

LAPPEENRANNAN-LAHDEN TEKNILLINEN YLIOPISTO

School of Engineering Science

Laskennallisen tekniikan koulutusohjelma

Kandidaatintyö

Touko Tikkanen

Gloaalien lämpötila-aikasarjojen tilastollinen tarkastelu

Ohjaaja: Apulaisprof. Lassi Roininen

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto

School of Engineering Science

Laskennallisen tekniikan koulutusohjelma

Touko Tikkanen

Gloaalien lämpötila-aikasarjojen tilastollinen tarkastelu

Kandidaatintyö

2021

22 sivua, 13 kuvaa, 1 taulukko

Ohjaaja: Apulaisprof. Lassi Roininen

Avainsanat: Aikasarja-analyysi, ARMA, Kalman-suodatin, DLM

Tutkimuksen tavoitteena on vertailla eri tapoja mallintaa lämpötila-aikasarjoja ja arvioida niiden sopivuutta ja luotettavuutta. Aikasarjojen tilastollisia ominaisuuksia pyritään selvittämään, ja tuloksia verrataan tämänhetkiseen tutkimustietoon ilmaston lämpenemisestä. Tutkimusmetodina sovitetaan kolme aikavälin 1850–2015 kattavaa globaalia lämpötila-aikasarjaa autoregressiiviseen moving average (ARMA) -malliin ja dynaamiseen lineaariseen malliin (DLM).

Saatujen tuloksien perusteella voidaan sanoa, että DLM sopii lämpötila-aikasarjojen mallinnukseen paremmin, sillä se sallii mallin parametrien muuttua ajan suhteen. Tämä antaa huomattavasti paremman kuvan aikasarjan taustalla esiintyvistä prosesseista. ARMA-mallin stationaarisuusvaatimuksen takia malliin sovitettiin 10 vuoden paloissa, jotka perustuvat vain edelliseen 50 vuoteen, mikä vähentää mallin luotettavuutta. Kaikissa saaduissa malleissa näkyi viimeistään 1900-luvun puolivälissä selvä lämpötilojen nousu, ja kasvava trendi jatkuu nykypäivään. Lähes jokainen malli päättyy vuodessa 2015 yli yhden asteen lämpötilan muutokseen verrattuna esiteolliseen aikaan, mikä on hieman suurempi kuin IPCC:n osoittama 1°C poikkeama vuonna 2017.

Sisällys

Symboli- ja lyhenneluettelo	4
1 JOHDANTO	5
1.1 Tausta	5
1.2 Tutkimuksen tavoite	6
1.3 Metodologiasta	6
1.4 Tutkimuksen rakenne	7
2 TILASTOLLISET AIKASARJAMALLIT	7
2.1 ARMA-malli	8
2.2 Dynaaminen lineaarinen malli	10
2.3 AR(1)-mallin ja Kalman-suotimen yhteys	11
3 TRENDIANALYYSI	12
3.1 Mallinnuksen tulokset	12
3.2 Mallien analyysi	14
4 YHTEENVETO	20
LÄHTEET	22

Symboli- ja lyhenneluettelo

AR	Autoregressiivinen prosessi tai malli
MA	Moving Average -prosessi tai -malli
ARMA	Autoregressiivinen Moving Average -prosessi tai -malli
DLM	Dynaaminen lineaarinen malli
MCMC	Markovin ketju Monte Carlo

1 JOHDANTO

1.1 Tausta

Aikasarja-analyysejä käytetään moneen tarkoitukseen, kuten talouden muutoksien tutkimisessa, tautien leviämisen tutkimisessa tai säätilojen ennustamisessa. Yleinen syy mallinnukselle on, että halutaan tunnistaa aikasarjan eri osia ja niiden ominaisuuksia. Mallista voidaan tunnistaa esimerkiksi aikasarjan pitkäaikaisen vaihtelun suunta, trendi, tai siitä voidaan huomata säännöllistä kausittaista vaihtelua. Aikasarjan mallia voidaan myös käyttää ennustamaan sitä, miten aikasarja jatkuu tulevaisuudessa (Harvey 1990, 1–2).

Mallintamismetodeja on useita, eikä aikasarjaan ja mallin käyttötarkoitukseen sopivan metodin valinta ole yksinkertaista. Lisäksi reaali maailman monimutkaisuudesta ja epäluotettavista mittauksista johtuen aikasarjan mallinnus ei voi kuvata todellista ilmiötä täydellisesti, vaan on aina vain arvio.

Kansainvälinen ilmastopaneeli (The Intergovernmental Panel on Climate Change, IPCC) raportoi globaalin keskilämpötilan nousseen. Viides arviointiraportti (Fifth Assessment Report) näyttää maan- ja merenpinnan keskivertolämpötilan nousseen $0,85^{\circ}\text{C}$ vuosien 1880–2012 välillä, kun dataan on sovitettu lineaarinen trendi (IPCC 2013, 161). Ihmisten toimien aiheuttaman lämpenemisen esiteolliseen aikaan verrattuna on arvioitu saavuttaneen 1°C vuonna 2017, kasvaen nopeudella $0,2^{\circ}\text{C}$ per vuosikymmen (IPCC 2018). Lineaariset sovitukset lämpötila-aikasarjoihin ovat olleet oleellisia näiden muutoksien huomaamisessa ja todistamisessa.

Globaalia lämpötila-aikasarjaa on rakentanut esimerkiksi EUSTACE (2017), jonka datasetit koostuvat paikallisesti mitatusta datasta ja suuremmille alueille luoduista ilmastomalleista. Käytettävissä on sekä paikan päällä mitattua että satelliittien avulla estimoituja maailmanlaajuisia maanpinnanlämpötiloja. Ilmastomallit ovat laskettu keskiarvoina ruudukolla, jonka koko vaihtelee ilmastomallista riippuen, ruudukon ilmeisesti tarkentuen teknologian kehittyessä. Varsinkin tarkkojen mittausten puuttuessa lämpötiloja on voitu estimoida statistisilla tekniikoilla, kuten ekstrapoloimalla taaksepäin ajassa perustuen saatavilla olevaan dataan. Kattavien datasettien luonnissa haasteena on näillä eri metodeilla tuotetun datan yhdistäminen.

1.2 Tutkimuksen tavoite

Tämän tutkimuksen tavoitteena on selvittää, millaisia tuloksia lämpötila-aikasarjojen sovitaminen eri menetelmillä tuottaa, ja kuinka luotettavia eri mallien tulokset ovat kuvaamaan todellista lämpötilan vaihtelua. Tutkimuskysymyksinä ovat mistä eri mallien tuloksien eroavaisuudet johtuvat ja millaisia tilastollisia ominaisuuksia aikasarjoista voidaan mallintamalla huomata. Lisäksi mallinnukset ovat osaltaan lisää todistetta globaalien lämpötilojen nousevasta trendistä.

Tutkimuksen lopputuloksena on kolme aikasarjaa samalta aikaväliltä sovitettuna kahteen erilaiseen malliin. Näiden sovitusten eroja vertaillaan jokaisen aikasarjan kohdalla. Tarkoituksena on hahmottaa niiden kykyä mallintaa todellista lämpötilan vaihtelua, pohtia, voivatko mallit olla sopivia eri tilanteissa, ja vastaavatko mallit muuta tämänhetkisiä tutkimustietoja globaalista lämpenemisestä.

1.3 Metodologiasta

Tutkimuksessa käytetään useampia eri toimijoiden tuottamia globaaleja lämpötila-aikasarjoja, jotka kattavat pääasiassa aikavälin 1850-2015. Aikasarjoja sovitetaan autoregressive moving average (ARMA) -mallilla sekä dynaamisella lineaarisella mallilla (DLM), joka hyödyntää Kalman-suodattimia. Malleihin sovitukset tehtiin MATLABin avulla. ARMA-sovitus onnistui MATLABin armax-funktiolla silloin, kun aikasarjat olivat saatettu stationaariseen muotoon. DLM-sovitukseen käytettiin tarkoitusta varten kehitettyä työkalupakkia (M. Laine et al. 2014).

