

Advanced Energy Solutions -maisteriohjelma

Sähköyhtiön lyhyen aikavälin kulutusennusteen kehittäminen

Arttu Niemelä



Copyright ©2022 Arttu Niemelä

Author Arttu Niemelä

Title of thesis Development of short-term electricity load forecasting for electric utility

Programme Advanced Energy Solutions

Major Sustainable Energy Systems and Markets

Thesis supervisor Professor Matti Lehtonen

Thesis advisor(s) M.Sc Mikael Mäkelä

Collaborative partner Vaasan Sähkö Oy

Date 23.09.2022

Number of pages 72

Language Finnish

Abstract

In this thesis, an effort was made to develop short-term load forecasting for an electricity company. Due to the structures of the electricity market, it is possible to achieve significant cost savings with accurate load forecasting, when the electricity company can minimize its exposure to imbalance energy price risk. Thus, electricity companies have a financial incentive to strive for the most accurate load forecasting.

The thesis investigated the electricity consumption of the Vaasa region for the years 2019-2021. The research methods used are, however, applicable for load forecasting in other areas as well. By dividing customers into smaller sub-groups according to small-scale production and electricity contract types, the aim was to achieve a lower forecasting error compared to a situation where load is forecasted for the entire customer group aggregated. The reasoning behind the segmentation was that the customers' electricity consumption depends on their own small-scale production and the type of contract in use. Based on the literature review, a multiple linear regression model was selected as the forecasting model used. By interpreting the past loads from 2019-2020 with respect to calendar and temperature variables a benchmark load forecasting model was created. Model was further developed by varying the number of explanatory variables. By interpreting the coefficients of the regression model, conclusions were drawn about the factors affecting electricity consumption. The results show that, for example, public holidays seem to have a great impact on electricity consumption. By segmenting the customers into smaller subgroups, only slightly better forecasting results were achieved compared to the aggregated forecast. Segmentation by small-scale production decreased the mean absolute percentage forecasting error by 0,04 percent points, whereas segmentation by contract type decreased the error by 0,05 percent points. However, if the amount of small-scale production increases significantly in the future, it is reasonable to assume that preparing a load forecast separately for customers with own small-scale production could produce better results compared to an aggregated forecast.

Keywords electricity markets, load forecasting, regression analysis

Tekijä Arttu Niemelä

Työn nimi Sähköyhtiön lyhyen aikavälin kulutusennusteen kehittäminen

Koulutusohjelma Advanced Energy Solutions

Pääaine Sustainable Energy Systems and Markets

Vastuopettaja/valvoja Prof. Matti Lehtonen

Työn ohjaaja(t) DI Mikael Mäkelä

Yhteistyötaho Vaasan Sähkö Oy

Päivämäärä 23.09.2022

Sivumäärä 72

Kieli Suomi

Tiivistelmä

Tässä diplomityössä pyrittiin kehittämään sähköyhtiön lyhyen aikavälin kulutusennustamista. Sähköyhtiön tulee tietää edellisen päivän aamupäivällä asiakkaidensa seuraavan päivän tunnittainen sähkönkulutus. Sähkömarkkinoiden rakenteista johtuen, tarkalla kulutuksen ja tuotannon ennustamisella on mahdollista saavuttaa merkittäviä kustannussäästöjä. Mikäli tuotanto- ja kulutusennusteet vastaavat toteutumia, voi sähköyhtiö minimoida altistumistaan tasesähkön hintariskille, kun tuotannon ja kulutuksen erotuksesta johtuva sähkön hankinta tai myynti voidaan toteuttaa vuorokausimarkkinalla tasesähkön sijaan. Näin ollen sähköyhtiöillä on taloudellinen kannustin pyrkiä mahdollisimman tarkkaan kulutuksen ennustamiseen.

Työssä tutkittiin Vaasan alueen sähkönkulutusta vuosilta 2019-2021. Käytetyt tutkimusmenetelmät ovat kuitenkin sovellettavissa muidenkin alueiden kulutuksen ennustamiseen. Jakamalla asiakkaat pienempiin osaryhmiin pientuottajuuden ja sähkösopimustyyppien mukaan, pyrittiin saavuttamaan pienempi ennustevirhe verrattuna tilanteeseen, jossa kulutusta ennustetaan koko asiakasjoukolle aggregoidusti. Ennustevirheen mittarina käytettiin keskimääräistä prosentuaalista absoluuttista ennustevirhettä. Segmentoinnin lähtöoletuksena oli, että kuluttajien sähkönkulutus on riippuvaista omasta pientuotannosta ja käytössä olevasta sopimustyyppistä. Kirjallisuuskatsauksen perusteella käytetyksi ennustemalliksi valikoitui usean muuttujan lineaarinen regressiomalli. Analysoimalla toteutuneita kulutuksia vuosilta 2019-2020 muun muassa kalenteri- ja lämpötilamuuttujien suhteen, luotiin benchmark-ennustemalli. Benchmark-mallia kehitettiin edelleen selittävien muuttujien määrää vaihtelemalla ja tarkastelemalla ennustevirheen käyttäytymistä. Regressiomallin kertoimia tulkitsemalla tehtiin puolestaan johtopäätöksiä sähkönkulutukseen vaikuttavista tekijöistä. Tulokset osoittavat, että muun muassa pyhäpäivillä vaikuttaisi olevan suuri merkitys sähkönkulutukseen. Lisäksi kulutus on vahvasti lämpötilariippuvaista, sillä pienilläkin lämpötilan muutoksilla näyttäisi olevan vaikutusta sähkönkulutuksen suuruuteen. Segmentoimalla asiakkaat pienempiin osaryhmiin saavutettiin vain hieman parempia ennustetuloksia verrattuna aggregoituun ennusteseen. Pientuottajuuden mukaan segmentointi tuotti 0,04 prosenttiyksikköä pienemmän ennustevirheen. Sopimustyypeittäin segmentointi pienensi ennustevirhettä puolestaan 0,05 prosenttiyksikköä. Mikäli pientuotannon määrä lisääntyy merkittävästi tulevaisuudessa, on perusteltua olettaa, että kulutusennusteen laatiminen erikseen

pientuottaja-asiakkaille voisi tuottaa parempia lopputuloksia verrattuna aggregoituun ennusteeseen.

Avainsanat sähkömarkkinat, sähkönkulutus, kulutusennustaminen, regressioanalyysi

Sisällys

Esipuhe	8
Symbolit ja lyhenteet.....	9
Symbolit	9
Operaattorit.....	9
Lyhenteet.....	9
1 Johdanto.....	10
2 Sähkömarkkinoiden toiminta ja kulutusennustamisen taloudellinen merkitys sähköyhtiölle	13
2.1 Sähkömarkkinoiden toiminta	13
2.1.1 Sähkömarkkinoiden fyysiset tuotteet	13
2.1.2 Sähkön johdannaistuotteet.....	15
2.1.3 Tasehallinta.....	16
2.2 Vaasan Sähkö markkinatoimijana ja yhtiön ennusteprosessi.....	17
2.2.1 Ennustamisen taloudellinen merkitys sähköyhtiölle	17
3 Kulutusennustemallit	23
3.1 Kulutusennustemalleista yleisesti	23
3.2 Ennustevirheen määritelmä.....	24
3.3 Ennustemallit	25
3.3.1 Usean muuttujan lineaarinen regressiomalli.....	25
3.3.2 Aikasarjamallit	26
3.3.3 Semiparametrinen additiivinen malli.....	27
3.3.4 Eksponentiaalisen tasoituksen malli	28
3.3.5 Neuroverkkomallit	28
3.3.6 Sumean logiikan regressiomalli.....	29
3.3.7 Tukivektorikonemalli	30
3.3.8 Gradienttitehostamismalli.....	31
3.4 Kulutusennustemallin valinta	31
4 Tutkimusaineisto ja -metodit.....	33
4.1 Käytetty aineisto	33
4.2 Kulutusennustemallin luominen.....	35
4.2.1 Usean muuttujan lineaarisen regressiomallin muuttujat	35
4.2.2 Benchmark-mallin luominen.....	37

4.2.3	Pientuottaja-segmentin benchmark-malli.....	44
4.2.4	Benchmark-mallin kehittäminen	45
4.2.5	Tutkimusmetodin kuvaus.....	46
5	Tulokset	48
5.1	Aggregoitu kulutusennuste	48
5.2	Segmentointi pientuottajuuden mukaan.....	51
5.3	Segmentointi sopimustyypeittäin	54
5.4	Tulosten tarkastelu	56
5.4.1	Kokonaiskulutuksen ennustevirheen ajallinen vaihtelu	57
5.4.2	Regressiokertoimien tulkinta	61
6	Yhteenveto.....	66
7	Lähteet.....	68

Esipuhe

Tämä diplomityö tehtiin Vaasan Sähkö Oy:n Sähkökauppa-yksikköön. Haluan kiittää työni valvojaa, professori Matti Lehtosta sekä ohjaajaani Mikael Mäkelää Vaasan Sähköltä neuvoista ja ohjauksesta läpi diplomityöprosessin. Suurkiitos myös Vaasan Sähkölle äärimmäisen mielenkiintoisesta tutkimusaiheesta, ja Mikaelille hyvistä sparraustuokioista. Muistan lukeneeni, kuinka energiasektori on vakaa ja ennalta-arvattava ala. Viime vuodet ovat kuitenkin osoittaneet, että myös energia-alalla sattuu ja tapahtuu. Olenkin iloinen siitä, että pääsin saattamaan opintoni loppuun juuri silloin, kun energiamarkkinoilla oli turbulenssia.

Tämä diplomityö päättää pitkäksi venyneen opiskelu-urani, ainakin toistaiseksi. Olen tyytyväinen siihen, että pystyin suorittamaan valtaosan tutkinnostani huippu-urheilu-urani aikana. Tästä kiitos kuuluu Aalto-yliopistolle, sekä kaikille professoreille, jotka ovat mahdollisuuksien mukaan tarjonneet joustoa opintojeni suorittamiseen.

Helsingissä 23.09.2022
Arttu Niemelä

Symbolit ja lyhenteet

Symbolit

β_i Regressiomallin beta-kerroin

Operaattorit

Σ_i summa indeksin *i* yli

Lyhenteet

MAPE Mean Absolute Percentage Error, absoluuttinen keskimääräinen virhe
MAE Mean Absolute Error, keskimääräinen absoluuttinen virhe
WRMSE Weighted Root Mean Square Error, painotettu keskineliövirheen neliöjuuri
EPAD Johdannaispörssin aluehintaerotuote
UMR Usean muuttujan lineaarinen regressiomalli

1 Johdanto

Sähkömarkkinaosapuolien, kuten sähkön myyjien ja tuottajien, verkkoyhtiöiden sekä tasevastaavien toiminnan kannalta ennustamisella on keskeinen rooli. Jotta osapuolet voivat toimia teknistaloudellisesti kannattavasti, tulee heidän jatkuvasti ennustaa sähkömarkkinoiden toimintaa sekä pitkällä, että lyhyellä aikavälillä. Oleellisia ennustettavia suureita ovat muun muassa sähkön hinta sekä tuotannon ja kulutuksen määrä kullakin ajanhetkellä. Pitkän aikavälin kulutusennusteet ovat olennaisessa roolissa, kun sähköyhtiö suunnittelee investointeja tuotantokapasiteettiinsa, kun taas lyhyen aikavälin ennustaminen ohjaa yhtiön toimintaa sähkömarkkinalla. Seuraavan päivän sähkönkulutuksen ennustaminen määrittelee sen, kuinka paljon sähköyhtiön tulee ostaa tai myydä sähköä sähkömarkkinalla.

Sähkömarkkinaosapuolien tehdessä tarjouksiaan sähköpörssin day-ahead markkinalle, heidän tulee pystyä arvioimaan seuraavan vuorokauden kullekin tunnille tuotannon ja kulutuksen määrä. Näiden tuotanto- ja kulutusennusteiden pohjalta markkinaosapuolet jättävät osto- ja myyntitarjouksensa sähköpörssiin, jossa hinta muodostuu lopulta kysynnän ja tarjonnan kohdatessa. Mikäli tuotanto- tai kulutusennusteet eivät vastaa lopullista toteumaa, joutuvat markkinaosapuolet tekemään tasapainottavaa kaupankäyntiä päivänsisäisillä markkinoilla. Tuotannon ja kulutuksen tasapaino voidaan pyrkiä saavuttamaan myös kysyntäjoustolla tai tuotantoa säätämällä. Mikäli toteuma poikkeaa ennusteista, eikä tasapainoa saavuteta edellä mainituilla keinoilla, tilaa kantaverkkoyhtiö säätömarkkinatoimijoilta ylös- tai alassäätöjä sähköverkon tehotasapainon ylläpitämiseksi. Sähkömarkkinoiden rakenteet on pyritty luomaan niin, että sähköyhtiölle on taloudellisesti kannattavaa pyrkiä mahdollisimman tarkkaan kulutuksen ja tuotannon ennustamiseen, sillä ennustevirheet altistavat sähköyhtiön merkittävillä kustannusriskeille.

Tyypillisesti sähkön tuotanto ja kulutus ovat olleet hyvin ennustettavissa. Sähköä on tuotettu keskitetysti isoilla voimalaitoksilla, jotka tuottavat sähköä tasaisesti ja ennustettavasti. Sähkön kulutus puolestaan vaihtelee muun muassa lämpötilan, vuodenajan, viikonpäivän ja päivän tuntien välillä. Lämpötila vaikuttaa lämmitystarpeeseen ja sitä kautta sähkönkulutukseen, kun taas ihmisten käyttäytyminen ja sen myötä sähkönkulutus vaihtelevat ajankohdan mukaan. Kuitenkin suuren sähkönkäyttäjäjoukon kulutus on ollut hyvin ennakoitavissa erilaisten kuormitus- ja ennustemallien avulla [1]. Tulevaisuus luo kuitenkin haasteita markkinatoimijoiden ennustamisprosesseihin. Ihmisten sähkönkulutus on murroksessa muun muassa sähköautojen lataamisen, pientuotannon ja kysyntäjoukon lisääntyessä. Tuotannon ennustamista vaikeuttaa puolestaan tuulivoiman voimakas lisääntyminen [2], sillä tuotannon määrä on yhä riippuvaisempi vaikeasti ennustettavista sääolosuhteista. Suomessa aurinkosähkön osuus on

vielä toistaiseksi marginaalinen, mutta sen on ennakoitu kasvavan nopeasti tulevaisuudessa. Aurinkosähkön roolin povataan kasvavan etenkin pientuotannossa, jossa tyypillinen sähkön kuluttaja ryhtyy myös sähkön tuottajaksi. [3] Tällaiset kuluttaja-tuottajat ovat kulutuksen ennustamisen kannalta haasteellisia, sillä heidän sähkönkulutuksen määrä ei enää riipu vain tyypillisistä kulutukseen vaikuttavista tekijöistä, vaan lopulliseen sähkönkulutukseen vaikuttaa myös heidän oma pientuotantonsa.

Sähkönkulutuksen ennustamista on aikaisemmassa kirjallisuudessa tutkittu laajasti jo vuosikymmenien ajan. Energia-alan ennustamiseen liittyvistä julkaisuista kulutuksen ennustaminen kattaa liki puolet julkaisujen määrästä. Yksinomaan kymmenessä suosituimmassa alan lehdessä on viimeisen kymmenen vuoden aikana sähkönkulutuksen ennustamiseen liittyviä tutkimuksia julkaistu useita satoja vuosittain. Muita suosittuja aiheita ovat olleet muun muassa sähkön hinnan, sekä tuuli- ja aurinkovoimatuotannon ennustamiseen liittyvät julkaisut. [4]

Tämä diplomityö keskittyy sähkönkulutuksen ennustamiseen lyhyellä aikavälillä. Sähkøyhtiön tulee tietää asiakkaidensa seuraavan päivän tunnittainen sähkönkulutus edellisenä aamupäivänä. Työn tavoitteena on pyrkiä ymmärtämään sähkönkulutukseen vaikuttavia tekijöitä, ja niiden avulla pyrkiä kehittämään sähkøyhtiön kulutusennusteen tarkkuutta. Tarkkuuden mittarina käytetään absoluuttista keskimääräistä prosentuaalista virhettä. Tyypillisesti sähkönkulutusta on tutkittu siten, että kaikkien kuluttajien sähkönkulutus on aggregoitu yhteen esimerkiksi jakeluverkkotasolla, ja ennuste on laadittu tälle kokonaiskulutukselle. Tässä diplomityössä tutkitaan, päästäänkö kulutusennusteessa parempaan lopputulokseen, mikäli asiakkaat jaetaan pienempiin osaryhmiin ja kunkin osaryhmän kulutusta ennustetaan erillään toisistaan. Täten tutkimuksessa voidaan huomioida muun muassa, poikkeako pientuottajien kulutuskäyttäytyminen muista kuluttajista, ja onko erilaisilla sähkösovimustyypeillä vaikutusta sähkönkulutukseen. Näin ollen tämä diplomityö pyrkii vastaamaan seuraaviin tutkimuskysymyksiin:

1. Kuinka minimoida day-ahead kulutusennusteen ennustevirhettä?
2. Saavutetaanko asiakkaita segmentoimalla parempia tuloksia kuin aggregoidulla ennusteella?

Tutkimuksen kohteena ovat Vaasan Sähköverkon alueen kuluttajat. Alue kattaa Vaasan kaupungin lisäksi lähikuntien alueita, yhteensä noin 75 000 asiakasta. Diplomityön rakenne on seuraava. Toisessa alaluvussa käsitellään sähkömarkkinoiden toimintaa, sekä kulutusennustamisen taloudellista merkitystä sähkøyhtiölle. Kolmas alaluku käsittelee aikaisemman kirjallisuuden pohjalta kulutusennustamisessa käytettyjä matemaattisia malleja. Kirjallisuuskatsauksen pohjalta valitaan tämän diplomityön kannalta sopivin ennustemenetelmä, jonka avulla tutkimuksen tavoitteisiin

pyritään. Neljännessä aluvussa kuvaillaan tutkimuksessa käytettyä aineistoa sekä tutkimusmetodeita, ja viides alaluku keskittyy tulosten analysointiin. Kuudes alaluku vetää yhteen tämän diplomityön.

2 Sähkömarkkinoiden toiminta ja kulutusennustamisen taloudellinen merkitys sähköyhtiölle

Kulutusennustamisen merkitys sähköyhtiölle perustuu sähkömarkkinoiden toimintaperiaatteisiin, jotka luovat taloudelliset kannustimet tarkalle kulutusennustamiselle. Tässä luvussa käydään läpi eurooppalaisen sähkömarkkinan toimintaa, sekä havainnollistetaan kulutusennustamisen taloudellisesta merkitystä sähköyhtiölle.

2.1 Sähkömarkkinoiden toiminta

Sähkömarkkinoiden tarkoituksena on toteuttaa sähkön kysynnän ja tarjonnan tasapainottaminen kustannustehokkaasti. Markkinoilla muodostuva hinta ohjaa markkinatoimijoiden tuotanto- ja kulutus päätöksiä lyhyellä aikavälillä sekä investointeja pitkällä aikavälillä. Sähkömarkkinat voidaan jakaa sähköpörssiin fyysisiin tuotteisiin, joiden kautta valtaosa sähkökaupasta tapahtuu, sähkön johdannaistuotteisiin sekä kantaverkkoyhtiön vastuulla oleviin tasapainottaviin toimenpiteisiin, joilla tuotannon ja kulutuksen tasapaino varmistetaan jokaisella ajanhetkellä.

2.1.1 Sähkömarkkinoiden fyysiset tuotteet

Sähkön vuorokausimarkkinalla, eli niin kutsutulla day-ahead markkinalla, käydään kauppaa seuraavan vuorokauden jokaiselle tunnille. Suomi on osa eurooppalaisia tukkusähkömarkkinoita, joiden Elspot-vuorokausimarkkinaa ylläpitää Nord Pool. Maantieteellisesti markkinat on jaettu 21 tarjousalueeseen. Markkinatoimijat arvioivat tuotanto- ja kulutusennusteidensa pohjalta seuraavan päivän tilannetta, ja jättävät osto- ja myyntitarjouksensa omalle tarjousalueelleen klo 13 mennessä Suomen aikaa. Tarjoukset yhdistämällä muodostetaan kullekin tunnille kysyntä- ja tarjontakäyrä, joiden leikkauspisteessä sähkölle muodostuu jokaiselle päivän tunnille markkina- eli systeemihinta. Systeemihinta muodostuu Pohjoismaissa jätettyjen tarjousten pohjalta ja se kuvastaa kalleinta tuotantotapaa, joka tarvitaan kysynnän tyydyttämiseksi ja toisaalta hintaa, joka energiasta ollaan valmiita maksamaan. Sähköpörssi toimii suljettuna huutokauppana, jolla varmistetaan markkinoiden tehokas toiminta siten, että edullisimmat tuotantomuodot käytetään ensin. Sähköpörssiin lisäksi kauppaa voidaan käydä myös kahdenvälisillä Over-the-Counter -markkinoilla (OTC), joissa kaupankäynnin ehdot sovitaan ostajan ja myyjän kesken. [1]

Day-ahead markkinalle asetettujen myynti- ja ostotarjousten perusteella muodostuva systeemihinta ei ota huomioon mahdollisia fyysisiä rajoitteita

siirtoverkossa, vaan ne huomioidaan tarjousalueiden aluehintoja laskettaessa. Jakamalla markkinat hinta-alueisiin, kuvaa aluehinta parhaiten sen hetkistä tuotanto- ja siirtokapasiteettia. Pullonkaulatilanteessa, eli kun riittävää siirtokapasiteettia ylituotantoalueilta alituotantoalueille ei ole käytössä, eriytyvät aluehinnat toisistaan. Alituotantoalueella sähkön kysynnän tyydyttämiseksi joudutaan ottamaan käyttöön kalliimpia tuotantomuotoja, jolloin hinta nousee, kun taas ylituotantoalueella hinta laskee. [1]

Vuorokausimarkkinan suljeuduttua ja sähkön toimitushetken lähestyessä tuotanto- ja kulutusennusteeseen saattaa syntyä poikkeamia. Tuotantopuolella säästä riippuvaisen tuotannon määrän lisääntyessä tuotantoennusteesta tulee epätarkempi, jolloin todellinen tuotannon määrä sähkön toimitushetkellä voi poiketa suunnitellusta sääolosuhteiden vaihtuessa ennustetusta. Lisäksi muun muassa voimalaitoksen tai siirtolinjan vaurioituminen saattaa aiheuttaa poikkeamia tuotannon määrässä. Myös sähkönkulutus on säästäriippuvaista, jolloin sääennusteiden epätarkkuus aiheuttaa ennustevirhettä myös kulutuspuolella. Lisäksi kulutusennustamisessa ihmisten sähkönkulutuksen satunnaisuus tekee virheettömästä ennustamisesta vaikeaa.

