräisessä sateisuudessa ei ole. Tällöin ennusteidenkin tekeminen on mielekästä.

## 5 Aineisto ja menetelmät

Tässä opinnäytetyössä on käytetty Ilmatieteen laitoksen eri puolilta Suomea keräämää avointa sadeaineistoa. Aineistoon valikoitiin kaikki sellaiset Ilmatieteen laitoksen havaintoasemat, joilla on mitattu kuukausittaista sadesummaa säännöllisesti vähintään tammi-kuusta 1965 joulukuuhun 2021. Aloitusvuodeksi valittiin 1965, koska tätä vanhempia havaintoja oli vähemmän saatavilla. Samasta kaupungista ei otettu kahta asemaa, jos molemmista oli havaintoja, vaan valittiin aakkosjärjestyksessä ensimmäinen aseman nimen mukaan. Näin vältetään jonkin alueen mittausten painottuminen havainnoissa. Havaintoasemia on aineistossa 25, ja niiden nimet ja koordinaatit on listattu liitteessä 1. Asemat jakautuvat melko tasaisesti eri puolille Suomea lukuunottamatta Lapin maakuntaa, josta saatiin valittua vain kolme asemaa (Inari, Salla ja Savukoski). Kuten kuvista 1 ja 2 nähdään, niin Lapin mittauspisteiden tulokset eivät kuitenkaan eroa merkittävästi eteläisempien asemien tuloksista, eikä asemien välillä ole muutenkaan yleisesti ottaen suuria eroja. Kuvista nähdään myös, että Suomessa sataa melko tasaisesti ympäri vuoden, mutta kuitenkin selvästi eniten kesällä ja syksyllä. (Kuva 1, kuva 2.)

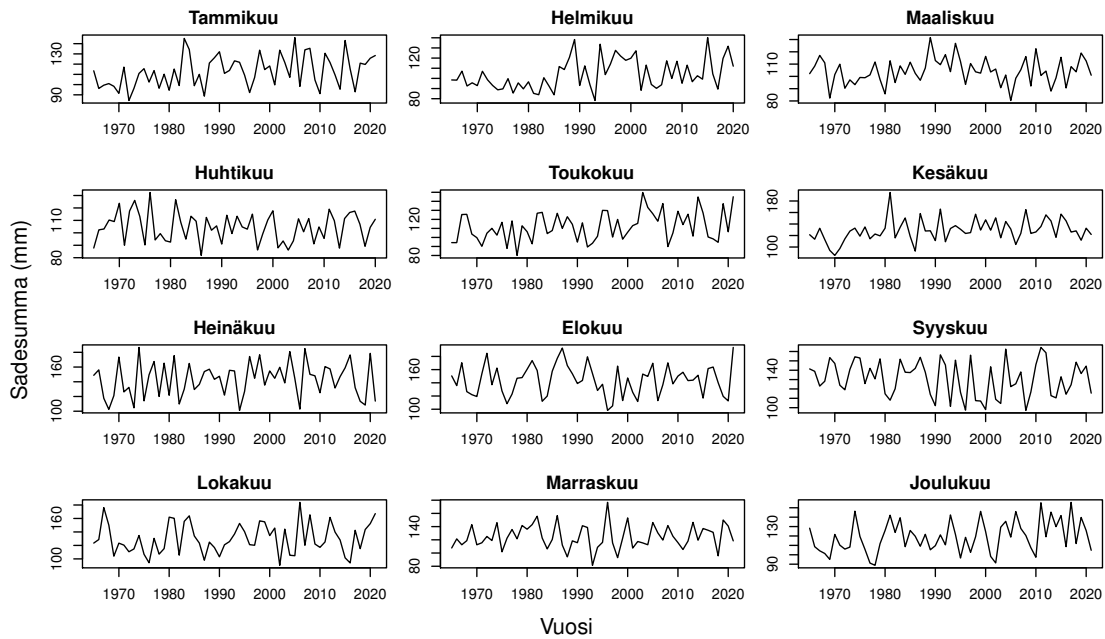


Kuva 1: Kuukausittaiset sadesummat havaintoaseman mukaan (ensimmäiset 13 asemaa).



Kuva 2: Kuukausittaiset sadesummat havaintoaseman mukaan (jatkoa kuvalle 1).

Analyyseja varten laskettiin kaikkien asemien havainnot kattava ”koko Suomen” keskiarvo jokaisen vuoden jokaiselle kuukaudelle, ja nämä eristettiin omiksi aikasarjoikseen (kuva 3). Keskiarvoja laskiessa ei otettu huomioon sellaisia asemia, joilta kyseisen kuukauden havainto puuttui. Analysoimalla aineistoa kuukausitasolla päästään eroon kausivaihtelusta, mutta nähdään myös, onko kuukausien välillä eroja stationaarisuudessa. Tämä lähestymistapa on perusteltu, sillä aiemmissa tutkimuksissa sateisuuden on ennustettu muuttuvan Suomessa eri tavoin eri vuodenaikoina [15]. Kuvaa 3 tarkastelemalla voikin jo nähdä joitakin eroja kuukausien välillä: esimerkiksi huhtikuun aikasarja näyttää varsin stationaariselta, kun taas helmikuussa varianssi kasvaa selvästi 80-luvun puolivälin jälkeen, mikä on yksi yksikköjuuriprosessille ominaisista piirteistä. Kuukausittaisten sademäärien vaihtelun onkin ennustettu kasvavan Suomessa [15]. Varianssin kasvu yksinään ei kuitenkaan tarkoita sitä, että kysessä olisi yksikköjuuriprosessi, eikä missään aikasarjassa toisaalta näytä selvästi esiintyvän yksikköjuuriprosessille tyypillistä ”vaeltelua”.



Kuva 3: Kuukausittaisen sadesumman vuosittaiset keskiarvot kuukauden mukaan.