Esimerkkinä lämpötila-aikasarjojen mallinnuksesta Mikkonen et al. (2015) käyttävät tutkimuksessaan samanlaista dynaamisen regression metodia ja Laineen DLM-koodia mallintamaan aikasarjaa keskilämpötiloista Suomessa. Aikasarja on kuukausittaista dataa väliltä 1847–2013. Dynaamisen mallin käyttöä tutkimuksessa perusteellaan sillä, että sen avulla voidaan arvioida sekä ajan suhteen vaihtelevaa trendiä että mallin parametreja, sekä huomioda aikasarjan todennäköisen epästationaarisuuden. Käytetty DLM on määritelty jakautuvan paikallisesti lineaariseen vaihtelevaan trendiin, kausivaihteluun ja autoregressiiviseen valkoiseen kohinaan. Mallissa trendi koostuu paikallisesta perustasosta sekä kahdesta piilotetusta tilasta, joista toinen kuvaa keskitasoa ja toinen tason vaihtelua ajan suhteen. Koska data on kuukausittaista, vuodenajan vaihtelut näkyvät todennäköisesti datassa, ja ne on huomioitu 11 tilamuuttujalla, vastaten eri kuukausia. Valkoisen kohinan autoregressio huomiodaan ensimmäisen asteen autoregression kertoimella. Tason, trendin, kausivaihtelun ja

autoregressiivisen osan epävarmuutta kuvataan variansseilla. Tason varianssi on oletettu nol-laksi, ja muut varianssit sekä autoregression kerroin on estimoitu datasta käyttämällä Marko-vin ketju Monte Carlo -metodia (MCMC). Mallinnuksen lopputuloksena on todettu Suomen keskilämpötilan nousseen noin $2,3^{\circ}\text{C}$ tutkitulla aikavälillä, muutoksen olleen $0,2$ ja $0,4^{\circ}\text{C}$ välillä per vuosikymmen. Tämä vastaa muuten aiemmin mainittuja globaaleja lukemia, jotka IPCC on julkaissut, mutta Suomen lämpötilan nousu koko aikavälillä on lähes kaksinkertai-nen globaaliin nousuun verrattuna. Tutkimuksessa tämän esitetään johtuvan ilmiöstä, jossa pohjoisemmilla leveyspiireillä lämpeneminen on suurempaa. Tulokset vaikuttavat järkeviltä, mikä viittaa DLM-metodin sopivuuteen lämpötila-aikasarjan mallintamiseksi.

1.4 Tutkimuksen rakenne

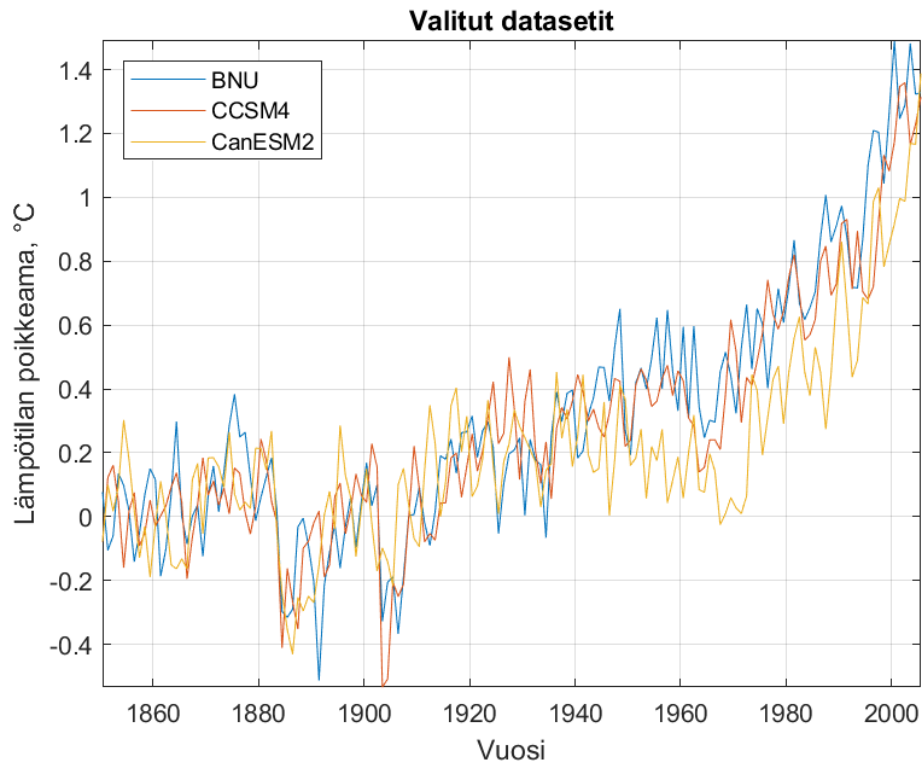
Tämän tutkimuksen alussa on selitetty sen tausta ja tarkoitus lyhyesti, ja kuvattu aikasarjo-jen muodostamista ja aiempien tutkimuksien käyttämiä lineaarisen mallinnuksen metodeja. Toinen osio kattaa tässä tutkimuksessa käytetyt aikasarjat sekä tilastollisten aikasarjamallien teorian ja metodologian. Kolmannessa kappaleessa esitetään mallinnuksien tulokset, vertail-laan malleja ja pohditaan niiden ominaisuuksia ja luotettavuutta, sekä käydään läpi tutki-muksen mahdollisia virheitä ja puutteita. Neljäs kappale on yhteenveto tuloksista.

2 TILASTOLLISET AIKASARJAMALLIT

Tässä työssä käytetään useita estimoituja historiallisia globaaleja lämpötila-aikasarjoja, jot-ka perustuvat simuloituihin ilmastomalleihin. Näitä eri toimijoiden kehittämiä ilmastomalle-ja on datasetissä yhteensä yksitoista, ja ne ovat GCESS Beijing Normal University:n tutkijoi-den yhteen keräämiä. Mallit käsittelevät laajasti erilaisia ilmaston piirteitä, kuten merijäätä ja ilmakehää, mutta tässä työssä käsiteltävä datasetti on koonnut yhteen vain maanpinnan lämpötilat. Työssä käytetyistä aikasarjoista kaikki noudattavat 365 päivän kalenteria ja kat-tavat vuodet 1850–2015 vuosittaisina datapisteinä. Tiedostot ovat .nc-muodossa.

Aikasarjoista on muutamia eri tavoin esikäsiteltyjä vaihtoehtoja. Tähän työhön valittiin tut-kittavaksi aikasarjoja, jotka kuvaavat historiallista lämpötilan poikkeamaa suhteessa esiteol-liseen tasoon. Datasetin ohessa on selitetty, että poikkeama on laskettu jokaisessa mallissa vähentämällä vuosien 1850–1899 lämpötilojen keskiarvo vuosittaisista lämpötiloista. Tämän selitteen vuoksi tämä datasetin osa on selkein, ja lisäksi sen sisältämät aikasarjat kattavat jo-kainen koko aikavälin ja noudattavat samaa kalenteria.

Työn aikarajoitteen vuoksi saatavilla olevista yhdestätoista mallista valittiin kolme ensimmäistä: BNU, joka on peräisin Beijing Normal University:ltä, CanESM2 Canadian Centre for Climate Modelling -instituutilta ja CCSM4 Yhdysvaltojen National Centre for Atmospheric Research -instituutilta. Valitut aikasarjat ovat kaikki saman pituisia, mutta muuten ne valittiin vaihtoehdoista satunnaisesti. Aikasarjat on esitetty kuvassa 1.



Kuva 1. Tutkimukseen valitut kolme aikasarjaa.