Päivänsisäinen Elbas-markkina, eli niin kutsuttu intraday-markkina tarjoaa markkinatoimijoille mahdollisuuden korjata tuotanto- ja kulutussuunitelmiaan lähempänä sähkön toimitushetkeä. Intraday-markkinat avautuvat vuorokausimarkkinan sulkeutumisen jälkeen, ja kauppaa voi käydä 30 min ennen toimitustuntia asti. Myös päivänsisäistä markkinapaikkaa ylläpitää Nord Pool. [5] Kaupankäynti tukkusähkömarkkinoilla johtaa aina sähkön fyysiseen toimitukseen, joten markkinaosapuolilla tulee olla yhteys sähköverkkoon [1].

Valtaosa Nord Pool-sähköpörssin kaupankäynnistä tapahtuu vuorokausimarkkinalla. Vuonna 2020 sähköpörssin vuorokausimarkkinan yhteenlaskettu myynti- ja ostovolyymi oli 995 TWh, kun päivänsisäisillä markkinoilla volyyymi oli 26 TWh [6]. Tyypillisesti sähkön keskihinta on vuorokausimarkkinoilla päivänsisäisillä markkinoille olevaa hintaa alhaisempi. Keskihintaa oleellisempaa on kuitenkin päivänsisäisen markkinan sähkönhinnan huomattavasti suurempi volatilitetti, joka johtuu pääosin siitä, että tehdyt kaupat toteutetaan niin kutsutulla pay-as-bid menetelmällä, jolloin hinta määräytyy yksittäisten osto- ja myyntitarjousten perusteella kullekin kaupalle erikseen. Tilanne siis eroaa vuorokausimarkkinasta, jossa osto- ja myyntitarjousten perusteella sähkölle muodostuu yksi hinta kullekin tunnille. Jotta volatiililta Elbas-hinnalta voidaan vältyä, on markkinatoimijoille taloudellisesti kannattavaa pyrkiä ennustamaan mahdollisimman tarkasti tuotantoa ja kulutusta, jolloin valtaosa sähkön hankinnasta ja myynnistä voidaan toteuttaa vuorokausimarkkinalla. [7]

2.1.2 Sähkön johdannaistuotteet

Sähkön johdannaistuotteiden käyttötarkoitus on pääasiallisesti riskienhallinta. Johdannaisten avulla voidaan suojautua sähkönhinnan epäsuotuisalta kehitykseltä kiinnittämällä tuotetun sähkön myyntihinta tai sähkön hankintahinta. Toisaalta johdannaisten avulla voidaan pyrkiä myös kasvattamaan toiminnan tuottoa, jonka lisäksi johdannaisten hinta kuvaa myös markkinoiden näkemystä tulevasta sähkön hintatasosta. Johdannaisilla käydään kauppaa Nasdaq OMX Commodities - finanssimarkkinoilla sekä kahdenvälisillä OTC-markkinoilla. Kaupankäynnin kohteena olevat finanssijohdannaiset toteutetaan nettoarvon tilityksenä, eikä kaupankäyntiin siten liity sähkön fyysistä toimitusta. Referenssihintana finanssimarkkinoilla toimii Elspot:n systeemihinta sekä kunkin hinta-alueen hintaero systeemihintaan nähden. Kauppaa käydään futuuri-, DS-futuuri- ja optiosopimuksilla sekä aluehintaerotuotteilla. [1]

Futuurit ja DS-futuurit tarkoittavat sopimuksia ostaa tai myydä hyödykettä tulevaisuudessa. Futuurisopimuksissa sopimusperiodi on päivä, viikko, kuukausi, vuosineljännes tai vuosi. DS-futuurisopimuksissa periodi on puolestaan kuukausi, vuosineljännes tai vuosi. Futuurit ja DS-futuurit eroavat toisistaan nettoarvon tilityksen suhteen. Futuureissa tilitys aloitetaan heti sopimuksen teon jälkeen, ja tilitys tehdään päivittäin sekä kaupankäyntiajan loppuun saakka että toimitusaikana. DS-futuurisopimuksissa tilitys tehdään toimitusaikana päivittäin, mutta ennen toimitusaikaa tilitys tehdään vain sopimuksen viimeisenä kaupankäyntipäivänä. [8] Aluehintaerotuotteilla (EPAD) puolestaan voidaan suojata aluehinteriskää, sillä fyysinen kaupankäynti käydään aluehintaena ja näin ollen aluehinnan erotessa systeemihinnasta syntyy fyysiselle sähkökaupalle avoin positio. Avoimella positiolla tarkoitetaan tilannetta, jossa sähköyhtiöllä on tiedossa tietty määrä sähkönmyyntiä tiettyyn hintaan, mutta osa sähkön hankinnasta on vielä avoinna määrän tai hinnan suhteen. Aluehinteriskiltä voidaan suojautua suojaamalla tarvittava tehomäärä esimerkiksi futuureilla, jonka jälkeen suojataan aluehintaero samalle tehomäärälle EPAD-tuotteella. Tämän jälkeen fyysinen sähkönhankinta suoritetaan oman toimitusalueen aluehinnalla. [1]

Optio on puolestaan sopimus tulevaisuudessa tehtävästä kaupasta ja se velvoittaa ainoastaan option myyjää. Osto-optiossa ostajalla on oikeus ostaa hyödykettä sovittuun hintaan ja myyjällä on velvollisuus myydä hyödyke sovitulla hinnalla. Myyntioptiossa ostajalla on oikeus myydä ja myyjällä velvollisuus ostaa hyödyke ennalta määrättyyn hintaan. Koska optio velvoittaa vain myyjää, maksaa ostaja tyypillisesti myyjälle riskipreemion korvauksena myyjän ottamasta riskistä. [1]

2.1.3 Tasehallinta

Sähkön tuotannon ja kulutuksen tasapainon ylläpitoa, sekä siihen liittyvän säätösähkökapasiteetin käyttöä ja ylläpitoa kutsutaan tasehallinnaksi [1]. Suomen kantaverkkoyhtiö Fingrid ylläpitää yhdessä muiden pohjoismaalaisten kantaverkkoyhtiöiden kanssa säätösähkömarkkinoita, joiden avulla tuotannon ja kulutuksen välinen tasapaino varmistetaan kullakin ajanhetkellä. Säätösähkömarkkinoille tarjouksia voi antaa resursseista, jotka pystyvät toteuttamaan 5 MW minimitarjouskoon mukaisen tehonmuutoksen 15 minuutissa. Tarjoukset annetaan Fingridille viimeistään 45 minuuttia ennen käyttötuntia. Ylös- ja alassäätötarjoukset asetetaan hintajärjestykseen. Hinnan määrittää ylössäädössä kunkin tunnin kallein käytetty ylössäätötarjouksen hinta, kun taas alassäädössä hinta määräytyy halvimman käytetyn alassäätötarjouksen mukaan. Ylössäätötarjouksista ensin käytetään halvin, kun taas alassäätötarjouksista käytetään ensin kallein. Näin ollen jokaiselle tunnille muodostuu sekä ylösettä alassäätöhinta, jotka toimivat myös tasesähkön hinnoittelun perusteena. [9] Lisäksi Fingrid ylläpitää muita reservimarkkinoita, joilla huolehditaan muun muassa sähköverkon taajuuden ylläpidosta sekä mahdollisiin häiriötilanteisiin varautumisesta. Nämä reservit ovat säätösähkömarkkinoilla toimivia reservejä nopeammin aktivoituvia voimallaitoksia, kulutuskohteita tai energiavarastoja. [10]

Sähkön toimituksen jälkeen suoritetaan taseselvitys, jossa selvitetään sähkömarkkinoilla toimivien osapuolten väliset sähköntoimitukset. Sähkömarkkinoilla toimivien osapuolten on jatkuvasti huolehdittava omasta sähkötaseestaan, eli tuotannon, hankinnan, kulutuksen ja myynnin välisestä tehotasapainosta. Käytännössä osapuolet eivät pysty tähän, vaan niillä tulee olla avoin toimittaja, joka huolehtii sähkötaseen tasapainottamisesta. Tasevastaavaksi kutsutaan sellaista osapuolta, jonka avoin toimittaja on Fingrid. [11]

Suomessa on käytössä hierarkkinen taseselvitysmalli, jossa ylimpänä on järjestelmävastaava Fingrid, joka vastaa taseselvityksestä valtakunnallisella tasolla, sekä Suomen ja muiden maiden välillä. Fingridin alapuolella hierarkiassa ovat tasevastaavat, jotka vastaavat kantaverkkotasolla taseselvityksestä avointen toimitustensa osalta. Alimpana hierarkiassa ovat jakeluverkonhaltijat, joiden vastuulla on taseselvityksen tekeminen oman jakeluverkkonsa osalta. [1]

Kun kukin taseselvityksen osapuoli on selvittänyt omat sähköntoimituksensa, käydään osapuolten toteutuneiden toimitusten ja hankintojen välisen tasepoikkeaman korjaamiseksi tasesähkökauppaa. Kaupankäynti tapahtuu tasevastaavan ja Fingridin tasesähköyksikön välillä. Aikaisemmin tasepoikkeamat hinnoiteltiin erikseen tuotantotasesähköön ja kulutustasesähköön. 1.11.2022 alkaen siirryttiin kuitenkin yhden taseen malliin, jossa tuotanto ja kulutus käsitellään yhdessä taseessa. Näin ollen

tasevastaavan tasepoikkeamalle sovelletetaan yksihintajärjestelmää, jossa osto- ja myyntihinnat ovat samansuuruiset. [12] Tasepoikkeaman hinnaksi tulee ylössäätötunnilla ylössäätöhinta ja alassäätötunnilla alassäätöhinta, ja mikäli tunnilla ei ole tehty säätöjä, käytetään hintana vuorokausimarkkinan Suomen aluehintaa. Ylössäätöhinta on vähintään ja alassäätöhinta korkeintaan vuorokausimarkkinoiden Suomen tarjousalueen hinta. [13]

2.2 Vaasan Sähkö markkinatoimijana ja yhtiön ennusteprosessi

Vaasan Sähkö Oy on sähkökauppaan, kaukolämpöön ja sähkönsiirtoon keskittyvä yritys. Yhtiöllä on noin 110 000 sähköasiakasta ympäri Suomen, jonka lisäksi yhtiö toimittaa kaukolämpöä Vaasassa. Tytäryhtiö Vaasan Sähköverkko Oy puolestaan vastaa Vaasan seudun sähkönsiirrosta. Yhtiö tuottaa sähköä omilla tuotantolaitoksilla ja saa kiinteitä toimituksia mankalayhtiö EPV Energia Oy:n kautta omistamalla tuotanto-osuuksilla. Lisäksi yhtiö käy sähkökauppaa sekä fyysisillä-, että johdannaistuotteilla. Se osa asiakkaiden kulutuksesta, jota ei pystytä kattamaan omalla tuotantoportfoliolla, hankintaan sähköpörssistä. Vaasan Sähkön sähkön hankinnasta ja myynnistä vastaa yhtiön Sähkökauppa-yksikkö.

Vaasan Sähkön tuotanto- ja kulutusennusteet toteutetaan yhtiön tasevastaavan EPV Tase Oy:n toimesta. EPV Tase kerää kaikkien tuotantoyksiköiden tuotantoennusteet ja rakentaa niistä tuotantosuunnitelman, josta kullekin EPV Taseen osakkaalle toimitetaan omistussuosituksen mukainen kiinteä toimitus sähköä. Tämän lisäksi tehdään kulutusennuste kullekin osakkaalle, ja kunkin osakkaan tuotanto- ja kulutusennusteen erotuksesta muodostuu tarve tehdä joko osto- tai myyntitarjouksia sähköpörssin vuorokausimarkkinalle. Kulutusennuste laaditaan noin kello 10:00 mennessä seuraavan päivän jokaiselle tunnille. Näin ollen, kulutusennuste luodaan 12-36 tunnin päähän ennustamishetkestä.

2.2.1 Ennustamisen taloudellinen merkitys sähköyhtiölle

Lähtökohtaisesti tuotannon ja kulutuksen ennusteiden poikkeamia toteutuneesta voidaan korjata neljällä tapaa: tekemällä korjaavia toimenpiteitä päivänsisäisellä Elbas-markkinalla, muuttamalla tuotantosuunnitelmaa, vaikuttamalla kulutukseen kysyntäjouston kautta ja viime kädessä tasesähkökaupalla. Mikäli ennusteissa on merkittävää poikkeamaa toteumiin nähden ja sähköyhtiö ei saa tasapainotettua tasetaan Elbas-toimenpiteillä, tuotantosuunnitelman muutoksilla tai kysyntäjoustolla, altistuu sähköyhtiö merkittäville tasesähkön hintariskeille kattaakseen oman poikkeaman. Vaikka tämä diplomityö keskittyy nimenomaan kulutuksen ennustamiseen, on tuotannon ennustamisella

myös merkittävä rooli, sillä tuotanto- ja kulutusennusteen välinen erotus määrittää sen, kuinka paljon sähköyhtiö joutuu hankkimaan tai myymään sähköä vuorokausimarkkinoilla. Mikäli ennusteiden perusteella sähköyhtiön tuotanto ylittää asiakkaidensa kulutuksen, myydään ylijäämä vuorokausimarkkinalla. Mikäli tuotanto ei pysty kattamaan kulutusta, paikataan alijäämä tekemällä ostoja vuorokausimarkkinalla. Taulukko 1 kuvaa mahdolliset skenaariot, mikäli kulutus- ja tuotantoennusteiden sekä tasapainottavien toimenpiteiden jälkeen sähköyhtiölle on syntynyt tasepoikkeamaa.

Taulukko 1: Tasepoikkeaman mahdolliset skenaariot.

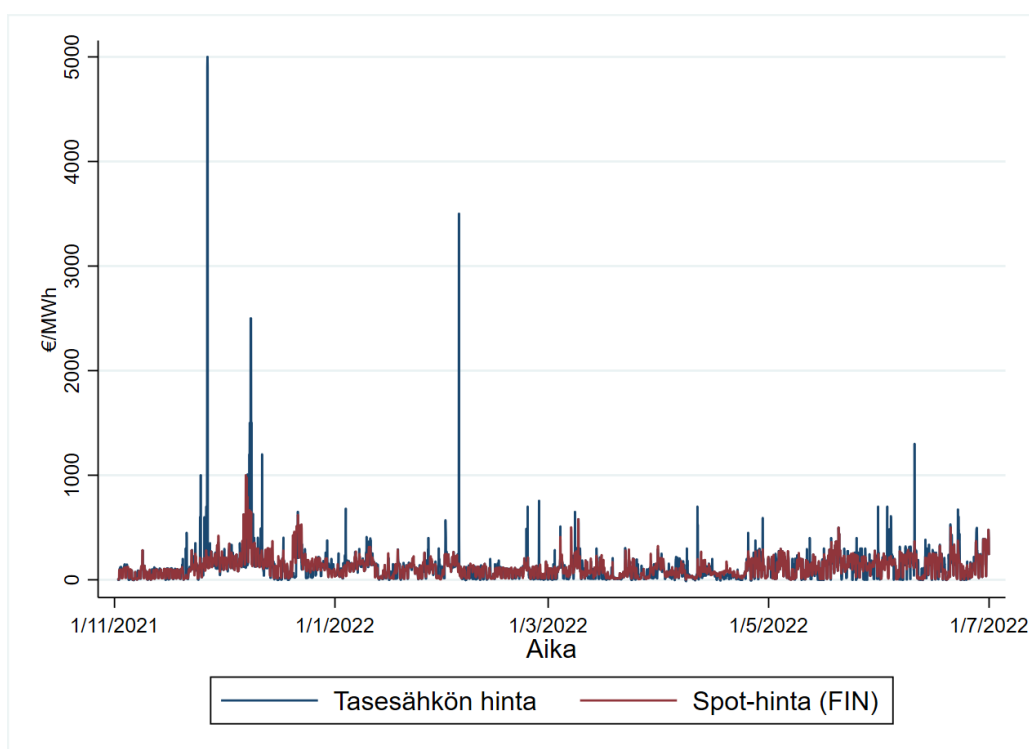
Säätötunti	Tasepoikkeama	Tasesähkön hinta	€	Tasesähkön kattohinta
Ylössäätötunti	Ylijäämä	min spot	hyvitys	10 000 €/MWh [14]
Ylössäätötunti	Alijäämä	min spot	veloitus	10 000 €/MWh [14]
Alassäätötunti	Ylijäämä	max spot	hyvitys	4000 €/MWh [15]
Alassäätötunti	Alijäämä	max spot	veloitus	4000 €/MWh [15]

Mikäli yhtiön tasepoikkeama on ylijäämäinen, saa yhtiö tasevastaavalta hyvitystä. Mikäli kyseessä on ylössäätötunti, hyvitys on vähintään spot-hinta, ja tällöin yhtiö tekee voittoa tasesähköllä verrattuna tilanteeseen, jossa ylijäämä oltaisiin myyty vuorokausimarkkinalla. Mikäli kyseessä on alassäätötunti, olisi yhtiölle ollut kannattavampaa saada myytyä ylijäämäsähkö vuorokausimarkkinoilla, sillä alassäätötunnilla tasesähkön hinta on korkeintaan spot-hinta. Mikäli yhtiön tasepoikkeama on alijäämäinen, tasevastaava veloittaa yhtiötä alijäämäisyydestä. Mikäli kyseessä on ollut ylössäätötunti, altistuu yhtiö tasesähkön hintariskille ja veloitettava hinta on vähintään spot-hinta, mutta voi korkeimmillaan nousta jopa tuhansiin euroihin. Mikäli tasepoikkeama on alijäämäinen ja kyseessä on alassäätötunti, on veloitus korkeintaan spot-hinta. Lisäksi tasevastaava maksaa Fingridille tuotannon ja kulutuksen volyymimaksuja, jotka se veloittaa tasevastuun alaisilta yhtiöiltä. Tasesähkön kattohinta on ollut 5000 €/MWh, mutta 1.11.2022 alkaen kattohinta nousee 10 000 euroon megawattitunnilta. [14]

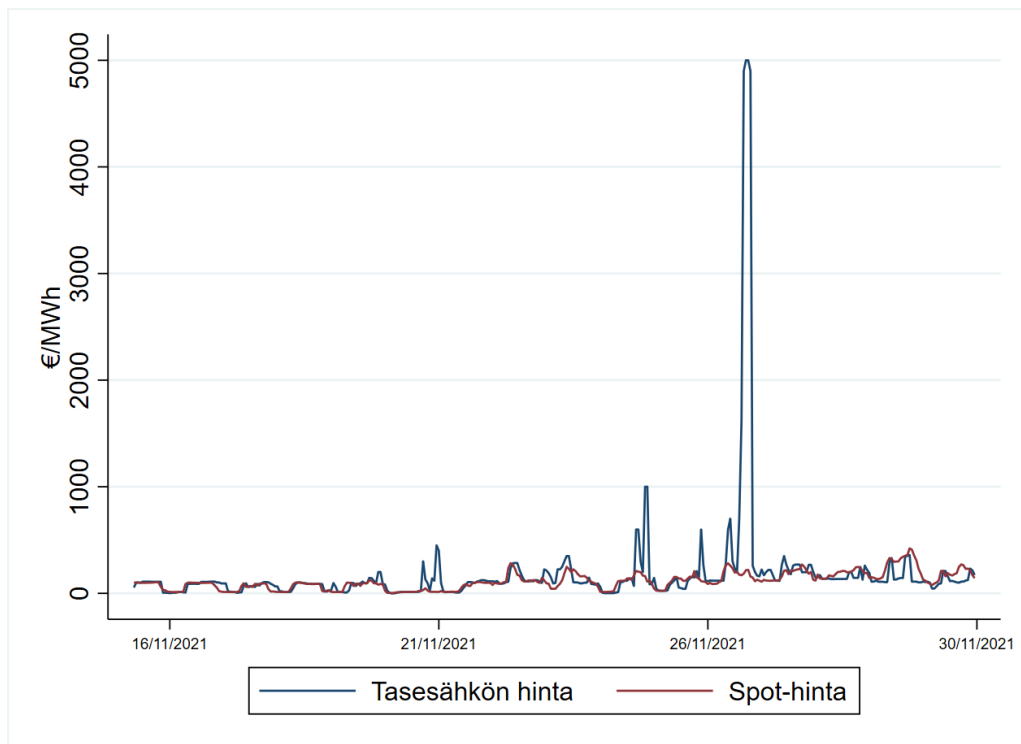
Teoriassa tasesähkökaupankäynnillä on mahdollista tehdä voittoa, mikäli tasesähköä voidaan ostaa tai myydä vuorokausimarkkinahintaa parempaan hintaan. Alijäämäinen tase alassäätötunnilla mahdollistaa tasesähkön ostamisen spot-hintaa edullisemmin, kun taas ylijäämäinen tase ylössäätötunnilla mahdollistaa sähkön myynnin spot-hintaa kalliimmalla. Kuitenkin tahallinen virheellinen ennustaminen ja siten tasepoikkeaman synnyttäminen on kielletty Fingridin ja tasevastaavan tasepalvelusopimuksessa [16]. Lisäksi tuotanto- ja kulutusennusteita laadittaessa ei tiedetä, mitkä seuraavan päivän tunneista ovat ylös- ja

alassätötunteja, joten voiton tavoittelu tasesähkön avulla on markkinoiden rakenteista johtuen vaikeaa. Onkin sähköyhtiön edun mukaista pyrkiä mahdollisimman tarkkaan tuotanto- ja kulutusennusteeseen ja näin ollen minimoimaan tasepoikkeaman suuruus.

Tasesähkön hintariskiä voidaan tarkastella vertailemalla toteutuneita vuorokausimarkkinan- ja tasesähkön hintoja. Kuva 1 esittää Suomen alueen spot-hinnan ja tasesähkön hinnan marraskuusta 2021 kesäkuuhun 2022. Kuvasta on huomattavissa muutamia merkittäviä hintapiikkejä muun muassa marraskuussa ja helmikuussa, jolloin tasesähkön hinta on noussut hetkellisesti jopa tuhansiin euroihin megawattitunnilta.