Kuukausikohtaisia aikasarjoja testattiin ADF-, PP- ja KPSS-testeillä. ADF-testiä käytettiin viiveparametrin arvoilla 0–3, sillä estimoitujen autokorrelaatiofunktioiden ja Box–Ljung-testien perusteella missään aikasarjassa ei ollut merkitsevää viivettä enää kolmen vuoden jälkeen. Myöskään taustalla olevan ilmiön (sateisuus) näkökulmasta kovin monen vuoden viiveen sisällyttäminen malliin ei välttämättä tunnu järkevältä. PP- ja KPSS-testeissä käytettiin R-ohjelmiston tseries-kirjaston `pp.test`- ja `kps.test`-funktioita, joissa vaihtoehtoina kaistanleveysparametrin  $l$  arvolle on joko  $l = \text{int}\{4(N/100)^{0.25}\} = \text{int}\{4(56/100)^{0.25}\} = 3$  tai  $l = \text{int}\{12(N/100)^{0.25}\} = \text{int}\{12(56/100)^{0.25}\} = 10$ . PP-testi tehtiin molemmilla  $l$ :n arvoilla, mutta KPSS-testi vain pienemmällä. KPSS-testillä haluttiin saada lisävarmistusta muiden testien tuloksille, johon pieni  $l$  sopii paremmin, sillä  $l$ :n ollessa suuri KPSS-testin voima heikkenee (ks. luku 3.3). ADF- ja KPSS-testit muodostettiin kaikilla kuukausilla sekä ilman trendikomponenttia että sen kanssa, jolloin voitiin testata mahdollista trendistationaarisuutta.

## 6 Tulokset

Kun ADF-testiin sisällytettiin pelkkä vakio, testi hylkäsi nollahypoteesin yksikköjuuri-prosessista suurimmassa osassa kuukausista käytetystä viiveestä riippumatta. Kuitenkin helmikuun aikasarjan tapauksessa nollahypoteesi jäi selkeästi voimaan viiveillä 2 ja 3, sekä kesäkuun aikasarjassa viiveellä 3. Tammikuu on 5%:n riskitasoa käyttäen tiukasti ottaen epästationaarinen ( $p=0,058$ ) viiveellä 3, mutta vahvaa näyttöä epästationaarisuuden toteamiseksi ei ole. Maaliskuun tulos on samankaltainen. Onkin loogista, että helmikuun aikasarjan ominaisuudet näkyvät jonkin verran myös ajallisesti vierekkäisissä aikasarjoissa. (Taulukko 1.)

| Kuukausi  | ADF-testin (vakio, ei trendiä) t-testisuureen arvot eri viiveillä ( $p_v$ ) |                     |                                       |  |
|-----------|---|---------------------|---------------------------------------|--|
|           | $p_v = 0$   | $p_v = 1$           | $p_v = 2$                             | $p_v = 3$                              |
| Tammikuu  | -7,149 ( $p<0,01$ )   | -5,206 ( $p<0,01$ ) | -3,826 ( $p<0,01$ )                   | -2,874 ( <b><math>p=0,058</math></b> ) |
| Helmikuu  | -5,977 ( $p<0,01$ )   | -3,609 ( $p<0,01$ ) | -2,356 ( <b><math>p=0,19</math></b> ) | -2,231 ( <b><math>p=0,23</math></b> )  |
| Maaliskuu | -7,260 ( $p<0,01$ )   | -4,651 ( $p<0,01$ ) | -3,576 ( $p<0,01$ )                   | -2,921 ( <b><math>p=0,051</math></b> ) |
| Huhtikuu  | -8,624 ( $p<0,01$ )   | -6,232 ( $p<0,01$ ) | -4,867 ( $p<0,01$ )                   | -3,642 ( $p<0,01$ )                    |
| Toukokuu  | -6,783 ( $p<0,01$ )   | -4,561 ( $p<0,01$ ) | -3,810 ( $p<0,01$ )                   | -3,453 ( $p=0,015$ )                   |
| Kesäkuu   | -7,161 ( $p<0,01$ )   | -4,858 ( $p<0,01$ ) | -3,267 ( $p=0,023$ )                  | -2,426 ( <b><math>p=0,16</math></b> )  |
| Heinäkuu  | -9,708 ( $p<0,01$ )   | -6,799 ( $p<0,01$ ) | -5,381 ( $p<0,01$ )                   | -3,789 ( $p<0,01$ )                    |
| Elokuu    | -5,979 ( $p<0,01$ )   | -5,353 ( $p<0,01$ ) | -4,794 ( $p<0,01$ )                   | -4,627 ( $p<0,01$ )                    |
| Syyskuu   | -7,115 ( $p<0,01$ )   | -6,365 ( $p<0,01$ ) | -4,297 ( $p<0,01$ )                   | -3,995 ( $p<0,01$ )                    |
| Lokakuu   | -6,563 ( $p<0,01$ )   | -4,885 ( $p<0,01$ ) | -3,877 ( $p<0,01$ )                   | -4,105 ( $p<0,01$ )                    |
| Marraskuu | -7,738 ( $p<0,01$ )   | -7,600 ( $p<0,01$ ) | -6,041 ( $p<0,01$ )                   | -4,494 ( $p<0,01$ )                    |
| Joulukuu  | -6,577 ( $p<0,01$ )   | -4,394 ( $p<0,01$ ) | -4,412 ( $p<0,01$ )                   | -4,114 ( $p<0,01$ )                    |

Taulukko 1: ADF-testin tulokset eri viiveillä, kun malliin on sisällytetty vakio, mutta ei trendiä. Suluissa p-arvo (korostettuna p-arvot, jotka ovat suurempia kuin 0,05).

Aikasarjoja testattiin myös ADF-testillä, johon sisällytettiin vakio ja trendi. Tällöin helmikuun aikasarja näyttää jälleen yksikköjuuri-prosessilta viiveillä 2 ja 3, sekä maaliskuu ja kesäkuu viiveellä 3. Muuten testitulokset puhuvat trendistationaarisuuden puolesta. (Taulukko 2.)



