

Veturina-ajojen määrien ennustamisesta

Sini Poikonen

Perustieteiden korkeakoulu

Kandidaatintyö
Espoo 8.12.2023

Vastuupettaja

Prof. Pauliina Ilmonen

Työn ohjaaja

DI Riitta Oja

Copyright © 2023 Sini Poikonen

The document can be stored and made available to the public on the open internet pages of Aalto University.
All other rights are reserved.

Tekijä Sini Poikonen

Työn nimi Veturina-ajojen määrien ennustamisesta

Koulutusohjelma Teknistieteellinen kandidaattiohjelma

Pääaine Matematiikka ja systeemitieteet **Pääaineen koodi** SCI3029

Vastuupettaja Prof. Pauliina Ilmonen

Työn ohjaaja DI Riitta Oja

Päivämäärä 8.12.2023 **Sivumäärä** 33+1 **Kieli** Suomi

Tiivistelmä

Raideliikenne on tärkeä osa Suomen taloutta ja suomalaisten elämää. Veturit vetävät raiteilla päivittäin satoja junia. Joskus veturi täytyy saada toiselle paikkakunnalle hakemaan juna tai esimerkiksi tankattavaksi, jos kyseessä on dieselveturi. Kun veturi liikkuu ilman vaunuja, kyseessä on veturina-ajo. Dieselvetureiden lisäksi on myös sähkövetureita. Eri vetureilla ajetaan erilaisia junia, joten veturina-ajoja saattaa olla järkevää ennustaa myös eri tavoilla.

Tämä työ tutkii eri veturisarjojen veturina-ajoja ja niiden ennustamista. Työssä selvitetään, kannattaako eri veturisarjoille kertyviä veturina-ajoja ennustaa eri tavoilla. Veturina-ajoja on suotavaa ennustaa, sillä ne vaikuttavat vetureille kertyviin kokonaiskilometreihin. Kokonaiskilometrit vaikuttavat siihen, kuinka monta veturinkuljettajaa tai veturia tarvitaan tulevaisuudessa. Tutkimuksessa hyödynnettiin aikasarja-analyysin menetelmiä. Työssä muodostettiin kolmen eri veturisarjan aikasarjoille ARIMA-mallit. Muodostettujen mallien parametreja ja ennustetuloksia vertailtiin. Veturisarjojen aikasarjoja tarkasteltiin myös visuaalisesti.

Tulosten perusteella pystytään sanomaan, että tutkittujen veturisarjojen veturina-ajoja on syytä ennustaa eri tavoilla. Visuaalisesti eri veturisarjoilla oli erilaiset trendit. Muodostettujen ARIMA-mallien kertaluvut ja parametrit olivat erilaiset. Tämä kertoo aikasarjojen korrelaatorakenteiden erilaisuudesta. Rakennettujen mallien antamat ennustearvot alkoivat tasaantumaan yli kahden kuukauden ennusteissa. Sähkövetureille oli tarkempaa ennustaa lyhyen aikavälin veturina-ajoja ARIMA-malleilla kuin laskemalla ennusteet aikasarjan keskiarvosta. Dieselveturisarjalle keskiarvosta lasketut ennusteet olivat hiukan tarkempia. Pidemmälle ajalle, esimerkiksi puolen vuoden päähän, ARIMA-mallit eivät juurikaan tuo lisäarvoa. Keskiarvollakin ennustaessa tulisi myös huomioida eri aikasarjojen kehitys ja trendit. Pitkälle ajalle ennustaminen on aina vaikeaa, ja virheiden mahdollisuus kasvaa.

Avainsanat rautatiet, veturina-ajo, aikasarjat, ennustaminen, ARIMA-malli

Author	Sini Poikonen	
Title	On the prediction of locomotive deadheadings	
Degree programme	Bachelor's Programme in Science and Technology	
Major	Mathematics and Systems Sciences	Code of major SCI3029
Teacher in charge	Prof. Pauliina Ilmonen	
Advisor	MSc (Tech.) Riitta Oja	
Date	Number of pages	Language
8.12.2023	33+1	Finnish

Abstract

Railroads are an important part of Finland's economy and the lives of citizens of Finland. Locomotives pull hundreds of trains every day on the tracks. Sometimes, the locomotive is moved to another town to collect a train or to be fueled if it is a diesel locomotive. This trip is called deadheading. Besides diesel locomotives, there are electric locomotives. Different types of locomotives pull different kinds of trains, so it might be beneficial to predict the deadheadings using different methods.

This thesis analyzes deadheadings for different types of locomotives and predicts them. The thesis examines whether it is beneficial to predict deadheadings differently for various locomotive types. Deadheadings should be predicted since they affect the total distance covered by locomotives, which then impacts the number of locomotives and locomotive drives needed in the future. In this study, ARIMA forecasting models were applied to three different types of locomotives. The parameters and forecast results of the developed models were compared. The time series of the locomotives were also visually inspected.

Based on the results of this study, it can be stated that the deadheadings of the examined locomotives should be predicted in different ways. Visually, different locomotive series exhibited different trends and developments. The developed ARIMA models had different orders and parameters, indicating differences in the correlation structure of the time series. Over three-month predictions, the predicted values started to converge. For a short period of a few weeks, the ARIMA model predictions for electric locomotives were more accurate. For diesel locomotives, predicting with average value of the time series was slightly more accurate than the ARIMA model. However, for a longer period, such as six months ahead, ARIMA forecasting does not significantly add value due to stabilization of the predicted values. Nevertheless, different time series developments and trends should be considered in forecasting for longer time periods. Predicting for an extended period is always challenging, and the magnitude of errors increases.

Keywords railways, deadheading, time series, forecasting, ARIMA model

Sisällys

Tiivistelmä	3
Tiivistelmä (englanniksi)	4
Sisällys	5
1 Johdanto	6
2 Aikaisempi tutkimus	8
2.1 Veturikaluston koon arviointi	8
2.2 Muiden ajoneuvolajien soveltaminen	9
3 Tutkimusaineisto ja -menetelmät	12
3.1 Veturisarjojen kuvailu	12
3.2 Aineiston kuvailu ja rajausta	14
3.3 Käytetyt menetelmät	17
4 Tulokset	20
4.1 Aikasarjojen ennustemalleille pääteltyt kertaluvut	20
4.2 Mallien estimoidut parametrit ja virheet malleille	22
4.3 Mallien antamat ennusteet ja virheet	25
4.4 Johtopäätökset	27
5 Yhteenveto ja tulevia tutkimusaiheita	29
A Saatujen ennusteiden arvot	34

1 Johdanto

Raideliikenne on Suomessa merkittävä osa taloutta ja suomalaisten elämää. Vuonna 2022 Suomessa liikkui raiteilla noin 30 miljoonaa tonnia rahtia (Väylävirasto, 2023 a), ja kaukoliikenteen matkoja tehtiin yli 13 miljoonaa kappaletta (Väylävirasto, 2023 b). Suurin osa junista muodostuu veturista ja veturiin kiinnitettävistä vaunuista.

Tässä kandidaatintyössä tutkitaan juniin liittyviä veturina-ajoja. Veturina-ajo on yleisesti rautatiealalla käytetty termi, jolla tarkoitetaan veturin ajamista toiselle paikkakunnalle ilman vaunuja. Tähän syynä voivat olla esimerkiksi epätasapainossa olevat kuljetusvirrat, jolloin paikkakunnalta toiselle kulkee enemmän kuormaa. Veturina-ajot eivät tuo rautatieyhtiöille liikevaihtoa, joten niitä pyritään välttämään. Veturina-ajon suurimpia kustannuksia ovat veturin kuluttama sähköenergia tai dieselpolttoaine sekä veturinkuljettajan palkka.

Veturit ovat suuri investointi rautatieyhtiöille. Yksi veturi voi Yleisradion (Yle) mukaan maksaa yli kolme miljoonaa euroa (Yle, 2022). Suomen valtion omistama logistiikkakonserni VR Group (VR), aiemmin Valtionrautatiet, omistaa yli neljäsataa veturia junaliikenteen havaintojärjestelmä Julian mukaan (Julia, 2018).

Tässä työssä tarkastellaan veturina-ajoista kertyvien kilometrien ennustamista. Veturille kertyviä kilometrejä halutaan ennustaa mahdollisimman tarkasti useista syistä. Rautatieyhtiöt haluavat tietää, kuinka moneen veturiin on kannattavaa investoida. Veturien kappalemäärän lisäksi vatureille kertyneet kilometrit vaikuttavat siihen, kuinka monta veturinkuljettajaa on kannattavaa palkata.

Vetureille kertyy eniten kilometrejä linja-ajosta, jolla tarkoitetaan matkustajatai tavarajunia. Matkustajaliikenteen kilometriarviot saadaan kysyntäennusteista. Rautatieyhtiöiden asiakkaat toimittavat kukin arvion kuljetustilausten määrästä, ja näiden pohjalta muodostetaan yhtiön tavaraliikenteen kuljetusten kokonaisennuste. Myös veturina-ajot ja ratapihoilla tehtävät vaihto- tai päivystystyöt kerryttävät kilometrejä.

Logistiikka-alalla suunnittelu on tyypillistä toteuttaa kolmessa eri vaiheessa: strategisessa, taktisessa ja operatiivisessa (Assad, 1980). Näistä strateginen suunnittelu saatetaan tehdä jopa vuoden päähän. Taktinen suunnittelu vastaa junien suunnittelusta vatureille muutamien viikkojen aikaikkunassa. Operatiivinen suunnittelu vastaa hyvin lyhyestä aikavälistä, esimerkiksi saman ja seuraavan päivän operatioiden toteutumisesta ja muutoksista. Veturina-ajot syntyvät eniten operatiivisessa suunnitteluvaiheessa. Kun aiemmin tehtyihin suunnitelmiin tulee loppuvaiheessa muutoksia, kuten peruuntunut juna, on yksinkertaisinta ajaa junalle suunniteltu veturi ja kuljettaja veturina-ajona peruuntuneen junan aikataululla. Näin varmistetaan jatkosuunnitelman eheys ja vältetään mahdolliset lisämuutokset.

Tässä työssä päätellään toteutuneiden veturina-ajojen perusteella, onko eri veturisarjoille tulevia veturina-ajot järkevää ennustaa eri menetelmillä. Erilaisia vatureita on useita. VR:llä on aktiivisessa käytössä neljä dieselveturisarjaa ja kolme sähköveturisarjaa. Veturisarjoille tulevia veturina-ajot voisi olla hyödyllistä ennustaa eri tavoilla, koska ajot ovat erityyppisiä. Dieselveturit ovat pääasiassa paikkakuntakohtaisia, ja käyvät hakemassa junan tai viemässä tyhjiä vaunuja lastaukseen läheiselle paikkakunnalle. Dieselveturit täytyy tankata säännöllisin väliajoin, joten

niitä voidaan säilyttää vain tietyillä paikkakunnilla. Niillä ajetaan lähinnä sähköistämättömillä rataosuuksilla. Suomen rataverkosta hieman alle 60% on sähköistettyä (Tilastokeskus, 2023). Sähkövetureilla ajetaan joka puolella Suomea. Niille syntyvä veturina-ajo on paikkakuntien välisen veturitasapainon ylläpitoa. Vähentämällä etenkin dieselvetureiden veturina-ajoja, vähennetään myös päästöjä. Sähkö, jota rautateillä käytetään, on VR:n vastuullisuusraportin mukaan tuotettu uusituvalla energialla (VR Group, 2023 a).

Työstä rajataan pois alueellinen veturina-ajojen tarkastelu. Veturina-ajot riippuvat myös ajettavien junien ja ennustettujen tavaratonniin määrästä, mutta tähän riippuvuuteen ei keskitytä tässä työssä. Riippuvuus ei ole välttämättä täysin suora niin, että junien vähentyessä myös veturina-ajot vähentyisivät. Junien avulla voidaan siirrellä vetureita moniajolla, jossa junaan kiinnitetään ylimääräisiä vetureita. Jos ajettavia junia on vähän, mahdollisuuksia moniajolle on myös vähemmän. Tästä syystä alhaisempi junien määrä saattaisi myös johtaa veturina-ajojen lisääntymiseen.

Työn Luvussa 2 esitellään veturina-ajoihin liittyvää tutkimuskirjallisuutta. Työssä käytettiin VR:n aineistoa ajetuista junista ja veturina-ajoista. Aineistoon ja käytettyihin menetelmiin perehdytään tarkemmin Luvussa 3. Tutkimuksen tulokset ja johtopäätökset esitetään Luvussa 4. Lopuksi Luvussa 5 työn tuloksista tehdään yhteenveto. Työn lähteinä on käytetty alan tutkimuskirjallisuutta ja suullisia keskusteluita VR:n suunnittelijoiden kanssa.