2.1 ARMA-malli

Klassinen tapa hahmotella aikasarjaa on jakaa se seuraaviin osiin: trendi, kausivaihtelu ja jäljelle jäävä epäsäännöllinen osa, jotka voivat noudattaa esimerkiksi additiivista tai multiplikaatiivista mallia (Harvey 1990, 1). ARMA perustuu klassiseen aikasarja-analyysiin, jossa aikasarjan oletetaan olevan sekä stationaarinen että stokastinen (Harvey 1990, 12). Se koostuu autoregressiivisestä (AR) ja moving average (MA) -prosessista. ARMA(p,q)-prosessia kuvaa kaava

$$T_t = \phi_1 T_{t-1} + \dots + \phi_p T_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}, \quad a_t \sim N(0, \sigma_t^2), \quad t \in \mathbb{Z}, \quad (1)$$

jossa T_t ovat havaitut arvot, ϕ_1, \dots, ϕ_p ovat AR-prosessia kuvaavia parametreja, $\theta_1, \dots, \theta_q$ kuvaavat MA-prosessia, ja a_t kuvaavat normaalijakautunutta valkoista kohinaa, jonka kovarianssi on σ_t^2 (Harvey 1990, 52). Tässä tutkimuksessa aikasarjojen oletetaan noudattavan parhaiten additiivista mallia.

Koska lämpötila-aikasarjojen ei voida olettaa olevan stationaarisia, vaan optimaaliset parametrit saattavat muuttua ajan suhteen, ARMA-mallin käyttäminen vaikeutuu. Mallin epäsäännöllisen osan stationaarisuutta voidaan testata esimerkiksi Kwiatkowski-Phillips-Schmidt-Shin-testillä (KPSS), jonka nollahypoteesi on sarjan stationaarisuus. Jos testi osoittaa hypoteesin vääräksi, epäsäännöllisen sarjan voi differentioida stationaarisuuden saavuttamiseksi (Harvey 1990, 49). Tässä tutkimuksessa ARMA-mallinnus tehtiin 10 vuoden pituisille paloille, jotka perustuvat edellisiin 50 vuoteen. Näin pyritään sallimaan mallin parametrien muuttuminen, jos taustaprosessi muuttuu ajan kuluessa. Mallinnuksesta puuttuvat ensimmäiset 50 vuotta, sillä niitä käytettiin sen jälkeisen 10 vuoden ennustamiseen.

Ensimmäiseksi valitusta 50 vuoden pituisesta aikasarjan osasta erotettiin sen trendi ratkaisemalla trendiviivan kertoimet lineaarisesta yhtälöstä. Vähentämällä saatu trendi alkuperäisestä datasta saadaan datasta trenditön versio. Seuraavaksi datasta erotettaisiin mahdolliset kausivaihtelut ja syklit, mutta koska aikasarja on vuosittaista dataa, vuodenaikoihin liittyvää vaihtelua ei ole. Selkeitä useamman vuoden pituisia syklejä ei myöskään löytynyt, ja huomioiden sen, että mahdollisten syklien tunnistaminen on usein hankalaa (viittaus), ne jätettiin huomiotta. Olettaen että syklejä ei ole, jäljellä on nyt epäsäännöllinen sarja, joka differentioitiin tarvittaessa.

ARMA-malia sovittaessa tarvittavat parametrit arvioitiin visuaalisesti epäsäännöllisen sarjan autokorrelaation (ACF) ja osittaisen autokorrelaation (partial autocorrelation, PACF) korrelogrammeista. ACF-korrelogrammin perusteella valittiin niin monta MA-parametria kuin merkittävästi autokorreloituneita aiempia arvoja korrelogrammissa voitiin huomata. PACF-korrelogrammissa valittiin samalla perusteella sopiva määrä AR-parametreja. Autokorrelaatioiden oikean määrän määrittäminen visuaalisesti on haastavaa, mutta korrelogrammien perusteella jokaisen aikasarjan pätkän voisi olettaa vastaavan ARMA(1,1) -mallia, jossa AR- ja MA-parametreja on molempia yksi. Paloittaisen mallinnuksen helpottamiseksi käytettiin siis aina ARMA(1,1) -mallia, jonka kaava perustuen ARMA-prosessin yleiseen kaavaan (1) on seuraava:

$$T_t = \phi_1 T_{t-1} + a_t + \theta_1 a_{t-1}, \quad a_t \sim N(0, \sigma_t^2), \quad t \in \mathbb{Z}. \quad (2)$$

Mallin sovittaminen tehtiin armax-funktiolla, johon epäsäännöllinen sarja annettiin paramet-

riksi. Mallin muodostamisen jälkeen sitä käytettiin ennustamaan seuraavien kymmenen vuoden lämpötilan vaihtelua. Ennustaminen tehtiin forecast-funktiolla. Saadulle ennustukselle tehtiin käänteinen differentiointi ja siihen lisättiin aiemmin erotettu trendi, ja näin saatiin yksittäinen kymmenen vuoden kattava ennuste perustuen sitä edeltävään dataan. Prosessi toistettiin, kunnes koko saatavilla oleva aikaväli oli käsitelty.

2.2 Dynaaminen lineaarinen malli

Toinen yleinen aikasarjojen mallinnusmetodi on tila-avaruusmalli, jossa systeemin tilaa kuvaavat useat tuntemattomat osat, kuten trendi, kausivaihtelu ja valkoinen kohina, ja Kalman-suodinta voidaan käyttää näiden tilojen päivittämiseen ajan suhteen (Harvey 1990, 10). Laine (2018) kertoo, että DLM tarkoittaa tila-avaruusmalliin perustuvaa dynaamista regressiota, jossa systeemin operaattorit ovat lineaarisia. Dynaamisuudella tarkoitetaan, että mallin kertoimet voivat muuttua ajan suhteen, mikä sallii myös aikasarjan taustaprosessien muuttumisen ajan suhteen eli epästationaarisuuden. Särkkä (2013) sanoo, että DLM ja Kalman-suotimen teorit ovat ekvivalentteja.

DLM koostuu kahdesta yhtälöstä. Ensimmäinen on havaittuja arvoja kuvaava yhtälö

$$T_t = F_t X_t + v_t, \quad v_t \sim N(0, V_t), \quad t \in \mathbb{Z}, \quad (3)$$

jossa T_t ovat havaitut arvot, jotka riippuvat systeemimatriista F . X_t ovat näkymättömät tilat, jotka noudattavat ensimmäisen asteen Markovin prosessin kaavaa

$$X_t = G_t X_{t-1} + w_t, \quad w_t \sim N(0, W_t), \quad t \in \mathbb{Z}, \quad (4)$$

jossa näkymättömät tilat voivat muuttua ajan suhteen ja riippuvat matriisista G . Muuttujat v_t ja w_t ilmaisevat systeemin normaalijakautuneen valkoisen kohinan, ja niiden kovarianssit ovat matriisit V_t ja W_t (Harvey 1990, 100–102; Laine 2018).

Kalman-suodin on rekursiivinen prosessi, jolla voidaan estimoida tila-avaruusmallin kertoimia perustuen saatavilla olevaan dataan. Kalman-suodin voidaan ilmaista todennäköisyysfunktiona, joka perustuu systeemin tilojen priorijakaumiin. Suotimelle annetaan jotkin alkuarvot, ja se laskee jakaumien parametrit ennustus- ja päivitysaskeleella (Harvey 1990, 104–109; Särkkä 2013, 56–57).

Tila-avaruusmallin erityisominaisuus on mallin sopeutuminen puuttuviin datapisteisiin (Harvey 1990 10). Tämä tarkoittaa, että datasta on mahdollista poistaa poikkeavia havaintoja. Usein on vaikea määrittellä, mikä arvo lasketaan poikkeavaksi, varsinkin jos datan keruutavasta ei ole tarkkaa tietoa, ellei datapiste ole äärimmäisen selkeästi erillään muista havainnoista. Tämän tutkimuksen aikasarjat sisältävät dataa 1800-luvulta ja 1900-luvun alusta, jonka voidaan olettaa olevan vähemmän luotettavaa kuin nykyaikaisempi data. Virheiden ja epätarkkuuden mahdollisuus on suurempi, sillä esimerkiksi mittalaitteet eivät olleet yhtä edistyneitä, joten puuttuvaa dataa on enemmän, mikä tarkoittaa puuttuvien lämpötilojen estimointia aikasarjoja muodostaessa (EUSTACE 2017). Nämä asiat huomioon ottaen aikasarjoista poistettiin muutamia pisteitä, jotka selkeästi poikkesivat muusta datasta. Datapiste määriteltiin poikkeavaksi, jos sen arvo oli alle $-0,45^{\circ}\text{C}$. Näitä pisteitä löytyi yhteensä kolme, ja ne kaikki esiintyivät ennen vuotta 1920.