|           | ADF-testin (vakio ja trendi) t-testisuureen arvot eri viiveillä ( $p_v$ ) |                 |                  |                  |
|-----------|---|-----------------|------------------|------------------|
| Kuukausi  | $p_v = 0$   | $p_v = 1$       | $p_v = 2$        | $p_v = 3$        |
| Tammikuu  | -8,249 (p<0,01)   | -6,450 (p<0,01) | -5,058 (p<0,01)  | -3,953 (p=0,018) |
| Helmikuu  | -6,792 (p<0,01)   | -4,337 (p<0,01) | -3,120 (p=0,12)  | -2,985 (p=0,18)  |
| Maaliskuu | -7,227 (p<0,01)   | -4,654 (p<0,01) | -3,636 (p=0,038) | -3,029 (p=0,16)  |
| Huhtikuu  | -8,658 (p<0,01)   | -6,359 (p<0,01) | -5,074 (p<0,01)  | -3,851 (p=0,023) |
| Toukokuu  | -7,394 (p<0,01)   | -5,067 (p<0,01) | -4,592 (p<0,01)  | -4,684 (p<0,01)  |
| Kesäkuu   | -7,692 (p<0,01)   | -5,349 (p<0,01) | -3,701 (p=0,033) | -2,611 (p=0,33)  |
| Heinäkuu  | -9,831 (p<0,01)   | -7,107 (p<0,01) | -5,743 (p<0,01)  | -3,806 (p=0,024) |
| Elokuu    | -5,898 (p<0,01)   | -5,291 (p<0,01) | -4,721 (p<0,01)  | -4,575 (p<0,01)  |
| Syyskuu   | -7,193 (p<0,01)   | -6,526 (p<0,01) | -4,496 (p<0,01)  | -4,321 (p<0,01)  |
| Lokakuu   | -6,581 (p<0,01)   | -4,915 (p<0,01) | -3,993 (p=0,016) | -4,349 (p<0,01)  |
| Marraskuu | -7,675 (p<0,01)   | -7,554 (p<0,01) | -5,988 (p<0,01)  | -4,442 (p<0,01)  |
| Joulukuu  | -7,324 (p<0,01)   | -5,087 (p<0,01) | -5,344 (p<0,01)  | -5,411 (p<0,01)  |

Taulukko 2: ADF-testin tulokset eri viiveillä, kun malliin on sisällytetty vakio ja trendi. Suluissa p-arvo (korostettuna p-arvot, jotka ovat suurempia kuin 0,05).

PP-testi antaa molemmilla käytetyillä kaistanleveysparametrin  $l$  arvoilla tulokset selvän stationaarisuuden kaikkien aikasarjojen kohdalla (Taulukko 3).

|           | PP-testin t-testisuureen arvot eri $l$ :n arvoilla |                  |
|-----------|--|------------------|
| Kuukausi  | $l = 3$  | $l = 10$         |
| Tammikuu  | -8,435 (p<0,01)                                    | -8,392 (p<0,01)  |
| Helmikuu  | -6,852 (p<0,01)                                    | -6,847 (p<0,01)  |
| Maaliskuu | -7,253 (p<0,01)                                    | -7,353 (p<0,01)  |
| Huhtikuu  | -8,927 (p<0,01)                                    | -9,072 (p<0,01)  |
| Toukokuu  | -7,391 (p<0,01)                                    | -7,708 (p<0,01)  |
| Kesäkuu   | -7,691 (p<0,01)                                    | -7,694 (p<0,01)  |
| Heinäkuu  | -10,561 (p<0,01)                                   | -10,988 (p<0,01) |
| Elokuu    | -5,769 (p<0,01)                                    | -5,658 (p<0,01)  |
| Syyskuu   | -7,213 (p<0,01)                                    | -7,262 (p<0,01)  |
| Lokakuu   | -6,567 (p<0,01)                                    | -6,839 (p<0,01)  |
| Marraskuu | -8,096 (p<0,01)                                    | -8,229 (p<0,01)  |
| Joulukuu  | -7,330 (p<0,01)                                    | -8,122 (p<0,01)  |

Taulukko 3: PP-testin tulokset eri  $l$ :n arvoilla. Suluissa p-arvo.

KPSS-testin tulokset (kun kaistanleveysparametri  $l = 3$ ) ovat suurimmalta osin linjassa muiden testien kanssa. Kun testattavassa mallissa ei ollut trendikomponenttia, niin tammi-, helmi-, touko-, kesä- ja joulukuun aikasarjat olivat epästationaarisia, tosin kesä- ja joulukuulla p-arvo oli hyvin lähellä nollahypoteesin hylkäämisrajaa. Kun malliin sisällytettiin trendi, eli nollahypoteesina oli trendistationaarisuus, niin kaikki valitut kuukaudet olivat trendistationaarisia. (Taulukko 4). Tuloksia verratessa tulee huomata, että KPSS-testissä

hypoteesinasettelu on päinvastainen kuin ADF- ja PP-testissä, eli pieni p-arvo tarkoittaa epästationaarisuutta. R-ohjelmiston kpss.test-funktio ei anna tarkkaa p-arvoa, kun p-arvo on suurempi kuin 0,1.

|                  | <b>KPSS-testin LM-testisuureen arvot (<math>l = 3</math>)</b> |                          |
|------------------|---|--------------------------|
| <b>Kuukausi</b>  | <b>Ilman trendiä</b>  | <b>Sisältäen trendin</b> |
| <b>Tammikuu</b>  | 0,756 ( <b>p&lt;0,01</b> )                                    | 0,110 (p>0,1)            |
| <b>Helmikuu</b>  | 0,573 ( <b>p=0,025</b> )                                      | 0,082 (p>0,1)            |
| <b>Maaliskuu</b> | 0,125 (p>0,1)   | 0,102 (p>0,1)            |
| <b>Huhtikuu</b>  | 0,131 (p>0,1)   | 0,083 (p>0,1)            |
| <b>Toukokuu</b>  | 0,567 ( <b>p=0,027</b> )                                      | 0,036 (p>0,1)            |
| <b>Kesäkuu</b>   | 0,466 ( <b>p=0,049</b> )                                      | 0,120 (p=0,099)          |
| <b>Heinäkuu</b>  | 0,252 (p>0,1)   | 0,105 (p>0,1)            |
| <b>Elokuu</b>    | 0,062 (p>0,1)   | 0,059 (p>0,1)            |
| <b>Syyskuu</b>   | 0,271 (p>0,1)   | 0,095 (p>0,1)            |
| <b>Lokakuu</b>   | 0,111 (p>0,1)   | 0,034 (p>0,1)            |
| <b>Marraskuu</b> | 0,099 (p>0,1)   | 0,076 (p>0,1)            |
| <b>Joulukuu</b>  | 0,498 ( <b>p=0,042</b> )                                      | 0,035 (p>0,1)            |

Taulukko 4: KPSS-testin tulokset trendillä ja ilman, kun  $l=3$ . Suluissa p-arvo (korostettuna p-arvot, jotka ovat pienempiä kuin 0,05).