Kuva 1: Tasesähkön ja Suomen alueen spot-hinnat marraskuusta 2021 kesäkuuhun 2022.



Kuva 2: Tasesähkön hinta ja vuorokausimarkkinan spot-hinta marraskuun kahdella viimeisellä viikolla vuonna 2021.

Kuva 2 puolestaan esittää marraskuun 2021 kahden viimeisen viikon spot- ja tasesähkön hinnat. Kuten kuvasta nähdään, hinnat ovat lähellä toisiaan suuren osan ajasta, mutta tasesähkön hinnassa on nähtävissä muutamia hintapiikkejä, jolloin hinnat kohoavat jopa 5000 euroon megawattitunnilta. Mikäli tuotanto- ja kulutusennusteen epätarkkuuden johdosta tällaisella tunnilla sähköyhtiön tasepoikkeama olisi alijäämäinen, jouduttaisiin poikkeama kattamaan erittäin kalliilla tasesähköllä. Tällaiset tasesähkön hintapiikit altistavat sähköyhtiön merkittävälle hintariskille. Näin ollen voidaan todeta, että pyrkimällä mahdollisimman tarkkaan ennustamiseen sekä tuotannon, että kulutuksen osalta, voidaan altistumista hintariskille minimoida. Ideaalitulanteessa tuotanto- ja kulutusennuste täsmäävät täydellisesti toteutumien kanssa, jolloin ennusteiden välinen erotus voidaan kattaa täysin vuorokausimarkkinalta ostetulla tai myydyllä sähköllä.

Tarkastelemalla tasevastaavan sähköyhtiölle toimittamia laskuja, voidaan analysoida millaisiin kustannussäästöihin päästäisiin, mikäli tuotanto- ja kulutusennusteet eivät sisältäisi ennustevirhettä. Tasevastaava veloittaa kuukausittain tasevastuun alaisuuteen kuuluvilta yhtiöiltään tehdyistä sähkömarkkinatoimenpiteistä. Taulukko 2 havainnollistaa yksinkertaistettua esimerkkiä tasevastaavan toimittaman kuukausilaskun sisällöstä. Luvut ovat peräisin marraskuun 2021 tasevastaavan Vaasan Sähkölle toimittamasta laskusta. Negatiivinen etumerkki tarkoittaa tuloa sähköyhtiölle, kun taas

positiivinen etumerkki kuvaa kustannusta. Luvut ovat skaalattu liikesalaisuussyistä siten, että spot-oston volyymi ja euromäärä toimivat vertailukohtana muihin lukuihin nähden. Kertoimien x ja y avulla spot-ostojen todellinen volyymi ja euromäärä skaalattiin niin, että ne saavat arvot 100x ja 100y. Kertoimien x ja y avulla on esitetty myös laskun muut komponentit.

Taulukko 2: Esimerkki tasevastaavan sähköyhtiölle toimittamasta kuukausilaskusta marraskuulta 2021.

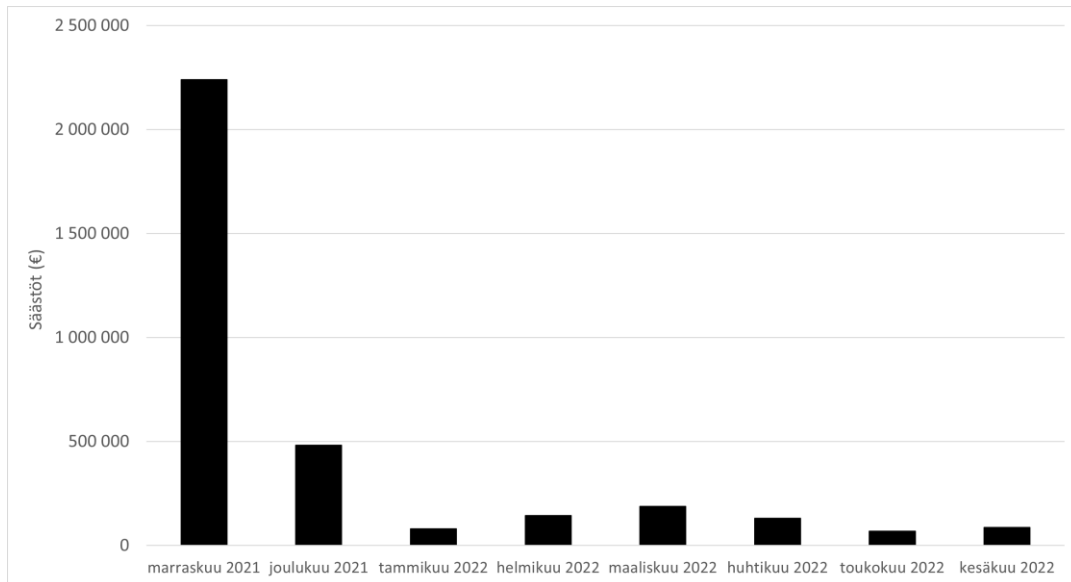
Laskutuserä	Skaalattu volyymi	€/MWh	Skaalattu €
Spot-osto	100x	121,94	100y
Spot-myynti	-19x	91,57	-14y
Tasesähkön osto	37x	362,93	111y
Tasesähkön myynti	-8x	104,14	-7y
Spot-hintaisen tasesähkön osto	21x	98,56	17y
Spot-hintaisen tasesähkön myynti	2x	107,83	-2y
Elbas-osto	13x	171,92	18y
Elbas-myynti	2x	159,02	-3y
Maksut	-	-	3y
SUMMA			223y

Spot-osto ja -myynti ovat tuotanto- ja kulutusennusteen erotuksesta johtuvat ostot ja myynnit vuorokausimarkkinalla. Tasesähkön osto ja -myynti ovat tasesähkön hinnalla ostettua tai myytyä tasesähköä. Spot-hintaisella tasesähköllä tarkoitetaan puolestaan tasevastaavan taseessa tehtyjä netoituksia. Elbas-ostot ja -myynnit ovat päivänsisäisillä markkinoilla tapahtuvia kauppvoja, jotka toteutuvat ennalta määritettyjen parametrien mukaan automaattisesti. Maksut-rivi sisältää erinäisiä volyymi-, tuotanto- ja kulutusmaksuja.

Kuten taulukosta 2 nähdään, marraskuussa 2021 sekä spot-ostoja, että tasesähkön ostoja jouduttiin tekemään paljon. Tasesähkön ostohinta oli keskimäärin korkea, joka johti siihen, että tasevastaava joutui veloittamaan sähköyhtiöltä merkittävän summan kuukausilaskulla. Tasesähkön korkeat ostohinnat näkyvät myös siinä, että euromääräisesti tasesähköstä maksettiin enemmän kuin spot-ostoista, vaikka tasesähkön ostovolyyymi oli alle puolet spot-ostojen volyymista.

Vähentämällä tasevastaavan avoimen toimituksen tunnittaisista kokonaisenergiämääristä kunkin tunnin spot-kaupat, jää jäljelle spot-kauppojen jälkeinen poikkeama, joka korjataan joko päivänsisäiseltä markkinalta hankitulla sähköllä, tai viime kädessä tasesähköllä. Kun spot-kauppojen jälkeisen poikkeaman energiamäärän kertoo kyseisten tuntien spot-hinnoilla, saadaan laskettua poikkeaman arvo spot-hinnalla. Vertaamalla tätä poikkeaman arvoa tasevastaavan avoimen toimituksen euromääräiseen laskuun tunneittain, josta on vähennetty kunkin tunnin

spot-kauppojen kulutus- ja tuotantomaksut, voidaan arvioida millaisiin säästöihin päästäisiin, mikäli koko tasepoikkeama pystyttäisiin kattamaan spot-hintaisella sähköllä päivänsisäisiltä markkinoilta tai tasesähkönä hankitun sähkön sijaan. Kuva 3 esittää kuukausittaiset säästöt kyseisessä ideaalitulanteessa marraskuusta 2021 kesäkuuhun 2022.



Kuva 3. Potentiaaliset kuukausittaiset säästöt, mikäli tasepoikkeama oltaisiin korvattu spot-hintaisella sähköllä.

Kuten aikaisemmin todettiin, marraskuussa 2021 tasesähkön hinta oli ajoittain poikkeuksellisen korkealla, mikä heijastuu myös kuvan 3 marraskuun potentiaalsiin säästöihin, mikäli tasepoikkeama oltaisiin onnistuttu kattamaan kokonaan spot-hintaisella sähköllä. Marraskuussa 2021 kustannussäästöt olisivat olleet yli kaksi miljoonaa euroa, ja vuoden 2022 alkupuoliskollakin jopa satoja tuhansia euroja kuukausittain. On siis perusteltua todeta, että pyrkimällä tarkempaan ennustamiseen tuotannon ja kulutuksen osalta, on sähköyhtiöillä mahdollisuus parantaa toimintansa kannattavuutta merkittävästi.

3 Kulutusennustemallit

Tässä luvussa tarkastellaan aikaisemman kirjallisuuden pohjalta sähkönkulutuksen ennustamisessa käytettyjä ennustemalleja, sekä käydään läpi ennustevirheen määritelmä ja kulutusennustamisen viimeaikaisia trendejä.

3.1 Kulutusennustemalleista yleisesti

Kulutusennustemallit voidaan jakaa kahteen ryhmään, tilastotieteellisiin sekä tekoälypohjaisiin malleihin. Raja tämän jaottelun välillä on kuitenkin hämärtynyt, kun poikkitieteelliset tutkimusryhmät ovat yhdistelleet ominaisuuksia molemmista kategorioista. [17] Lisäksi jaottelua tehdään myös ennusteen aikajänteen perusteella. Ennustemallit voidaan jakaa erittäin lyhyen aikavälin, lyhyen aikavälin, keskipitkän aikavälin ja pitkän aikavälin ennusteisiin [18]. Tässä diplomityössä tarkasteltavana oleva day-ahead ennustaminen luokitellaan lyhyen aikavälin ennusteeksi.

Tilastotieteellisistä malleista tyypillisimmän käytettyjä ovat usean muuttujan lineaariset regressiomallit, autoregressiiviset liukuvan keskiarvon ARMA-mallit, semiparametriset additiiviset mallit sekä eksponentiaaliseen tasoitukseen perustuvat mallit. Tekoälypohjaisista malleista usein käytettyjä ovat muun muassa neuroverkko-, sumean logiikan regressio-, tukivektorikone- ja gradienttitehostamismallit. [17]

Hong et. al. [4] kuvailevat tutkimuksessaan kulutusennustamisen viimeaikaisia trendejä. Tekoälyä ja koneoppimista on hyödynnetty ennustamisessa jo kolmen vuosikymmenen ajan, mutta niiden suosio on viimeaikoina lisääntynyt datan määrän ja tietokoneiden laskutehon kasvun myötä. Lisäksi usean ennusteen yhdistäminen on kirjallisuuden perusteella hyväksi havaittu tapa parantaa ennustetarkkuutta. [4] Tekemällä useita ennusteita eri ennustemalleilla ja yhdistämällä ennusteet joko yksinkertaisesti aritmeettisella keskiarvolla, tai vaihtoehtoisesti esimerkiksi pienimmän neliösumman menetelmällä, ollaan tutkimuksissa saavutettu parempia lopputuloksia, kun luomalla yhdellä ennustemallilla yksi ennuste [19].

Kolmas merkittävä tutkimusalue kulutusennustamisen viimeaikaisessa kehityksessä on ollut hierarkinen ennustaminen. Sen sijaan, että ennuste laadittaisiin yhdelle isolle joukolle kuluttajia, kuten koko sähköverkkoyhtiön hallinnoimalle alueelle, kuluttajia jaetaan pienempiin osaryhmiin, joille laaditaan omat ennusteensa. Tälläisen segmentoinnin pitäisi ainakin teoriassa johtaa parempaan ennustetarkkuuteen verrattuna aggregoidun kulutuksen ennustamiseen. [4] Hierarkisesta ennustamisesta on osaltaan tullut mahdollista, kun etäluettavat sähkömittarit ovat johtaneet käytettävissä olevan kulutusaineiston räjähdysmäiseen kasvuun.

Kulutuspaikkakohtainen data tarjoaa uudenlaisen mahdollisuuden segmentoida samankaltaisia asiakkaita pienempiin osaryhmiin ja pyrkiä siten säilyttämään tiettyjä tärkeitä kulutuskäyttäytymisen piirteitä mukana ennustamisessa, jotka aggregoidussa ennustamisessa muutoin häviäisivät. Lisääntynyt datan määrä tarjoaa kulutusennustamisen lisäksi myös muita mahdollisuuksia sähköyhtiöille. Ymmärtämällä asiakkaidensa sähkönkulutusta aikaisempaa paremmin, voi sähköyhtiö tarjota entistä yksilöidympiä palveluita asiakkailleen, ja löytää muun muassa potentiaalisia kohteita, joita pyrkiä hyödyntämään kysyntäjoustossa. [20]

Neljäntenä trendinä tutkijat mainitsevat todennäköisyyspohjaisen ennustamisen, jossa ennusteista lasketaan todennäköisyysjakaumia tai kvanttileja sen sijaan, että ennusteista ilmoitettaisiin vain pistearvoja, tyypillisimmin odotusarvoja. Todennäköisyyspohjainen ennustaminen on ollut suosittua erityisesti tuulivoiman tuotantoennusteissa, josta se on sittemmin lähtenyt leviämään myös kulutusennustamisen piiriin. [4] Viimeistään vuonna 2014 järjestetyssä Global Energy Forecasting Competition kilpailussa todennäköisyyspohjainen kulutusennustaminen sai laajaa huomiota, kun yhteensä 581 osanottajaa 61:stä maasta osallistui kilpailuun, jossa todennäköisyyspohjaista ennustamista tuli soveltaa sähkönkulutuksen, sähkön hinnan, tuulivoimatuotannon ja aurinkosähkötuotannon ennustamiseen [21].

3.2 Ennustevirheen määritelmä

Ennustevirheen avulla erilaisia kulutusennustemalleja voidaan vertailla keskenään. Ennustevirheellä kuvataan, kuinka lähellä kulutusennuste on toteutunutta sähkönkulutusta. Tässä diplomityössä ennustevirheen mittarina käytetään keskimääräistä absoluuttista prosentuaalista virhettä (Mean Absolute Percentage Error, MAPE), joka lasketaan kaavan 1 mukaisesti:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (1)$$

jossa

A_t on toteutunut kulutus

F_t on ennustettu kulutus

n on tarkasteltavan jakson tuntien lukumäärä

$\sum_{i=1}^n$ on summa, jossa i käy 1:stä n :ään

GEFCom 2012 -ennustekilpailussa [22] ennustevirheen mittarina käytettiin puolestaan painotettua keskineliövirheen neliöjuurta (Weighted Root Mean Square Error WRMSE), joka voidaan laskea kaavan 2 mukaisesti:

$$WRMSE = \sqrt{\frac{\sum_i w_i (A_i - P_i)^2}{\sum_i w_i}} \quad (2)$$

jossa

A_i on i :n:n havainnon toteutunut kulutus

P_i on i :n:n havainnon ennustettu kulutus

w_i on i :n:n havainnon painoarvo

Kilpailussa ennusteet laadittiin erikseen systeemitasolle, sekä pienemmille vyöhykkeille, ja molemmille annettiin eri painoarvot. [22] Lisäksi muun muassa keskimääräistä absoluuttista virhettä (Mean Absolute Error, MAE) on käytetty kirjallisuudessa [23]. MAPE:n valinta käytetyksi ennustevirheen mittariksi on perusteltua, sillä se on yleisesti käytössä Vaasan Sähkön ennusteprosessissa. Lisäksi MAPE:a on käytetty laajalti aikaisemmassa kirjallisuudessa, mm. [24], [25], [26], [27]. Selkeyden vuoksi tässä diplomityössä ennustevirheestä puhuttaessa tarkoitetaan nimenomaan keskimääräistä absoluuttista prosentuaalista virhettä, ellei erikseen muuta mainita.

3.3 Ennustemallit

Seuraavat alaluvut käsittelevät aikaisemman kirjallisuuden pohjalta yleisimmin käytettyjä kulutusennustemalleja. Mallien taustalla olevaa teoriaa kuvaillaan lyhyesti yleisellä tasolla, jonka lisäksi lukijalle tarjotaan muutamia paljon viitattuja tutkimuksia, joissa mallien hyödyntämistä käytännön sovelluksissa käsitellään tarkemmin.

3.3.1 Usean muuttujan lineaarinen regressiomalli

Usean muuttujan lineaarinen regressiomalli pyrkii selittämään toisistaan riippumattomien selittävien muuttujien (esim. lämpötila, kuukausi, kellonaika) avulla selitettävän muuttujan (sähkönkulutus) arvoja [17]. Riippuvuuden oletetaan olevan lineaarinen, jolloin esimerkiksi kahden selittävän muuttujan malli voidaan esittää:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad (3)$$

jossa Y on selitettävä muuttuja (sähkönkulutus), x_1 ja x_2 ovat selittäviä muuttujia, β_i :t ovat estimoitavia regressiokertoimia ja ϵ on virhetermi.

Huomionarvoista on termin ”lineaarinen” harhaanjohtavuus. Vaikka esimerkiksi sähkönkulutuksen ja lämpötilan riippuvuus olisi epälineaarinen, voidaan lineaarista regressiomallia käyttää sähkönkulutuksen ennustamiseen, sillä lineaarisuudella viitataan nimenomaan lineaarisiin yhtälöihin, joilla mallin regressiokertoimet estimoidaan, ei selitettävän ja selitettävien muuttujien riippuvuussuhteeseen [17].

Usean muuttujan lineaarisia regressiomalleja ovat tarkemmin tutkineet muun muassa Papalexopoulos ja Hesterberg [28], jotka kehittivät regressiopohjaisen lyhyen aikavälin kulutusennustemallin Yhdysvaltalaiselle Pacific Gas and Electric Company:lle. Mallin parametrit estimoitiin vuosien 1979-1984 toteutuneiden kulutusten ja lämpötilojen pohjalta, jonka jälkeen mallilla ennustettiin seuraavalle päivälle tunnittainen kulutus vuodelle 1985. Ramanathan et. al. [29] tekivät vastaavanlaisen tutkimuksen, jossa he loivat 24 regressiomallia, yhden jokaiselle päivän tunnille. Lisäksi Hong [25] loi väitöskirjassaan systemaattisen prosessin usean muuttujan regressiomallin kehittämiseksi. Ennustaessaan seuraavan päivän sähkönkulutusta keskikokoiselle Yhdysvaltalaiselle sähköyhtiölle, jonka huippukulutus oli noin 800 megawattia, päästiin parhaalla ennustemallilla noin kolmen prosentin keskimääräiseen absoluuttiseen prosentuaaliseen ennustevirheeseen. [25]

Hong [25] kehitti väitöskirjassaan myös yksinkertaisen usean muuttujan lineaariseen regressioon perustuvan benchmark-kulutusennustemallin. Benchmark-mallia on sittemmin käytetty vertailukohtana useissa tutkimuksissa, muun muassa [30] ja [31]. Lisäksi benchmark-malli toimi myös kansainvälisen Global Energy Forecasting Competition 2012 (GEFCom 2012) -kulutusennustamiskilpailun vertailumallina [22].