2 Aikaisempi tutkimus

Suurin osa alan tutkimuskirjallisuudesta keskittyy Euroopan, Pohjois-Amerikan tai Aasian rautateille, joista Suomen olosuhteet poikkeavat paljon. Suomen rautatiet eroavat muun Euroopan rautateistä maantieteellisesti suurten etäisyyksien vuoksi sekä kilpailun määrällä. Suomen raideleveys on kapeampi kuin muualla Euroopassa. Koska muita operaattoreita VR:n lisäksi on hyvin vähän, vetureita ei voi vuokrata muilta toimijoilta. Tämä saattaisi olla kannattavampaa kuin veturina-ajo ([Teichmann et al., 2015](#)). Suomessa junissa on korkeintaan kolme veturia kuten Euroopassa, kun taas esimerkiksi Yhdysvalloissa vetureita voi olla jopa kaksitoista yhdessä junassa ([Frisch et al., 2019](#)). Näiden eroavaisuuksien vuoksi aiheeseen liittyvää tutkimuskirjallisuutta on rajallisesti. Osa tutkimuksesta on sovellettavissa veturina-ajoihin ja veturikaluston koon arviointiin.

2.1 Veturikaluston koon arviointi

Veturikierto, eli veturin vedettäväksi suunniteltujen junien järjestys, on yksi vaikuttava tekijä veturina-ajojen syntymiseen. Veturikiertojen optimointimalleja on kehitetty useita erilaisilla rajauksilla ja laajennuksilla. [Ahuja et al. \(2005\)](#) ovat kehittäneet mallin, joka huomioi muun muassa erilaiset veturityypit, niiden hevosvoimat, veturina-ajot ja hinaukset. Malli pyrkii minimoimaan kustannukset ja vaunujen irrottamiset vetureista. Optimointiongelman ratkaisuun kuluva aika venyi pitkäksi, koska malliin liittyi paljon rajoitteita ja huomioitavia muuttujia. Rautatieliikennettä on mallinnettu myös useilla erilaisilla simulointimenetelmillä ([Assad, 1977](#)).

Veturi- ja vaunukaluston kokojen vaikutusta junien myöhästymisiin on tutkittu kirjallisuudessa. [Powell et al. \(2012\)](#) kehittivät approksimoivan dynaamisen optimointimallin veturien käytön suunnitteluun. Malli huomioi myös veturien aiemmat reitit suunnitelmassa. Mallin avulla he pystyivät simuloimaan rautatiekaluston kokoa ja sen vaikutuksia. He laskivat junien myöhästymiset, jotka aiheutuivat kaluston koon muutoksesta. Myöhästymisten määrä pieneni, kun kaluston kokoa kasvatettiin.

[Godwin et al. \(2008\)](#) tekemä tutkimus keskittyi Intian rautatieliikenteeseen, jossa ei käytetä etukäteen tehtyjä aikatauluja. Tutkimuksessa kehitettiin simulointimalli, jonka avulla oli mahdollista päätellä sopiva veturikaluston koko huomioiden veturina-ajopolitiikka. Tutkimuksessa motivaatio veturina-ajolle oli minimoida lastattujen vaunujen odotusaika asemalla. Tutkimuksessa jaettiin veturina-ajot kolmeen eri tyyppiin: ei odotusta sekä tarkka tai mukautettu odotusaika eri asemilla. Veturien kappalemäärän lisääntyessä vaunujen odotusaika laski ja investointikulut nousivat korkeiksi. Työn johtopäätös oli rajoittaa veturien määrää ja löytää sopiva odotusaika vetureille eri asemilla, jotta vaunut eivät joudu odottamaan.

Veturikaluston koon suurentamisen on huomattu laskevan veturina-ajojen kustannuksia ja määrää ([Kuo ja Nicholls, 2007](#); [Frisch et al., 2019](#)). [Kuo ja Nicholls \(2007\)](#) tavoitteena oli maksimoida veturien käyttöaste ilman, että veturina-ajoista syntyvät kustannukset kasvavat. Maksimointi tapahtui optimointimallin avulla. He keskittyivät allokoidaan veturit oikeisiin paikkoihin. Tutkimus tehtiin 16 veturilla ja investoimalla yhteen veturiin lisää onnistuttiin säästämään veturina-ajojen

kustannuksista vuosittain noin 57 000 dollaria. Veturiin tarvittava investointi oli tässä tutkimuksessa 1,5 miljoonaa dollaria. Vuosittaista säästöä ei pidetty riittävän suurena oikeuttamaan investointia uuteen veturiin. Frisch et al. (2019) tutkivat huoltokiertojen vaikutusta veturikiertoihin ja arvioivat, että veturien huolloista syntyy myös veturina-ajoja, kun tietty veturiyksilö tulee saada huoltoon tietylle paikkakunnalle. He totesivat että veturina-ajojen määrä laskee, kun veturikaluston kokoa kasvatetaan.

Yksi suurimmista syistä veturina-ajoihin ovat epätasaiset tavaravirrat (Hall, 1999). Hall (1999) kehitti stokastisen mallin, jonka avulla voidaan mallintaa epätasaisuutta. Työ keskittyi logistiikka-ajoneuvoihin yleisesti. Tavoitteena oli hyödyntää mallia kalustonhallinnassa. Mallissa tyhjänäajoa mallinnettiin kahdesta eri näkökulmasta: väliaikaisena poikkeamana ja palaamisena omalle lähtöpaikalle. Hän huomasi että pitkän aikavälin epätasaisuudet ovat merkittävämpiä tyhjänäajojen kannalta kuin lyhyen aikavälin epätasaisuudet.

Veturina-ajojen avulla on onnistuttu kasvattamaan rautatieyhtiöiden tuottoja (Wagenaar et al., 2017). Wagenaar et al. (2017) keskittyivät Alankomaiden matkustajaliikenteeseen. Tutkimuksessa veturina-ajoja ei ajettu, ellei se ollut taloudellisesti kannattavaa. Veturina-ajoilla onnistuttiin vähentämään peruttujen junien määrää häiriöiden tapahtuessa. Suomessa tavaraliikenteeseen liittyvät veturina-ajot ovat tyyppillisiä häiriötilanteissa. On siis mahdollista, että veturina-ajot lisäävät epäsuorasti tuottoja myös Suomessa.

Rautateiden tavaraliikenteen strategiseen suunnitteluun on kehitetty erilaisia operaatiotutkimuksen malleja. Strategisia päätöksiä ovat esimerkiksi kaluston koko, mihin kalusto pääasiallisesti sijoitellaan ja mitä kalustoa laitetaan sivuun. Crainic ja Laporte (1997) esittelevät useita malleja suunnitteluprosessin eri vaiheisiin rautatieliikenteessä. Strategiseen suunnitteluun he esittelivät kolmen tyyppisiä suunnittelumalleja: sijaintimalleja, verkostomalleja sekä alueellisia moniosaisen suunnittelun malleja. Näihin malleihin esiteltiin tarkemmat optimointimallit, joiden avulla strategisia päätöksiä voidaan tehdä. Erityisesti moniosaisessa suunnittelussa tärkeää on ennustaa tulevaisuuden kuljetusvolyymejä. Tutkimus ei kuitenkaan keskittynyt ennustamiseen. Pääperiaatteena tässä mallissa oli muodostaa monen eri alan, tässä tapauksessa matkustaja- ja tavaraliikenteen, verkosto. Verkoston kuljetusvirroista syntyvät kustannukset minimoitiin.

Veturikaluston kokoa ja koon muutoksien vaikutuksia on siis tutkittu ja huomattu että mitä enemmän on vetureita, sitä vähemmän on tarvetta veturina-ajoille. Muilta osin tutkimus painottuu optimointimalleihin, joilla minimoidaan kustannukset. Veturina-ajojen ennustamiseen on hyvin vähän kirjallisuutta.

2.2 Muiden ajoneuvolajien soveltaminen

Veturina-ajoihin voidaan myös soveltaa paljeluajoneuvojen, eli esimerkiksi bussien ja taksien, tyhjänäajoista tehtyä tutkimusta. Jopa puolet taksien ajosta tapahtuu ilman asiakasta (Nair et al., 2020). Tämä on suurempi osuus kuin rautatieliikenteessä (Poikonen, 2023).

Taksien tyhjänäajojen vähentämiseen on kehitetty erilaisia malleja. Tähän on

pyrityt esimerkiksi ennustamalla seuraavan asiakkaan haku- tai jättöpaikkaa (Nair et al., 2020; Kontou et al., 2020; Carson-Bell et al., 2021). Tyhjänäajon vähentämistä varten Nair et al. (2020) kehittivät matemaattisen mallin, jolla ennustettiin seuraavan asiakkaan hakupaikkaa, kun ajoneuvon alkuperäinen paikka on tiedossa. He huomioivat myös, että osa tyhjänäajosta on taksikuskihenkilökohtaisten asioiden hoitamista. Tutkimuksessa käytettiin Ride-Austin -yhtiön julkista aineistoa. Samaa aineistoa käyttivät osittain myös Kontou et al. (2020). Tutkimuksessa käytettiin takaisinkytkettyjä neuroverkkoja ennustamaan tulevaisuuden matkojen kysyntää. He totesivat, että taksikuskit siirtyvät usein uudelle paikalle odottamaan asiakasta, mikä lisää tyhjänäajoa. Carson-Bell et al. (2021) käyttivät aineistona taksimatkoja Chicagon alueelta. He ennustivat asiakkaiden jättö- ja hakupaikkoja tietolouhintamenetelmien avulla. Tutkimuksessa onnistuttiin vähentämään tyhjänäajojen määrää kuskien paremmilla sijainneilla, jotka voitiin laskea heidän mallillaan.

Bussit viedään toisinaan yöksi eri paikkoihin, mistä reitit alkavat seuraavana päivänä. Säilytyspaikkoja on useita, eikä lähin säilytyspaikka ole aina optimaalisin, jos seuraavan päivän bussireitti alkaa kauempaa (Prakash et al., 1999). Busseilla tyhjänäajoa tulee siis vuorokauden vaihteessa. Prakash et al. (1999) optimoivat säilytyspaikkoja minimoiden tyhjänäajot uusien reittien alkaessa. Vastaava ongelma esiintyy joskus myös vetureilla, jos veturia ei voida jättää tietylle paikkakunnalle yön yli. Raideverkko on rajatumpi kuin tieverkosto, joten vaihtoehtoja veturien säilytykselle on vähemmän ja päätöksenteko on siten myös yksinkertaisempaa.

Toinen esimerkki sovellettavasta tutkimuskirjallisuudesta on maanteiden logistiikkaliikenne. Kuljetustapa on maanteillä samanlainen kuin rautateillä: haetaan tai viedään kuormaa paikkakunnalta toiselle. Kuljetusajoneuvoja voidaan ajaa alitäytettyinä, ja tällä tavoin on onnistuttu vähentämään tyhjänäajojen kustannuksia (Özkaya et al., 2010). Tutkimuksessa muodostettiin regressiomalli, jolla mallinnettiin kuljetusmarkkinoiden hintatasoa. Johtopäätöksenä tutkimuksessa todettiin, että ajamalla alitäytettyjä kuormia, voidaan nostaa kuljetusten hintaa ja tuottoja. Tässä siis tyhjänäajoja korvattiin alitäytetyillä kuljetuksilla. Alitäyttämällä junia voitaisiin mahdollisesti vähentää myös veturina-ajojen kuluja.

Maantielogistiikan kaluston, ja erityisesti vetoajoneuvokaluston kokoa, on tutkittu kirjallisuudessa. Amjath et al. (2022) tutkivat vetoajoneuvokaluston koon muutoksien vaikutusta kysyntään vastaamiseen käyttäen jonotusverkkomalleja. Muita mahdollisia tapoja määritellä kaluston kokoa olivat simulointi, analyttiset mallit tai hybridimallit. Tutkimuksen lopussa todettiin, että vuokrasopimuksilla voi saada kaluston kokoon joustoa ilman suuria investointeja.

Laivaliikenteessä matkat ovat kestoltaan pidempiä kuin muilla aloilla. Laivojen kuvataankin kulkevan ympyrän muotoista reittiä niin, ettei laiva ole koskaan tyhjänä. Rahtiliikenne merellä on kansainvälistä, ja esimerkiksi lakot eri maissa vaikuttavat laivojen liikkeisiin. Reittivaihtoehtoja on lähes loputtomasti. (Ronen, 1983)

Lentoliikenteen rahdin kysynnän määrää on arvoitu tutkimuskirjallisuudessa. Esimerkiksi Li et al. (2020) esittelevät rahdin määrän ennustamiseen useita eri tapoja, kuten ARIMA-mallin ja takaisinkytketyn neuroverkon. Koska lentoliikenne on laivaliikenteen tavoin kansainvälistä, kerrotaan esimerkiksi maan bruttokansantuotteen vaikuttavan lentorahdin määrään. Tutkimuksen mallit eivät sovellu tähän työhön,

sillä siinä ennustettiin rahtitavaran määrää.