## 7 Pohdinta

Yleisesti ottaen mikään käytetyistä yksikköjuuritestistä ei antanut kovin poikkeavaa tulosta, vaan kaikki testit luokittelivat suurimman osan aikasarjoista stationaariseksi. Muutamia ristiriitaisia tuloksia kuitenkin ilmeni. Helmikuu on ADF-testien perusteella yksikköjuuriprosessi viiveillä 2 ja 3 riippumatta siitä, otetaanko malliin mukaan trendikomponentti vai ei. Kuitenkin viiveillä 0 ja 1 molemmat ADF-testit luokittelevat aikasarjan stationaariseksi. KPSS-testien tulos taas viittaisi trendistationaarisuuteen. Visuaalisesti tarkasteltuna helmikuun aikasarjassa esiintyy selviä pienemmän ja suuremman varianssin ajanjaksoja, mutta yksikköjuuriprosessille tyypillistä vaeltelevaa käyttäytymistä ei ole ainakaan selvästi havaittavissa. Näiden tulosten perusteella ei siis voi varmasti sanoa, onko helmikuun aikasarjassa yksikköjuuria vai ei. Boxin, Jenkinsin ja Reinsel (2008) mukaan stationaarisuuden suhteen epäselvissä tapauksissa epästationaarisuus voi kuitenkin olla jatkoanalyysien kannalta parempi valinta [1]. Siis jos tälle aineistoille käytetään sellaisia tilastollisia menetelmiä, joissa oletuksena on aikasarjan stationaarisuus, niin helmikuun aikasarja kannattaa ehkä differensoida.

Mielenkiintoinen kuukausi on myös tammikuu, jolle KPSS-testi ilman trendikomponenttia antaa hyvin pienen p-arvon. Vastoin muita testejä KPSS-testin perusteella tammikuun aikasarja olisi siis selvästi yksikköjuuriprosessi. Jos KPSS-testin nollahypoteesiksi asetetaan trendistationaarisuus, niin se jää voimaan ( $p > 0,1$ ). Tämän ja visuaalisen arvioinnin perusteella tammikuun aikasarjaa voisikin pitää trendistationaarisena loivalla nousevalla trendillä.

Maalis- ja kesäkuu ovat ADF-testien perusteella yksikköjuuriprosesseja viiveellä 3 riippumatta siitä, sisällytetäänkö testiin trendikomponenttia vai ei. Muilla käytetyillä viiveillä nämä kuukaudet ovat kuitenkin stationaarisia. PP- ja KPSS-testit tukevat maaliskuun stationaarisuutta, mutta kesäkuun osalta KPSS-testin p-arvo jää hieman alle 0,05:n rajan ilman trendiä ja on noin 0,1 trendin kanssa. Tiukasti ottaen KPSS-testi voisi siis viitata trendistationaarisuuteen, mutta ilman trendiä saatu p-arvo on kuitenkin niin lähellä kriittistä 5%:n riskitasoa, ettei selviä päätelmiä voi tästä tehdä. Myös PP-testit tukevat kesäkuun stationaarisuutta, eikä aikasarjassa näytä kuvankaan perusteella olevan trendiä.

KPSS-testi antaa myös touko- ja joulukuun aikasarjoille tulokseksi epästationaarisuuden ilman trendiä, mutta stationaarisuuden trendin kanssa. Tässäkään tapauksessa tulokset eivät ole aivan niin selviä kuin tammikuulle, sillä p-arvot ilman trendiä ovat suurempia (noin 0,03 ja 0,04). ADF-testit eivät touko- tai joulukuun osalta viittaa selvästi suuntaan tai toiseen, PP-testit taas tukevat jälleen stationaarisuutta. Visuaalisen tarkastelun perusteella erityisesti toukokuun aikasarjassa voisi olla nouseva trendi, eikä se joulukuun osaltakaan ole poissuljettu.

Yhteenvetona voisi sanoa, että tarkastelluista aikasarjoista ainoastaan helmikuun osal-

ta löytyi näyttöä yksikköjuuren mahdollisesta olemassaolosta, mutta tämäkään tulos ei ole kiistaton. Tammikuun aikasarjaa voisi perustellusti pitää trendistationaarisena touko- ja joulukuun jäädessä hieman epäselviksi. Loput kuukaudet voi tulkita näiden testien perusteella tason suhteen stationaariseksi. Kiinnostava jatkotutkimuksen aihe olisi ainakin tammikuun aikasarjan (mahdollisen) trendin kulmakertoimen estimointi.

Erityisen kiinnostavaa tämän työn tuloksissa on se, että ne ovat vain muutaman kuukauden osalta linjassa simuloituihin ilmastomalleihin perustuvien tutkimusten kanssa, joiden mukaan sademäärät ovat kasvaneet ja tulevat todennäköisesti kasvamaan Suomessa [8, 15]. Tutkimusten mukaan kasvua tapahtuisi erityisesti talvella, johon tammikuun (ja mahdollisesti joulukuun) tulos sopiikin hyvin. Useimpien kuukausien osalta näyttöä minkäänlaisesta trendistä ei kuitenkaan löytynyt. Toisaalta tulokset ovat hyvin samanlaisia kuin suurimmassa osassa luvussa 4 mainituista tutkimuksista, joissa käytettiin yksikköjuuritestejä eri maissa havaittuihin sadeaikasarjoihin. Näistä parhaiten Suomen ilmastoon vertautuu Clarkin ym. (2000) Kanadassa tehty tutkimus, jossa siinäkin ei löydetty näyttöä yleiselle trendille sademäärissä [3]. Vaikuttaisi siis siltä, että tulkinnalliset erot ilmastonmuutoksen vaikutuksesta sademääriin johtuvat jossain määrin käytetyistä menetelmistä ja aineistosta (suoraan havaittu vs. simuloitu, paikallinen vs. maailmanlaajuinen).