3.3.2 Aikasarjamallit

Tyypillisiä kulutusennustamisessa käytettyjä aikasarjamalleja ovat niin kutsutut ARMA-mallit. Kun usean muuttujan regressiomallissa ennuste laaditaan hyödyntämällä selittävien ja selitettävän muuttujan menneitä arvoja ja estimoimalla niiden pohjalta regressiokertoimet, autoregressiivisissa AR-malleissa ennuste laaditaan hyödyntäen selitettävän muuttujan menneitä arvoja. Näin ollen termi autoregressiivinen viittaa selitettävän muuttujan regressioon itsensä kanssa. Liukuvan keskiarvon MA-malleissa puolestaan hyödynnetään menneitä ennustevirheitä ennustetta laadittaessa. Yhdistämällä autoregressiivisen- ja liukuvan keskiarvon mallin päädytään autoregressiiviseen liukuvan keskiarvon ARMA-malliin. Mikäli aikasarja ei ole stationaarinen, hyödynnetään autoregressiivista integroitua liukuvan keskiarvon ARIMA-mallia. [32]

Muun muassa Huang ja Shih [33] hyödynsivät ARMA-mallia lyhyen aikavälin kulutusennustamisessa Taiwanissa. Heidän kehittämänsä ARMA-malli saavutti parempia tuloksia verrattuna tekoälypohjaiseen

neuroverkkomalliin. ARMA-mallin keskimääräinen absoluuttinen prosentuaalinen ennustevirhe oli noin 1,7 prosentin luokkaa, kun neuroverkkomallilla päästiin noin 2,2 prosentin ennustevirheeseen. Tutkimuksessa käytetyn kulutusdatan huippukulutus oli noin 15 000 megawattia. [33] Elamin ja Fukushige [24] puolestaan tutkivat ARMA-mallin suorituskykyä Japanin kulutusdatan pohjalta. Malliin otettiin mukaan myös eksogeenisiä muuttujia, kuten lämpötila- ja aikamuuttujia, joilla pyrittiin huomioimaan sähkönkulutuksen kausittaista vaihtelua ja lämpötilariippuvuutta. Ottamalla ennustemalliin mukaan eksogeenisiä muuttujia saavutettiin parempia tuloksia, kuin perinteisellä ARMA-mallilla. Eksogeenisiä muuttujia sisältäneen aikasarjamallin tuloksia verrattiin myös Hongin [25] kehittämän regressiopohjaisen ennustemallin tuottamiin tuloksiin. Tulosten vertailun lisäksi tutkimuksessa eksogeenisten muuttujien valinta mukaili Hongin [25] väitöskirjassaan kehittämää systemaattista muuttujien valinnan prosessia. Eksogeenisiä muuttujia sisältänyt ARMA-malli saavutti käytetyllä datalla hieman parempia tuloksia, kuin regressiopohjainen ennustemalli. Molempien mallien keskimääräinen absoluuttinen prosentuaalinen ennustevirhe oli kuitenkin erinomainen, vain noin yhden prosentin luokkaa. Ennusteet laadittiin hyödyntäen Tokyo Electric Power Companyn (TEPCO) kulutusdataa, jossa huippukulutus oli jopa 50 000 megawattia. [24]

3.3.3 Semiparametrinen additiivinen malli

Semiparametrinen additiivinen-malli toimii regressiomallin tavoin, mutta mahdollistaa epälineaarisen riippuvuuden huomioonottamisen selittävän ja selitettävien muuttujien välillä. [26] Kun lineaarisessa regressiossa selittävät muuttujat kuvataan lineaarisilla yhtälöillä, semiparametrissa additiivisessä mallissa lineaariset termit korvataan jollain tuntemattomalla sileällä funktiolla, jota estimoidaan. Lisäksi mallissa virhetermien oletetaan autokorreloivan, eli uudet havainnot riippuvat jollain tavalla edellisistä havainnoista, kun taas lineaarisessa regressiomallissa virhetermien oletetaan olevan toisistaan riippumattomia. [34]

Muun muassa Nedellec et. al. [35] käyttivät semiparametrista additiivista kulutusennustemallia GEFCom2012 -kilpailussa. Kilpailussa käytetty kulutusaineisto oli peräisin Yhdysvalloista, ja aineiston huippukulutus oli noin 3000 megawattia. Tutkimuksessa luotiin kolme komponenttia, yksi pitkälle, keskipitkälle ja lyhyellä aikavälille, ja lopulta komponentit yhdistettiin yhdeksi ennustemalliksi. Semiparametrista additiivista mallia hyödynnettiin pitkän ja keskipitkän aikavälin komponenteissa. Mikäli riittävän tarkka sääennuste on tarjolla, on tutkimuksessa luodulla ennustemallilla mahdollisuus päästä erittäin hyviin tuloksiin, mutta malli on herkkä sääennusteen epätarkkuuksille. [35] Goude et. al. [36] hyödynsivät vastaavaa mallia tutkiessaan sähkönkulusta Ranskassa. Tutkimuksessa

todetaan, että käytetyn ennustamallin tulokset ovat helposti tulkittavissa, mutta myös heidän mallinsa on herkkä tarkalle lämpötilaennusteelle. Ennustettaessa seuraavan päivän sähkönkulutusta yhteensä noin 2000 jakeluverkon sähköasemalle päästiin ennustemallilla noin viiden prosentin keskimääräiseen absoluuttiseen prosentuaaliseen ennustevirheeseen. [36] Hyndman ja Fan [26] ennustivat puolestaan semiparametrisella additiivisella mallilla kulutusta Victorian osavaltiossa Australiassa. Ennustaessaan seuraavan päivän sähkönkulutusta, päästiin heidän mallillaan alle kahden prosentin keskimääräiseen absoluuttiseen prosentuaaliseen ennustevirheeseen. [26]

3.3.4 Eksponentiaalisen tasoituksen malli

Eksponentiaalinen tasoitus antaa kulutuksen historiahavainnoille painoarvon, joka vähenee eksponentiaalisesti yli ajan. Näin ollen eksponentiaalisen tasoituksen mallissa ennusteet ovat painotettuja keskiarvoja menneistä havainnoista, jossa viimeisimmillä havainnoilla painoarvo on suurempi. [32] Mallilla on käytettävän datan suhteen pienemmät vaatimukset verrattuna esimerkiksi usean muuttujan regressiomalliin tai neuroverkkomalleihin, sillä eksponentiaalisessa tasoituksessa selittäviä muuttujia ei käytetä [17]. Koska sähkönkulutus on vahvasti säästäriippuvaista, eikä säähavaintoja hyödynnetä eksponentiaalisessa tasoituksessa, malli soveltuu melko huonosti käytännön sovelluksiin. Tästä kertoo muun muassa se, ettei kansainvälisessä GEFCom2012 -kilpailussa yksikään finalisti käyttänyt kyseistä mallia. [4]

Eksponentiaalinen tasoitus on tutkimusasetelmissä kuitenkin suoriutunut hyvin niiden ennustemallien joukossa, jotka eivät ota huomioon lämpötilaan liittyviä muuttujia. Muun muassa Taylorin ja McSharryn [37] tutkimuksessa eksponentiaalista tasoitusta hyödyntänyt malli saavutti parempia tuloksia kuin ARMA-malli käytettäessä kymmenen Euroopan maan kulutusaineistoa. Ennustettaessa sähkön kulutusta 24 tunnin päähän, eksponentiaalisella tasoituksella kymmenen maan keskimääräisten absoluuttisten prosentuaalisten ennustevirheiden keskiarvo oli noin 1,5 prosenttia, kun ARMA-mallilla ennustevirhe nousi yli kahteen prosenttiin. [37] Tutkimusasetelmaa on kuitenkin kritisoitu siitä, ettei vertailuun otettu mukaan malleja, jotka huomioisivat myös lämpötilan vaikutuksen sähkönkulutukseen, kuten regressiopohjaisia ennustamalleja. [17]

3.3.5 Neuroverkkomallit

Neuroverkkomallit (ANN, Artificial Neural Networks) ovat tekoälypohjainen ennustetekniikka, joille tyypillinen piirre on, ettei ennustemallin luoja tarvitse määritellä mallissa käytettäviä muuttujia. Tämä voi osaltaan helpottaa ennustamisprosessia, sillä sähkönkulutukseen vaikuttavien

tekijöiden tuntemista ei vaadita mallia rakennettaessa. Malli pyrkii oppimaan historiahavaintojen pohjalta sähkönkulutuksen käyttäytymistä muun muassa lämpötilan funktiona, ja sen pohjalta ennustamaan tulevaisuuden kulutusta. Neuroverkot soveltuvat kulutusennustamiseen hyvin kahdesta syystä. Ensinnäkin, neuroverkot pystyvät approksimoimaan numeerisesti kaikkia jatkuvia funktioita ja siten ne pystyvät mallintamaan sähkönkulutuksen monimutkaista käyttäytymistä muun muassa lineaarisia malleja paremmin. Toiseksi, neuroverkkomallit oppivat niille syötetystä datasta, eikä mallin kehittäjän siten tarvitse tehdä oletuksia käytettävän mallin suhteen tai arvioida mallin parametreja. [38]

Neuroverkkomallien hyödyntämistä kulutusennustamisessa on tutkittu aktiivisesti 90-luvulta lähtien [17], ja yhdessä usean muuttujan regressiomallien kanssa neuroverkkomalleja on tutkittu ennustemalleista eniten [39]. Muun muassa Taylor ja Buzza [27] tutkivat neuroverkkomallin suorituskykyä ennustaessaan Englannin ja Walesin sähkönkulutusta. Kulutusaineistossa huippukulutus oli noin 45 000 megawatin luokkaa. Heidän neuroverkkomallinsa pääsi noin 1,7 prosentin ennustevirheeseen, kun kulutusta ennustettiin seuraavalle päivälle. [27] Neuroverkkopohjaista ennustemallia on hyödynnetty myös Pohjois-Amerikassa 35 sähköyhtiön kulutusennusteissa. Valtaosassa kirjallisuudessa tutkituista ennustemalleista lämpötila-muuttujat ovat olleet ainoat säähän liittyvät muuttujat, mutta kyseisessä neuroverkkomallissa myös suhteellinen ilmankosteus ja tuulen nopeus otettiin huomioon. Tutkimuksessa ennustemallia testattiin yhteensä kymmenen sähköyhtiön kulutusaineistoon ja keskimääräinen absoluuttinen prosentuaalinen ennustevirhe oli välillä 1,57-2,72 prosenttia sähköyhtiöstä riippuen. Huomionarvoista on, että käytetyssä neuroverkkomallissa hyödynnettiin kahden mallin yhdistämistä. Toinen malleista ennusti peruskuormaa ja toinen huippukuormaa, ja lopullinen ennuste saavutettiin yhdistämällä mallit. [40]

3.3.6 Sumean logiikan regressiomalli

Sumean logiikan regressiomallit pyrkivät vastaamaan lineaarisen regression ongelmiin, kuten selittävän ja selitettävien muuttujien epämääräiseen suhteeseen, historiahavaintojen riittämättömään määrään sekä hankalasti tulkittaviin virhejakauksiin. Sumean logiikan regressiomalleissa ennustevirheiden ajatellaan riippuvan tarkasteltavan systeemin epämääräisyydestä. Sumean logiikan regressiomallissa regressiokertoimet eivät saa yksittäisiä arvoja, vaan kertoimet muodostavat sumean joukon, jossa kertoimet voivat saada arvoja tietyltä väliltä. [41]

Sumean logiikan regressiomallia ovat tutkineet muun muassa Al-Kandari et. al. [42], jotka kehittivät erikseen ennustemallit kesälle ja talvelle ennustaessaan seuraavan 24 tunnin kulutusta Kanadassa [42]. Hong ja Wang [41] kehittivät Al-Kandar et. al. [42] tutkimuksen pohjalta sumean

logiikan regressiomallia edelleen, lisäten malliin kalenterimuuttujia sekä interaktiotermejä. Mallia sovellettiin New Englandin alueen kulutusaineistoon Pohjois-Amerikassa ja mallilla saavutettiin tarkempia ennustetuloksia verrattuna usean muuttujan regressiomalliin. Sumean logiikan regressiomallilla päästiin 3,7 prosentin keskimääräiseen absoluuttiseen prosentuaaliseen ennustevirheeseen. Huomionarvoista kuitenkin on, että selittävien muuttujien yhdysvaihtelua huomioivia interaktiotermejä sisältänyttä sumean logiikan regressiomallia verrattiin usean muuttujan regressiomalliin, joka ei sisältänyt interaktiotermejä. Vertailtaessa keskenään malleja, jotka sisälsivät samat muuttujat, saavutettiin usean muuttujan regressiomallilla parempia tuloksia. [41]

3.3.7 Tukivektorikonemalli

Tukivektorikoneet ovat ohjatun koneoppimisen menetelmä, joita käytetään usein luokitteluun, hahmontunnistukseen sekä aikasarjojen ennustamisessa regressioanalyysiin. Aikasarjojen ennustamiseen käytetty tukivektoriregressio on metodi, jossa ennustefunktio estimoidaan käyttämällä havaittua dataa, joka puolestaan kouluttaa tukivektorikonetta. Prosessi poikkeaa perinteisistä aikasarjamalleista siten, ettei varsinaista ennustamallia ole, vaan data ohjaa ennustetta. [43]

Tukivektorikonepohjaista ennustemallia ovat hyödyntäneet muun muassa Pai ja Hong [44], jotka vertasivat tukivektorikonemalliaan neuroverkko- ja regressiomalleihin Taiwanin kulutusdatan pohjalta. Malleilla ennustettiin neljän eri maantieteellisen alueen vuosittaista huippukulutusta, ja tukivektorikoneen keskimääräinen absoluuttinen prosentuaalinen ennustevirhe oli eri alueiden ennusteissa välillä 0,75-1,9 prosenttia, kun neuroverkkoon pohjautuvan mallin ennustevirhe oli välillä 1,1-3,6 prosenttia ja regressiomallin 2,5-8,5 prosenttia. Neuroverkko- ja regressiomalli perustuivat samalla kulutusdatalla Hsun ja Chenin [45] tekemään aikaisempaan tutkimukseen. Huomionarvoista on, että aikaisemmassa tutkimuksessa muuttujien valinnat tehtiin vain neuroverkkomallia silmällä pitäen, eikä regressiomallia pyritty kehittämään tarkemmaksi valitsemalla mukaan lisää selittäviä muuttujia. [45]

Chen et. al. [46] loivat puolestaan tukivektorikonemallin, jolla he voittivat EUNITE -ennustuskilpailun vuonna 2001. Kilpailussa tehtävänä oli laatia kulutusennuste kuukaudeksi eteenpäin. Ennusteet tuli laatia annetun kulutusaineiston pohjalta, jonka huippukulutus oli noin 800 megawattia. Tukivektorikoneella päästiin noin 2 prosentin keskimääräiseen absoluuttiseen prosentuaaliseen ennustevirheeseen, kun neuroverkkomallilla ennustevirhe nousi 3,6 prosenttiin. Mielenkiintoinen havainto tutkimuksessa oli myös se, ettei ennustevirheen osalta ollut merkitystä laadittiinko ennuste lämpötilaennusteen vai toteutuneen lämpötilan pohjalta. [46] Vaikka tukivektorikonemallia käytettiin

keskipitkän aikavälin ennustamiseen [46], johti mallin toimivuus EUNITE -kilpailussa tukivektorikonemallien yleistymiseen myös lyhyen aikavälin ennustamisessa [17].

3.3.8 Gradienttitehostamismalli

Tehostaminen on yksi koneoppimisen tehokkaimpia oppimistekniikoita, joita on kehitetty viimeisen 20 vuoden aikana. Vaikka tehostamista on käytetty alunperin luokitteluongelmiin, soveltuu se myös regressio-ongelmien ratkaisemiseen. Gradienttitehostaminen tuottaa ennustemallin yhdistämällä useita ”heikkoja” oppijoita, joilla tarkoitetaan vain hieman arvausta parempaa oppijaa. Gradienttitehostamisessa uusi malli sovitetaan siten, että se pyrkii parantamaan mallia niiden havaintojen osalta, joita edellinen malli ei kyennyt ennustamaan hyvin. [47]

Muun muassa Taieb ja Hyndman [48] hyödynsivät gradienttitehostamista GEFCom 2012 -kilpailuun kehittämässään mallissa, jolla he sijoittuivat kilpailussa viidenneksi. Ennustemallit luotiin jokaiselle päivän tunnille erikseen, ja tehostamista käytettiin malliin sisällytettävien komponenttien valintaan, jolloin kunkin päivän tunnin ennustemalliin saatiin sisällytettyä parhaat muuttujat. Esimerkiksi päivän ensimmäisten tuntien (01:00-05:00) malleissa lämpötilamuuttujat olivat kalenterimuuttujia merkityksellisempiä. [48] Myös Lloyd [49] osallistui GEFCom 2012 -kilpailuun ennustemallilla, joka hyödynsi gradienttitehostamista.

3.4 Kulutusennustemallin valinta

Hong ja Fan [17] vertailivat tutkimuksessaan kahdeksaa yllä kuvailtua ennustemallia. Tutkijat toteavat, että parhaan mallin valinta riippuu sovelluskohteesta, ja ennusteiden tulokset vaihtelevat eri yhtiöiden, sijainnin ja aikajaksojen välillä. Täten yksiselitteistä parasta mallia ei ole olemassa, vaan mallin valinnan tulee pohjautua kunkin yhtiön tarpeisiin. Tutkijat myös huomauttavat, että usein tutkimuksissa käytettyä aineistoa muokataan liikaa ja käytettyjä metodeja ei avata yksityiskohtaisesti, joka johtaa siihen, että saavutettujen tutkimustulosten toisintaminen muulla aineistolla on vaikeaa. [17] Tämä hankaloittaa edelleen parhaan mallin löytämistä, kun tutkimuksista ja käytössä olevasta tietoaineistosta riippuen parhaan tuloksen tuottava ennustemalli vaihtelee.

Kullakin ennustemallilla on kuitenkin ominaisia heikkouksia ja vahvuuksia. Tekoälypohjaiset mallit ovat ns. ”black-box” -malleja, jotka eivät tarjoa lisätietoa sähkönkulutukseen vaikuttavista tekijöistä [17]. Näin ollen tekoälypohjaiset ennustemallit eivät anna mahdollisuutta tulosten tulkitsemiselle. Toisaalta tekoälypohjaisten mallien soveltaminen on yksinkertaista, sillä niiden hyödyntäminen ei vaadi tilastotieteellistä ymmärrystä. Vastakohtana ”black-box”-malleille ovat puolestaan mallit,

jotka tarjoavat lisätietoa sähkönkulutukseen vaikuttavista tekijöistä. Näistä yleisimmin käytetty on usean muuttujan lineaarinen regressiomalli, jossa regressiokertoimet kuvaavat, kuinka suuri vaikutus kullakin selittävällä muuttujalla on sähkönkulutuksen ennusteeseen. Tällaisten mallien heikkoudeksi voidaan mainita muun muassa se, että ennusten laatijan tulee tehdä valintoja esimerkiksi käytettävien muuttujien suhteen [17].

Lisäksi saatavilla oleva tietoaaineisto määrittää osaltaan ennustamallin valitsemista. Mikäli saatavilla ei ole tarkkoja ja luotettavia säähavaintoja, tulee käytettävän ennustamallin olla yhteen muuttujaan, eli sähkönkulutukseen perustuva malli. Tyypillisimpiä yhteen muuttujaan perustuvia malleista ovat eksponentiaalinen tasoitus sekä aikasarjamallit. Useaan muuttujaan pohjautuvat mallit, kuten neuroverkko- ja usean muuttujan lineaarinen regressiomalli puolestaan voivat hyödyntää käytössä olevaa säädettä, ja koska sähkönkulutus on tyypillisesti vahvasti säästäriippuvaista, näillä malleilla saavutetaan yleensä parempia tuloksia. [17]

Koska tämän diplomityön tavoitteena ei ole yksinomaan löytää parasta ennustemallia ennustevirheen minimoimiseksi, vaan tavoitteena on myös ennustemallin tulosten tulkitseminen, on perusteltua valita käytettäväksi malliksi sellainen, joka mahdollistaa tämän. Näin ollen tekoälypohjaisten mallien käyttö ei palvele tämän diplomityön tarkoituksia. Usean muuttujan lineaarinen regressiomalli on yleisimmin käytetty ennustemalli, joka tarjoaa mahdollisuuden tulosten tulkitsemiseen. Lisäksi usean muuttujan regressiomalli on laskennallisesti yksinkertainen ja tehokas toteuttaa, joten tässä diplomityössä sovellettavaksi ennustemenetelmäksi valikoitui usean muuttujan lineaarinen regressiomalli.

4 Tutkimusaineisto ja -metodit

Tässä luvussa käsitellään tämän diplomityön metodologiset valinnat ja niiden rajoitteet, sekä kuvaillaan käytettyä tutkimusaineistoa. Lisäksi luku keskittyy usean muuttujan lineaarisen regressiomallin luomiseen.

4.1 Käytetty aineisto

Käytetty tutkimusaineisto kattaa Vaasan Sähkön Vaasan Sähköverkon alueella olevien asiakkaiden tunnittaisen kulutuksen vuosilta 2019-2021. Tutkimusaineiston rajaaminen Vaasan alueeseen yksinkertaistaa suoritettavia laskelmia, sillä muun muassa lämpötilan huomioimiseksi ennustemallissa riittää, että laskelmissa käytetään vain Vaasan lämpötilahavaintoja ja -ennusteita. Tutkimusaineiston rajaaminen Vaasan Sähköverkon alueeseen on perusteltua, sillä kyseisen alueen asiakkaat kattavat noin puolet koko Vaasan Sähkön sähkönmyynnistä. Käytetyt tutkimusmenetelmät ovat kuitenkin sovellettavissa myös muiden alueiden kuluttajiin. Tässä diplomityössä käytetyn kulutusaineiston huippukulutus oli noin 160 megawattia, joka on merkittävästi pienempi kuin luvun 3 kirjallisuuskatsauksessa läpi käydyissä tutkimuksissa.

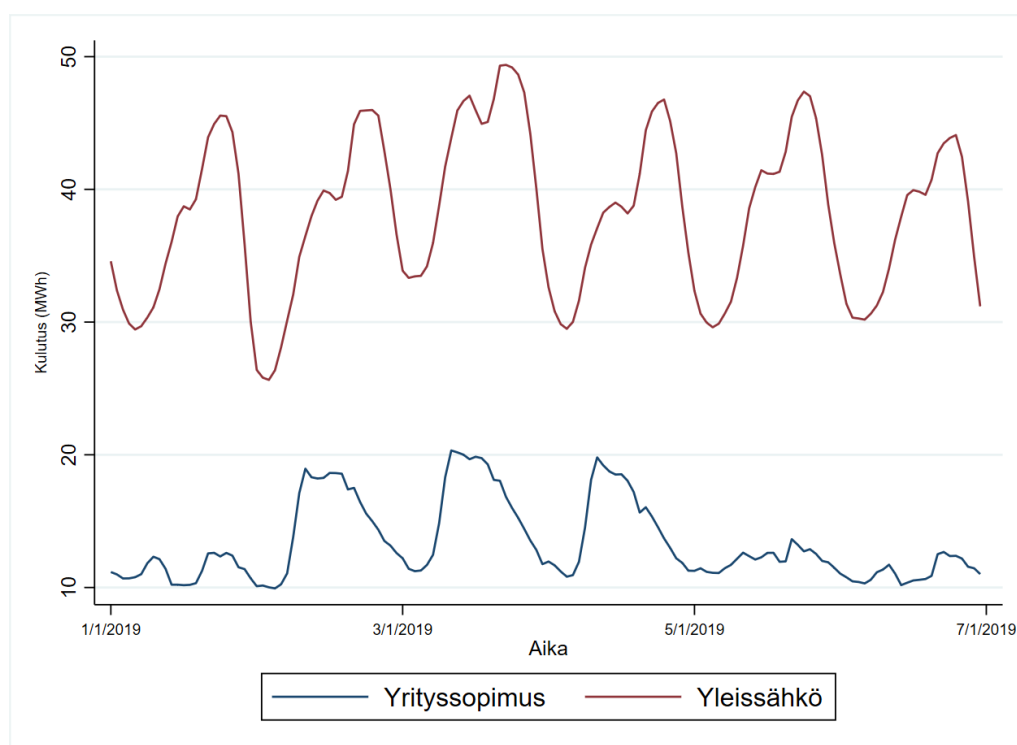
Kaikilla kuluttajilla, joilla on sähkönmyyntisopimus Vaasan Sähkön kanssa sekä omaa pientuotantoaan, tulee solmia pientuotannon ostosopimus yhtiön kanssa. Sopimus velvoittaa yhtiötä ostamaan asiakkaidensa ylijäämätuotannon tilanteessa, jossa asiakkaan oma pientuotanto ylittää sen hetkisen sähkönkulutuksen. Yhtiön sopimustietojen avulla kulutusdatasta saatiin eroteltua niiden asiakkaiden sähkönkulutus, joilla on vuosien 2019-2021 aikana ollut omaa pientuotantoa. Tyypillisesti oma pientuotanto on aurinkosähköä. Näin ollen voidaan tutkia, päästäänkö ennustevirheen osalta parempaan lopputulokseen, mikäli pientuottaja-asiakkaiden kulutusta ennustetaan erillään muista asiakkaista.