Logistiikkaliikenteen haasteet ovat samantyyppisiä alasta riippumatta. Logistiikkaajoneuvojen reittejä optimoidaan kustannuksien minimoimiseksi. Uudemmissa tutkimuksissa 2020-luvulla on sovellettu koneoppimismenetelmiä taksien tyhjänäajon vähentämiseen. Tämä saattaakin olla myös maanteiden ja raideliikenteen seuraava askel. Veturina-ajojen määrää ei kuitenkaan ole juurikaan tutkittu tai pyritty ennustamaan.

3 Tutkimusaineisto ja -menetelmät

3.1 Veturisarjojen kuvailu

Kuten työn Luvussa 1 todettiin, VR omistaa useita eri veturisarjoja. Veturit ovat joko sähkö- tai dieselvetureita, ja eri veturisarjoilla on erilaisia ominaisuuksia. Ominaisuuksien vuoksi veturisarjoja käytetään erilaisissa junissa ja tehtävissä. Osa vetureista soveltuu paremmin raskaisiin vaihto- ja päivystystöihin ratapihoilla, ja osa ajamaan nopeita matkustajajunia.

VR:llä on myös vetureita, joilla liikennöidään hyvin harvoin museoliikennettä, kuten dieselveturimalli Dr13 ja höyryveturimalli Hr1. Näitä veturimalleja ei tarkasteltu työssä ollenkaan. Tässä luvussa kuvaillut veturimallit ovat dieselveturimallit Dv12, Dr14, Dr16, Dr19 ja sähköveturimallit Sr1, Sr2 ja Sr3. Taulukossa 1 kuvataan tarkemmin veturien ominaisuuksia veturisarjoittain. Eri käyttötarkoituksia ovat matkustajaliikenne (M), tavaraliikenne (T) sekä vaihto- ja päivystystyö (PV).

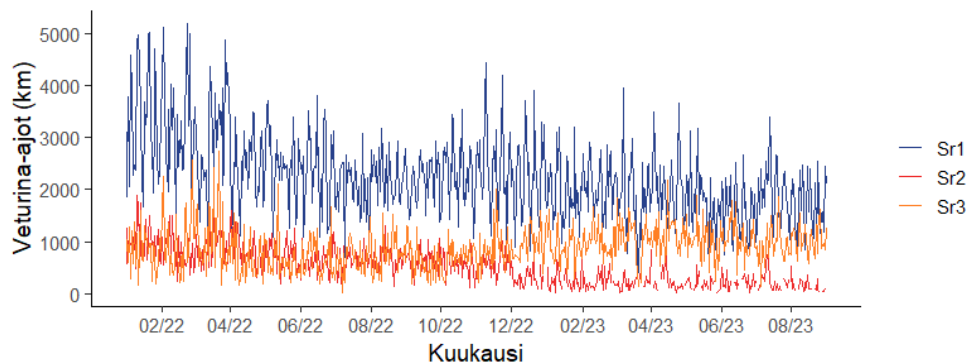
Taulukko 1: Veturisarjojen tärkeimmät ominaisuudet ja pääasialliset käyttötarkoitukset

Malli	Teho (kW)	Käyttö	Huippunopeus (km/h)
Sr1	3280	T	140
Sr2	6000	M	210
Sr3	6400	M, T	200
Dv12	1000	T, PV	125
Dr14	875	PV	75
Dr16	1600	M	140
Dr19	2×950	T, PV	120

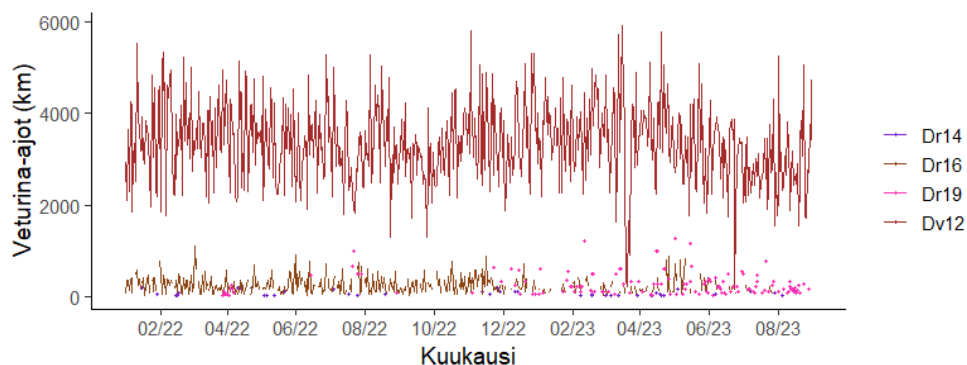
Taulukosta 1 nähdään, että sähkövetureita käytetään pääasiallisesti linja-ajoon, kun taas dieselvetureita käytetään tavaraliikenteessä tai vaihto- ja päivystystöissä. Dieselvetureista Dv12 ajaa myös tavaraliikenteen junia. Dr19 ajaa myös tavaraliikennettä, joskin huomattavasti vähemmän kuin Dv12. Dr16-veturilla ajetaan matkustajaliikenne sähköistämättömällä rataosuudella Kemin ja Kolarin välillä. Yllä olevasta taulukosta nähdään myös veturien tehot. Dr19-veturissa on kaksi dieselmoottoria, ja se on tehokkain dieselveturi. Sähkövetureista tehokkain on Sr3. Tehokkaimmat veturisarjat ovat myös uusimmat. Sr3-vetureita on otettu käyttöön vuodesta 2017 alkaen, ja Dr19-vetureita vuodesta 2023. Sr2-vetureista osa on otettu käyttöön 2000-luvun alussa. Muut veturisarjat on otettu käyttöön viime vuosikymmenen puolella. (VR Group, 2023 b)

Kuvasta 1 nähdään että vain tavaraliikennettä ajavalla Sr1-veturisarjalla on huomattavasti enemmän veturina-ajoa kuin Sr2- ja Sr3-veturisarjoilla. Määrällisesti Sr1-vetureita on liikenteessä enemmän kuin muita sähkövetureita (VR Group, 2023 c). Kuvasta nähdään myös, että Sr2-veturien veturina-ajot ovat laskussa ja Sr3-veturien veturina-ajot nousussa.

Kuvasta 2 nähdään myös Taulukossa 2 näkyvä ero dieselveturisarjojen veturina-ajojen määrissä. Dr14- ja Dr19-sarjojen veturina-ajoa oli todella vähän tarkastel-



Kuva 1: Sähköveturien päivittäiset veturina-ajot vuoden 2022 alusta syyskuuhun 2023



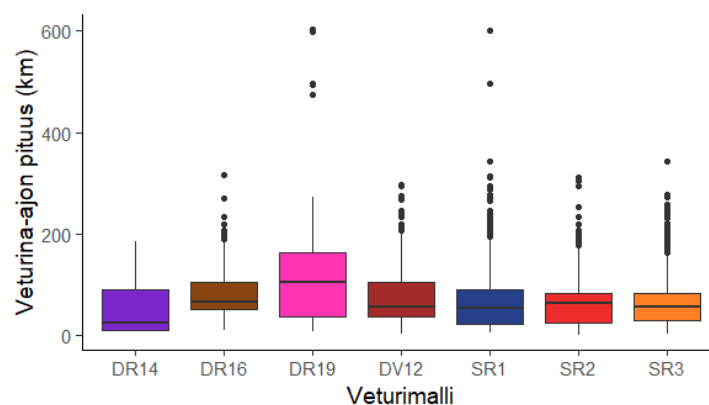
Kuva 2: Dieselveturien päivittäiset veturina-ajot vuoden 2022 alusta syyskuuhun 2023

lulla ajanjaksolla, joten ne on merkitty alla olevaan kuvaan yksittäisillä pisteillä. Dv12-veturisarjalla on Kuvan 2 perusteella huomattavasti enemmän veturina-ajoja kuin muilla dieselveturisarjoilla. Tämä johtuu siitä, että kappalemäärällisesti Dv12-vetureita on enemmän ja tavaraliikenteeseen liittyy enemmän veturina-ajoja kuin vaihtotöihin (VR Group, 2023 c).

Taulukko 2: Veturina-ajojen määrät ja keskipituudet (km) veturisarjoittain

Malli	Määrä	Mediaanipituus	Pituuden keskiarvo
Sr1	22055	53	62
Sr2	4697	64	63
Sr3	7991	56	66
Dv12	28049	55	73
Dr14	57	24	50
Dr16	1176	66	88
Dr19	269	105	131
Sähköt yht.	34743	54	63
Dieselit yht.	29551	56	73

Taulukosta 2 nähdään, että sähköveturien veturina-ajot ovat keskimäärin lyhyempiä, mutta niitä on hiukan enemmän. Dieselvepureilla on hiukan pidempiä veturina-ajoja mutta kappalemäärällisesti vähemmän kuin sähkövetureilla. Yllä oleva taulukko vahvistaa ajatusta siitä, että sähköveturit tasapainottavat lähipaikkakunnan veturipuutetta, ja dieselvehuri taas saattaa hakea junan paikkakunnalta, missä ei ole sähköistettyjä raiteita. Yhteenlasketut veturina-ajot sähkövetureilla olivat 2 390 000 kilometriä ja dieseleillä 2 360 000 kilometriä. Kokonaiskilometrimäärät ovat lähellä toisiaan.



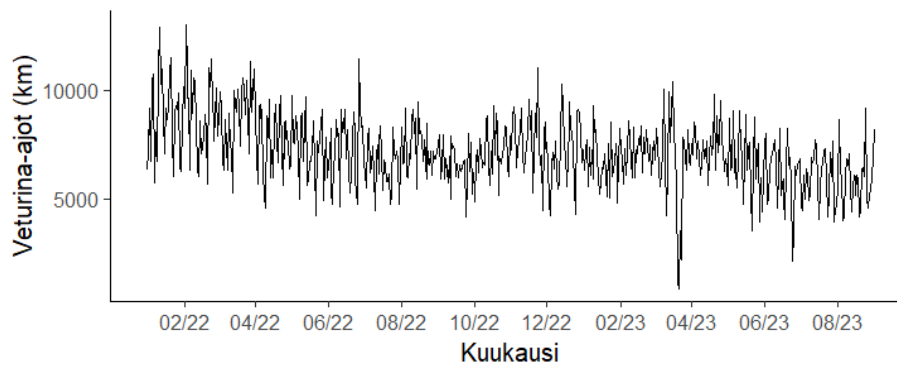
Kuva 3: Veturina-ajojen pituuksien jakautuminen veturisarjoittain

Dv12-, Sr1- ja Sr3-veturisarjojen veturina-ajoja oli määrällisesti eniten. Kuvasta 3 nähdään, että veturina-ajojen määrien kasvaessa niiden jakautuminen alkaa muistuttaa toisiaan. Kuvassa mediaani on ilmoitettu mustalla viivalla laatikon keskellä, laatikon alareuna on ensimmäinen neljännes, ja yläraja kolmas neljännes. Kuvassa 3 nähdään yksittäiset, erityisen pitkät veturina-ajot Sr1- ja Dr19-veturisarjoilla. Nämä yli 600 kilometrin pituiset ajot ovat saattaneet syntyä, kun tietty veturi on täytynyt saada huoltoon tietylle paikkakunnalle. Suurin osa matkoista on kuitenkin alle sadan kilometrin pituisia kaikilla veturisarjoilla.

3.2 Aineiston kuvailu ja rajaus

Aineistosta otettiin käyttöön vain varsinaiset veturina-ajot. Muita veturina-ajon tyyppisiä tilanteita ovat testiajot ja veturin siirtämiset vaihtotyönä. Testiajoja ajetaan esimerkiksi rataosuuksilla, joihin on juuri valmistunut sähköistys. Nämä ovat harvinaisia ja poikkeustapauksia tavalliseen liikenteeseen. Veturin siirtämiset vaihtotyönä tapahtuvat ratapihoilla, joissa välimatkat ovat suuria ratapihoiksi ja siirtyminen toiselle osalle ratapihaa vaatii virallisen aikataulun. Siirtymiset ovat hyvin lyhyitä suhteessa varsinaisiin veturina-ajoihin. Testiajoja tai veturin siirtämisä vaihtotyönä ei otettu aineistoon mukaan.