M. Laine et al. (2014) kehittämän MATLAB-työkalupakin `dlnfit`-funktiota käytettiin DLM-sovituksen tekemiseen ja `dlnplotdiag`-funktiota sovituksen autokorrelaation ja normaali-jakautumisen selvittämiseen. Funktioiden toiminta on selitetty Laine (2018) nettisivulla osiossa ”DLM Tutorial”, ja niiden käyttö tässä tutkimuksessa perustuu tähän tutoriaaliin. Funktio `dlnfit` vaatii toimiakseen ainakin havaintojen keskihajonnan sekä mallin eri osien, kuten trendin ja kausivaihtelun keskihajonnat. Tässä tutkimuksessa tarkkoja keskihajontoja ei ollut tiedossa, joten havaintojen keskihajonta laskettiin suoraan datasta ja trendin vaihtelun keskihajonta optimoitiin funktion sisällä maksimitodennäköisyyden funktiolla Kalman-suotimen avulla. Tämän lisäksi malli määritettiin sisältämään yhden autokorreloivan parametrin, joka optimoitiin `dlnfit`-funktion palauttamalla todennäköisyysarvolla ja sijoitettiin malliin. Vaihtoehtona olisi ollut parametrien estimointi MCMC-menetelmällä.

2.3 AR(1)-mallin ja Kalman-suotimen yhteys

AR(1)-malli voidaan ilmaista kaavalla

$$T_t = \phi_1 T_{t-1} + a_t, \quad a_t \sim N(0, \sigma_a^2), \quad t \in \mathbb{Z}, \quad (5)$$

joka vastaa autoregressiivista osaa ARMA-mallin kaavassa (1). Diskretisoimalla tämä pystytään johtamaan Kalman-suotimen yhtälö, mikä tarkoittaa, että AR(1) on käytännössä jatkuva muoto Kalman-suotimesta. Diskretisointiin käytetään Euler-Maruyama-metodia. Diskretisoinnissa tarvittava kovarianssi esitetään Fourier-tasossa tekemällä käänteinen Fourier-muunnos:

$$\begin{aligned} cov_t &\propto \exp(-|t|), \\ \mathcal{F}^{-1}(cov_t) &= \frac{1}{\xi^2 + 1} = S(\xi). \end{aligned}$$

Käänteisestä Fourier-muunnoksesta saatu spektritiheys $\sqrt{S(\xi)}$ toimii lineaarisena aikainvarianttina suotimena (LTI), jonka avulla valkoinen kohina W muutetaan T :ksi, joka on jakautunut samalla tavalla kuin W . Spektritiheyden sisältämästä neliöjuuresta valitaan tässä stabiili muoto.

$$W \rightarrow [T(\xi) = \sqrt{S(\xi)}] \rightarrow T$$

Tämä prosessi voidaan ilmaista myös siirtomuodossa, josta saadaan Fourier-muunnospari

$$\mathcal{F}(T) = \sqrt{S(\xi)} \mathcal{F}(W) = \frac{1}{i\xi + \lambda} \iff (i\xi + \lambda) \mathcal{F}(T) = \mathcal{F}(W).$$

Ajan suhteen yhtälö kirjoitetaan seuraavaan muotoon ja diskretisoidaan, jolloin saadaan

$$\left(\frac{d}{dt} - \lambda \right) T = W \implies \frac{d}{dt} T = \lambda T + W,$$

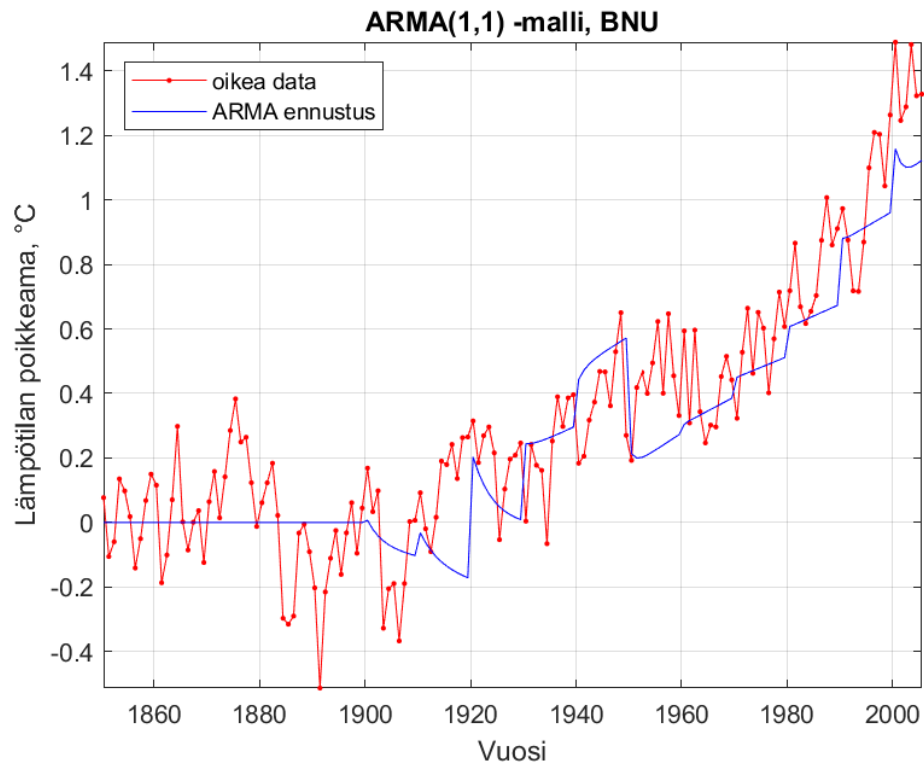
jolloin ollaan johdettu Kalman-suotimen stokastinen differentiaaliyhtälömuoto

$$\begin{aligned} \frac{d}{dt} T &= -\lambda T + w \\ T_k &= HT + e_k. \end{aligned} \tag{6}$$