Lisäksi kulutusdata jaettiin segmentteihin sopimustyyppin mukaan. Tarkasteltavat sopimustyypit olivat:

- Yrityisasiakkaiden sopimukset
- Yksityisasiakkaiden yleissähkösopimukset
- Yksityisasiakkaiden yösähkösopimukset
- Yksityisasiakkaiden kausisähkösopimukset

Eri sopimukset hinnoitellaan sähkøyhtiössä eri tavoin, joten erilaisten asiakkaiden olisi optimaalista valita heidän kulutukseensa sopivin sopimustyyppi. Esimerkiksi yösähkösopimuksessa sähkön hinta on matalampi yöaikaan, jolloin varaavan sähkölämmityksen omistavan asiakkaan kannattaisi valita yösähkösopimus, ja hyödyntää yön

edullisempaa sähkön hintaa asuntonsa lämmittämisessä. Lisäksi muun muassa yritysasiakkaiden kulutusprofiilit poikkeavat merkittävästi yksityisasiakkaista, joten yritysasiakkaiden sopimukset otettiin omaksi segmentikseen. Kuva 4 havainnollistaa yritysasiakkaiden ja suosituimman yksityisasiakkaiden sopimustyyppin, yleissähkön, poikkeavaa kulutuskäyttäytymistä vuoden 2019 ensimmäiseltä viikolta.



Kuva 4: Yritysasiakkaiden ja yleissähköasiakkaiden tunnitainen sähkönkulutus tammikuun ensimmäisellä viikolla 2019.

Kuten kuvasta 4 nähdään, uudenvuodenpäivän ja loppiaisen vaikutus yritysasiakkaiden sähkönkulutukseen on merkittävä. Yritysasiakkaiden kulutus putoaa pyhäpäivinä huomattavasti arkipäivien kulutuksesta, kun taas yleissähkösopimusasiakkaiden kulutukseen pyhäpäivien vaikutus on pienempi ja kulutusprofiili käyttäytyy tasaisemmin läpi viikon. Jakamalla asiakkaat segmentteihin, voidaan tarkastella saavutetaanko parempi lopputulos ennustevirheen osalta, mikäli kulutusennusteet luodaan jokaiselle segmentille erikseen. Ennustamalla sähkönkulutusta segmenteittäin, voidaan ennusteprosessissa huomioida eri segmenttien kulutukselle ominaisia piirteitä paremmin, ja siten tavoitella ennustevirheen osalta parempaa lopputulosta.

Vertailu segmenteittäin luodun ja aggregoidun ennusteen välillä tehdään kahdessa osassa. Ensin pientuottaja-asiakkaiden ja muiden asiakkaiden kulutusennusteita verrataan kokonaiskulutuksen ennusteeseen. Toisessa

osassa verrataan sopimustyypeittäin laadittuja ennusteita kokonaiskulutuksen ennusteeseen.

4.2 Kulutusennustemallin luominen

Usean muuttujan lineaarisen regressiomallin toimivuuden kannalta olennainen osa on löytää selittävät muuttujat, joilla selitettävää muuttujaa eli sähkönkulutusta pyritään ennustamaan. Hong [25] kehitti väitöskirjassaan kulutusennustamisen benchmark-mallin, johon selittävät muuttujat valittiin tutkimalla kulutusdatan käyttäytymistä lämpötilan ja kalenterimuuttujien funktiona. Benchmark-mallia kehitettiin edelleen ottamalla malliin mukaan yksitellen lisää selittäviä muuttujia, jonka jälkeen ennustevirhettä tarkastelemalla valittiin paras ennustemalli. Lopputuloksena syntyi useita eri muuttujia sisältäviä ennustemalleja. [25] Vastaavaa prosessia sovellettiin myös tässä diplomityössä. Tutkimalla vuosien 2019-2020 kulutusaineistoa valittiin benchmark-malliin mukaan otettavat muuttujat, jonka jälkeen benchmark-malliin lisättiin selittäviä muuttujia yksitellen Hongin [25] kehittämän prosessin mukaisesti.

4.2.1 Usean muuttujan lineaarisen regressiomallin muuttujat

Usean muuttujan lineaarinen regressiomalli voidaan kuvata yhtälöllä:

$$Y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \dots \beta_{p-1} x_{i,p-1} + \epsilon_i \quad (4)$$

jossa $\beta_0 \dots \beta_{p-1}$ ovat regressiokertoimet, $x_{i1}, \dots, x_{i,p-1}$ ovat tunnettuja vakioita ja ϵ_i on normaalijakautunut virhetermi, jonka keskiarvo on 0 [25]. Tyypillisesti mallin parametrit estimoidaan pienimmän neliösumman menetelmällä, jolloin virhetermin neliöiden summaa minimoidaan [17]. Vastefunktio, jolla selitettävän muuttujan Y ennustettu arvo $E[Y]$ lasketaan, voidaan kirjoittaa:

$$E[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \beta_{p-1} x_{p-1} \quad (5)$$

jossa x_1, \dots, x_{p-1} ovat $p - 1$ selittävää muuttujaa. [25] Regressiokerroin β_0 on vakiotermin, joka kuvaa kuinka suuri sähkönkulutus olisi, mikäli kaikki selittävät muuttujat olisivat nollia. Regressiokertoimien $\beta_1, \dots, \beta_{p-1}$ arvoja voidaan tulkita seuraavasti. Mikäli esimerkiksi selittävän muuttujan x_1 arvoa kasvatetaan yhdellä yksiköllä, nousee selitettävän muuttujan arvo β_1 yksikköä, mikäli muut termit pysyvät samana. [50]

Usean muuttujan regressioanalyysissä selitettävää muuttujaa, eli tämän diplomityön tapauksessa tunnittaista sähkönkulutusta, voidaan selittää useilla erilaisilla muuttujilla. Usein regressioanalyysissä selittävät muuttujat

ovat määrällisiä, eli kvantitatiivisia muuttujia. Sähkönkulutusta ennustettaessa muun muassa lämpötila on kvantitatiivinen muuttuja, joka saa numeerisia arvoja. Selittävät muuttujat voivat olla myös laadullisia eli kvalitatiivisia muuttujia, joita kutsutaan myös kategorisiksi- tai dummy-muuttujiksi. Esimerkki kvalitatiivisesta muuttujasta on viikonpäivä, joka voi saada arvoja maanantaista sunnuntaihin. Toisin sanoen, viikonpäivä on kategorinen muuttuja, joka voi saada seitsemän eri luokkaa. Näin ollen viikonpäivää voidaan kuvata dummy-muuttujilla, joita luodaan 6, yksi vähemmän kun luokkien lukumäärä. Viikonpäivät maanantaista sunnuntaihin voidaan nimetään muuttujiksi X_1, \dots, X_7 , jolloin kvalitatiivinen viikonpäivä-muuttuja voidaan esittää seuraavasti:

$$\left\{ \begin{array}{l} X_1 = 1, \text{ jos viikonpäivä on maanantai} \\ \quad X_1 = 0, \text{ muutoin} \\ X_2 = 1, \text{ jos viikonpäivä on tiistai} \\ \quad X_2 = 0, \text{ muutoin} \\ \quad \dots \\ X_6 = 1, \text{ jos viikonpäivä on lauantai} \\ \quad X_6 = 0, \text{ muutoin} \end{array} \right. \quad (6)$$

Mikäli muuttujat X_1, \dots, X_6 saavat arvon nolla, on kyseessä sunnuntai. [25]

Vaikka kyseessä on lineaarinen regressiomalli, voidaan malliin sisällyttää myös epälineaarisia polynomitermejä. Mikäli sähkönkulutusta selittävinä muuttujina ovat esimerkiksi lämpötila T ja lämpötilan neliö T^2 , voidaan malli kirjoittaa:

$$Y = \beta_0 + \beta_1 T + \beta_2 T^2 + \epsilon \quad (7)$$

Kyseessä on vain kaavan 4 erikoistapaus, koska voidaan merkitä $T = x_1$ ja $T^2 = x_2$, jolloin yhtälö vastaa kaavan 4 mukaista yhtälöä.

Mikäli selittävien muuttujien arvot riippuvat jonkin toisen selittävän muuttujan arvoista, voidaan regressiomalliin lisätä niin kutsuttuja interaktiotermiä, jotka ovat kahden tai useamman selittävän muuttujan tulo muodostamia muuttujia. [25] Esimerkiksi kahden selittävän muuttujan x_1 ja x_2 , sekä muuttujien välisen interaktiotermin $x_1 x_2$ muodostama malli voidaan kirjoittaa :

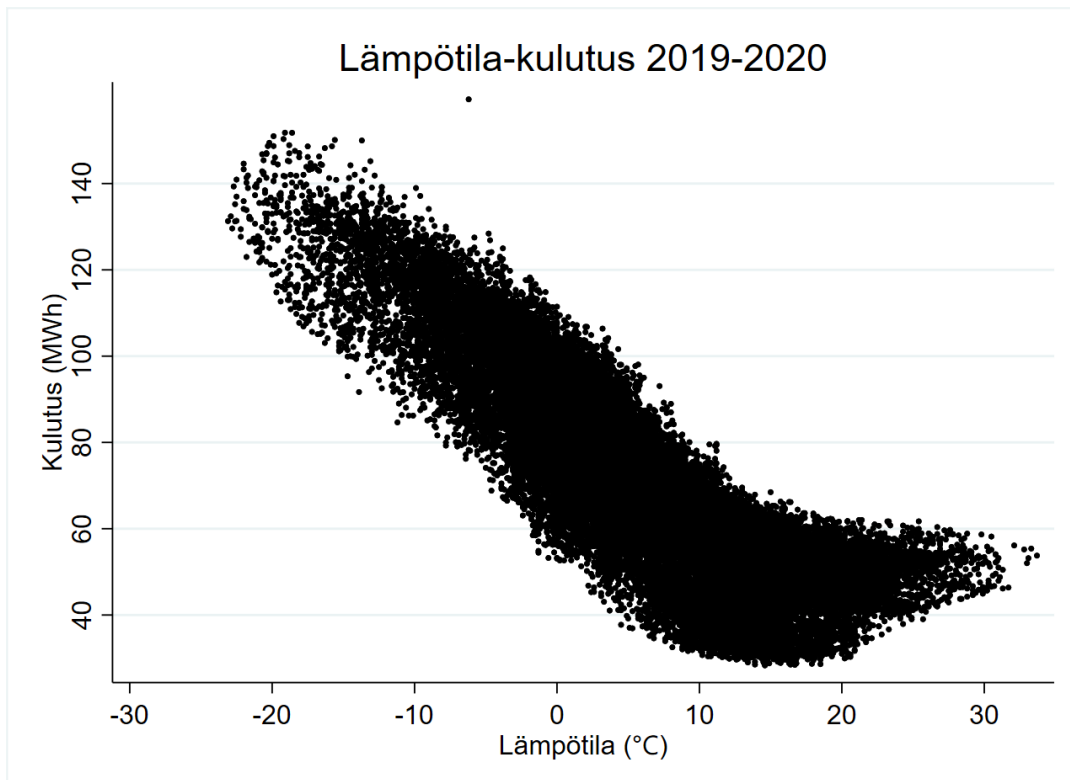
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \quad (8)$$

Jälleen voidaan kirjoittaa $x_1 x_2 = x_3$, jolloin yhtälö vastaa kaavan 4 mukaista lineaarista regressiomallia. Esimerkki interaktiostermistä on päivän tunnin ja viikonpäivän välinen yhteys. Eri viikonpäivinä, kuten viikonloppuna ja arkipäivinä, samaan kellonaikaan sähkönkulutus voi poiketa toisistaan, jolloin ennustemalliin tulee sisällyttää päivän tunnin ja

viikonpäivän välinen interaktiotermit. Interaktiotermit siis kuvaavat kahden muuttujan välistä yhdysvaihtelua.

4.2.2 Benchmark-mallin luominen

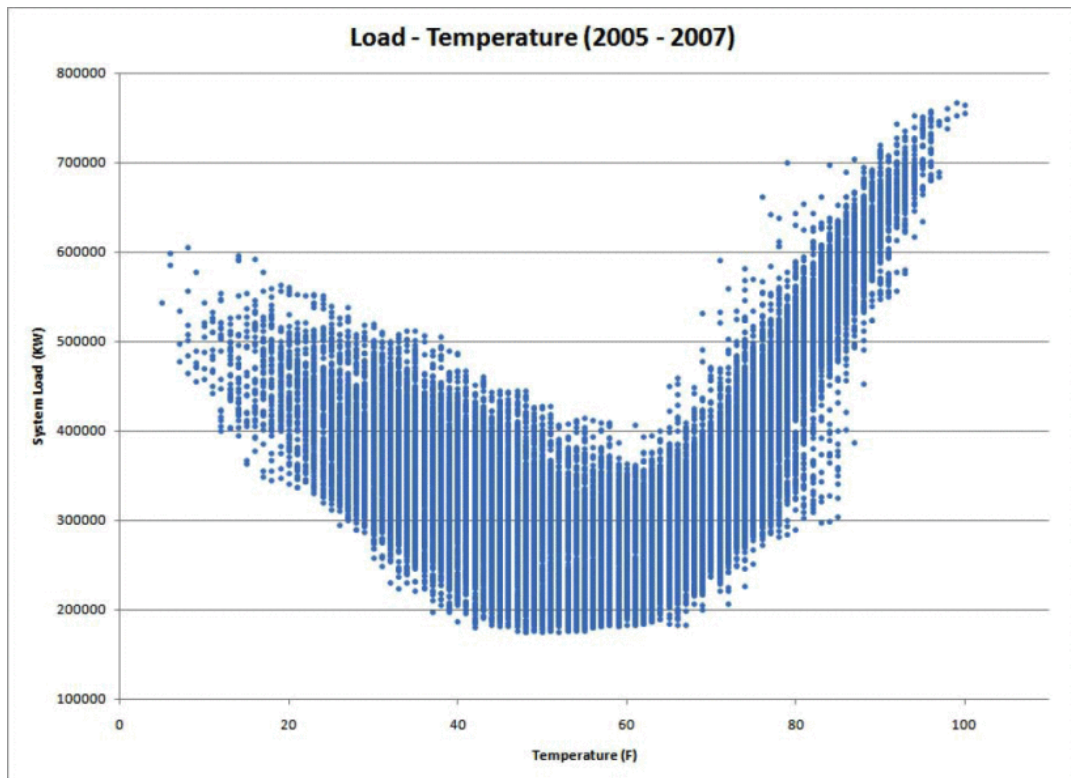
Kuva 5 havainnollistaa sähkönkulutuksen lämpötilariippuvuutta pistekuvion avulla. Kukin kuvan piste kuvaa tunnittaista sähkönkulutusta ja sen tunnin aikaista keskilämpötilaa Vaasassa. Kuten kuvasta nähdään, sähkönkulutus on vahvasti lämpötilariippuvaista. Kulutus on suurimmillaan lämpötilojen ollessa alhaisia, kun taas lämpötilan ollessa korkeampi on kulutus huomattavasti vähäisempää. Koska kulutuksen ja lämpötilan välinen suhde vaikuttaa kuvan perusteella käyttäytyvän toiseen asteen polynomin tavoin, otettiin ennustemalliin selittäväksi muuttujiksi sekä tunnittainen lämpötila T , että tunnittaisen lämpötilan neliö T^2 .



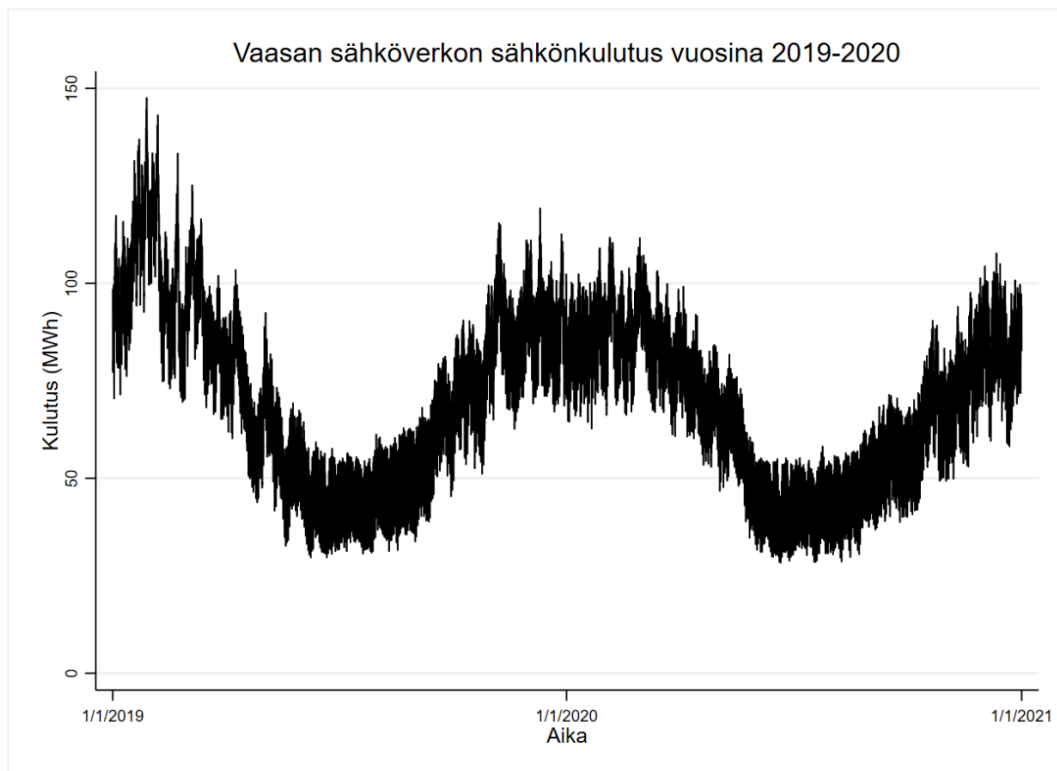
Kuva 5: Lämpötila-kulutus pistekuvio vuosilta 2019-2020.

Kuva 6 havainnollistaa puolestaan lämpötila-kulutus pistekuviota Hongin [25] tutkimuksessaan käyttämästä kulutusdatasta Yhdysvalloista. Poiketen tämän diplomityön kulutusdatasta, kuvan 6 pistekuviossa sähkönkulutus on suurta myös korkeilla lämpötiloilla. Kuvien 5 ja 6 poikkeavuus selittynee osaltaan ilmastoinnilla, sillä Yhdysvalloissa kesäisin sähkönkulutus on korkea asuntojen huomattavan viilentämistarpeen takia [25]. Kuvan 6 perusteella sähkönkulutus näyttäisi käyttäytyvän kolmannen asteen

funktiona lämpötilan suhteen, jonka takia Hongin väitöskirjassa [25] ennustemalliin sisällytettiin myös lämpötilan kuutiollinen termi T^3 . Koska kuvan 5 perusteella tämän diplomityön kulutusaineiston sähkönkulutus taas näyttäisi käyttäytyvän toisen asteen funktiona lämpötilan suhteen, oli tässä diplomityössä käytettäviin ennustemalleihin riittävää ottaa mukaan lämpötilan neliöllinen termi T^2 .



Kuva 6. Lämpötila-kulutus pistekuvio Yhdysvalloissa. [25]

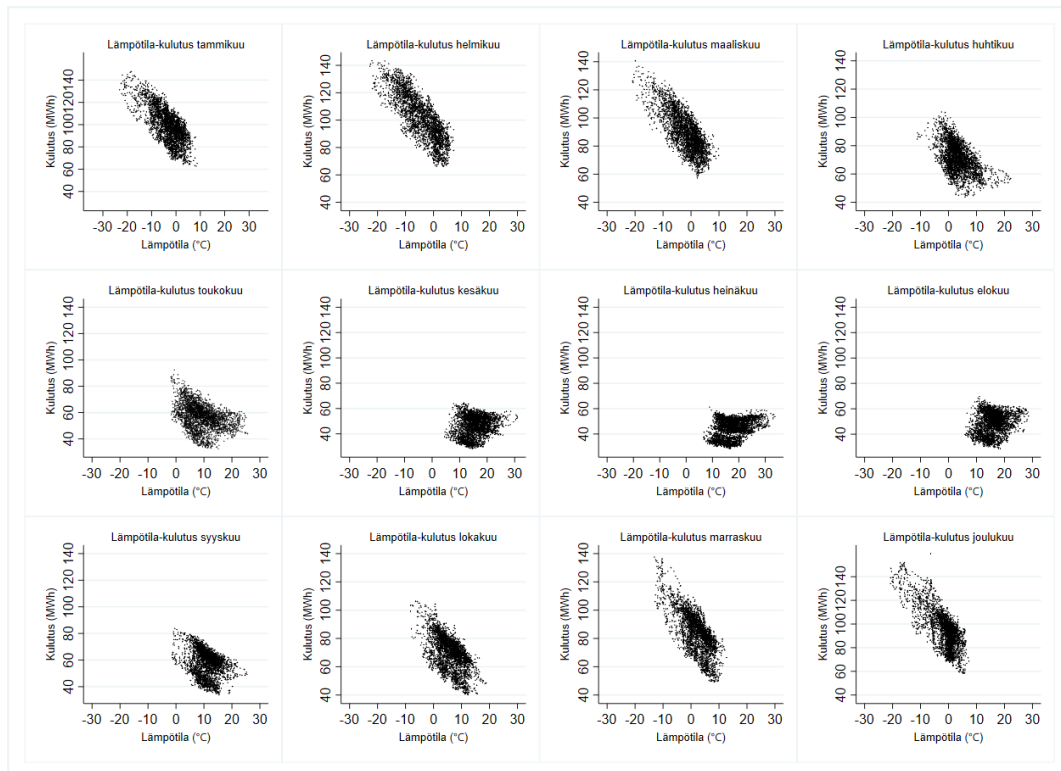


Kuva 7. Vaasan Sähköverkon sähkönkulutus vuosina 2019-2020.