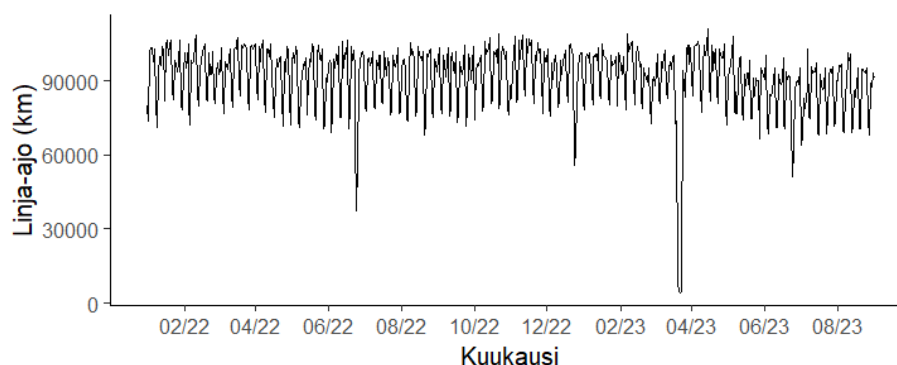
Raideliikenteessä tapahtuu jatkuvasti poikkeuksia. Ajettavien junien määriin vaikuttavia poikkeuksia ovat ratatyöt, raideliikenteen työntekijöiden tai asiakkaiden lakot tai tehdasseisokit. Näistä poikkeuksista työssä huomioitiin veturinkuljettajien lakko, jolloin lähes kaikki junat jäivät ajamatta (Poikonen, 2023). Kuvista 4 ja 5



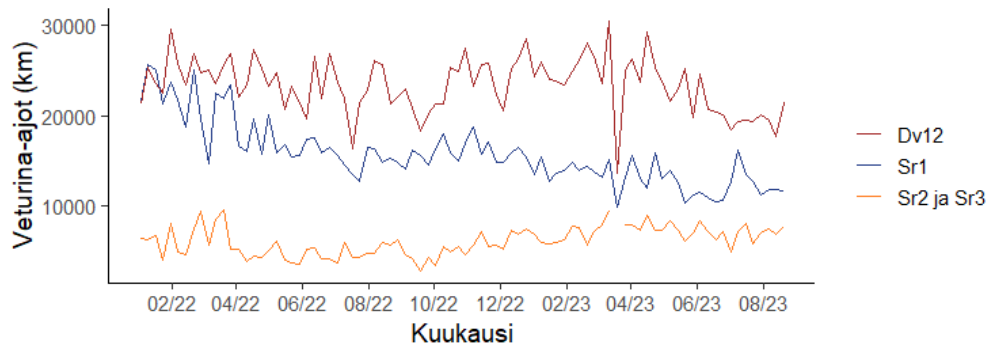
Kuva 4: Kaikkien veturisarjojen yhteenlasketut päivittäiset veturina-ajot vuoden 2022 alusta syyskuuhun 2023

nähdään lakon aiheuttama alaspäin suuntautuva piikki vuoden 2023 maaliskuussa. Vähäiset ajomäärät olivat satunnaisvaihtelua, eikä vastaavia lakkoja ole vuosittain. Lakon aikaiset alhaiset veturina-ajomäärät tasattiin lopulliseen aineistoon. Tasaa-minen tapahtui selvittämällä lakkoa edeltävän ja seuraavan viikon veturina-ajojen määrät, ja laskemalla niistä keskiarvo lakkoviikolle.

Poikkeuksista säännöllisempiä ovat tehtaiden seisokit ja ratatyöt. Tehdasseisokit ovat yleisiä, sillä suurille tehtailla täytyy tehdä vuosittain huolto- ja puhdistustöitä. Seisokit voivat kestää useita päiviä. Ne vaikuttavat koko Suomen liikenteeseen, sillä tehtailla ajetaan junia joka puolelta Suomea (Poikonen, 2023). Tehdasseisokkeja pidettiin myös satunnaisvaihteluna. Tässä työssä oletettiin, että niiden vaikutus tasaantuu koko Suomen tasolla. Ratatöitä on myös ympäri vuoden, kesällä hieman enemmän kuin talvella. Esimerkiksi juhannuksena on usein suuria ratatöitä. Juhannuksina 2022 ja 2023 on mahdollista nähdä pieni lasku kaikissa ajetuissa junissa Kuvas-sa 5. Kuvassa 4 näkyy myös alaspäin suuntautuva piikki veturina-ajoissa vuoden 2023 juhannuksen aikaan. Juhannuksen ratatöiden aiheuttamat piikit tasaantuivat viikkotason tarkastelussa, joten niitä ei tasattu erikseen.

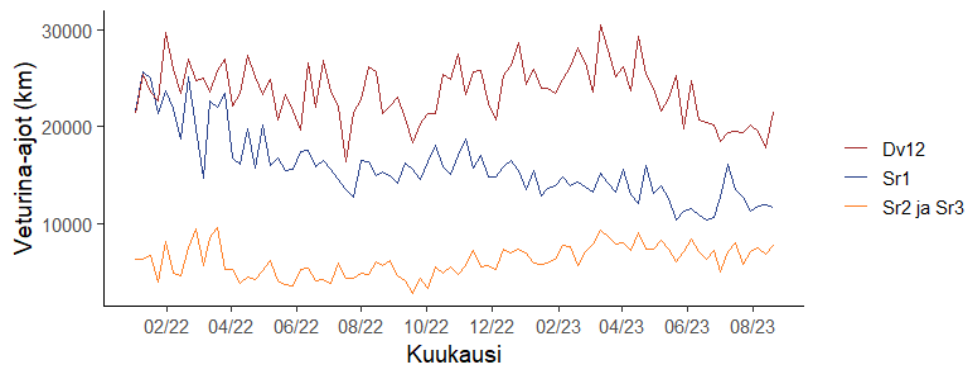


Kuva 5: Matkustaja- ja tavaraliikenteen junien kilometrit päivätasolla vuodesta 2022 alkaen



Kuva 6: Viikoittaiset veturina-ajot valituilla veturisarjoilla, veturinkuljattajien lakon aiheuttama piikki näkyvissä

Lopullisessa aineistossa olivat käsiteltävinä veturisarjojen Sr1 ja Dv12 aikasarjat ja yhdistetty aikasarja Sr2- ja Sr3-veturisarjoista. Sr2-veturisarjan veturina-ajoja ei ollut mielekästä ennustaa erikseen niiden vähyyden vuoksi joten lopulliseen aineistoon päätettiin yhdistää Sr2- ja Sr3-veturien veturina-ajot. Kaikilla veturisarjoilla ajettavat junat ovat erityyppisiä, joten niiden ennustamista on mielekästä vertailla. Lopulliseen jakoon ja tarkasteltaviin veturisarjoihin päädyttiin Alaluvussa 3.1 esiin tulleiden käyttötarkoituks- ja määräerojen vuoksi. Sr1-aikasarja kuvaa sähköllä ajettavaan tavaraliikenteeseen liittyviä veturina-ajoja. Sr2 ja Sr3 -aikasarja kuvaa tavara- ja matkustajaliikennettä ajaville vetureille tulevia veturina-ajoja. Dv12-aikasarja kuvaa dieselveturilla ajettaviin tavarajuniin liittyviä veturina-ajoja. Muilla dieselvetureilla ajettujen junien osuus on vain pieni osa kaikista junista, joten tässä työssä ei keskitytty niiden ennustamiseen. Kuvasta 6 nähdään, että valituista veturisarjoista Dv12:lla on eniten veturina-ajoja, seuraavaksi eniten Sr1:lla ja vähiten Sr2- ja Sr3-vetureilla.



Kuva 7: Viikoittaiset veturina-ajot valituilla veturisarjoilla, lakon aiheuttama piikki tasattu

Kuvasta 7 nähdään lopullinen aineisto. Mallin rakennusta varten aineistosta otettiin 70 ensimmäistä viikkoa. Seuraavat viisi ja kymmenen viikkoa toimivat mallin validoinnissa. 70 viikkoa aikasarjan alusta laskettuna päättyy vuoden 2023 toukokuun alkupuolelle. Validointiaineiston viimeinen viikko on heinäkuun 2023 alussa.

3.3 Käytetyt menetelmät

Työssä käytetty aineisto on aikasarjadataa. Aikasarjassa datapisteet on kerätty ajanhetkittäin ja usein tasavälein. Työn lopullinen aineisto muodostettiin laskemalla ensin päivittäiset veturina-ajot yhteen veturisarjoittain. Päivittäiset veturina-ajot summattiin viikoittaisiksi aikasarjoiksi. Tässä työssä tehtiin aikasarja-analyysi eri veturisarjojen veturina-ajoille. Yksi aikasarja-analyysissä yleisesti käytetty ja tunnettu malli on ARMA-malli. ARMA-malli voidaan jakaa kahteen erilliseen malliin: AR- ja MA-malliin. Veturina-ajojen määrät ovat viikoittaisellakin tasolla vaihtelevia ja satunnaisia, kuten Kuvasta 7 nähdään. ARMA-mallin avulla on mahdollista muodostaa lähitulevaisuudelle tarkka ennustemalli. ARMA-mallia rakentaessa on mahdollisuus arvioida, onko eri veturisarjojen veturina-ajoilla esimerkiksi kausittaisia eroja tai trendieroja.

ARMA-mallin ensimmäinen osa tarkoittaa autoregressiivistä mallia (engl. autoregressive, AR). Autoregressiivinen malli huomioi aikasarjan edellisiä termejä erilaisilla parametreilla, jotka määritetään aikasarjasta estimoimalla. Malli voidaan esittää matemaattisesti seuraavalla tavalla:

$$X_t = \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i},$$

missä ϵ_t on mallin virhetermi ja ϕ_1, \dots, ϕ_p ovat mallin parametrit. Kertaluku p kertoo, kuinka monta edellistä termiä X_{t-p}, \dots, X_{t-1} mallissa huomioidaan. Mallista käytetään merkintää AR(p).

MA-malli ARMA-mallissa on liukuvan keskiarvon malli (engl. moving average, MA). MA-mallissa aikasarjan muuttujaa X_t selittävät edellisten ennustusten virhetermit. Matemaattisesti MA-malli voidaan esittää seuraavasti:

$$X_t = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i},$$

missä $\epsilon_{t-q}, \dots, \epsilon_t$ ovat mallin satunnaisvirheet ja $\theta_1, \dots, \theta_q$ ovat mallin parametrit. Mallista käytetään merkintää MA(q). Mallin kertaluku q kertoo, kuinka monta edellistä virhetermiä mallissa huomioidaan.

ARMA(p, q)-malli saadaan, kun yhdistetään autoregressiivinen AR(p)-malli ja liukuvan keskiarvon MA(q)-malli. Matemaattisesti malli voidaan esittää seuraavasti:

$$X_t = \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i}.$$

Mallissa lasketaan yhteen kertalukujen p ja q mukaiset määrät AR- ja MA-mallien termejä.

Jotta aiemmin esitelty ARMA-malli voidaan sovittaa aikasarjaan, aikasarjan täytyy täyttää mallin oletukset. Aikasarjan tulee olla stationaarinen ja kääntyvä. Stationaarisuudella tarkoitetaan sitä, että aikasarjan arvot eivät riipu ajasta. Aikasarjan odotusarvo tai varianssi ei saa vaihdella aikasarjan eri kohdissa, vaan niiden tulee olla vakioita. Tämä tarkoittaa sitä, että aikasarjan kuvaajassa ei näy trendiä tai

jaksottaisia vaihteluita. Kääntyvyydellä tarkoitetaan sitä, että MA-osuus sarjasta on ilmaistavissa $AR(\infty)$ -mallina. Lisää stationaarisuudesta ja kääntyvyydestä voi lukea esimerkiksi kirjasta *Time series: Theory and Methods* (Brockwell ja Davis, 2009).

Joskus aikasarjan arvot kasvavat tai pienenevät lineaarisesti ajan edetessä. Tällöin aikasarjassa näkyy nouseva tai laskeva lineaarinen trendi, eikä se ei ole stationaarinen. Tällaisesta aikasarjasta saadaan stationaarinen differoimalla se. Differointi tapahtuu vähentämällä edellinen arvo tarkasteltavasta arvosta. Differoituun aikasarjaan sovitettu ARMA-malli ilmaistaan ARIMA(p, d, q)-mallina, jossa d ilmoittaa differointien määrän. Usein riittää ottaa yksi differenssi.