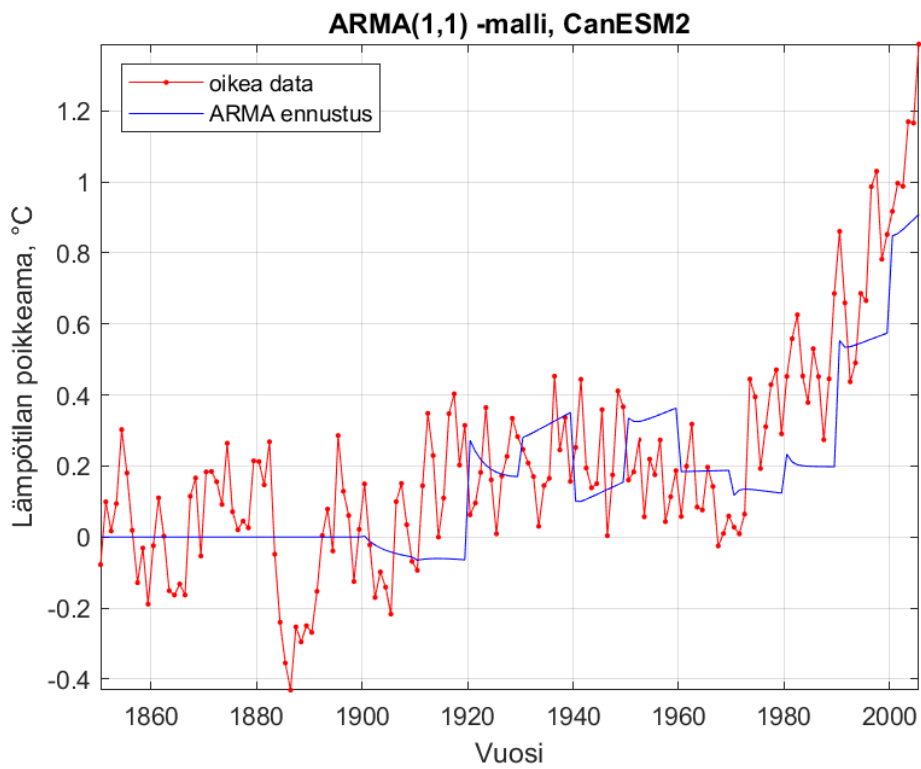
3 TRENDIANALYYSI

3.1 Mallinnuksen tulokset

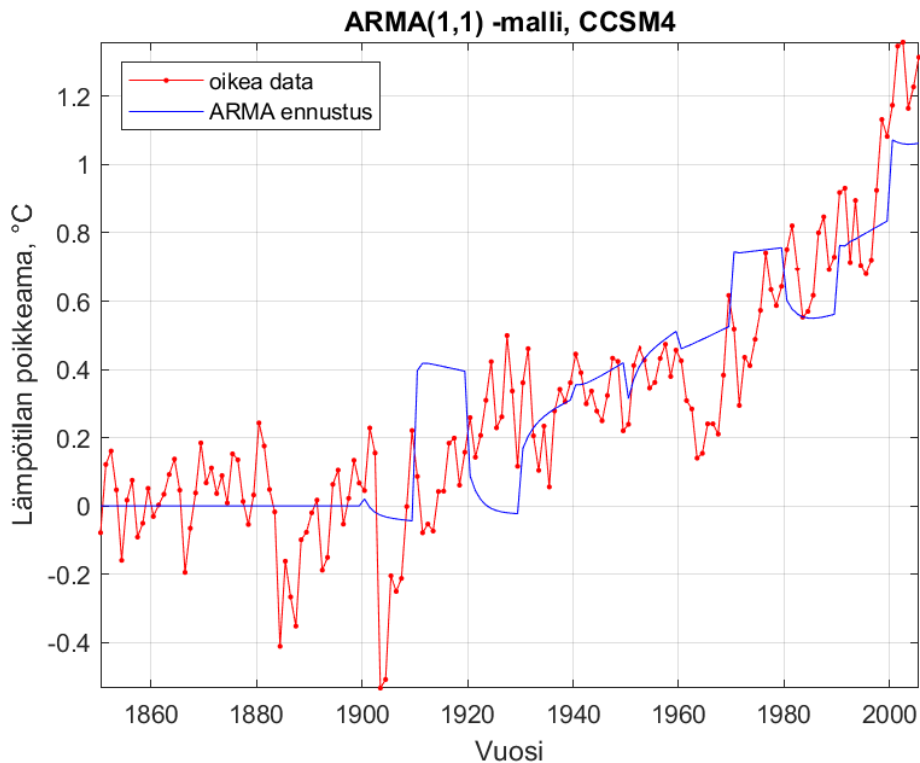
ARMA-mallinnuksen tuloksena saatujen mallien kuvaajat ovat kuvissa 2, 3 ja 4. Kuvaajat ovat piirretty alkuperäisen aikasarjan päälle vertailun helpottamiseksi. Malli alkaa vuodesta 1900, jonka jälkeen ennustukset ovat erillisiä kymmenen vuoden osia.



Kuva 2. Sovitettu paloittainen ARMA(1,1) -malli BNU-aikasarjalle.



Kuva 3. Sovitettu paloittainen ARMA(1,1) -malli CanESM2-aikasarjalle.



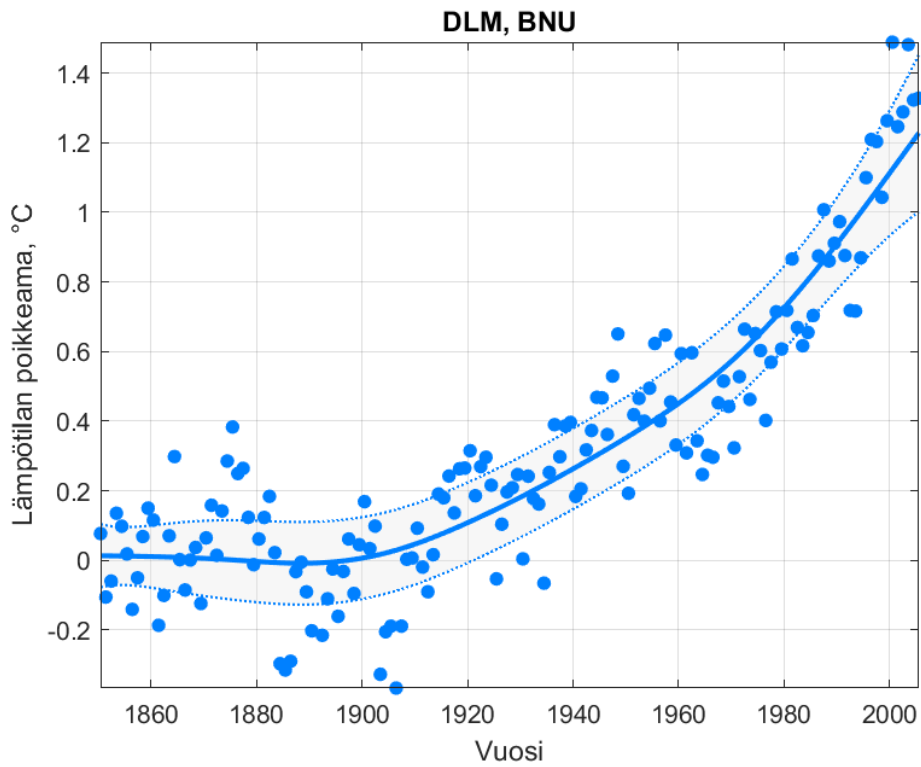
Kuva 4. Sovitettu paloittainen ARMA(1,1) -malli CCSM4-aikasarjalle.

Kuvat 5, 7 ja 9 ovat DLM-mallinnuksen tuloksena saatujen mallien kuvaajat. Kuvaajissa siniset pisteet ovat alkuperäiset datapisteet, tasainen sininen viiva on varsinainen malli, ja katkoviivat ilmaisevat virherajat. Lisäksi kuvissa 6, 8 ja 10 on esitetty valmiiden mallien korrelogrammit ja normaalijakautumisen todennäköisyydet. Näiden perusteella voidaan arvioida mallien sopivuutta.

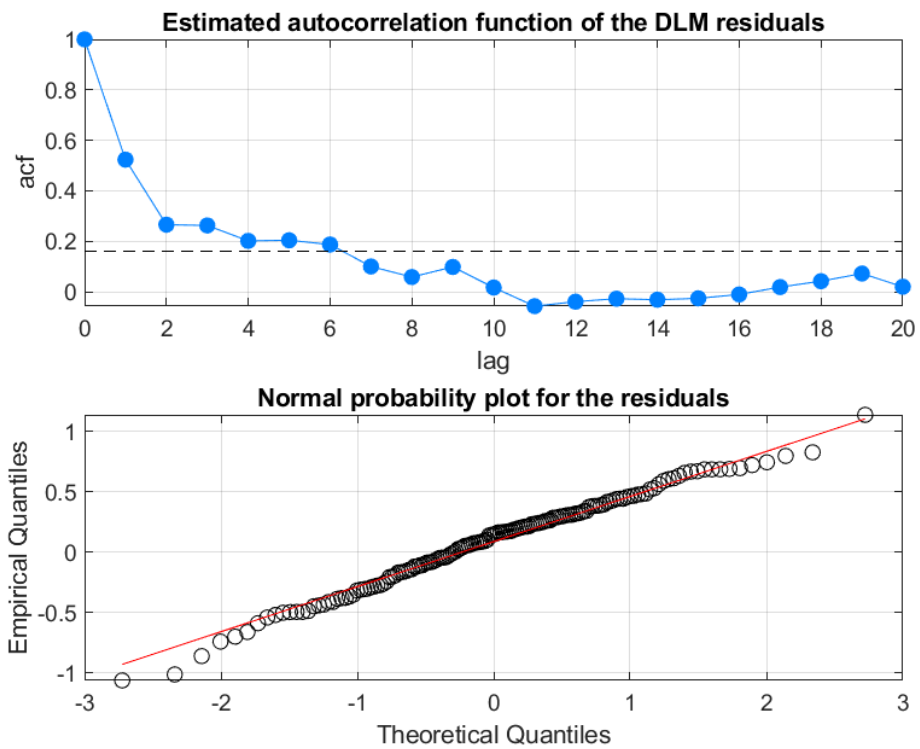
3.2 Mallien analyysi

Paloittaisesta mallinnuksesta johtuen ARMA-mallit eivät ole jatkuvia. Varsinkin ennen 1950-lukua kuvaajassa 2 ilmenee suuria eroja eri palojen välillä, mikä on odotettavissa, kun aikasarjat ovat simulaatioita, joissa menneisyyden data perustuu saatavilla olevan datan ekstrapolointiin. Kuvaajassa 4 sama ilmenee ennen vuotta 1930, ja kuvaajassa 3 eri palojen erot ovat jatkuvasti suuria. Tästä johtuen mallinnusten ei voida sanoa antavan luotettavia arvoja eri vuosien keskilämpötiloista.