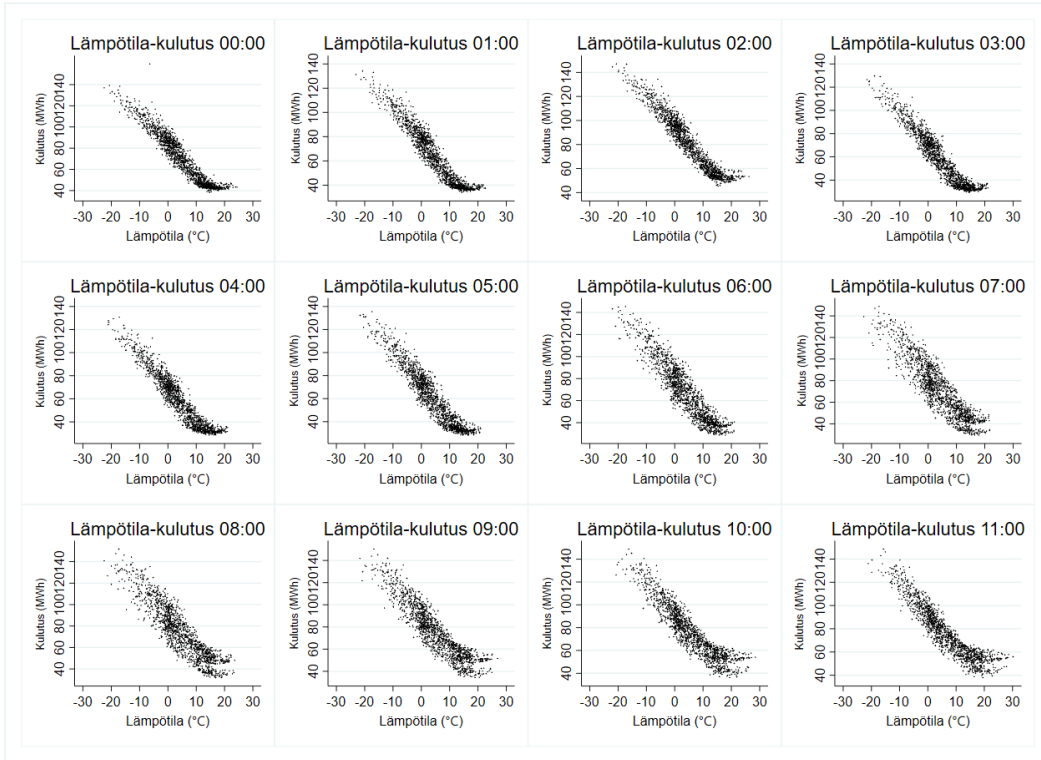
Kuva 7 havainnollistaa sähkönkulutuksen kausittaista vaihtelua. Kuvasta näkyy selkeästi talvikuukausien huomattavasti korkeampi kulutus kesäkuukausiin verrattuna. Vuodenajan aiheuttaman vaihtelun lisäksi sähkönkulutuksessa on tyypillisesti vaihtelua päivän tuntien ja viikonpäivien välillä [25]. Täten benchmark-malliin otettiin mukaan kolme kategorista kalenterimuuttujaa: *Tunti*, *Viikonpäivä* ja *Kuukausi*. Nämä kategoriset muuttujat saavat 24, 7 ja 12 luokkaa, jotka kuvaavat päivän 24 tuntia, seitsemää viikonpäivää ja kahtatoista kuukautta. Näin ollen ennustemallin regressiokertoimia estimoitaessa jokaiselle päivän tunnille, viikonpäivälle ja kuukaudelle saadaan oma kerroin, joiden avulla kulutusta voidaan ennustaa. Lisäksi malliin lisättiin Hongin [25] tavoin muuttuja *Trendi*, joka saa arvon 1 tarkasteltavan ajanjakson ensimmäiselle alkavalle tunnille (1.1.2019 00:00), arvon 2 toiselle tunnille ja niin edelleen. *Trendi*-muuttuja pyrkii huomioimaan sähkönkulutuksen trendinomaisen kasvun [25].

Kuva 8 puolestaan esittää sähkönkulutuksen ja lämpötilan välistä riippuvuutta kuukausitasolla. Kuvasta huomataan, että sama lämpötila eri kuukausina johtaa erisuuruiseen sähkönkulutukseen. Esimerkiksi lämpötilan ollessa 0°C , kulutus vaihtelee tammikuussa noin 60-110 megawattitunnin välillä, kun taas huhtikuussa samassa lämpötilassa sähkönkulutus vaihtelee noin 50-90 megawattitunnin välillä. Tämä tarkoittaa, että lämpötilamuuttujan T vaikutus selitettävään muuttujaan, eli sähkönkulutukseen, riippuu muuttujan *Kuukausi* arvoista. Jotta tätä

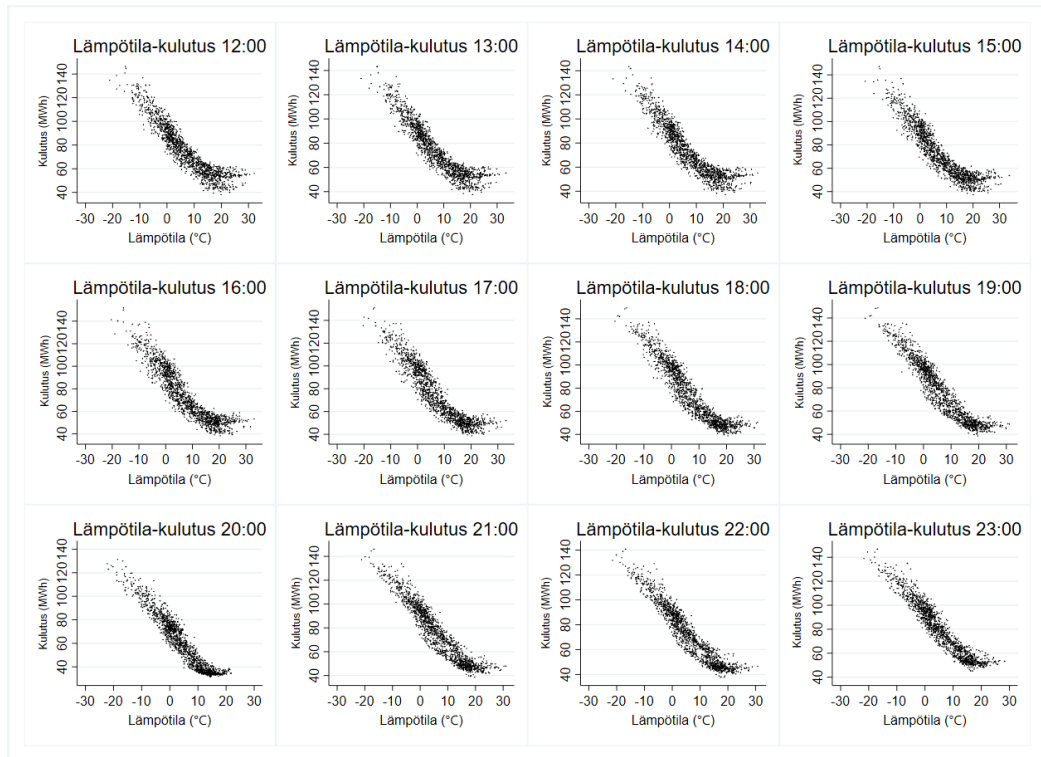
yhdysvaihtelua voidaan huomioida kulutusennusteessa, sisällytettiin regressiomalliin interaktiotermit $T^*Kuukausi$ ja $T^2*Kuukausi$. Interaktiotermi mahdollistaa lämpötilaan liittyvien regressiokertoimien vaihtelun kuukausien mukaan. Yhdysvaihtelua voidaan selittää ihmisten erilaisella kulutuskäyttäytymisellä eri kuukausina [25], [24].



Kuva 8. Kuukausittaiset lämpötila-kulutus pistekuviot.



Kuva 9. Tuntien 00:00-11.00 lämpötila-kulutus pistekuviot.



Kuva 10. Tuntien 12:00-23:00 lämpötila-kulutus pistekuviot.

Vastaavat lämpötila-kulutus pistekuviot jokaiselle päivän tunnille on esitetty kuvissa 9 ja 10. Yhtäläillä kun kuukausien kohdalla, sähkönkulutuksen lämpötilariippuvuus näyttäisi poikkeavan päivän tuntien välillä. Esimerkiksi päivän ensimmäisten tuntien 00:00-02:00 aikana lämpötilan ollessa 20°C sähkönkulutusta kuvaavat pisteet ovat hyvin pienellä alueella noin 40 megawattitunnin kohdalla, kun taas samassa lämpötilassa kello 08:00 pisteet hajaantuvat laajemmalla alueelle noin 30-50 megawattitunnin kohdalle. Myös eri tuntien erilainen sähkönkulutus selittyy ihmisten kulutuskäyttäytymisellä [25], ja jotta tätä yhdysvaihtelua voidaan huomioida ennustemallissa, lisättiin benchmark-malliin interaktiotermit $T * Tunti$ ja $T^2 * Tunti$.

Lisäksi päivän tunnint eri viikonpäivinä saattavat johtaa erilaiseen kulutuskäyttäytymiseen. Esimerkiksi viikonloppuaamujen kulutus on tyypillisesti pienempää kuin arkipäiväaamujen. Täten interaktiotermit päivän tunnin ja viikonpäivän otettiin mukaan benchmarking-malliin. Huomionarvoista on, että kun kategorisen muuttujan (esim. *Kuukausi*) ja kvantitatiivisen muuttujan (esim. lämpötila T) interaktiotermit lisätään malliin, voidaan kyseinen kvantitatiivinen muuttuja jättää mallista pois. Mikäli puolestaan kahden kategorisen muuttujan välinen interaktiotermit sisällytetään malliin, molemmat kategoriset muuttujat voidaan jättää pois mallista. [25] Näin ollen benchmarking-malli sisältää seuraavat selittävät muuttujat:

- *Kuukausi* (kategorinen muuttuja)
- $Tunti * T$ (interaktio)
- $Tunti * T^2$ (interaktio)
- $Kuukausi * T$ (interaktio)
- $Kuukausi * T^2$ (interaktio)
- $Päivä * Tunti$ (interaktio)
- *Trendi* (kvantitatiivinen muuttuja)

Benchmark-malli voidaan yhtälömuodossa kirjoittaa:

$$Y = \beta_0 + \beta_1 Kuukausi + \beta_2 Tunti * T + \beta_3 Tunti * T^2 + \beta_4 Kuukausi * T + \beta_5 Kuukausi * T^2 + \beta_6 Päivä * Tunti + \beta_7 Trendi \quad (9)$$

Kun regressiokertoimet estimoitiin ensin vuosien 2019-2020 aineiston pohjalta, voitiin niiden avulla laskea selitettävän muuttujan arvoja vuodelle 2021. Käytännön sovelluksissa, kun kulutusennustetta laaditaan seuraavalle päivälle, käytetään ennustetta laskettaessa lämpötilamuuttujan arvona

lämpötilaennusteita. Näin ollen, mikäli seuraavan vuorokauden tunnitaiset kulutusennusteet laaditaan kello 10:00 mennessä edellisenä päivänä, tulee käytössä olla 36 tunnin päähän ulottuvat tunnitaiset lämpötilaennusteet. Tässä diplomityössä käytetyt lämpötilaennusteet olivat päivittäin kello 10:00 talteen otettuja lämpötilaennusteita seuraavan päivän tunneille. Näin ollen, vaikka ennustemalleja ei päivitetä jokaisen ennusteen laatimisen jälkeen, kuvastaa laaditut ennusteet seuraavan päivän day-ahead kulutusennustamista, koska ennustamisessa käytetyt lämpötilaennusteet vastaavat käytännön tilannetta.

Myös pyhäpäivien vaikutus sähkönkulutukseen tulee huomioida ennustemalleissa. Tyypillisesti pyhäpäivinä kulutus poikkeaa normaalista, sillä ihmiset eivät mene töihin, ja muun muassa koulut ja päiväkodit ovat suljettuina. Tältä osin pyhäpäivät muistuttavat sähkönkulutukseltaan viikonloppuja. [25], [51] Ennustemallissa pyhäpäivien vaikutusta huomioitiin luomalla dummy-muuttuja *Pyhä*, joka sai arvon 1 mikäli kyseessä on pyhäpäivä ja arvon 0 muuten. Näin ollen tehdään oletus, että jokaisen pyhäpäivän vaikutus sähkönkulutukseen on samankaltainen. Tämä on yksinkertaistava oletus, sillä esimerkiksi arkipäivälle sattuvan pyhäpäivän vaikutus kulutukseen voi riippua siitä, mille arkipäivälle pyhäpäivä sattuu. Lisäksi toiset pyhäpäivät vaikuttavat kulutuskäyttämiseen enemmän kuin toiset. Esimerkiksi aina lauantaisin vietettävä pyhäinpäivä vaikuttanee kulutukseen eri tavoin kun vuosittain eri viikonpäiville sattuva vappu. Ennustemalleihin otettiin mukaan seuraavat viralliset sekä epäviralliset pyhäpäivät:

- Uudenvuodenpäivä
- Loppiainen
- Pitkäperjantai
- Pääsiäispäivä
- 2. pääsiäispäivä
- Vappu
- Äitienpäivä
- Helatorstai
- Helluntai
- Juhannusaatto
- Juhannuspäivä
- Pyhäinpäivä
- Isänpäivä
- Itsenäisyyspäivä
- Jouluaatto
- Joulupäivä
- Tapaninpäivä

Kun *Pyhä*-muuttuja lisättiin benchmark-malliin, voidaan lopullinen benchmark-malli kirjoittaa yhtälömuodossa:

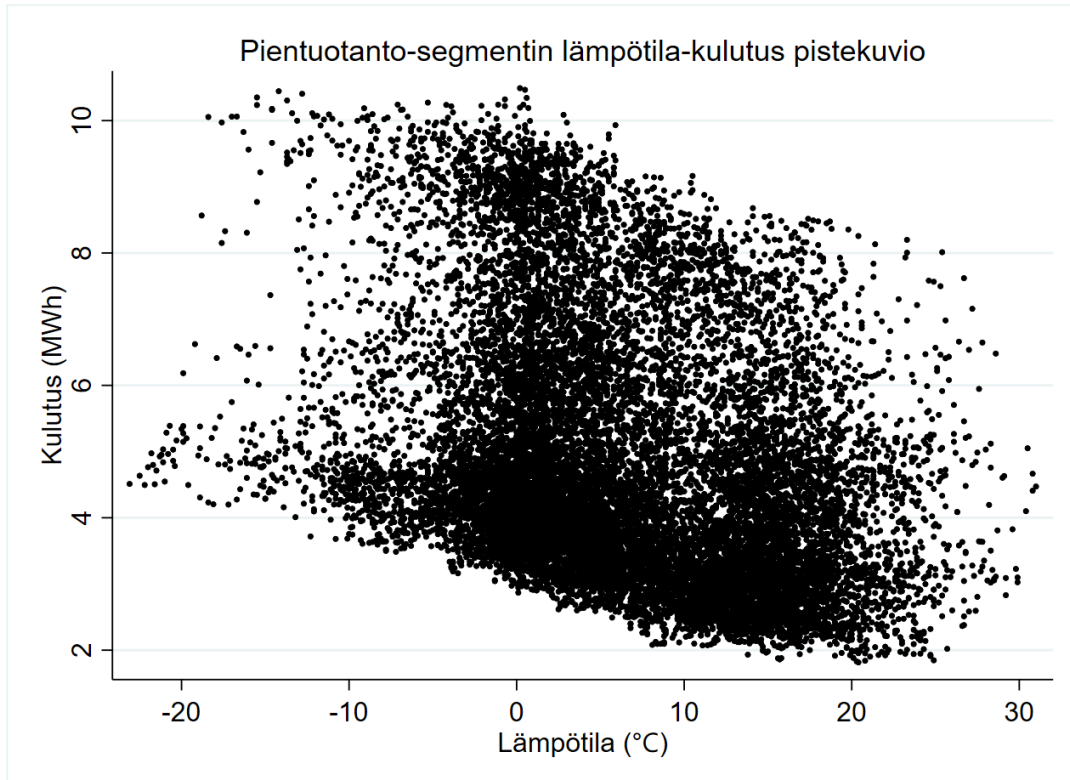
$$Y = \beta_0 + \beta_1 \text{Kuukausi} + \beta_2 \text{Tunti} * T + \beta_3 \text{Tunti} * T^2 + \beta_4 \text{Kuukausi} * T + \beta_5 \text{Kuukausi} * T^2 + \beta_6 \text{Päivä} * \text{Tunti} + \beta_7 \text{Trendi} + \beta_8 \text{Pyhä} \quad (10)$$

Benchmark-malliin, joka sisältää muuttujan *Pyhä*, viitataan jäljempänä nimellä UMR-B. Kaavan 10 mukaista benchmark-mallia käytetään kaikkien muiden kulutussarjojen, paitsi pientuottaja-segmentin ennustamiseen, jolle luotiin oma benchmark-malli.

4.2.3 Pientuottaja-segmentin benchmark-malli

Koska pientuottajien tuotantomuoto on tyypillisesti aurinkosähköä ja aurinkosähkön tuotannon määrä on vahvasti riippuvainen auringon säteilystä [52], sisällytettiin pientuottaja-segmentin ennustemalliin auringon kokonaissäteilyä kuvaava muuttuja *Aurinko*. Muuttujan arvot olivat mitattuja kokonaissäteilymääriä (W/m²) lähimpänä Vaasaa sijaitsevalta kokonaissäteilyä mittaavalta Ilmatieteenlaitoksen mittausasemalta Seinäjoelta [53]. Kuva 11 esittää pientuottaja-asiakkaiden lämpötila-kulutus pistekuvion. Verrattuna kokonaiskulutuksen lämpötila-kulutus pistekuvioon, pientuottajien kohdalla kulutuksen ja lämpötilan välisessä riippuvuudessa ei ole havaittavissa selkää epälineaarisuutta, joten ennustemalleja testattiin sekä sisällyttämällä lämpötilamuuttuja *T*, että lämpötilan neliöllinen muuttuja *T*² ennustemalleihin. Sisällyttämällä vain muuttuja *T* saavutettiin parempia lopputuloksia, joten muuttuja *T*² jätettiin pois ennustemalleista. Muilta osin pientuottajasegmentin ennustemalli ja sen kehittäminen vastasi aikaisemmin luvussa 4.2.1 kuvattua prosessia. Pientuottajien sähkönkulutusta ennustava benchmark-malli voidaan yhtälömuodossa kirjoittaa:

$$Y = \beta_0 + \beta_1 \text{Kuukausi} + \beta_2 \text{Tunti} * T + \beta_3 \text{Kuukausi} * T + \beta_4 \text{Päivä} * \text{Tunti} + \beta_5 \text{Aurinko} + \beta_6 \text{Pyhä} \quad (11)$$



Kuva 11: Pientuotanto-segmentin lämpötila-kulutus pistekuvio.

4.2.4 Benchmark-mallin kehittäminen

Benchmarking-mallia lähdettiin kehittämään Hongin [25] väitöskirjan tavoin, jossa regressiomalliin lisättiin yksitellen selittäviä muuttujia ja ennusteen tarkentumista tutkittiin ennustevirheen avulla. Benchmark-mallissa oletus oli, että jokaisen tunnin kulutus on riippuvainen vain kyseisen tunnin lämpötilasta. Kuitenkin on perusteltua olettaa, että myös tarkasteltavaa tuntia edeltävien tuntien lämpötiloilla on vaikutusta sähkön kulutukseen [25], [30]. Edeltävien tuntien lämpötiloja otettiin ennustemallissa huomioon seuraavan prosessin mukaisesti [25]:

1. Lisätään selittäväksi muuttujaksi tarkasteltavaa tuntia edeltävän 24 tunnin yksinkertainen liukuva keskiarvo. Yksinkertainen liukuva keskiarvo lasketaan:

$$T_{avg}(t) = \sum_{k=1}^{24} \frac{T(t-k)}{24} \quad (12)$$

Nimetään yksinkertaisen liukuvan keskiarvon sisältämä malli UMR-o:ksi. Mikäli yksinkertaisen liukuvan keskiarvon lisääminen malliin ei tuota tarkempia tuloksia, jatketaan seuraavan vaiheeseen benchmark-mallilla.

2. Lisätään edellisessä vaiheessa valittuun malliin yksitellen edeltävien tuntien lämpötiloja, aloittaen edeltävän tunnin lämpötilasta. Edeltävien tuntien lämpötiloja lisätään niin kauan kun ennustevirhe paranee, tai kun kolmen edeltävän tunnin lämpötilat on lisätty. Näin saadaan mallit UMR-1, UMR-2 ja UMR-3, jossa numerot 1-3 viittaavat siihen, kuinka monen edeltävän tunnin lämpötilat on otettu malliin mukaan. Valitaan näistä kolmesta mallista paras, ja siirrytään seuraavaan vaiheeseen.
3. Korvataan kohdan 1 yksinkertainen liukuva keskiarvo painotetulla liukuvalla keskiarvolla. Edellisten 24 tunnin lämpötilojen painotettu liukuva keskiarvo saadaan:

$$T_w(t) = \frac{\sum_{k=1}^{24} \alpha^{k-1} T(t-k)}{\sum_{k=1}^{24} \alpha^{k-1}} \quad (13)$$

Tasoituskertoimelle α kokeiltiin arvoja väliltä 0.95 – 0.80 käyttäen 0.05 suuruisia askelia Hongin [25] tavoin. Nimetään mallit UMR-4 - UMR-7, jossa UMR-4 mallin tasoituskerron saa arvon 0.95 ja UMR-7 mallin tasoituskerron arvon 0.80. Tasoituskertoimen saadessa arvon 1, vastaa painotettu liukuva keskiarvo yksinkertaista liukuvaa keskiarvoa. Mallien UMR-4 – UMR-7 lähtötietona käytettiin kahden desimaalin tarkkuudella parhaan ennustevirheen tuottanutta mallia malleista UMR-B – UMR-3. Mikäli ennustevirhe oli malleilla sama, valittiin malleista yksinkertaisin eli se, jossa selittäviä muuttujia oli vähiten.

Kun ennustemallissa halutaan huomioida jokin yllä kuvatuista lämpötilatermeistä, lisätään ennustemalliin neljä interaktiitermiä jokaista uutta lämpötilatermiä kohden. Näin ollen, jos mallin halutaan huomioida edeltävän tunnin lämpötila T_{t-1} , lisätään ennustemalliin interaktiitermit $Tunti * T_{t-1}$, $Tunti * T_{t-1}^2$, $Kuukausi * T_{t-1}$ ja $Kuukausi * T_{t-1}^2$. Pientuottaja-segmentin ennustemallissa interaktiitermejä lisätään kaksi, sillä segmentin benchmark-malli ei sisältänyt lämpötilan neliöllistä muuttujaa.