ARIMA(p, d, q)-mallin kertaluvut p ja q voidaan päätellä aikasarjan autokorrelaatiofunktioiden avulla. Korrelaatio kertoo kahden muuttujan lineaarisesta riippuvuudesta. Autokorrelaatiofunktio (engl. autocorrelation function, ACF) kertoo aikasarjan arvojen korrelaatiosta sarjan aiempiin arvoihin eri viiveillä (engl. lag). ARIMA(p, d, q)-malliin tulevien MA-termien määrä, eli kertaluku q päätellään aikasarjan autokorrelaatiofunktion avulla. Jos tietyllä viiveellä on tilastollisesti merkittävä autokorrelaatio, se on mukana mallissa. Tällöin näillä viiveillä on vaikutusta mallin seuraaviin arvoihin, joten niitä voidaan käyttää apuna ennustamaan tulevia arvoja. Myös osittaisautokorrelaatiofunktio (engl. partial autocorrelation function, PACF) mittaa korrelaatiota tietyllä viiveellä. PACF jättää kuitenkin aikaisemmat viiveet huomioimatta. ARIMA(p, d, q)-malliin tulevien AR-termien määrä, eli kertaluku p , päätellään osittaisautokorrelaatiofunktion kuvaajassa näkyvien piikkien perusteella.

Mallin eri termien parametrit ϕ_i ja θ_i estimoidaan aikasarjan arvojen avulla, sovittamalla malli aikasarjaan. Estimointi on vaativa matemaattinen prosessi, eikä tämä työ keskity sen avaamiseen. Eri tapoihin estimoida parametreja voi tutustua kirjan *Time series: Theory and Methods* (Brockwell ja Davis, 2009) Luvuista 5.1 ja 5.2. Tässä työssä estimointi on tehty Rstudio-ohjelman Arima-funktiolla forecast-paketista (Hyndman et al., 2023). Funktio käyttää suurimman todennäköisyyden arviota (engl. maximum likelihood estimation) parametrien estimointiin.

Parametrien estimoinnin jälkeen mallin hyvyttä voidaan arvioida ennen mallin ennusteita ja ennusteiden jälkeen. Ennen varsinaisia ennusteita mallin hyvyttä voidaan arvioida esimerkiksi Ljung-Box -testillä. Testi mittaa, ovatko virhetermien ϵ_t autokorrelaatiot riittävällä tarkkuudella nolaa, eli onko niiden välillä vielä autokorrelaatiota. Nollahypoteesi on että virhetermit mallissa ovat itsenäisesti jakautuneita, eivätkä riipu ajasta. Jos testin p -arvo on suurempi kuin 0.05, nollahypoteesi voidaan jättää voimaan. Tässä työssä mallien arvioinnissa tarkasteltiin visuaalisesti virhetermien autokorrelaatiofunktioita eri viiveillä ja laskettiin Ljung-Box -testin p -arvot kaikille malleille.

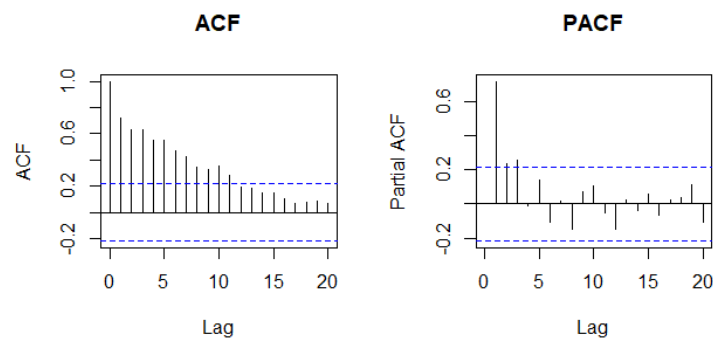
Mallien sopivuutta aikasarjaan ja mallin antamien ennusteiden hyvyttä voidaan mitata usealla erilaisella virheellä. Tässä työssä käytettiin kolmea yleistä tapaa arvioida virhettä. Lasketut virheet olivat keskiarvoinen absoluuttinen virhe (engl. mean absolute error, MAE) sekä keskiarvoinen absoluuttinen prosenttinvirhe (engl. mean absolute percentage error, MAPE). Lyhyen aikavalin ennusteille laskettiin myös keskiarvoinen neliövirhe (engl. mean squared error, MSE). Neliövirhe painottaa enemmän suuria virheitä. Keskimääräinen absoluuttinen virhe taas kertoo virheiden absoluuttisen suuruuden keskiarvon. Koska veturisarjoilla oli eri määrät veturina-

ajoja, mallien tarkkuutta voitiin verrata MAPE:n avulla. MAPE kertoo suhteellisen virheen toteutuneiden ja ennustettujen arvojen välillä. Suhteellisuus auttaa hahmottamaan mallien hyvyttä yleisesti, sillä veturina-ajojen määrät ovat suuruusluokaltaan jokseenkin vaikeasti hahmotettavissa.

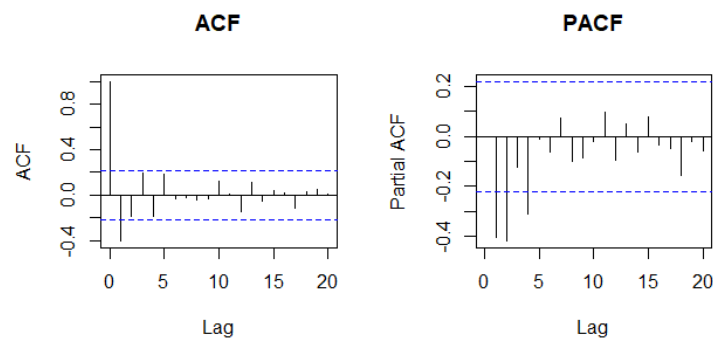
4 Tulokset

4.1 Aikasarjojen ennustemalleille päätelty kertaluvut

Ennustemallit rakennettiin jokaiselle aikasarjalle erikseen. Ensin tarkasteltiin mahdollisia trendejä Kuvasta 7. Sr1-aikasarjassa näkyy alaspäin laskeva lineaarinen trendi, joten aikasarja differoitiin. Myös Sr2 ja Sr3 -aikasarjassa nähdään loivaa nousevaa trendiä, joten myös tämä aikasarja differoitiin. Tämän jälkeen aikasarjoille piirrettiin ACF- ja PACF-kuvaajat viiveeseen 20 asti.



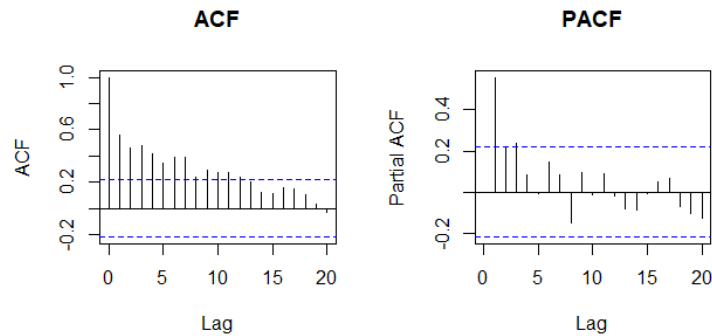
Kuva 8: Sr1-aikasarjan ACF- ja PACF-kuvaajat viiveeseen 20 asti



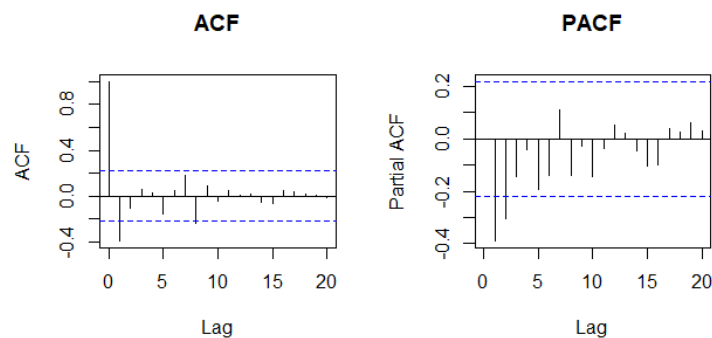
Kuva 9: Sr1-aikasarjan ACF- ja PACF-kuvaajat differoinnin jälkeen viiveeseen 20 asti

Kuvassa 8 nähdään Sr1-aikasarjan ACF- ja PACF-kuvaajat ennen differointia, ja Kuvassa 9 nähdään ACF- ja PACF-kuvaajat differoinnin jälkeen. Jälkimmäisessä kuvassa ACF-kuvaajassa nähdään tilastollisesti merkitsevä piikki kohdassa 1, joten ennustemallin MA-järjestysluku q oli 1. PACF-kuvaajassa nähdään tilastollisesti merkitsevät piikit kolmessa kohdassa, kertalukujen 1, 2 ja 4 kohdalla. AR-osan kertaluvuksi p tuli siis 4, mutta kertaluvun 3 parametri asetettiin nolllaksi, sillä se ei ole tilastollisesti merkitsevä. Jos piikki menee selvästi yli kuvissa näkyvän sinisen katkoviivan, sitä pidetään tilastollisesti merkitseväenä. Rajatapauksia ei otettu

huomioon, sillä tieteen käytäntöjen mukaisesti pyrittiin muodostamaan mahdollisimman yksinkertainen malli. Lopullinen malli Sr1 veturina-ajojen ennustamiselle oli ARIMA(4,1,1), sillä differointeja d tehtiin myös yksi.



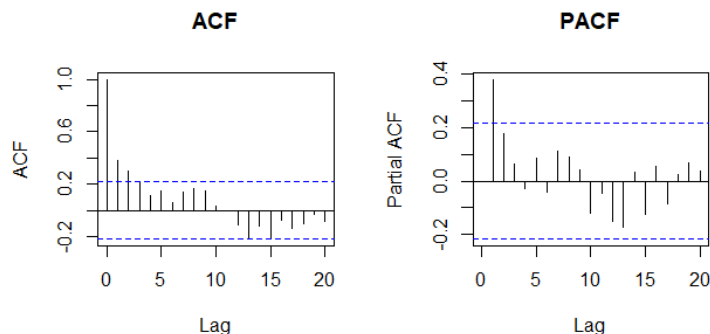
Kuva 10: Yhdistetyn Sr2 ja Sr3 -aikasarjan ACF- ja PACF-kuvaajat viiveeseen 20 asti



Kuva 11: Differoidun yhdistetyn Sr2 ja Sr3 -aikasarjan ACF- ja PACF-kuvaajat viiveeseen 20 asti

Kuvasta 10 nähdään yhdistetyn Sr2 ja Sr3 -aikasarjan ACF- ja PACF-funktioiden kuvaajat ennen differointia. Nähdään, että ne muistuttavat Sr1-aikasarjan vastaavia kuvaajia. Kuvassa 11 esitetään differoinnin jälkeiset kuvaajat, jotka eivät muistuta enää Sr1-aikasarjan kuvaajia. Differoinnin jälkeisessä ACF-kuvaajassa vasemmalla nähdään tilastollisesti merkitsevä piikki kohdassa 1. MA-osan q kertaluku oli siis 1. AR-osan kertaluvuksi p pääteltiin PACF-kuvaajasta 2, sillä tilastollisesti merkitsevät piikit nähdään viiveillä 1 ja 2. ACF-kuvaajan piikkiä viiveellä 8 ei huomioitu, koska se oli lähellä tilastollisen merkitsevyyden rajaa, ja myös hyvin korkea viive, joka monimutkaistaisi mallia. Lopullinen malli oli ARIMA(2,1,1).

Dv12-aikasarjan ennustusmallin kertaluvut pääteltiin Kuvan 12 kuvaajista. Vasemmalla ACF-kuvaajassa piikit kohdissa 1 ja 2, joten MA-osan kertaluvun q pääteltiin olevan 2. Kohdan 3 piikki on tilastollisen merkitsevyyden rajalla, mutta se on jätetty pois samoin perustein kuin aiempien aikasarjojen rajatapaukset. AR-osan ker-



Kuva 12: Dv12-sarjan ACF- ja PACF-kuvaajat viiveeseen 20 asti

taluvuksi p päätettiin 1 oikeanpuoleisen PACF-kuvaajan perusteella, koska kuvaajassa näkyy yksi tilastollisesti merkitsevä piikki kohdassa 1. Malliksi tuli ARMA(1,2).