Tässä aineistossa esiin tuli varsin selvästi ADF-testin riippuvuus viiveparametrin valinnasta: kun  $p_v$  asetettiin suuremmaksi kuin 4 (ei mukana raportoiduissa tuloksissa), niin monilla kuukausilla nollassa nollahypoteesi epästationaarisuudesta jäikin voimaan. Tähän työhön otettiin mukaan kuitenkin vain ADF-testin viiveillä 0–3 saadut tulokset, kuten luvussa 5 on perusteltu. PP- ja KPSS-testien tulokset tukevat myös useimpien aikasarjojen stationaarisuutta. Kuitenkin tiedetään, että myös näiden testien kaistanleveysparametrin  $l$  valintaan liittyy riskiä myöhemmän päättelyn kannalta. Erityisesti KPSS-testi olisikin mielenkiintoista tehdä vielä pienemmällä  $l$ :n arvolla, jotta tulos stationaarisuudesta olisi mahdollisimman luotettava, kuten Kočenda ehdottaa (ks. luku 3.3) [11]. KPSS-testin tuloksista (ilman trendiä) erityisesti kesä- ja joulukuu jäivät 5%:n riskitason tuntumaan, joten eri  $l$ :n valinnalla tulokset olisivat voineet olla päättelyn kannalta selkeämpiä.

Aina voidaan pohtia myös aikasarjan pituutta: onko 56 vuotta tarpeeksi pitkä ajanjakso, jotta sademäärien käyttäytymisestä voidaan tehdä luotettavia päätelmiä? Tässä työssä kiinnostuksen kohteena on erityisesti teollistumisen jälkeinen ilmastonmuutos, joten ideaalitulanteessa aineisto kattaisi myös 1900-luvun alun tai jopa osan 1800-lukua, kuten esimerkiksi Estradan ja Perronin (2014, 2019) lämpötila-analyyseissa [5, 6]. Pidemmällä havaintojaksolla esimerkiksi mahdolliset trendit olisivat saattaneet tulla selvemmin näkyviin. Tiedetään kuitenkin, että ilmastonmuutos on kiihtynyt 1900-luvun puolivälin jälkeen verrattuna sitä edeltäneisiin vuosiin [8], joten ehkäpä kiinnostavin ajanjakso onkin juuri viimeisten 50–70 vuoden sisällä.

## 8 Viitteet

- [1] Box, G.E.P., Jenkins, G.M. ja Reinsel, G.C., 2008. *Time Series Analysis. Forecast and Control*. (4. painos.) John Wiley & Sons, Inc., Hoboken, New Jersey, s. 8–10, 22–25, 29, 79–80, 95–97, 220–221. 1, 2, 7
- [2] Castle, J. L. ja Hendry, D.F., 2020. *Climate Econometrics: An Overview*. *Foundations and Trends in Econometrics*, 10:3–4, s. 145–322. 1
- [3] Clark, J. S., Yiridoe, E. K., Burns, N. D. ja Astatkie, T., 2000. *Regional Climate Change: Trend Analysis of Temperature and Precipitation Series at Selected Canadian Sites*. *Canadian Journal of Agricultural Economics*, 48:1, s. 27–38. 2.4, 4, 7
- [4] Dickey, D. A. ja Fuller, W. A., 1981. *Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root*. *Econometrica*, 49, s. 1057–1072. 3.1.1
- [5] Estrada, F. ja Perron, P., 2014. *Detection and Attribution of Climate Change Through Econometric Methods*. *Boletín de la Sociedad Matemática Mexicana*, 2014:20, s.107–136. 2.1, 2.3, 2.4, 4, 7
- [6] Estrada, F. ja Perron, P., 2019. *Breaks, Trends and the Attribution of Climate Change: A Time-Series Analysis*. *Economía*, 42:1, s. 1–31. 4, 7
- [7] Hall, A., 1994. *Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection*. *Journal of Business & Economic Statistics*, 12:4, s. 461–470. 3.1.2
- [8] IPCC, 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneve, Sveitsi. 1, 4, 7
- [9] Jarvis, D., Stoeckl, N. ja Chaiechi, T., 2013. *Applying Econometric Techniques to Hydrological Problems in a Large Basin: Quantifying the Rainfall–Discharge Relationship in the Burdekin, Queensland, Australia*. *Journal of Hydrology*, 496, s. 107–121. 1, 2.4, 4
- [10] Kelly, D.L., 2000. *Unit Roots in the Climate: Is the Recent Warming Due to Persistent Shocks?* Työpaperi. Department of Economics, University of Miami. 1, 2.4, 4
- [11] Kočenda, E., 2015. *Elements of Time Series Econometrics: An Applied Approach*. (3. painos.) Karolinum Press, Praha, Tsekki, s. 25, 66–76. 2.2, 3, 3.1.2, 3.2, 3.3, 4, 7

- [12] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., ja Shin, Y. 1992. *Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root*. Journal of Econometrics, 54:1–3, s. 159–178. 3.3
- [13] Maddala, G. S. ja Kim, I., 1999. *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press, Cambridge, s. 77–78. 3.1.2
- [14] Rahman, M. J., Syeda, J. A. ja Nasser, M., 2013. *Time Series Properties of Some Climatic Variables in Dinajpur District*. Journal of Environmental Science & Natural Resources, 6:2, s. 127–137. 2.3, 2.4, 4
- [15] Ruosteenoja, K., Räisänen, J., Jylhä, K., Mäkelä, H., Lehtonen, I., Simola, H., Luomaranta, A. ja Weiher, S., 2013. *Maailmanlaajuisiin CMIP3-malleihin perustuvia arvioita Suomen tulevasta ilmastosta*. Ilmatieteen laitoksen Raportteja 2013:4. 1, 5, 7
- [16] Saidi, H., Ciampittiello, M., Dresti, C. ja Ghiglieri, G., 2014. *Assessment of Trends in Extreme Precipitation Events: A Case Study in Piedmont (North-West Italy)*. Water Resources Management, 29:1, s. 63–80. 4
- [17] Schwert, G. W., 1989. *Tests for Unit Roots: A Monte Carlo Investigation*. Journal of Business & Economic Statistics, 7, s. 147–159.