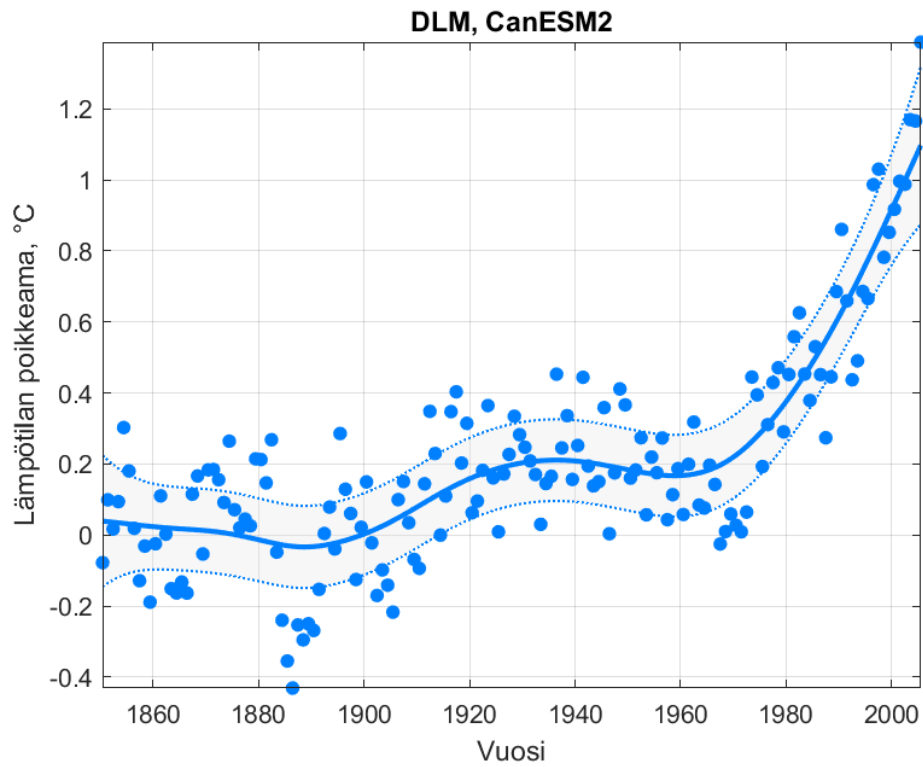
Mallit voivat kuitenkin kertoa keskilämpötilojen muutoksista koko aikavälillä. Vuoden 1950 jälkeen kuvaaja 2 nousee melko tasaisesti, mikä näyttää selkeästi keskilämpötilan kasvun edelliseen vuosisataan verrattuna. Kuvaajassa 4 melko tasaista nousua näkyy jo alkaen vuo-



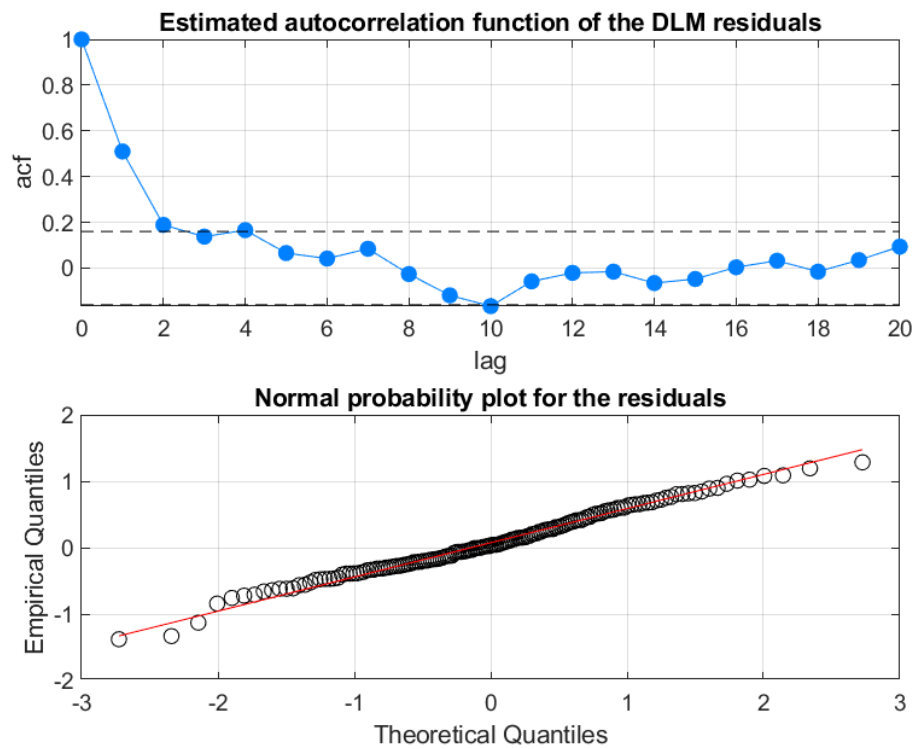
Kuva 5. Sovitettu DLM BNU-aikasarjalle.



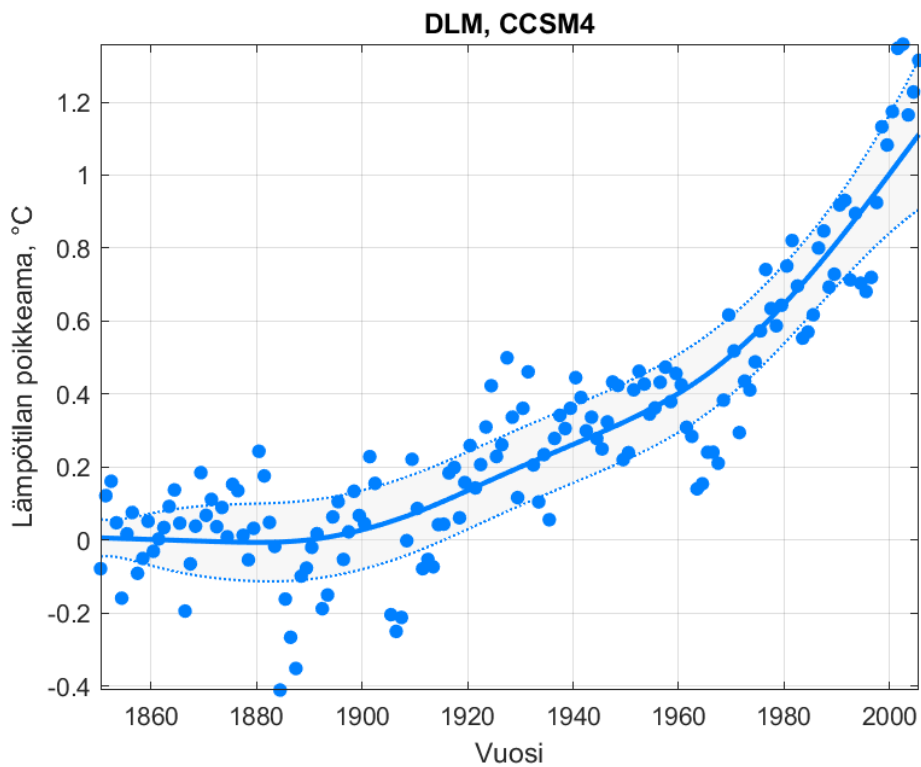
Kuva 6. BNU-aikasarjan DLM-sovituksen autokorrelaatio ja normaalijakautuminen.