Taulukko 3: Mallien kertaluvut ja Ljung-Box -testin p-arvo

Sarja	Ennustemalli	P-arvo
Sr1	ARIMA(4,1,1)	0.7507
Sr2 ja Sr3	ARIMA(2,1,1)	0.9182
Dv12	ARMA(1,2)	0.9986

Taulukossa 3 esitetään ACF- ja PACF-kuvaajista päätellyt kertaluvut ja tehdyt differoinnit. Sr1- ja Dv12-ennustemallien kertaluvut eroavat toisistaan eniten. Kaikkien mallien kertaluvut eroavat toisistaan, paitsi sähkövetureilla MA-mallin parametri q on sama. Lisäksi vain sähköveturimallien aikasarjoja differoitiin. Taulukossa 3 näkyy myös malleille laskettujen Ljung-Box -testien p-arvot. Jokaisen mallin p-arvo on korkea ja yli 0.05, joten niitä voidaan pitää Ljung-Box -testin perusteella hyvinä. Myös mallien virhetermien ϵ_t ACF-kuvaajista näkyi, että kaikki arvot olivat riittävällä tarkkuudella nollaa. Näiden tulosten perusteella voidaan todeta, etteivät minkään mallin virhetermit riipu toisistaan.

4.2 Mallien estimoidut parametrit ja virheet malleille

Kertalukujen päättelyn jälkeen mallit sovitettiin aikasarjoihin. Sovituksessa käytettiin aiemmin mainittua Arima-funktiota Rstudion forecast-paketista. Taulukossa 4 nähdään Sr1-aikasarjan ennustemallin estimoidut parametrit. Lopullinen malli oli siis

$$Z_t = \epsilon_t - 0.7136Z_{t-1} - 0.6965Z_{t-2} - 0.3089Z_{t-4} - 0.033\epsilon_{t-1},$$

jossa $Z_t = X_t - X_{t-1}$ on differoitu termi. Nähdään että suurin painoarvo on edellisellä termillä, pienin edellisen ennusteen virhetermillä.

Taulukko 4: Sr1-ennustemallin estimoidut parametrit

Parametri	Estimaatti
ϕ_1	-0.7136
ϕ_2	-0.6965
ϕ_4	-0.3089
θ_1	-0.0330

Sr2 ja Sr3 -aikasarjan ennustemallin estimoidut parametrit nähdään Taulukossa 5. Lopulliseksi malliksi saatiin siis

$$Z_t = \epsilon_t + 0.1704Z_{t-1} + 0.0188Z_{t-2} - 0.7696\epsilon_{t-1},$$

jossa $Z_t = X_t - X_{t-1}$ on differoitu termi. Nähdään että suurin painoarvo mallissa on edellisen muuttujan virhetermillä. Vähiten mallissa merkitsee termi Z_{t-2} , jonka parametrin estimaatti on hyvin pieni.

Taulukko 5: Sr2 ja Sr3 -ennustemallin estimoidut parametrit

Parametri	Estimaatti
ϕ_1	0.1704
ϕ_2	0.0188
θ_1	-0.7696

Dv12-ennustemallin estimoidut parametrit nähdään Taulukossa 6. Lopullinen malli oli

$$X_t = \epsilon_t + 0.9067X_{t-1} - 0.6588\epsilon_{t-1} - 0.1085\epsilon_{t-2}.$$

Huomataan että edellisellä termillä X_{t-1} on todella suuri painoarvo. Suuri painoarvo viittaa siihen, ettei aikasarja ole välttämättä stationaarinen. Kun parametrin estimaatti on lähellä numeroa yksi, on syytä harkita differointia. Korkea AR(1)-arvo viittaa siihen että seuraavat termit muistuttavat paljon edellistä termiä. Tutkimuksessa toteutettiin myös differoitu malli, mutta tulokset eivät olleet yhtä tarkkoja kuin differoimattomalla mallilla. Lisäksi visuaalisen tarkastelun perusteella aikasarjassa ei nähdä merkittävää trendiä Kuvassa 2. Ljung-Box -testin perusteella differoimattomassa mallissa ei ollut huonoutta, joka viittaisi differoinnin tarpeeseen.

Taulukko 6: Dv12-ennustemallin parametrit

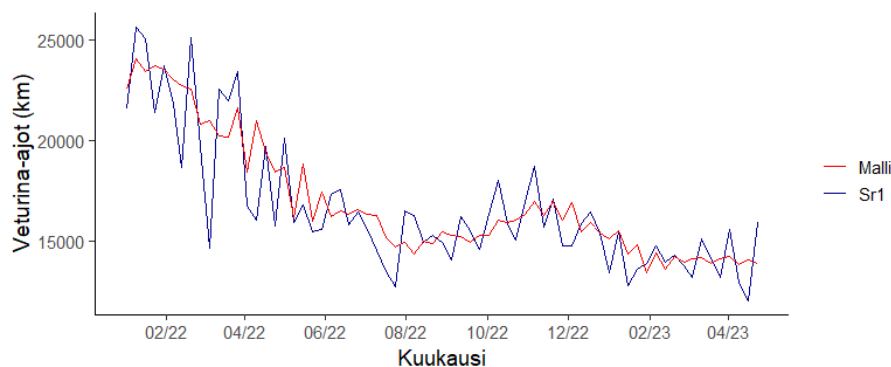
Parametri	Estimaatti
ϕ_1	0.9067
θ_1	-0.6588
θ_2	-0.1085

Parametrien estimoinnin yhteydessä saadaan tietoa siitä, kuinka hyvin estimoidut parametrit sovittavat mallin koko aikasarjaan. Taulukkoon 7 on koottu jokaisen ennustemallin sovitusvirheet. MAPE-arvoista nähdään, että Sr2 ja Sr3 -aikasarjaan

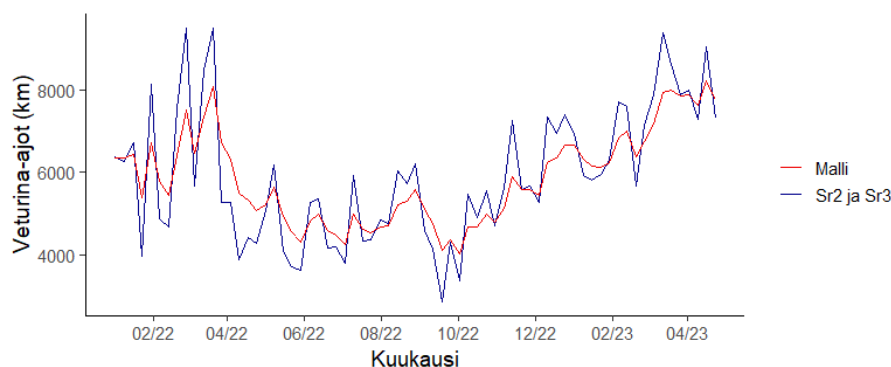
sovitettu ennustemalli oli suhteellisesti huonoin, vaikka absoluuttinen virhe on pienin. Dv12-aikasarjan mallin suhteellinen sopivuus on paras, mutta absoluuttinen virhe on suurin. Sr1-sarjan malli putoaa näiden kahden väliin absoluuttisen virheen puolesta. MAPE Sr1-mallilla on lähes yhtä matala kuin Dv12:lla.

Taulukko 7: Ennustemallien sovituksien virheparametrit

Sarja	MAPE (%)	MAE
Sr1	9.05	1483.101
Sr2 ja Sr3	18.36	1026.005
Dv12	8.40	1993.325

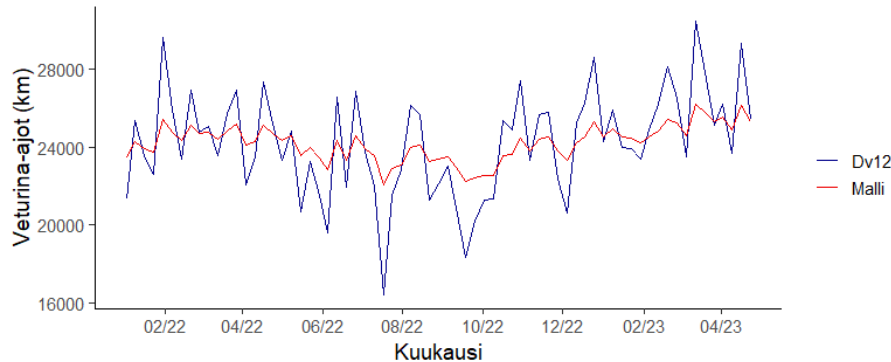


Kuva 13: Sr1-aikasarjan ennustemallin sovitus viikoittaiseen dataan



Kuva 14: Sr2 ja Sr3 -aikasarjan ennustemallin sovitus viikoittaiseen dataan

Kuvissa 13, 14 ja 15 nähdään jokainen aikasarja ja siihen sovitettu ennustemalli. Huomataan, että kaikissa sarjoissa ennustemalli mukailee aikasarjaa, mutta suurimpia vaihteluita se ei pysty toistamaan. Sr2 ja Sr3 -aikasarjassa vaikuttaa silmämääräisesti olevan suurimpia vaihteluita, jotka saattavat selittää suurimman suhteellisen virheen. Sr1- ja Dv12-aikasarjojen kuvaajissa on myös joitain suurempia piikkejä, mutta myös enemmän keskimääräisiä arvoja.



Kuva 15: Dv12-aikasarjan ennustemallin sovitus viikoittaiseen dataan

4.3 Mallien antamat ennusteet ja virheet

Mallien avulla tehtiin ennusteet viidelle ja kymmenelle viikolle. Viiden viikon ennusteista laskettiin virheet käyttäen MAPE:a, MAE:ta ja MSE:tä. Kymmenen viikon ennusteista laskettiin vain MAPE, jotta malleja voitiin verrata myös hieman pidemmän aikavälin ennusteissa. Saatuja ennusteita vertailtiin MAPE-arvoon, joka saatiin laskemalla vastaavat ennusteet aikasarjan keskiarvosta. Tarkat ennustearvot ovat nähtävissä Liitteestä A.

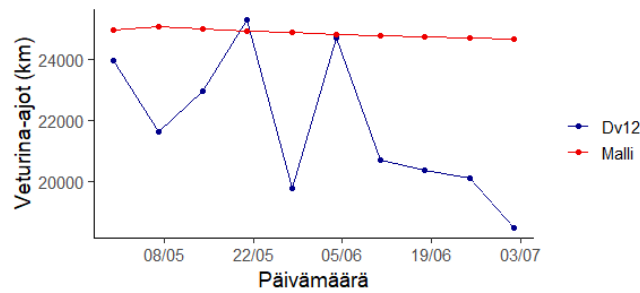
Taulukko 8: Mallien viiden viikon ennustevirheet ja ennustevirhe keskiarvolla lasketuista ennusteista

Sarja	MAPE (%)	MAE	MSE	MAPE, keskiarvo (%)
Sr1	17.34	1970.389	4730969	35.03
Sr2 ja Sr3	11.42	799.0369	832518	17.20
Dv12	10.41	2196.52	8320477	8.67

Ennustemallien ennustevirheet viiden viikon ennusteille esitetään Taulukossa 8. Taulukossa nähdään myös MAPE, jos vastaavat ennusteet olisi laskettu keskiarvosta. Huomataan että sähköveturisarjojen ennustemalleilla tehtyjen ennusteiden suhteelliset virheet ovat useita prosenttiyksiköitä pienemmät kuin keskiarvoennusteiden.

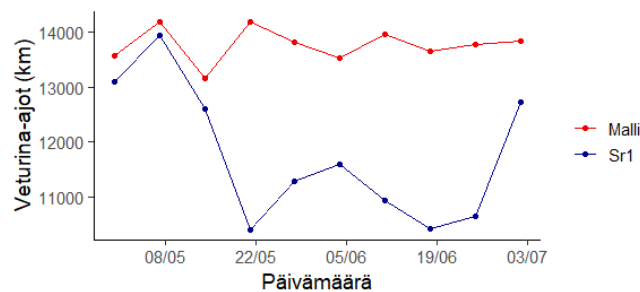
Yllä olevasta taulukosta nähdään, että Dv12-aikasarjan ARMA-mallin ennusteiden MAPE-arvot ovat noin kaksi prosenttiyksikköä huonommat kuin keskiarvosta laskettujen ennusteiden. Kuvasta 16 nähdään, että ennusteiden arvo on hieman yli 24000 kilometriä. Nähdään, että suurin osa toteutuneista arvoista on alle 24000 kilometrin. Aikasarjan keskiarvo oli noin 23800 kilometriä viikossa. Keskiarvoennusteet ovat siis matalammat kuin ennustemallin antamat ennusteet, ja Kuvan 16 perusteella lähempänä toteutuneita arvoja.