3.2

## 9 Liitteet

### Liite 1. Havaintoasemien nimet ja koordinaatit

| Havaintoasema                   | Latitudi (desimaaliasteita) | Longitudi (desimaaliasteita) |
|---------------------------------|-----------------------------|------------------------------|
| Äänekoski Kalaniemi             | 62,87574                    | 25,86171                     |
| Enonkoski Simanala              | 62,05563                    | 29,04407                     |
| Heinävesi Palokki               | 62,57053                    | 28,59248                     |
| Huittinen Sallila               | 61,02353                    | 22,69863                     |
| Inari Raja-Jooseppi Kontiojärvi | 68,47067                    | 28,31619                     |
| Joensuu Huhtilampi              | 62,42668                    | 30,40736                     |
| Jyväskylä Muuratjärvi           | 62,15685                    | 25,42166                     |
| Kalajoki Mehtäkylä              | 64,27381                    | 24,20831                     |
| Kangasniemi kirkonkylä          | 61,98111                    | 26,64228                     |
| Kaustinen Tastula               | 63,59353                    | 23,71761                     |
| Kiuruvesi Korpijoki             | 63,72868                    | 26,36981                     |
| Kokemäki Rausenkulma            | 61,26002                    | 22,22752                     |
| Nokia Tottijärvi                | 61,39648                    | 23,33637                     |
| Outokumpu Viuruniemi            | 62,61751                    | 28,94241                     |
| Salla Kellosekä                 | 66,94435                    | 28,97693                     |
| Savukoski kirkonkylä            | 67,28611                    | 28,17691                     |
| Siikalatva Kestilä              | 64,35544                    | 26,30033                     |
| Sulkava Halttula                | 61,92365                    | 28,31944                     |
| Sysmä Joutsjärvi                | 61,51086                    | 25,81147                     |
| Teuva Kauppilankylä             | 62,49206                    | 21,75847                     |
| Toholampi Oravala               | 63,76627                    | 24,28492                     |
| Utajärvi Särkijärvi             | 64,92733                    | 27,19236                     |
| Uusikaupunki Itätulli           | 60,80183                    | 21,42478                     |
| Vieremä Kaarakkala              | 63,84090                    | 27,22061                     |
| Vihti Hiiskula                  | 60,51583                    | 24,52403                     |

## Liite 2. R-koodi

#Tuodaan aineisto (Huom. aineisto ladattavissa Excel-tiedostoina Ilmatieteenlaitoksen sivuilta)

```
library(readxl)
Aanekoski <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Aanekoski.xlsx", sheet = "Havainnot")
Enonkoski <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Enonkoski.xlsx", sheet = "Havainnot")
Heinavesi <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Heinavesi.xlsx", sheet = "Havainnot")
Huittinen <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Huittinen.xlsx", sheet = "Havainnot")
Inari <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Inari.xlsx", sheet = "Havainnot")
Joensuu <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Joensuu.xlsx", sheet = "Havainnot")
Jyvaskyla <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Jyvaskyla.xlsx", sheet = "Havainnot")
Kalajoki <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Kalajoki.xlsx", sheet = "Havainnot")
Kangasniemi <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Kangasniemi.xlsx", sheet = "Havainnot")
Kaustinen <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Kaustinen.xlsx", sheet = "Havainnot")
Kiuruvesi <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Kiuruvesi.xlsx", sheet = "Havainnot")
Kokemaki <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Kokemaki.xlsx", sheet = "Havainnot")
Nokia <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Nokia.xlsx", sheet = "Havainnot")
Outokumpu <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Outokumpu.xlsx", sheet = "Havainnot")
Salla <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Salla.xlsx", sheet = "Havainnot")
Savukoski <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Savukoski.xlsx", sheet = "Havainnot")
Siikalatva <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Siikalatva.xlsx", sheet = "Havainnot")
Sulkava <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Sulkava.xlsx", sheet = "Havainnot")
Sysma <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Sysma.xlsx", sheet = "Havainnot")
Teuva <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Teuva.xlsx", sheet = "Havainnot")
Toholampi <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Toholampi.xlsx", sheet = "Havainnot")
Utajarvi <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Utajarvi.xlsx", sheet = "Havainnot")
Uusikaupunki <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Uusikaupunki.xlsx", sheet = "Havainnot")
Vierema <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Vierema.xlsx", sheet = "Havainnot")
Vihti <- read_excel("C:/Users/charl/Desktop/Gradu/Sadedata/Vihti.xlsx", sheet = "Havainnot")
```