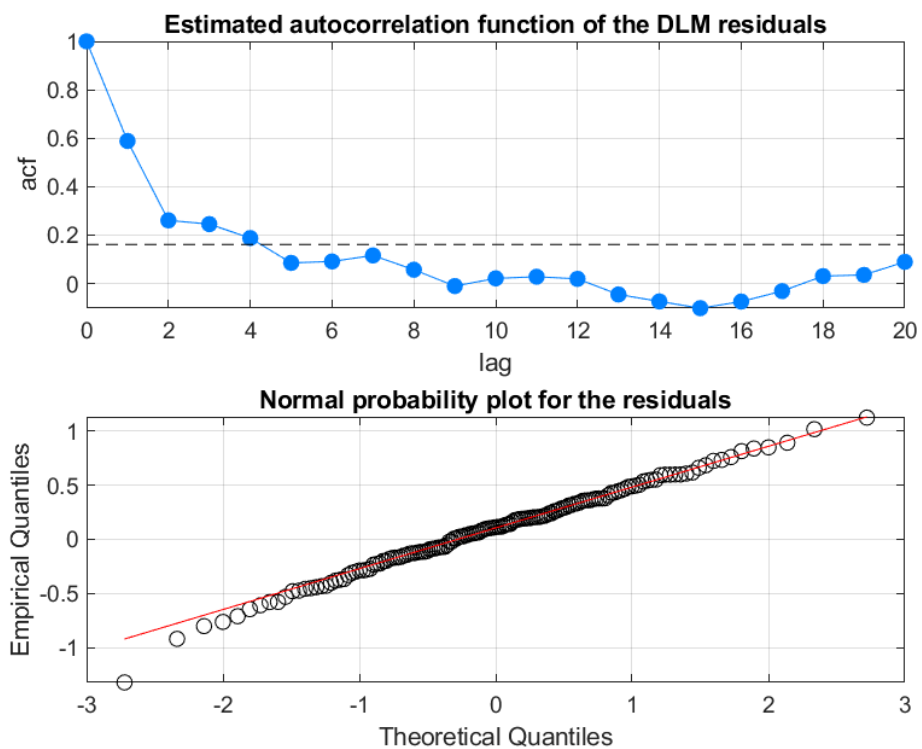
Kuva 7. Sovitettu DLM CanESM2-aikasarjalle.



Kuva 8. CanESM2-aikasarjan DLM-sovituksen autokorrelaatio ja normaalijakautuminen.



Kuva 9. Sovitettu DLM CCSM4-aikasarjalle.



Kuva 10. CCSM4-aikasarjan DLM-sovituksen autokorrelaatio ja normaalijakautuminen.

desta 1930. Kuvaajassa 3 selkeä kasvu näkyy vasta vuonna 1990.

Jos aikasarjat olisi mallinnettu kokonaisina, nykyisten kuvaajien perusteella voidaan sanoa, että malleista olisi tullut hyvin yksinkertaisia. ARMA ei olisi pystynyt huomioimaan aikasarjan taustaprosessin muuttumista ajan suhteen, jolloin on todennäköistä, että viimeisillä vuosikymmenillä ilmenevä voimakas nousu olisi mallissa tasoittunut, ja nousu olisi ilmennyt hitaampana. Malli ei olisi pystynyt kuvaamaan trendin muutosta täsmällisesti, vaan lämpötila, johon malli päättyy, olisi ollut huomattavasti matalampi kuin IPCC:n ilmoittamat tulokset.

Mahdolliset poikkeavat arvot aikasarjoissa voivat myös vaikuttaa mallinnukseen, sillä ARMA ei sopeudu puuttuviin datapisteisiin. Ne voivat vääristää kuvaajaa, ja koska jokainen mallin pala perustuu vain 50 vuoden dataan, muutamakin poikkeava arvo on jo suuri osa käytettävissä olevasta datasta.

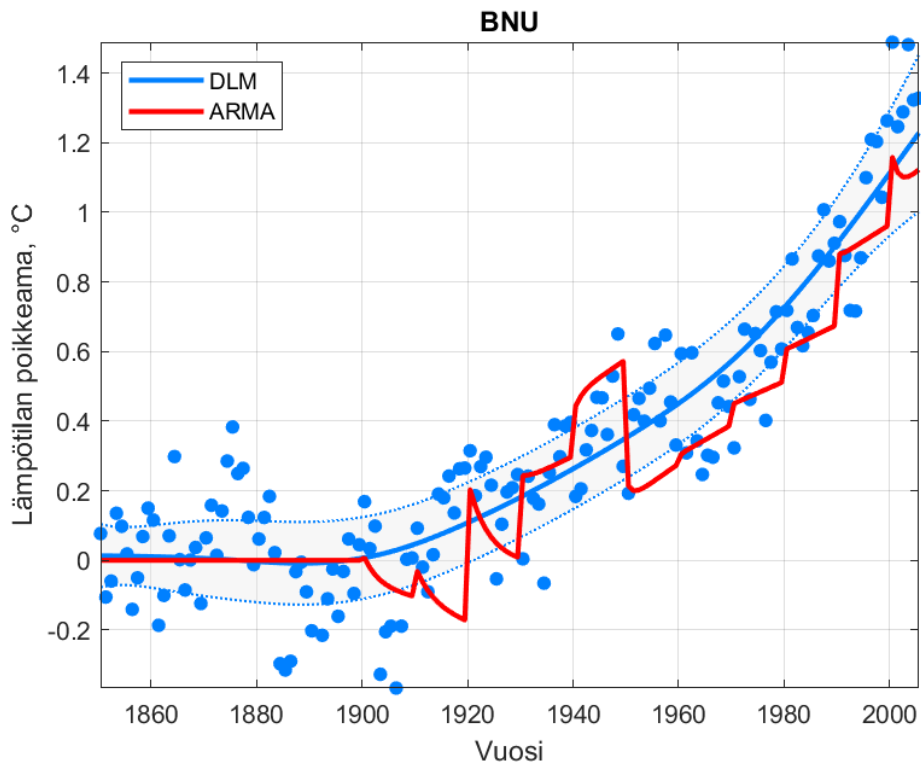
DLM-kuvaajat ovat ARMA-sovituksia tasaisempia. Kuvaajissa on myös näkyvissä mallin virherajat, joiden ulkopuolelle osa datapisteistä sijoittuu, mutta suurin osa datapisteistä on virherajojen sisällä. Virherajojen ulkopuolella olevien pisteiden määrä näyttäisi pienenevän ajan suhteen. Kuvaajissa 5 ja 9 keskilämpötilan nousun voidaan huomata alkavan noin vuonna 1900. Kuvaajassa 7 nousu alkaa lähes samaan aikaan, mutta tasaantuu vuosien 1920–1960 välillä, mitä ei esiinny kahdessa muussa kuvaajassa.

DLM-metodilla todennäköisesti poikkeavat havainnot voitiin poistaa, mikä oletettavasti parantaa tuloksien luotettavuutta hieman. Sovitusten residuaalikuviosta 6, 8 ja 10 voidaan huomata, että jokaisen mallin residuaaleissa näkyy vielä autokorrelaatiota, minkä johdosta tulokset eivät ole yhtä luotettavia. Oletettavasti tämä johtuu huomioimattomista sykleistä aikasarjoissa. On mahdollista, että aikasarjoissa esiintyy esimerkiksi muutaman vuoden pituinen sykli, joka olisi pitänyt huomioida mallinnusvaiheessa.