Sr1-sarjan aikasarjan keskiarvo oli noin 16000 kilometriä viikossa. Kuvasta 17 nähdään, että keskiarvo on 2000 kilometriä suurempi kuin suurin validointisarjan arvo. Vaikka ARIMA-mallin prosentuaalinen virhe oli viiden viikon ennusteissa noin 17 prosenttia, keskiarvolla ennustettaessa virhe oli huomattavasti suurempi, noin 35 prosenttia. Sr1-aikasarjassa oli nähtävissä laskeva trendi viikoittaisella tasolla. Onkin



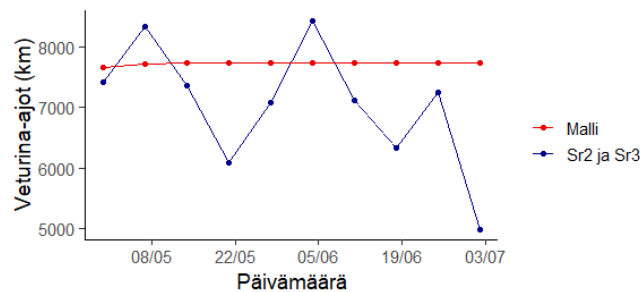
Kuva 16: Dv12-aikasarjan mallin ennusteet 10:lle seuraavalle viikolle

loogista, että keskiarvoennusteet ovat huomattavasti suurempia kuin toteutuneet arvot ja ennustus, joissa trendi on otettu huomioon.



Kuva 17: Sr1-aikasarjan mallin ennusteet 10:lle seuraavalle viikolle

Myös yhdistetyn Sr2 ja Sr3 -aikasarjan viiden viikon ARIMA-mallin ennusteet olivat tarkemmat kuin keskiarvolla lasketut ennusteet. Taulukosta 8 nähdään että ARIMA-mallin ennusteiden MAPE-arvo on noin kuusi prosenttiyksikköä parempi kuin keskiarvoennusteiden. Ero ei ole yhtä merkittävä kuin Sr1-sarjalla. Sr2 ja Sr3 -aikasarjan keskiarvo oli noin 6090 kilometriä viikossa. Keskiarvo on siis matalampi kuin ennustetut arvot, kuten Kuvasta 18 nähdään. Tämä selittyy aikasarjan nousevalla trendillä.



Kuva 18: Sr2 ja Sr3 -aikasarjan mallin ennusteet 10:lle seuraavalle viikolle

Kymmenen viikon ennustevirheiden ja keskiarvoennusteiden MAPE-arvot esitetään Taulukossa 9. Huomataan että ennustemalleilla tehtyjen ennusteiden virheet

Taulukko 9: Mallien MAPE-arvot kymmenen viikon ennusteista ja keskiarvolla tehdyistä ennusteista

Sarja	MAPE (%)	MAPE, keskiarvo (%)
Sr1	19.53	35.20
Sr2 ja Sr3	15.83	15.65
Dv12	17.65	14.50

kasvavat, eniten Dv12-aikasarjalla. Keskiarvolla ennustaessa Dv12-aikasarjan ennusteiden virhe kasvoi eniten verrattuna viiden viikon keskiarvoennusteisiin. Lisäksi keskiarvoennusteen virhe oli pienempi kuin ARMA-mallilla tehtyjen ennusteiden. Sr1-aikasarjan keskiarvoennusteiden prosentuaalinen virhe ei juuri kasvanut verrattuna viiden viikon keskiarvon ennustevirheeseen. Sr2 ja Sr3 -aikasarjan keskiarvolla tehdyt ennusteet olivat hieman parempia kuin viidelle viikolle tehdyt ennusteet. Taulukoista 8 ja 9 huomataan siis, että Sr1-aikasarja, jolla oli vahvin trendi, sai vielä kymmenen viikon kohdalla tarkempia ennustearvoja ARIMA-mallilla. Sr2 ja Sr3- sekä Dv12-aikasarjojen keskiarvoennusteet olivat hiukan tarkempia kuin ennustemallilla lasketut ennusteet.

Tutkimuksessa poikkeuksellisen huomion kiinnitti Sr2 ja Sr3 -aikasarjaan sovitettun mallin ja mallin ennusteiden virhearvojen erot. Taulukoista 7 ja 8 nähdään että ennusteiden MAPE-arvo on noin seitsemän prosenttiyksikköä parempi kuin sovitteessa. Myös MAE-arvo on hieman pienempi viiden viikon ennusteilla kuin sovituksella. Tämä on erikoista, sillä yleensä ennusteiden virheet ovat suuremmat kuin sovitteen. Tämä saattaa johtua aikasarjan suurista vaihteluista, joihin malli ei ole pystynyt sopeutumaan. Myös viisi ennustettua arvoa ovat saattaneet olla sattumalta tasaisemmat kuin aikasarja aiemmin.

4.4 Johtopäätökset

Rakennettujen ennustemallien ja niillä tehtyjen ennusteiden perusteella voidaan sanoa, että eri veturisarjoille kertyviä veturina-ajoja kannattaa ennustaa eri tavoilla. Visuaalisesti tarkasteltuna tutkittujen veturisarjojen aikasarjat erosivat toisistaan. Suurin ero oli eri aikasarjojen kehitys. Sr1-aikasarjan trendi oli laskeva, Sr2 ja Sr3 -aikasarjalla taas nouseva. Dv12-aikasarjassa ei havaittu trendiä.

Visuaalisen tarkastelun lisäksi aikasarjojen pohjalta rakennettujen ennustemallien kertaluvut erosivat toisistaan ja myös parametrien estimaatit olivat erilaiset. Kertalukujen erilaisuus kertoo aikasarjojen alla olevien prosessien ja arvojen riippuvuusrakenteiden erilaisuudesta. Eri veturisarjojen veturina-ajojen määrät olivat tarkastellulla aikajaksolla erilaiset niin kehityksiltään kuin rakenteiltaan. Tämä antaa syyn ennustaa tulevia veturina-ajoja eri tavoilla veturisarjasta riippuen.

Etenkin Dv12- ja Sr2 ja Sr3 -aikasarjojen ennustekuvista 16 ja 18 nähdään että ennustetut arvot ovat hyvin tasaisia. ARIMA-mallien ennusteet painuvatkin kohti tiettyä lukua, joten ne eivät sovellu pidemmän aikavälin tarkkoihin ennusteisiin erityisen hyvin. Tämän vuoksi rakennetuilla malleilla on yhtä hyvä ennustaa strategisen aikavälin määriä kuin laskemalla keskiarvo.

Kun vertailtiin keskiarvosta otettujen ennusteiden ennustevirheitä ARIMA-mallien ennusteisiin, nähtiin että sähköveturisarjoilla ARIMA-ennusteet olivat parempia viidelle seuraavalle viikolle. Tämä johtui aikasarjojen trendeistä. Erityisesti Sr1-ennustemallin ennusteista nähtiin että lyhyellä aikavälillä keskiarvo saattaa olla hyvinkin paljon pienempi tai, kuten tässä tapauksessa, suurempi kuin toteutuneet arvot. ARIMA-mallin ennustus osasikin jatkaa toteutunutta aikasarjaa paremmin kuin keskiarvo myös kymmenen viikon ennusteissa. Sr2 ja Sr3 -aikasarjan tapauksessa trendi ei Kuvassa 7 näyttänyt niin suurelta kuin Sr1-sarjalla. Tämän vuoksi keskiarvon ero ARIMA-mallin ennusteisiin oli pienempi viiden viikon ennusteissa, mutta silti huomattava. Kymmenen viikon ennusteissa keskiarvolla ennustaminen oli yhtä hyvä kuin ennustemalli.

Dv12-sarjalle ennustusten laskeminen viidelle viikolle oli hiukan tarkempaa keskiarvolla kuin ARMA-mallilla. Kymmenen viikon ennusteissa keskiarvo oli kolme prosenttiyksikköä tarkempi. Siispä Dv12-sarjalle keskiarvolla ennustaminen vaikuttaa olevan myös hyvä vaihtoehto strategisen ja taktisen tason ennusteille ja yksinkertaisena menetelmänä suositeltavampi kuin monimutkaisempi ARMA-malli.

Aikasarjoja analysoidessa työssä jätettiin huomioimatta useita veturina-ajoihin vaikuttavia tekijöitä. Tästä syystä saatuja tuloksia ei voida pitää aukottomina. Alaluvussa 2.1 tuli ilmi, että veturina-ajojen määrään vaikuttaa suuresti veturien kappalemäärä. Tätä ei huomioitu työssä, ja voikin olla, että muutokset sähköveturien kappalemäärissä ovat aiheuttaneet aikasarjoissa havaitut trendit. Toinen mahdollisesti vaikuttava tekijä on junakilometrien muutos. Työn alussa avattiin syitä, miksi tätä ei otettu huomioon ennusteissa. Kolmas mahdollinen vaikutus veturina-ajoihin on poikkeamat rautatieliikenteessä, eli ratatyöt ja tehdasseisokit. Työssä oletettiin niiden tasaantuvan, mutta on myös mahdollista, että todellisuudessa nämä poikkeamat ovat kausittaisia, eli niitä tapahtuu enemmän esimerkiksi kesällä. Tämä saattaa aiheuttaa pidempää kausittaista vaihtelua veturina-ajoihin, joita ei huomioitu tässä työssä. Viimeinen vaikuttava asia on aikasarjan pituus. Luotettavampia tuloksia olisi saanut pidemmällä aikasarjalla, josta olisi voinut saada tietoa vuosittaisista vaihteluista esimerkiksi kesän ja talven välillä.

5 Yhteenveto ja tulevia tutkimusaiheita

Tämän työn tavoitteena oli tutkia veturina-ajojen määrien ennustamista. Työssä pyrittiin päättämään toteutuneiden veturina-ajojen pohjalta, onko eri veturisarjoille kertyviä veturina-ajoja viisasta ennustaa eri tavoilla. Työssä tutkittiin alan tutkimuskirjallisuutta ja selvitettiin, onko jo olemassa joku sovellettava ennustusmalli, jota voisi hyödyntää veturina-ajojen ennustamisessa.

Työn alussa tutkittiin alan tutkimuskirjallisuutta aiheeseen liittyen. Tutkittiin erikseen veturikaluston kokoon liittyviä lähteitä ja muiden ajoneuvojen, kuten takkien, tyhjänäajojen arviointia. Muihin ajoneuvolajeihin käytetyt menetelmät olivat pääosin edistyneitä koneoppimismenetelmiä, joita voidaan mahdollisesti soveltaa myös raideliikenteeseen tulevaisuudessa. Ennustemallia veturina-ajoille ei tämän työn kirjallisuuslähteistä löytynyt. Tutkimuskirjallisuuden pohjalta voitiin päätellä, että veturina-ajojen määrään vaikuttaa ainakin veturien kappalemäärä. Monet lähteet pyrkivät vähentämään veturina-ajoja optimointimallein, jotka minimoivat kustannukset. Veturina-ajot ovat vain pieni osa kaikesta raideliikenteestä, joten on ymmärrettävää ettei tutkimus keskity siihen niin paljon.

Työssä tutkittiin visuaalisesti ja aikasarja-analyysin keinoin veturisarjoja, joille kertyy eniten veturina-ajoja. Malleille rakennettiin ARIMA-mallit, joiden parametrejä vertailtiin. Myös mallien sovitteiden ja ennusteiden hyvyttä arvioitiin. Saaduista tuloksista tehtiin johtopäätökset. Tuloksien luotettavuutta arvioitiin.

Tärkein tulos työssä oli se, että eri veturisarjoille kertyviä veturina-ajoja on hyödyllistä ennustaa eri tavoin. Tämä johtuu siitä, että veturina-ajojen kehitys on eri sarjoilla erilaista. Myös veturina-ajojen aikasarjojen alla olevat olevien prosessit ja arvojen korrelaatiot ovat erilaisia, mikä pääteltiin rakennettujen ennustusmallien parametrien erilaisuudesta.

Viiden viikon aikavälille rakennetut ennustemallit antoivat sähköveturisarjoille parempia tuloksia kuin laskemalla ennusteet aikasarjojen keskiarvoista. Tutkitulle dieselveeturisarjalle Dv12:lle keskiarvolla ennustaminen antoi hieman parempia tuloksia. Sähkövetureiden aikasarjoissa oli havaittavissa trendiä, mutta dieselveeturilla ei. Erot tarkkuuksissa johtuvat todennäköisimmin siitä, sillä ennustemallit huomioivat trendin, mutta keskiarvo ei. Kymmenen viikon ennusteissa vain Sr1-aikasarjan ennustemallin ennusteet olivat tarkempia kuin keskiarvoennusteet.

Ennustemallien antamat arvot tasaantuvat kohti tiettyä arvoa pidemmällä aikavälillä. Tästä syystä strategisen tason ennusteet eri veturisarjoille voi yhtä hyvin tehdä keskiarvon avulla, jos siinä huomioidaan aikasarjojen trendit. ARIMA-mallien sopivuus riippuu täysin menneestä datasta. Tämän takia niillä on vaikea ennustaa tarkasti pidemmälle tulevaisuuteen, sillä tulevia muutoksia ja vaihteluita ei tiedetä. Tulevaisuuden vaihteluita ei tosin kykene ennustamaan mikään matemaattinen malli. Oli siis malli mikä tahansa, ennustaminen pitkälle tulevaisuuteen tarkasti on hyvin vaikeaa.