#Siistitaan dataa ennen yhdistamista

```
Aanekoski2 <- Aanekoski[, -3:-5]
names(Aanekoski2)[names(Aanekoski2) == 'Kuukauden sadesumma (mm)'] <- 'Aanekoski'
Enonkoski2 <- Enonkoski[, -3:-5]
names(Enonkoski2)[names(Enonkoski2) == 'Kuukauden sadesumma (mm)'] <- 'Enonkoski'
Heinavesi2 <- Heinavesi[, -3:-5]
names(Heinavesi2)[names(Heinavesi2) == 'Kuukauden sadesumma (mm)'] <- 'Heinavesi'
Huittinen2 <- Huittinen[, -3:-5]
names(Huittinen2)[names(Huittinen2) == 'Kuukauden sadesumma (mm)'] <- 'Huittinen'
Inari2 <- Inari[, -3:-5]
names(Inari2)[names(Inari2) == 'Kuukauden sadesumma (mm)'] <- 'Inari'
Joensuu2 <- Joensuu[, -3:-5]
names(Joensuu2)[names(Joensuu2) == 'Kuukauden sadesumma (mm)'] <- 'Joensuu'
Jyvaskyla2 <- Jyvaskyla[, -3:-5]
names(Jyvaskyla2)[names(Jyvaskyla2) == 'Kuukauden sadesumma (mm)'] <- 'Jyvaskyla'
Kalajoki2 <- Kalajoki[, -3:-5]
names(Kalajoki2)[names(Kalajoki2) == 'Kuukauden sadesumma (mm)'] <- 'Kalajoki'
Kangasniemi2 <- Kangasniemi[, -3:-5]
names(Kangasniemi2)[names(Kangasniemi2) == 'Kuukauden sadesumma (mm)'] <- 'Kangasniemi'
Kaustinen2 <- Kaustinen[, -3:-5]
names(Kaustinen2)[names(Kaustinen2) == 'Kuukauden sadesumma (mm)'] <- 'Kaustinen'
Kiuruvesi2 <- Kiuruvesi[, -3:-5]
names(Kiuruvesi2)[names(Kiuruvesi2) == 'Kuukauden sadesumma (mm)'] <- 'Kiuruvesi'
Kokemaki2 <- Kokemaki[, -3:-5]
names(Kokemaki2)[names(Kokemaki2) == 'Kuukauden sadesumma (mm)'] <- 'Kokemaki'
Nokia2 <- Nokia[, -3:-5]
names(Nokia2)[names(Nokia2) == 'Kuukauden sadesumma (mm)'] <- 'Nokia'
Outokumpu2 <- Outokumpu[, -3:-5]
names(Outokumpu2)[names(Outokumpu2) == 'Kuukauden sadesumma (mm)'] <- 'Outokumpu'
Salla2 <- Salla[, -3:-5]
names(Salla2)[names(Salla2) == 'Kuukauden sadesumma (mm)'] <- 'Salla'
Savukoski2 <- Savukoski[, -3:-5]
names(Savukoski2)[names(Savukoski2) == 'Kuukauden sadesumma (mm)'] <- 'Savukoski'
Siikalatva2 <- Siikalatva[, -3:-5]
names(Siikalatva2)[names(Siikalatva2) == 'Kuukauden sadesumma (mm)'] <- 'Siikalatva'
Sulkava2 <- Sulkava[, -3:-5]
names(Sulkava2)[names(Sulkava2) == 'Kuukauden sadesumma (mm)'] <- 'Sulkava'
Sysma2 <- Sysma[, -3:-5]
names(Sysma2)[names(Sysma2) == 'Kuukauden sadesumma (mm)'] <- 'Sysma'
Teuva2 <- Teuva[, -3:-5]
names(Teuva2)[names(Teuva2) == 'Kuukauden sadesumma (mm)'] <- 'Teuva'
Toholampi2 <- Toholampi[, -3:-5]
names(Toholampi2)[names(Toholampi2) == 'Kuukauden sadesumma (mm)'] <- 'Toholampi'
Utajarvi2 <- Utajarvi[, -3:-5]
```



```

names(Utajarvi2)[names(Utajarvi2) == 'Kuukauden sadesumma (mm)'] <- 'Utajarvi'
Uusikaupunki2 <- Uusikaupunki[, -3: -5]
names(Uusikaupunki2)[names(Uusikaupunki2) == 'Kuukauden sadesumma (mm)'] <- 'Uusikaupunki'
Vierema2 <- Vierema[, -3: -5]
names(Vierema2)[names(Vierema2) == 'Kuukauden sadesumma (mm)'] <- 'Vierema'
Vihti2 <- Vihti[, -3: -5]
names(Vihti2)[names(Vihti2) == 'Kuukauden sadesumma (mm)'] <- 'Vihti'

#Yhdistetaan eri havaintoasemien data
Sadesummat<-Reduce(function(x,y) merge(x, y, all.x=TRUE, by = c("Vuosi", "Kk")), list(Aanekoski2, Enonkoski2,
Heinavesi2, Huittinen2, Inari2, Joensuu2, Jyvaskyla2, Kalajoki2, Kangasniemi2, Kaustinen2, Kiuruvesi2,
Kokemaki2, Nokia2, Outokumpu2, Salla2, Savukoski2, Siikalatva2, Sulkava2, Sysma2, Teuva2, Toholampi2,
Utajarvi2, Uusikaupunki2, Vierema2, Vihti2))

#Lasketaan kuukausittainen keskiarvo, puuttuvat havainnot jätetään huomiotta
Keskiarvo <- rowMeans(Sadesummat, na.rm=TRUE)
Sadesummat <- cbind(Sadesummat, Keskiarvo)

#Tarkastellaan kuukausitasolla
library(dplyr)
tammikuu <- filter(Sadesummat, Kk=="1")
helmikuu <- filter(Sadesummat, Kk=="2")
maaliskuu <- filter(Sadesummat, Kk=="3")
huhtikuu <- filter(Sadesummat, Kk=="4")
toukokuu <- filter(Sadesummat, Kk=="5")
kesakuu <- filter(Sadesummat, Kk=="6")
heinakuu <- filter(Sadesummat, Kk=="7")
elokuu <- filter(Sadesummat, Kk=="8")
syyskuu <- filter(Sadesummat, Kk=="9")
lokakuu <- filter(Sadesummat, Kk=="10")
marraskuu <- filter(Sadesummat, Kk=="11")
joulukuu <- filter(Sadesummat, Kk=="12")

#Katsotaan kuvista, eroaako jokin asema paljon toisista
#Ensimmäinen kuva
par(mfrow=c(4,4), mar=c(3,3,2,0.1), oma=c(2.5,2.5,1,1))
plot(Aanekoski2$Kk, Aanekoski2$Aanekoski, ylim=c(0,210), xlab="", ylab="", main="i@i@enkoski")
plot(Enonkoski2$Kk, Enonkoski2$Enonkoski, ylim=c(0,210), xlab="", ylab="", main="Enonkoski")
plot(Heinavesi2$Kk, Heinavesi2$Heinavesi, ylim=c(0,210), xlab="", ylab="", main="Heini@vesi")
plot(Huittinen2$Kk, Huittinen2$Huittinen, ylim=c(0,210), xlab="", ylab="", main="Huittinen")
plot(Inari2$Kk, Inari2$Inari, ylim=c(0,210), xlab="", ylab="", main="Inari")
plot(Joensuu2$Kk, Joensuu2$Joensuu, ylim=c(0,210), xlab="", ylab="", main="Joensuu")
plot(Jyvaskyla2$Kk, Jyvaskyla2$Jyvaskyla, ylim=c(0,210), xlab="", ylab="", main="Jyvi@skyl@")
plot(Kalajoki2$Kk, Kalajoki2$Kalajoki, ylim=c(0,210), xlab="", ylab="", main="Kalajoki")
plot(Kangasniemi2$Kk, Kangasniemi2$Kangasniemi, ylim=c(0,210), xlab="", ylab="", main="Kangasniemi")
plot(Kaustinen2$Kk, Kaustinen2$Kaustinen, ylim=c(0,210), xlab="", ylab="", main="Kaustinen")
plot(Kiuruvesi2$Kk, Kiuruvesi2$Kiuruvesi, ylim=c(0,210), xlab="", ylab="", main="Kiuruvesi")
plot(Kokemaki2$Kk, Kokemaki2$Kokemaki, ylim=c(0,210), xlab="", ylab="", main="Kokemi@ki")
plot(Nokia2$Kk, Nokia2$Nokia, ylim=c(0,210), xlab="", ylab="", main="Nokia")
mtext("Sadesumma (mm)", side=2, line=0, outer=TRUE, cex=1, las=0)
mtext("Kuukausi", side=1, line=0, outer=TRUE, cex=1)

#Toinen kuva
par(mfrow=c(4,3), mar=c(3,3,2,0.1), oma=c(2.5,2.5,1,1))
plot(Outokumpu2$Kk, Outokumpu2$Outokumpu, ylim=c(0,210), xlab="", ylab="", main="Outokumpu")
plot(Salla2$Kk, Salla2$Salla, ylim=c(0,210), xlab="", ylab="", main="Salla")
plot(Savukoski2$Kk, Savukoski2$Savukoski, ylim=c(0,210), xlab="", ylab="", main="Savukoski")
plot(Siikalatva2$Kk, Siikalatva2$Siikalatva, ylim=c(0,210), xlab="", ylab="", main="Siikalatva")
plot(Sulkava2$Kk, Sulkava2$Sulkava, ylim=c(0,210), xlab="", ylab="", main="Sulkava")
plot(Sysma2$Kk, Sysma2$Sysma, ylim=c(0,210), xlab="", ylab="", main="Sysmi@")
plot(Teuva2$Kk, Teuva2$Teuva, ylim=c(0,210), xlab="", ylab="", main="Teuva")
plot>Toholampi2$Kk, Toholampi2$Toholampi, ylim=c(0,210), xlab="", ylab="", main="Toholampi")
plot(Utajarvi2$Kk, Utajarvi2$Utajarvi, ylim=c(0,210), xlab="", ylab="", main="Utaji@rvi")
plot(Uusikaupunki2$Kk, Uusikaupunki2$Uusikaupunki, ylim=c(0,210), xlab="", ylab="", main="Uusikaupunki")
plot(Vierema2$Kk, Vierema2$Vierema, ylim=c(0,210), xlab="", ylab="", main="Vieremi@")
plot(Vihti2$Kk, Vihti2$Vihti, ylim=c(0,210), xlab="", ylab="", main="Vihti")
mtext("Sadesumma (mm)", side=2, line=0, outer=TRUE, cex=1, las=0)
mtext("Kuukausi", side=1, line=0, outer=TRUE, cex=1)

#Muutetaan aikasarjaksi
par(mfrow=c(1,1))
Sade_ts <- ts(Sadesummat$Keskiarvo, start=c(1965,1), freq=12)
ts.plot(Sade_ts)

#Eristetään jokainen kuukausi omaksi aikasarjaiseen
tamm_i_ts <- ts(tammikuu$Keskiarvo, start=c(1965,1))
helmi_ts <- ts(helmikuu$Keskiarvo, start=c(1965,1))