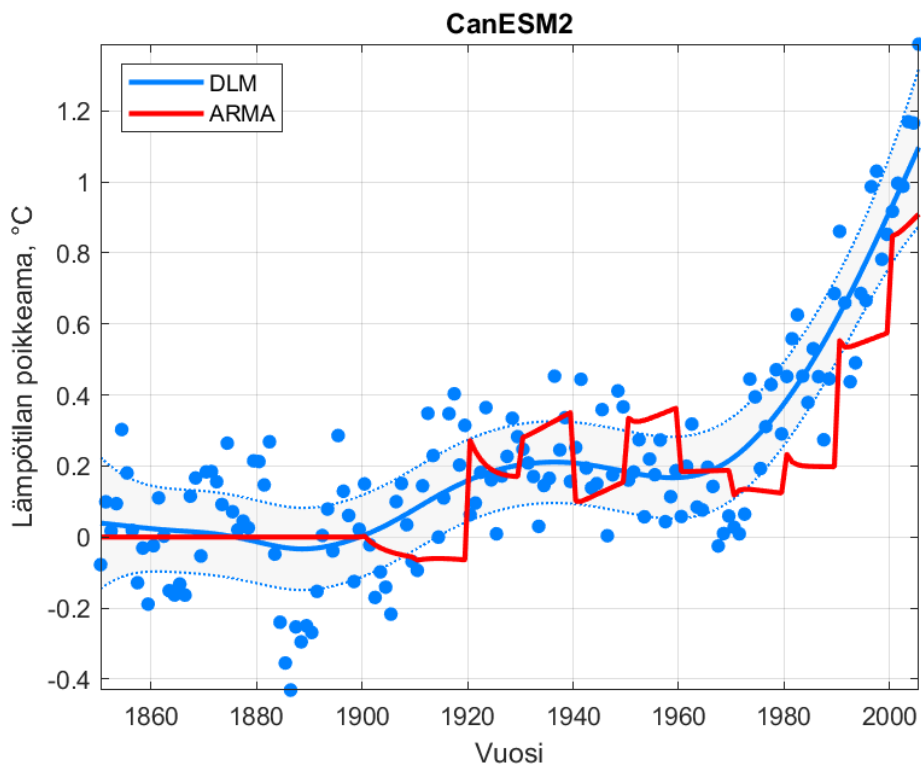
Kuvissa 11, 12 ja 13 esitetään samalle aikasarjalle piirretyt kaksi mallia samassa kuvassa. Näistä kuvista voidaan selkeästi huomata mallien eroavaisuudet.

Kuvaajassa 11 voidaan huomata, että 1950-luvun jälkeinen tasaisempi osa ARMA-mallinnuksesta antaa matalempia lämpötiloja kuin DLM-kuvaaja. Tämä näkyy myös muissa kuvaajissa. Kuten ARMA-sovitusten analyysissä kerrottiin, mallin parametrien pysyminen vakiona saattaa näkyä kuvaajassa hitaampana nousuna. DLM pystyy mukautumaan taustaprosessien muutoksiin ja käyttämään koko saatavilla olevaa aikasarjaa, joten se on herkempi kuvaajissa näkyvälle lämpötilan yllättävälle nousulle.

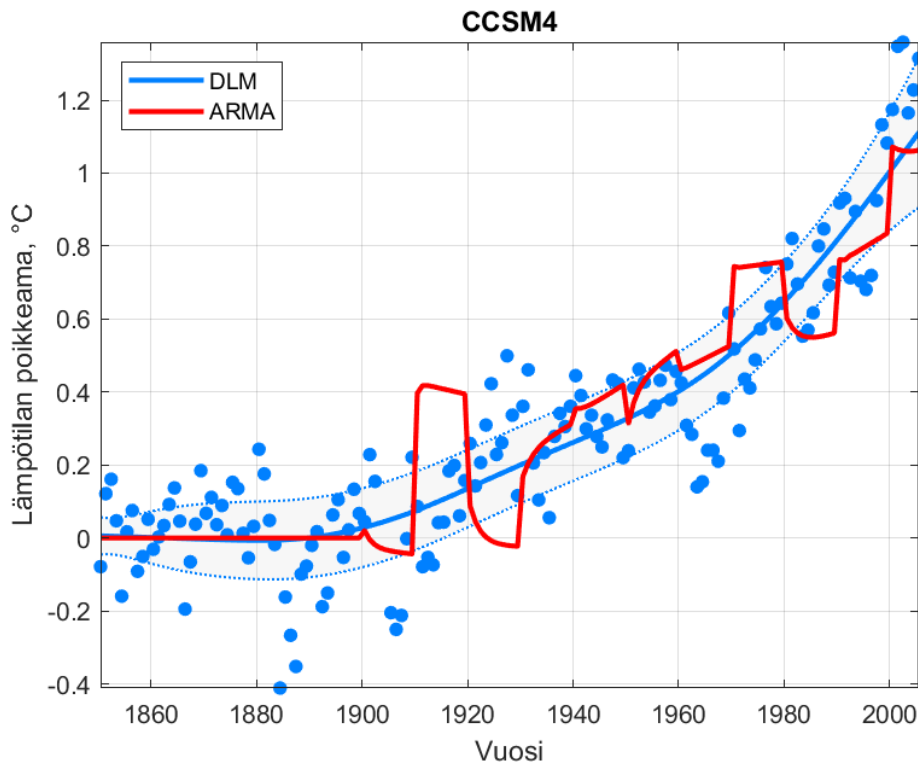
Eri mallit eri aikasarjoille päättyvät tutkittavan ajan lopussa vuotena 2015 lähelle toisinaan. Taulukko 1 esittää lämpötilojen poikkemat tässä pisteessä kolmen merkitsevän nu-



Kuva 11. BNU-aikasarjan DLM ja ARMA-mallinnus päällekkäin piirrettynä.



Kuva 12. CanESM2-aikasarjan DLM ja ARMA-mallinnus päällekkäin piirrettynä.



Kuva 13. CCSM4-aikasarjan DLM ja ARMA-mallinnus päällekkäin piirrettynä.

meron tarkkuudella. Tässä voidaan myös huomata ARMA-sovitusten antavan matalempia lämpötiloja. Lähes jokainen malli päättyy vuodessa 2015 yli yhden asteen lämpötilan muutokseen verrattuna esiteolliseen aikaan. Tämä johtaa hieman suurempaan lämpötilan nousuun kuin IPCC:n tutkimusten osoittama yhden asteen poikkeama globaaleissa keskilämpötiloissa vuonna 2017.

Taulukko 1. Eri mallien ennustama lämpötila vuodelle 2015

Aikasarja	ARMA (°C)	DLM (°C)
BNU	1,12	1,23
CanESM2	0,909	1,10
CCSM4	1,06	1,11

4 YHTEENVETO

Tutkimuksen tavoitteena oli verrata DLM- ja ARMA-sovituksia lämpötila-aikasarjoille ja pohtia niiden sopivuutta tarkoitukseen. Saatujen tuloksien perusteella voidaan sanoa, että

DLM sopii lämpötila-aikasarjojen mallinnukseen paremmin, sillä se sallii mallin parametrien muuttua ajan suhteen. Tämä antaa huomattavasti paremman kuvan aikasarjan taustalla esiintyvistä prosesseista. ARMA-mallin rajoitteisuus tekee sovituksista vähemmän luotettavia. Mallien virheiden suuruuksia ei verrattu laskennallisesti, mutta eroja voidaan huomata visuaalisesti. Kaikissa saaduissa malleissa näkyi viimeistään 1900-luvun puolivälissä selvä lämpötilojen nousu, ja kasvava trendi jatkuu nykypäivään. Mallit päätyvät myös lähelle IPCC:n laajempien tutkimusten esittämiä globaaleja keskilämpötiloja.

Lähteet

- EUSTACE (2017). *Using EUSTACE temperature data for calibration or validation in climate model research*. URL: <https://www.eustaceproject.org/users/use-if-temperature-data/high-resolution-climate-modelling/> (viitattu 09.04.2021).
- Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- IPCC (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Toim. T.F. Stocker et al. Cambridge, United Kingdom ja New York, NY, USA: Cambridge University Press.
- (2018). *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Toim. V. Masson-Delmotte et al. URL: <https://www.ipcc.ch/sr15>.
- Laine, M (2018). *DLM Matlab Toolbox*. URL: <https://mjllaine.github.io/dlm/> (viitattu 28.02.2021).
- Laine, M., Latva-Pukkila, N. ja Kyrölä, E. (2014). “Analyzing time-varying trends in stratospheric ozone time series using state the space approach”. *Atmospheric Chemistry and Physics* 18.14.
- Mikkonen, S., Laine, M., Mäkelä, H. M., Gregow, H., Tuomenvirta, H., Lahtinen, M. ja Laaksonen, A. (2015). “Trends in the average temperature in Finland, 1847–2013”. *Stochastic environmental research and risk assessment* 6.29, s. 1521–1529.
- Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press.