4.2.5 Tutkimusmetodin kuvaus

Vuosia 2019 ja 2020 käytettiin usean muuttujan regressiomallin parametrien estimoimiseen (ns. in-sample), ja vuotta 2021 käytettiin ennustemallin testaamiseen (out-of-sample). Näin ollen ajanhetki 1.1.2021 00:00 toimi ennusteprosessin lähtöpisteenä, origona (t_0). Kun ennustemallia testataan out-of-sample -menetelmällä, voidaan käyttää yhtä tai useampaa origoa. Yhden origon tapauksessa ennusteet generoidaan ajanjaksoille $t_0+1, t_0+2, \dots, t_0+n$. Tällöin koko vuoden 2021 ennuste laaditaan kerralla. Usean origon tapauksessa origosta t_0 tehdään ennusteet ensin

ajanjaksoille $t_0+1, t_0+2, \dots, t_0+n$, jonka jälkeen origo päivitetään kohtaan t_0+1 . Tästä uudesta origosta ennustetaan ajanjaksoja $t_0+2, t_0+3, \dots, t_0+n$, jonka jälkeen origo jälleen päivitetään. [54] Usean origon, eli niin kutsutun liukuvan origon-menetelmä (rolling-origin) siis päivittää ennustemallia ajan kuluessa. Muun muassa Hong [25] hyödynsi tutkimuksessaan liukuvan origon-menetelmää eri päivitysväleillä [25].

Tässä diplomityössä liukuvan origon-menetelmää sovellettiin siten, että aikavälin 1.1.2019-31.12.2020 toteutuneen sähkönkulutuksen ja lämpötilojen perusteella estimoitiin selittävien muuttujien regressiokertoimet. Regressiokertoimien estimoinnin jälkeen ennustettiin vuoden 2021 tammikuulle tunnittainen sähkönkulutus jokaisella ennustemallilla. Tämän jälkeen tammikuun 2021 toteutunut kulutus lisättiin aineistoon, jonka pohjalta ennustemallien regressiokertoimet estimoitiin uudelleen aikaväliltä 1.1.2019-31.1.2021. Sen jälkeen helmikuun 2021 kulutusta ennustettiin uusien regressiokertoimien avulla. Prosessia jatkettiin, kunnes joulukuun 2021 ennuste saatiin laskettua. Näin ollen ennustemallien päivitysväliksi muodostui yksi kuukausi. Vertailun vuoksi myös yhden origon mallia testattiin, jossa koko vuoden 2021 ennusteet tehtiin kerralla.

Sekä liukuvan origon, että yhden origon menetelmällä ennustetuille kulutuksille laskettiin ennustevirhe (MAPE) ottamalla keskiarvo vuoden 2021 jokaisen tunnin prosentuaalisesta absoluuttisesta ennustevirheestä kaavan 1 mukaisesti. Eri ennustemallien ja -menetelmien tuottamia tuloksia analysoidaan tarkemmin luvussa 5.

5 Tulokset

Tässä luvussa tarkastellaan luvussa 4 kuvailtujen ennustemallien toimivuutta Vaasan Sähkön sähkönkulutuksen ennustamiseen. Ensin luodaan aggregoitu kulutusennuste koko Vaasan Sähköverkon alueen kulutuskohteille. Tämän jälkeen tarkastellaan, saavutetaanko parempi ennustetarkkuus segmentoimalla asiakkaat pientuottajuuden ja sähkösovimustyyppin mukaan.

5.1 Aggregoitu kulutusennuste

Tässä diplomityössä laskennat suoritettiin Stata 17 -ohjelmistolla. Regressiokertoimet estimoitiin havaintoaineiston pohjalta Statan regressiokomennolla, ja ennusteet laadittiin käyttäen predict-komentoa. [55] Havaintoaineistossa jokaiselle tunnille on merkattu toteutunut sähkönkulutus, kalenterimuuttujien luokat, tieto siitä onko kyseessä pyhäpäivä ja lämpötilaan liittyvien muuttujien arvot. Kun regressiokertoimet on estimoitu, voidaan esimerkiksi tammikuun 2021 ensimmäisen tunnin sähkönkulutus ennustaa seuraavan yhtälön mukaisesti, kun käytössä on malli UMR-1:

$$Y = \beta_0 + \beta_1 Kuukausi + \beta_2 Tunti * T(t) + \beta_3 Tunti * T(t)^2 + \beta_4 Tunti * T_{avg}(t) + \beta_5 Tunti * T_{avg}(t)^2 + \beta_6 Kuukausi * T(t) + \beta_7 Kuukausi * T(t)^2 + \beta_8 Kuukausi * T_{avg}(t) + \beta_9 Kuukausi * T_{avg}(t)^2 + \beta_{10} Päivä * Tunti + \beta_{11} Pyhä + \beta_{12} Tunti * T(t - 1) + \beta_{13} Tunti * T(t - 1)^2 + \beta_{14} Kuukausi * T(t - 1) + \beta_{15} Kuukausi * T(t - 1)^2 \quad (14)$$

Kategoristen muuttujien kohdalla yksi kategorian luokka toimii vertailukohtana, jonka regressiokerroin on 0. Näin ollen muiden kategorian luokkien regressiokertoimet kertovat, kuinka paljon enemmän tai vähemmän sähkönkulutus on kullakin luokalla vertailuluokkaan nähden. *Kuukausi*-muuttujassa tammikuu toimii vertailuluokkana, joten kaavassa 14 kerroin β_1 saa arvon 0. Samoin interaktiotermi *Päivä***Tunti* kohdalla ensimmäisen viikonpäivän (maanantai) ja ensimmäisen tunnin (00:00) välinen interaktio toimii vertailukohtana, jolloin regressiokerroin saa arvon 0. Koska tammikuun 2021 ensimmäinen päivä on uudenvuodenpäivä (perjantai), saa muuttuja *Pyhä* arvon 1. Näin ollen tammikuun 2021 ensimmäisen tunnin sähkönkulutuksen ennustemalli UMR-1 voidaan kirjoittaa:

$$Y = \beta_0 + \beta_2 T_{\text{unti}} * T(t) + \beta_3 T_{\text{unti}} * T(t)^2 + \beta_4 T_{\text{unti}} * T_{\text{avg}}(t) + \beta_5 T_{\text{unti}} * T_{\text{avg}}(t)^2 + \beta_6 K_{\text{uukausi}} * T(t) + \beta_7 K_{\text{uukausi}} * T(t)^2 + \beta_8 K_{\text{uukausi}} * T_{\text{avg}}(t) + \beta_9 K_{\text{uukausi}} * T_{\text{avg}}(t)^2 + \beta_{10} P_{\text{äivä}} * T_{\text{unti}} + \beta_{11} P_{\text{yhä}} + \beta_{12} T_{\text{unti}} * T(t - 1) + \beta_{13} K_{\text{uukausi}} * T(t - 1) \quad (15)$$

Ja sijoittamalla estimoitujen regressiokertoimien sekä muuttujien arvot saadaan:

$$Y = 89,72 - 0,73 * 0,4^{\circ}\text{C} + 0,024 * (0,04^{\circ}\text{C})^2 - 2,20 * 0,49^{\circ}\text{C} + 0,042 * (0,49^{\circ}\text{C})^2 - 0,53 * 0,04^{\circ}\text{C} - 0,038 * (0,04^{\circ}\text{C})^2 + 0,85 * 0,49^{\circ}\text{C} - 0,039 * (0,49^{\circ}\text{C})^2 + 1,62 - 3,79 * 1 - 0,48 * 0,5^{\circ}\text{C} + 0,06 * (0,5^{\circ}\text{C})^2 + 0,68 * 0,5^{\circ}\text{C} - 0,004 * (0,5^{\circ}\text{C})^2 = 88,14 \text{ MWh}$$

Toteutunut sähkönkulutus tammikuun 2021 ensimmäisellä tunnilla oli 85,96 megawattituntia, joten tämän yhden tunnin ennustevirhe voidaan laskea:

$$MAPE = 100\% * \frac{|85,96 - 88,14|}{85,96} = 2,5\%$$

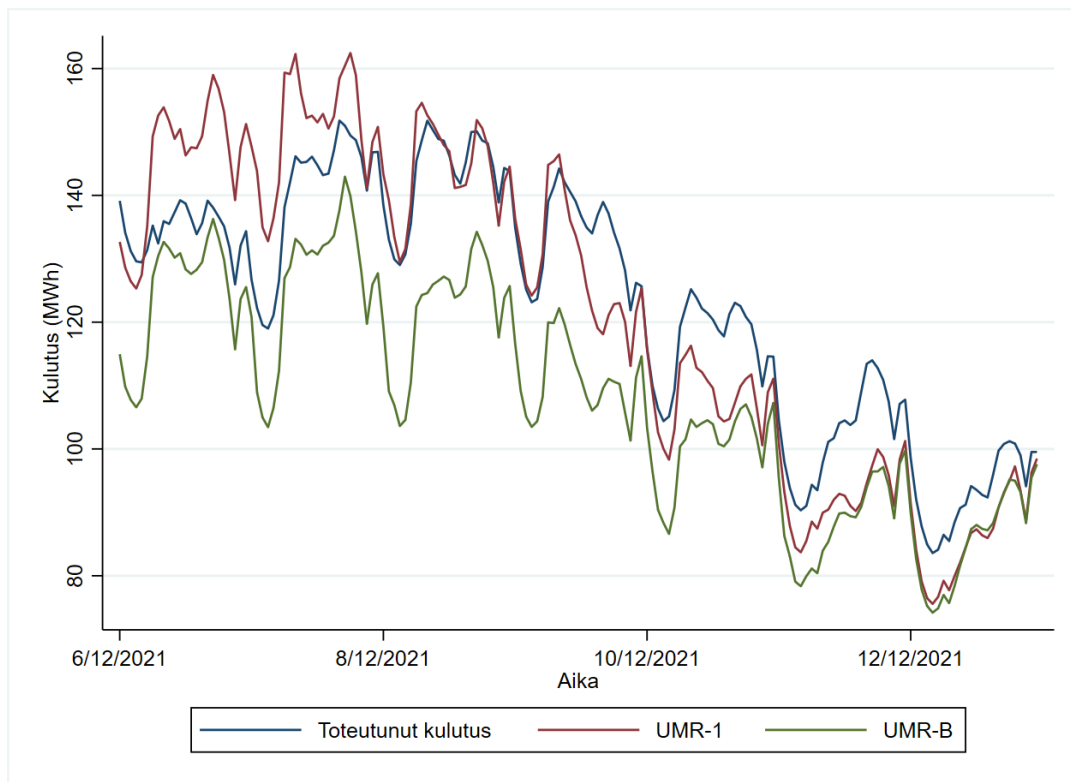
Vastaavan prosessin mukaisesti laskenta suoritettiin vuoden 2021 jokaiselle tunnille. Taulukko 3 esittää Vaasan Sähköverkon asiakkaiden muodostaman kokonaiskulutuksen ennustevirheet. Malleista UMR-B – UMR-3 pienimpään ennustevirheeseen päästiin mallilla UMR-1, joten tätä mallia sovellettiin liukuvan keskiarvon malleihin UMR-4 – UMR-7. Ennustetarkkuus ei parantunut, kun edeltävien 24 tunnin yksinkertainen keskiarvo korvattiin liukuvilla keskiarvoilla eri tasoituskertoimilla, joten malli UMR-1 tuotti parhaan lopputuloksen sekä kuukausittain päivitetyllä liukuvan origon menetelmällä, että yhden origon menetelmällä. Huomionarvoista on, etteivät ennustevirheet parantuneet merkittävästi kuukausittain päivitetyllä mallilla. Joidenkin mallien kohdalla tulokset olivat parempia, kun koko vuosi 2021 ennustettiin kerralla. Tämä on osaltaan ristiriidassa muun muassa Hongin [25] tulosten kanssa, joissa ennustemallin päivittäväläin lyhentäminen paransi ennustetarkkuutta kaikilla malleilla [25].

Taulukko 3. Vaasan Sähköverkon alueen kokonaiskulutuksen keskimääräiset absoluuttiset prosentuaaliset ennustevirheet eri ennustemalleilla.

Ennustemalli	Tasoiuskerroin	Liukuva origo	Yksi origo
UMR-B	1	5,32	5,31
UMR-0	1	4,59	4,57
UMR-1	1	4,57*	4,56*
UMR-2	1	4,57	4,56
UMR-3	1	4,57	4,56
UMR-4	0.95	4,57	4,56
UMR-5	0.90	4,60	4,59
UMR-6	0.85	4,68	4,67
UMR-7	0.80	4,76	4,75

*käytetty mallien UMR4-UMR7 lähtötietona. Tummennettuna parhaan lopputuloksen tuottanut malli.

Kokonaiskulutuksen ennuste tarkentui huomattavasti kun UMR-B -mallista vaihdettiin UMR-1 malliin. Ottamalla ennustemalliin mukaan sekä edeltävän 24 tunnin lämpötilan keskiarvo, että edeltävän tunnin lämpötila pieneni ennustevirhe liukuvan origon ja yhden origon menetelmillä 0,75 prosenttiyksikköä. Kuva 12 havainnollistaa toteutunutta kulutusta, sekä UMR-B ja UMR-1 mallien tuottamaa ennustetta vuoden 2021 viikolta 49. Kuten kuvasta nähdään, UMR-B malli ennustaa jatkuvasti liian alhaisia sähkönkulutuksen arvoja, kun taas UMR-1 malli myötäilee toteutunutta kulutusta huomattavasti tarkemmin, päästen täten alhaisempaan ennustevirheeseen.



Kuva 12: Toteutunut kulutus sekä UMR-B- ja UMR-1 -mallien tuottamat ennusteet vuoden 2021 viikolta 49.

5.2 Segmentointi pientuottajuuden mukaan

Ensin asiakkaat jaettiin kahteen segmenttiin pientuottajuuden perusteella. Ensimmäisen segmentin muodosti ne asiakkaat, kenellä ei ole omaa pientuotantoa ("ei-pientuotanto"). Segmentti muodostettiin vähentämällä kokonaiskulutuksesta pientuottaja-asiakkaiden kulutus. Toinen segmentti sisältää niiden asiakkaiden kulutusdatan, joilla on omaa pientuotantoa ("pientuotanto"). Taulukko 4 sisältää eri ennustemallien ennustevirheet ei-pientuotanto-segmentille. Parhaaseen lopputulokseen päästiin mallilla UMR-2 liukuvan origon menetelmällä ja mallilla UMR-3 yhden origon menetelmällä.

Taulukko 4: Ei-pientuotantoa-segmentin keskimääräiset absoluuttiset prosentuaaliset ennustevirheet eri ennustemalleilla.

Ennustemalli	Tasoiusterroin	Liukuva origo	Yksi origo
UMR-B	1	5,38	5,37
UMR-0	1	4,61	4,60
UMR-1	1	4,60	4,59
UMR-2	1	4,59*	4,59
UMR-3	1	4,59	4,58*
UMR-4	0.95	4,60	4,59
UMR-5	0.90	4,64	4,61
UMR-6	0.85	4,70	4,64
UMR-7	0.80	4,77	4,67

*käytetty mallien UMR4-UMR7 lähtötietona. Tummennettuna parhaan lopputuloksen tuottanut malli.

Taulukko 5 esittää puolestaan pientuotanto-segmentin ennustevirheet. Ennuste tarkentui huomattavasti, kun malliin lisättiin lämpötilatermejä. Kuukauden päivitysvälillä parhaaseen tulokseen päästiin mallilla UMR-5, joka ottaa huomioon edeltävän 24 tunnin liukuvan keskiarvon tasoiustertoimella 0,9. Yhden origon mallilla parhaaseen lopputulokseen päästiin mallilla UMR-6. Jälleen poiketen Hongin [25] tuloksista, ennustamalla koko vuoden 2021 kulutus kerralla saavutettiin parempia tuloksia kun päivittämällä ennustemallia kuukausittain [25].

Taulukko 5: Pientuotanto-segmentin keskimääräiset absoluuttiset prosentuaaliset ennustevirheet eri ennustemalleilla.

Ennustemalli	Tasoiusterroin	Liukuva origo	Yksi origo
UMR-B	1	9,16	9,13
UMR-0	1	8,89	8,88
UMR-1	1	8,86	8,86
UMR-2	1	8,85*	8,84*
UMR-3	1	8,86	8,84
UMR-4	0.95	8,83	8,81
UMR-5	0.90	8,80	8,79
UMR-6	0.85	8,80	8,78
UMR-7	0.80	8,80	8,79

*käytetty mallien UMR4-UMR7 lähtötietona. Tummennettuna parhaan lopputuloksen tuottanut malli.

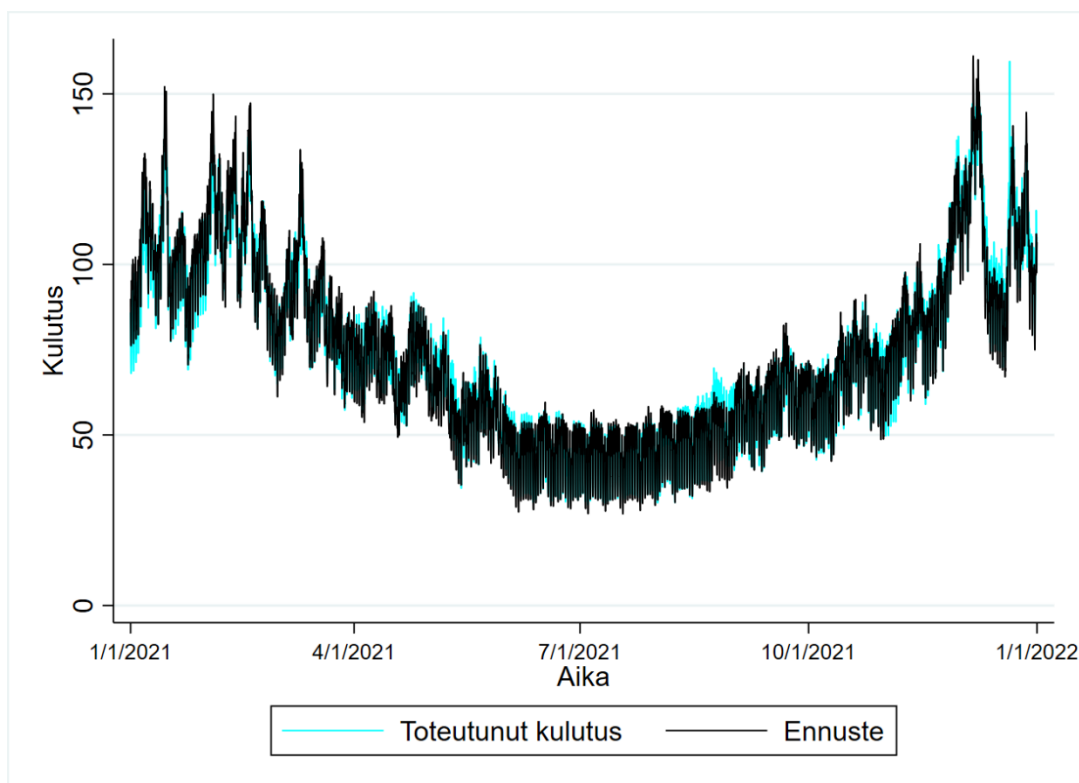
Kun molemmille segmenteille löydettiin paras ennustemalli, summattiin segmenttien kullekin tunnille lasketut kulutusennusteet yhteen ja tätä ennustesummaa verrattiin asiakaskunnan toteutuneeseen kokonaiskulutukseen. Taulukko 6 esittää ennustevirheet segmenttien muodostamalle ennustesummalle. Verrattuna kokonaiskulutuksen ennusteeseen, ennustamalla segmentit erikseen saavutettiin 0,03 prosenttiyksikköä pienempi ennustevirhe liukuvan origon -menetelmällä.

Kun koko vuosi ennustettiin kerralla, päästiin 0,04 prosenttiyksikköä pienempään ennustevirheeseen. Näin ollen, segmentoimalla asiakkaat pientuottajuuden mukaan ei saavutettu merkittävästi parempia tuloksia verrattuna aggregoituun ennusteeseen.

Taulukko 6: Segmenttien muodostaman ennustesumman keskimääräiset absoluuttiset prosentuaaliset ennustevirheet eri ennustemenetelmillä.

Ennustemalli	Liukuva origo	Yksi origo
UMR-ennustesumma	4,54	4,52

Kuva 13 havainnollistaa segmenttien muodostamaa ennustesummaa ja toteutunutta kulutusta vuodelta 2021, kun koko vuosi ennustettiin kerralla yhden origon mallilla. Ennuste myötäilee toteutunutta kulutusta varsin tasaisesti läpi vuoden, eikä ennustevirheessä ole havaittavissa merkittävää kausittaista vaihtelua.



Kuva 13: Toteutunut kulutus ja segmenttien muodostama ennustesumma vuodelle 2021.

Lopuksi testattiin, onko *Aurinko*-muuttujan lisäämisellä vaikutusta ei-pientuotantoa-segmentin ja kokonaiskulutuksen ennustemalleihin. Muuttujan lisääminen ei kahden merkitsevän numeron tarkkuudella vaikuttanut näiden mallien ennustevirheisiin, joten *Aurinko*-muuttuja sisällytettiin vain pientuotanto-segmentin ennustemalleihin. Tämän lisäksi

Trendi-muuttujan poistaminen ennustemalleista johti pienempään ennustevirheeseen sekä molemmissa segmenteissä, että kokonaiskulutusta ennustettaessa, joten kyseinen muuttuja jätettiin pois kaikista ennustemalleista.

5.3 Segmentointi sopimustyypeittäin

Sopimustyytit luokiteltiin neljään segmenttiin: yritys-, yleis-, kausi- ja yö sähköopimuksiin. Tutkimusaineistossa olevien puutteiden vuoksi sopimustyyppien yhteenlaskettu tunnittainen kulutus ei täsmännyt koko Vaasan Sähköverkon alueen tunnittaisen kokonaiskulutuksen kanssa. Näin ollen, segmenttien muodostamaa ennustesummaa verrattiin segmenttien yhteenlasketun tunnittaisen kulutuksen ennusteeseen. Jäljempänä tähän segmenttien yhteenlaskettuun tunnittaiseen kulutukseen viitataan nimellä Summa. Muutoin ennustamisprosessi vastasi luvussa 5.1 kuvailtua.