```

```

maalis_ts <- ts(maaliskuu$Keskiarvo, start=c(1965,1))
huhti_ts <- ts(huhtikuu$Keskiarvo, start=c(1965,1))
touko_ts <- ts(toukokuu$Keskiarvo, start=c(1965,1))
kesa_ts <- ts(kesakuu$Keskiarvo, start=c(1965,1))
heina_ts <- ts(heinakuu$Keskiarvo, start=c(1965,1))
elo_ts <- ts(elokuu$Keskiarvo, start=c(1965,1))
syys_ts <- ts(syyskuu$Keskiarvo, start=c(1965,1))
loka_ts <- ts(lokakuu$Keskiarvo, start=c(1965,1))
marras_ts <- ts(marraskuu$Keskiarvo, start=c(1965,1))
joulu_ts <- ts(joulukuu$Keskiarvo, start=c(1965,1))

#Kuvat
par(mfrow=c(4,3))
ts.plot(tammi_ts, xlab="", ylab="", main="Tammikuu")
ts.plot(helmi_ts, xlab="", ylab="", main="Helmikuu")
ts.plot(maalis_ts, xlab="", ylab="", main="Maaliskuu")
ts.plot(huhti_ts, xlab="", ylab="", main="Huhtikuu")
ts.plot(touko_ts, xlab="", ylab="", main="Toukokuu")
ts.plot(kesa_ts, xlab="", ylab="", main="Kesikuu")
ts.plot(heina_ts, xlab="", ylab="", main="Heinikuu")
ts.plot(elo_ts, xlab="", ylab="", main="Elokuu")
ts.plot(syys_ts, xlab="", ylab="", main="Syyskuu")
ts.plot(loka_ts, xlab="", ylab="", main="Lokakuu")
ts.plot(marras_ts, xlab="", ylab="", main="Marraskuu")
ts.plot(joulu_ts, xlab="", ylab="", main="Joulukuu")
mtext("Sadesumma (mm)", side=2, line=0, outer=TRUE, cex=1, las=0)
mtext("Vuosi", side=1, line=0, outer=TRUE, cex=1)

#Autokorrelaatioiden tarkastelu ja Box-Ljung-testi (toistetaan eri laeille ja kuukausille)
acf(tammi_ts)
pacf(tammi_ts)
Box.test(tammi_ts, lag=1)

#-----YKSIKKOJUURITESTIT-----#

#ADF-testi, sisallytetaan vakio, mutta ei trendia (ajetaan kaikille kuukausille erikseen laeilla 0-3).
library(fUnitRoots)
adfTest(tammi_ts, lag=3, type="c")

#ADF-testi, sisallytetaan vakio ja trendi (ajetaan kaikille kuukausille laeilla 0-3).
adfTest(tammi_ts, lag=0, type="ct")

#PP-testi 1-arvolla trunc(4*(n/100)^0.25), ajetaan kaikille kuukausille
library(tseries)
pp.test(tammi_ts, type="Z(t_alpha)", lshort=TRUE)

#PP-testi 1-arvolla trunc(12*(n/100)^0.25), ajetaan kaikille kuukausille
pp.test(tammi_ts, type="Z(t_alpha)", lshort=FALSE)

#KPSS-testi ilman trendia 1-arvolla trunc(4*(n/100)^0.25), ajetaan kaikille kuukausille
kpss.test(tammi_ts, null = c("Level"), lshort=TRUE)

#KPSS-testi trendilla 1-arvolla trunc(4*(n/100)^0.25), ajetaan kaikille kuukausille
kpss.test(tammi_ts, null = c("Trend"), lshort=TRUE)

```