Työssä jätettiin huomioimatta useita eri tekijöitä. Sopivia laajennuksia työlle olisikin huomioida nämä tekijät, joita osaa sivuttiin Alaluvussa 4.4. Merkittävimpinä tekijöinä on veturien kappalemäärien muutokset ja ajettavien junien suhde veturina-ajoihin. Veturien kappalemäärien muutokset ovat merkittävässä asemassa,

kun otetaan käyttöön uusia veturisarjoja. Kun joku toinen veturisarja vanhenee ja poistuu käytöstä, sillä ajettu liikenne siirtyy uudelle veturisarjalle asteittain, jolloin uutta veturisarjaa on kappalemäärällisesti enemmän ja vanhaa vähemmän. Tässä työssä sähköveturisarjojen trendit saattaisivat selittyä sillä, että Sr3-vetureita olisi esimerkiksi tullut lisää, ja Sr1-vetureita poistunut käytöstä. Tätä ei kuitenkaan voida osoittaa, ja aihetta olisi syytä tutkia tarkemmin.

Tulevaisuudessa erityisesti Dv12-veturisarjan kohdalla on huomioitava myös Suomen rataverkon sähköistyminen. Mitä enemmän on sähköistettyä rataa, sitä enemmän dieselvetureilla ajettava liikenne vaihtuu sähkövetureilla ajettavaan. Jos esimerkiksi rataosuus, jossa on aiemmin ajettu paljon dieselvetureilla, sähköistetään, tulee tämä muutos huomioida ennustuksessa. Todennäköisesti tällöin dieselvetureilla liikennöinti kyseisellä rataosuudella siirtyy sähkövetureille.

Jo työn alussa mainittiin, että ajettujen junien määrällä on todennäköisesti vaikutusta veturina-ajojen määriin. Ajettujen junien määrän suhdetta veturina-ajojen määrään olisikin hyvä tutkia tulevaisuudessa lisää. Määrittämällä suhde ajetuille junille ja veturina-ajoille, veturina-ajojen määrä voitaisiin laskea suoraan sen perusteella, paljonko junia arvioidaan ajettavan.

Tavoite veturina-ajojen ennustamisessa oli saada parempi kokonais käsitys vetureille kertyvistä kilometreistä, joka taas edelleen vaikuttaa tarvittavaan veturinkuljettajien määrään. Työn aihetta voisi tulevaisuudessa tarkentaa paikkakuntaakohtaiselle tasolle junista kertyvät kilometrit huomioiden, sillä veturinkuljettajat palkataan paikkakunnittain. Veturinkuljettajien määrää olisi hyvä tarkastella paikkakunnittain liikennetarpeen mukaan. Veturina-ajojen alueelliseen määrään vaikuttavia asioita ovat myös ratatyöt ja tehdasseisokit. Näiden kausittaisuutta arvioitiin jo Alaluvussa [4.4](#). Poikkeamien alueellisuutta voisi tarkastella enemmän. Tarkastelusta voisi olla hyötyä, jotta voidaan tarkastella tarkemmin paikkakuntaakohtaisia eroja vetureiden määrissä ja junakilometreissä.

Viitteet

- Ravindra K. Ahuja, Jian Liu, James B. Orlin, Dushyant Sharma, ja Larry A. Shughart. Solving real-life locomotive-scheduling problems. *Transportation Science*, 39(4): 503–517, 2005.
- Mohamed Amjath, Laoucine Kerbache, James MacGregor Smith, ja Adel Elomri. Fleet sizing of trucks for an inter-facility material handling system using closed queueing networks. *Operations Research Perspectives*, 9:100245, 2022.
- Arjang Assad. Analytical models in rail transportation: an annotated bibliography. 1977.
- Arjang Assad. Modelling of rail networks: Toward a routing/makeup model. *Transportation Research Part B: Methodological*, 14(1-2):101–114, 1980.
- Peter J. Brockwell ja Richard A. Davis. *Time series: Theory and Methods*. Springer science & business media, 2009.
- Divine Carson-Bell, Mawutor Adadevoh-Beckley, ja Kendra Kaitoo. Demand prediction of ride-hailing pick-up location using ensemble learning methods. *Journal of Transportation Technologies*, 11(2):250–264, 2021.
- Teodor Gabriel Crainic ja Gilbert Laporte. Planning models for freight transportation. *European journal of operational research*, 97(3):409–438, 1997.
- Sarah Frisch, Philipp Hungerländer, Anna Jellen, ja Dominic Weinberger. A mixed integer linear program for optimizing the utilization of locomotives with maintenance constraints. Teoksessa *Operations Research Proceedings 2018: Selected Papers of the Annual International Conference of the German Operations Research Society (GOR), Brussels, Belgium, September 12-14, 2018*, pages 103–109. Springer, 2019.
- Tennyson Godwin, Ram Gopalan, ja T.T. Narendran. Tactical locomotive fleet sizing for freight train operations. *Transportation Research Part E: Logistics and Transportation Review*, 44(3):440–454, 2008.
- Randolph W. Hall. Stochastic freight flow patterns: implications for fleet optimization. *Transportation Research Part A: Policy and Practice*, 33(6):449–465, 1999.
- Rob Hyndman, George Athanasopoulos, Christoph Bergmeir, Gabriel Caceres, Leanne Chhay, Mitchell O’Hara-Wild, Fotios Petropoulos, Slava Razbash, Earo Wang, ja Farah Yasmien. *forecast: Forecasting functions for time series and linear models*, 2023. R package version 8.21.1. Viitattu 9.11.2023. Saatavissa: <https://pkg.robjhyndman.com/forecast/>.
- Julia, 2018. Junaliikenteen havaintojärjestelmä Julia, Kalusto. Päivitetty 9/2018. Viitattu 11.10.2023. Saatavissa: <https://juliadata.fi/rollingstock>.

- Eleftheria Kontou, Venu Garikapati, ja Yi Hou. Reducing ridesourcing empty vehicle travel with future travel demand prediction. *Transportation Research Part C: Emerging Technologies*, 121:102826, 2020.
- Ching-Chung Kuo ja Gillian M. Nicholls. A mathematical modeling approach to improving locomotive utilization at a freight railroad. *Omega*, 35(5):472–485, 2007.
- Hongtao Li, Juncheng Bai, Xiang Cui, Yongwu Li, ja Shaolong Sun. A new secondary decomposition-ensemble approach with cuckoo search optimization for air cargo forecasting. *Applied Soft Computing*, 90:106161, 2020.
- Gopindra S. Nair, Chandra R. Bhat, Irfan Batur, Ram M. Pendyala, ja William H.K. Lam. A model of deadheading trips and pick-up locations for ride-hailing service vehicles. *Transportation Research Part A: Policy and Practice*, 135:289–308, 2020.
- Evren Özkaya, Pinar Keskinocak, V. Roshan Joseph, ja Ryan Weight. Estimating and benchmarking less-than-truckload market rates. *Transportation Research Part E: Logistics and Transportation Review*, 46(5):667–682, 2010.
- Sini Poikonen. Haastattelu VR Transpointin ja VR Groupin junaliikennöinnin henkilöstön kanssa: Nieminen, Juha (Strategisen suunnittelun päällikkö); Salo, Erkki (Strateginen suunnittelija). Syyskuu, 2023.
- Warren B. Powell, Belgacem Bouzaiene-Ayari, Clark Cheng, Ricardo Fiorillo, Sourav Das, ja Coleman Lawrence. Strategic, tactical and real-time planning of locomotives at norfolk southern using approximate dynamic programming. Teoksessa *ASME/IEEE Joint Rail Conference*, volume 44656, pages 491–500. American Society of Mechanical Engineers, 2012.
- Satya Prakash, B.V. Balaji, ja Deepak Tuteja. Optimizing dead mileage in urban bus routes through a nondominated solution approach. *European Journal of Operational Research*, 114(3):465–473, 1999.
- David Ronen. Cargo ships routing and scheduling: Survey of models and problems. *European Journal of Operational Research*, 12(2):119–126, 1983.
- Dušan Teichmann, Michal Dorda, Karel Golc, ja Helena Bínová. Locomotive assignment problem with heterogeneous vehicle fleet and hiring external locomotives. *Mathematical Problems in Engineering*, 2015, 2015.
- Tilastokeskus, 2023. Rautateilla kuljetetun rahdin määrä laski 22% vuonna 2022. Päivitetty 2023. Viitattu 9.11.2023. Saatavissa: <https://www.stat.fi/julkaisu/cl8mr34d5z4ni0dutl8dxq8u6>.
- VR Group, 2023 a. Vastuullisuusraportti 2022. Päivitetty 7.3.2023. Viitattu 20.10.2023. Saatavissa: https://vrgroup.studio.crasman.cloud/file/dl/i/HWSWzQ/eYR8XOVqzb0Gn_fjZWwkeg/VR_Group_Vastuullisuusraportti_2022.pdf.

- VR Group, 2023 b. Veturityypit. Päivitetty 2023. Viitattu 31.10.2023. Saatavissa: <https://www.vrgroup.fi/fi/vrgroup/yriytyksemme/liiketoiminta/vr-tranpoint/veturityypit/>.
- VR Group, 2023 c. VR:n veturistrategia perustuu investointeihin, varaosien uudelleen käyttöön ja kierrätykseen. Päivitetty 13.11.2023. Viitattu 2.12.2023. Saatavissa: <https://www.vrgroup.fi/fi/vrgroup/uutiset/vr-n-veturistrategia-perustuu-investointeihin-varaosien-uudelleenkayttöön-ja-kierrätykseen-131120230941/>.
- Väylävirasto, 2023 a. Rautateiden tavaraliikenne, 1990-2022. Päivitetty 9.5.2023. Viitattu 20.10.2023. Saatavissa: https://vayla.fi/documents/25230764/55126781/Rautateiden+tavaraliikenne_1990-2022.pdf/c2796998-2d9d-acae-14b5-307836b3599d/Rautateiden+tavaraliikenne_1990-2022.pdf?t=1683633056357.
- Väylävirasto, 2023 b. Rautateiden henkilöliikenne, 1990-2022. Päivitetty 9.5.2023. Viitattu 20.10.2023. Saatavissa : https://vayla.fi/documents/25230764/55126781/Rautateiden+henkil%C3%B6liikenne_1990-2022.pdf/6fded58d-3c62-ad14-166e-ee128629be7b/Rautateiden+henkil%C3%B6liikenne_1990-2022.pdf?t=1683633000316.
- Joris Wagenaar, Leo Kroon, ja Ioannis Fragkos. Rolling stock rescheduling in passenger railway transportation using dead-heading trips and adjusted passenger demand. *Transportation Research Part B: Methodological*, 101:140–161, 2017.
- Yle, 2022. Täysin Suomen oloihin tehty dieselveturi saatiin testiin Tampereelle – tällainen on nappeja myöten kuljettajien kanssa suunniteltu investointi. Yle Uutiset, Clarissa Jääri. Päivitetty 9.2.2022. Viitattu 20.10.2023. Saatavissa: <https://yle.fi/a/3-12309655>.

A Saatujen ennusteiden arvot

Tässä liitteessä ovat tarkat arvot saaduille viidelle ja kymmenelle ennustusaskeleelle tutkimussarjoittain. Ennustukset tehtiin aikasarjojen kohdille 71-80 eli vuoden 2023 viikoille 19-31.

Taulukko A1: Sr1-aikasarjan ennusteiden arvot

Parametri	Estimaatti
X_{71}	13567.89
X_{72}	14190.69
X_{73}	13170.00
X_{74}	14182.84
X_{75}	13818.05
X_{76}	13523.40
X_{77}	13962.78
X_{78}	13664.17
X_{79}	13782.87
X_{80}	13849.57

Taulukko A2: Sr2 ja Sr3 -aikasarjan ennusteiden arvot

Parametri	Estimaatti
X_{71}	7665.002
X_{72}	7711.113
X_{73}	7723.866
X_{74}	7726.903
X_{75}	7727.660
X_{76}	7727.846
X_{77}	7727.892
X_{78}	7727.903
X_{79}	7727.906
X_{80}	7727.906

Taulukko A3: Dv12-aikasarjan ennusteiden arvot

Parametri	Estimaatti
X_{71}	24956.34
X_{72}	25045.51
X_{73}	24979.18
X_{74}	24919.04
X_{75}	24864.52
X_{76}	24815.09
X_{77}	24770.27
X_{78}	24729.63
X_{79}	24692.79
X_{80}	24659.38