Luvun 5.2 ennusteissa *Trendi*-muuttujan poistaminen johti pienempään ennustevirheeseen, mutta sopimustyypeittäin segmentoiduissa ennusteissa *Trendi*-muuttujan mukaan ottaminen yleis- ja kausisähkösegmentteihin, sekä sopimustyyppien yhteenlaskettuun Summa-kulutussarjaan tuotti paremman lopputuloksen. Näin ollen näiden kolmen kulutussarjan ennustemalleihin sisällytettiin sähkönkulutuksen kasvua huomioon ottava trendimuuttuja. Yö sähkö- ja kausisähkösegmenttien ennuste puolestaan tarkentui kun *Trendi*-muuttuja jätettiin malleista pois, joten muuttujaa ei sisällytetty näiden segmenttien ennustemalleihin.

Taulukot 7 ja 8 esittävät segmenttien ennustevirheet yhden- ja liukuvan orion menetelmillä laskettuna, ja taulukko 9 esittää segmenttien muodostaman ennustesumman ennustevirheet. Samoin kun pientuottajuuden mukaan segmentoidussa ennusteessa, sopimustyypeittäin segmentoidussa ennusteessa kunkin segmentin parhaan lopputuloksen tuottanutta ennustemallia hyödynnettiin ennustesummaa laskettaessa. Huomionarvoista on, että erityisesti kausisähkö-segmentin ennustevirheet olivat merkittävästi muita segmenttejä suurempia. Tämä selittynee kausisähköopimusten pienellä lukumäärällä. Kulutuksen ennustamiselle on tyypillistä, että mikäli kulutuskohteiden lukumäärä on pieni, on satunnaisvaihtelun merkitys suuri [1], jolloin ennustetarkkuus kärsii.

Taulukko 7: Sopimustyypeittäin segmentoidun aineiston keskimääräiset absoluuttiset prosentuaaliset ennustevirheet eri ennustemalleilla yhden origon menetelmällä.

Ennuste-malli	Tasoi-tus-kerroin	Yritys-sopimukset	Yleis-sähkö	Kausi-sähkö	Yösähkö	Summa
UMR-B	1	6,50	5,90	15,69	7,61	5,83
UMR-0	1	6,48	5,12	15,41	5,94	4,68
UMR-1	1	6,47*	5,10*	15,39*	5,91	4,66
UMR-2	1	6,47	5,10	15,39	5,90*	4,65*
UMR-3	1	6,48	5,10	15,39	5,90	4,65
UMR-4	0.95	6,47	5,13	15,33	5,89	4,66
UMR-5	0.90	6,46	5,17	15,27	5,93	4,70
UMR-6	0.85	6,44	5,23	15,24	6,00	4,77
UMR-7	0.80	6,43	5,31	15,23	6,09	4,86

*käytetty mallien UMR4-UMR7 lähtötietona. Tummennettuna parhaan lopputuloksen tuottanut malli.

Taulukko 8: Sopimustyypeittäin segmentoidun aineiston keskimääräiset absoluuttiset prosentuaaliset ennustevirheet eri ennustemalleilla liukuvan origon menetelmällä.

Ennuste-malli	Tasoi-tus-kerroin	Yritys-sopimukset	Yleis-sähkö	Kausi-sähkö	Yösähkö	Summa
UMR-B	1	6,48	6,11	15,62	7,51	5,43
UMR-0	1	6,44*	5,38	15,37	6,06	4,54
UMR-1	1	6,44	5,35	15,34	6,03*	4,52
UMR-2	1	6,44	5,33	15,32*	6,03	4,51*
UMR-3	1	6,45	5,32*	15,32	6,03	4,51
UMR-4	0.95	6,44	5,33	15,27	6,01	4,52
UMR-5	0.90	6,42	5,35	15,22	6,06	4,56

UMR-6	0.85	6,41	5,39	15,19	6,16	4,61
UMR-7	0.80	6,41	5,44	15,18	6,30	4,67

*käytetty mallien UMR4-UMR7 lähtötietona. Tummennettuna parhaan lopputuloksen tuottanut malli.

Taulukko 9: Segmenttien muodostaman ennustesumman keskimääräiset absoluuttiset prosentuaaliset ennustevirheet eri ennustemenetelmillä.

	Liukuva origo	Yksi origo
UMR-ennustesumma	4,48	4,46

Taulukoista 7 ja 8 nähdään, että sopimustyyppisegmenttien ennusteissa oli eroja sen suhteen, saavutettiin pienempi ennustevirhe yhden- vai liukuvan origon menetelmällä. Muun muassa yleissähkösegmentin ennustevirhe oli merkittävästi pienempi, kun ennuste laadittiin yhden origon menetelmällä. Lisäksi eri segmenteille eri ennustemallit tuottivat parhaita lopputuloksia. Esimerkiksi yrityssojimussegmentissä parhaan lopputuloksen tuotti ennustemalli UMR-6 liukuvan origon menetelmällä, kun taas yleissähkösegmentissä pienimmän ennustevirheen tuotti malli UMR-1 yhden origon menetelmällä. Tarkasteltaessa taulukon 9 tuloksia huomataan, että ennustamalla jokainen sopimustyyppisegmentti erikseen saavutettiin yhden origon mallilla 4,46 prosentin ennustevirhe, joka on 0,19 prosenttiyksikköä parempi, kuin taulukon 7 Summa-kulutussarjan ennustevirhe. Käytettäessä liukuvaa origoa saavutettiin segmentoidulla ennusteella 0,03 prosenttiyksikköä tarkempi lopputulos. Yhden origon menetelmällä tehty segmentoitu ennuste oli 0,05 prosenttiyksikköä parempi verrattuna Summa-kulutussarjan parhaan ennusteen tuottaneeseen malliin.

5.4 Tulosten tarkastelu

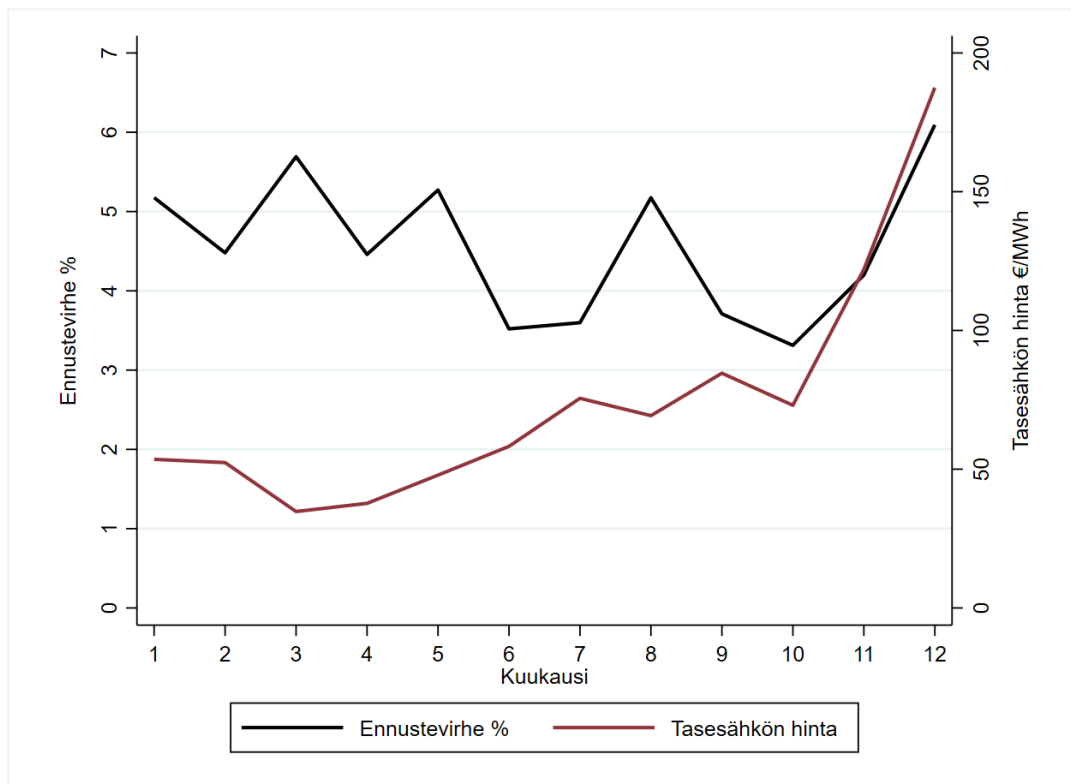
Segmentoimalla asiakkaat pientuottajuuden mukaan ei saavutettu tässä diplomityössä merkittävää parannusta ennustevirheeseen. Pientuottajien kulutusprofiilin pitäisi teoriassa erota normaalista kulutusprofiilista huomattavasti etenkin kuukausina, jolloin auringonsäteily on voimakasta. Tällöin kulutuskohteen oma aurinkosähkötuotanto kattaa merkittävän osan sähkönkulutuksesta. Pientuottajien toistaiseksi vähäinen lukumäärä selittää sen, ettei ennustevirheen osalta päästy merkittävästi parempaan lopputulokseen segmentoimalla asiakkaat pientuottajuuden mukaan. Tarkasteltavassa aineistossa pientuottajien lukumäärä oli hieman yli 1000 asiakasta, kun koko tarkasteltava asiakaskunta kattoi noin 75 000 asiakasta. Kuitenkin tulevaisuudessa, mikäli pientuottajien lukumäärä kasvaa, on perusteltua olettaa, että segmentoidulla ennusteella saavutettaisiin parempia tuloksia. Huomionarvoista on, ettei vuoden 2021 kulutusta

ennustettaessa ollut juurikaan merkitystä sillä, että ennustettiin koko vuoden ennuste kerrallaan päivittäen ennustemallia, vai tehtiinkö koko vuoden ennuste kerralla. Myöskään sopimustyypeittäin segmentoimalla ennustevirhettä ei saatu merkittävästi parannettua, sillä parhaalla ennustemallilla ennustevirhettä saatiin pienennettyä 0,05 prosenttiyksikköä.

5.4.1 Kokonaiskulutuksen ennustevirheen ajallinen vaihtelu

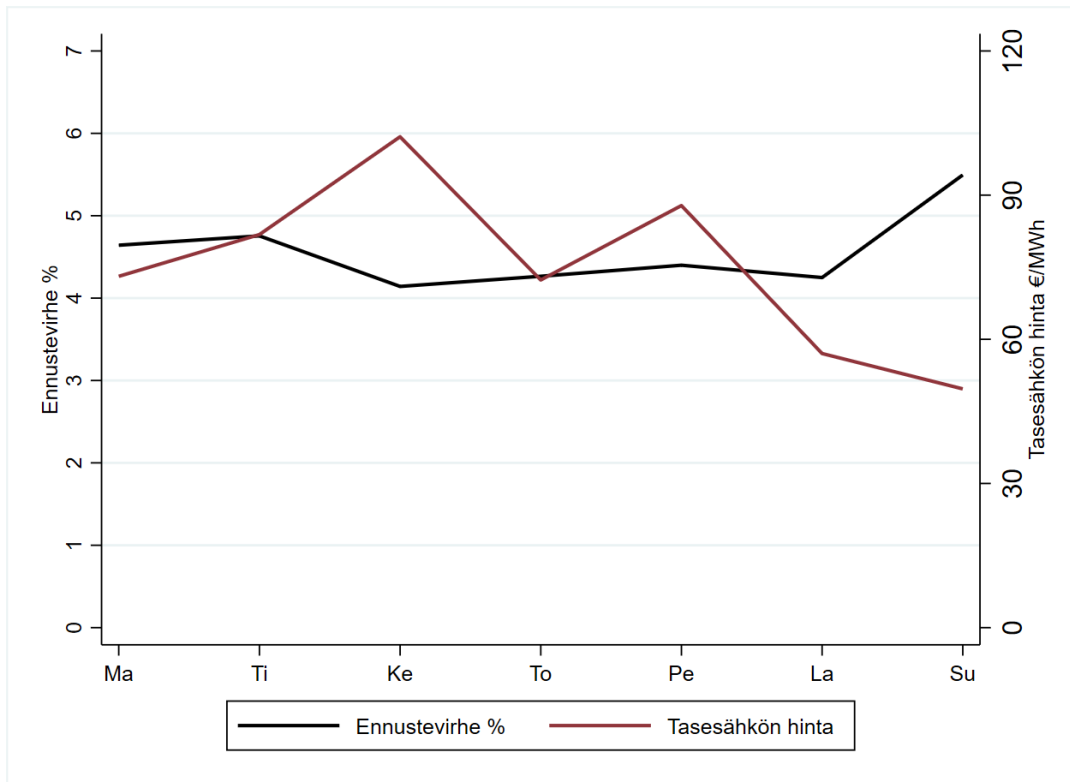
Vaasan Sähköverkon kokonaiskulutuksen ennusteessa parhaan lopputuloksen tuotti ennustemalli UMR-1 yhden origon menetelmällä. Ennustemallin tunnittainen keskimääräinen absoluuttinen prosentuaalinen ennustevirhe oli koko vuoden osalta oli 4,56 prosenttia. Vaikka sopimustyypeittäin segmentoimalla päästiin hieman alhaisempaan keskimääräiseen ennustevirheeseen, on perusteltua tarkastella kokonaiskulutuksen ennustevirhettä, sillä sopimustyypeittäin segmentoidusta aineistosta puuttui kulutuskohteita. Tarkastelemalla ennustevirheen keskiarvoja eri aikamuuttujien suhteen voidaan analysoida, ovatko tietyt ajanhetket ennustamisen kannalta haastavampia kuin toiset. Lisäksi tutkimalla miten tasesähkön keskihinnat ovat käyttäytyneet eri ajanhetkinä, voidaan tarkastella onko tasesähkön hinnan ja ennustevirheen välillä havaittavissa selkeitä trendejä.

Kuva 14 esittää ennustevirheen ja tasesähkön hinnan kuukausittaiset keskiarvot. Eri kuukausien välillä ennustevirhe vaihtelee. Muun muassa joulukuun 2021 ennustevirhe oli yli kuuden prosentin luokkaa, kun taas kesä- ja heinäkuussa ennustevirhe oli noin 3,5 prosenttia. Kuitenkaan selkeää kausittaista vaihtelua ennustevirheiden kuukausikeskiarvoissa ei ole nähtävissä.



Kuva 14: Keskimääräinen absoluuttinen prosentuaalinen ennustevirhe ja tasesähkön keskihinta kuukausittain vuodelta 2021.

Tasesähkön kuukausikeskihinta oli marras- ja joulukuussa huomattavasti muuta vuotta korkeampi. Kun joulukuussa myös ennustevirhe oli suuri, heijastuu tämä sähköyhtiölle korkeampina tasesähkökustannuksina, mikäli ennustevirhettä joudutaan korjaamaan tasesähkökaupalla. Tasesähkökustannusten osalta on kuitenkin todettava, että mikäli yksittäisten tuntien korkea tasesähkön hinta osuu sellaisille tunneille, jolloin myös ennustevirhe on suuri, voivat kustannukset kohota merkittäviksi. Näin ollen pelkkien keskiarvojen tarkastelu ei anna riittävän tarkkaa kuvaa toteutuneista tasesähkökustannuksista. Keskiarvoja tarkastelemalla voidaan kuitenkin pyrkiä havaitsemaan ennustevirheen ja tasesähkön hinnan välisiä trendejä.



Kuva 15: Keskimääräinen absoluuttinen prosentuaalinen ennustevirhe ja tasesähkön keskihinta viikonpäivittäin vuodelta 2021.

Kuva 15 puolestaan esittää ennustevirheen ja tasesähkön hinnan keskiarvot viikonpäivittäin. Maanantaisin, tiistaisin ja sunnuntaisin ennustevirheet ovat hieman muita viikonpäiviä korkeampia. Viikonpäivien välillä vaihtelu ennustevirheessä on kuitenkin melko pientä, ja kaikkien viikonpäivien keskimääräinen ennustevirhe on varsin lähellä koko vuoden tunnittaisten ennustevirheiden keskiarvoa 4,56 prosenttia. Tasesähkön hinnassakin on nähtävissä vaihtelua viikonpäivien välillä, mutta selkää yhteyttä ennustevirheen ja tasesähkön hinnan välillä ei kuvan perusteella ole havaittavissa.

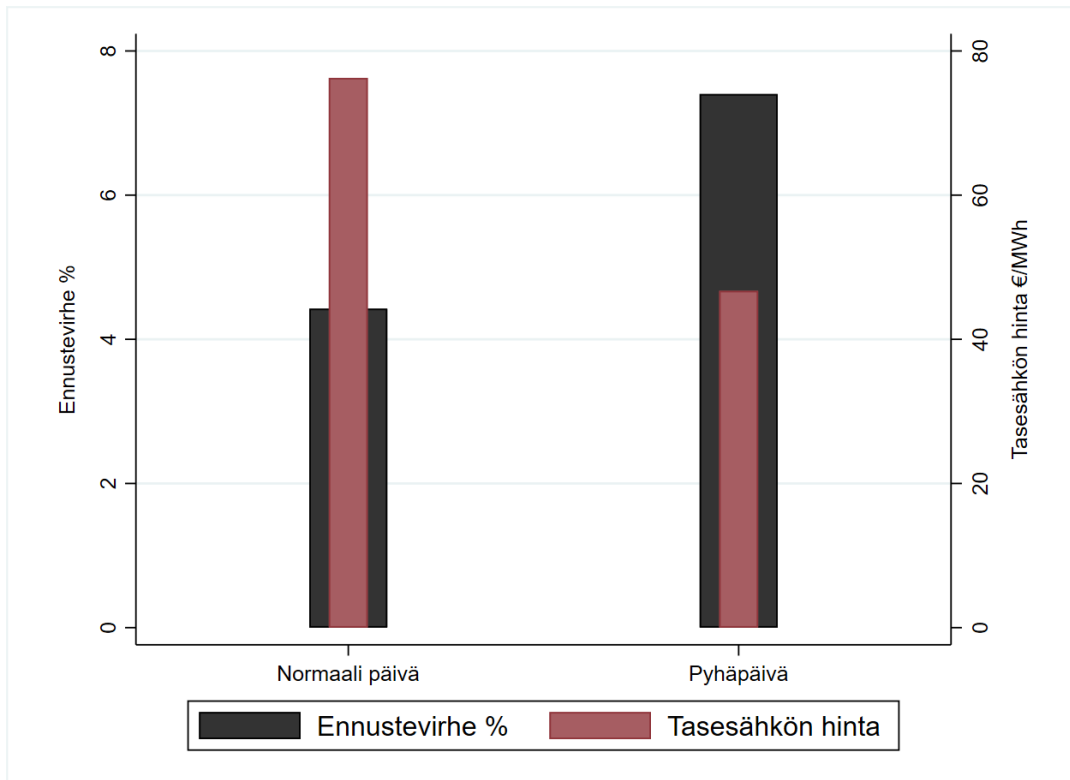
Kuva 16 havainnollistaa ennustevirheen ja tasesähkön hinnan keskiarvoja päivän tunneittain. Kuvasta on havaittavissa kaksi selkeää piikkiä ennustevirheessä, kun 07:00-08:00 alkavien tuntien kohdalla ennustevirhe kohoaa noin viiteen prosenttiin ja 16:00-19:00 alkavien tuntien kohdalla yli 5,5 prosenttiin. Pienimmät ennustevirheet saavutetaan puolestaan päivän viimeisten tuntien 22:00-23:00 kohdalla, jolloin ennustevirhe on alle neljä prosenttia.



Kuva 16: Keskimääräinen absoluuttinen prosentuaalinen ennustevirhe ja tasesähkön keskihinta päivän tunneittain vuodelta 2021.

Myös tasesähkön hinnan keskiarvo on korkeampi päiväsaikaan kuin yöllä. Näin ollen, erityisesti 16:00-18:00 alkavat tunnit ovat sähköyhtiön kannalta kriittisiä, kun sekä ennustevirheen, että tasesähkön hinnan keskiarvot ovat korkeita.

Kuva 17 havainnollistaa ennustevirheen käyttäytymistä pyhäpäivinä. Pyhäpäivien sähkönkulutuksen ennustaminen osoittautui varsin haastavaksi, sillä pyhäpäivien ennustevirhe kohosi yli seitsemään prosenttiin. Tämä selittyy osaltaan sillä, että ennustemalleissa jokaista pyhäpäivää käsiteltiin samanlaisena, riippumatta siitä oliko kyseessä arkipäivälle vai viikonloppuun ajoittuva pyhäpäivä. Tasesähkön keskihinta pyhäpäivinä on kuitenkin normaaleja päiviä alhaisempi, joten keskimäärin pyhäpäivien ennustevirhettä pystytään paikkaamaan halvemmalla tasesähköllä verrattuna normaalien päivien ennustevirheeseen.



Kuva 17: Keskimääräinen absoluuttinen prosentuaalinen ennustevirhe ja tasesähkön keskihinta pyhinä ja normaaleina päivinä vuodelta 2021.

5.4.2 Regressiokertoimien tulkinta

Usean muuttujan lineaarisen regressiomallin hyödyntäminen kulutusennustamisessa mahdollistaa sähkönkulutukseen vaikuttavien tekijöiden arvioimisen regressiokertoimia tulkitsemalla. Taulukossa 10 on listattuna luvussa 5.1 lasketun tammikuun 2021 ensimmäisen tunnin lämpötilaan ja pyhäpäivään liittyvät regressiokertoimet, joita tulkitsemalla saadaan lisätietoa siitä, millainen merkitys lämpötilalla ja pyhäpäivällä on sähkönkulutukseen.

Taulukko 10: Tammikuun 2021 ensimmäisen tunnin kokonaiskulutuksen ennustemallin UMR-1 regressiokertoimia.

Muuttuja	Regressiokerroin
$Tunti * T(t)$	-0,7**
$Tunti * T(t)^2$	-0,02
$Tunti * T_{avg}(t)$	-2,2***
$Tunti * T_{avg}(t)^2$	0,04***
$Kuukausi * T(t)$	-0,5
$Kuukausi * T(t)^2$	-0,04**
$Kuukausi * T_{avg}(t)$	0,9***
$Kuukausi * T_{avg}(t)^2$	-0,04***

$Tunti * T(t - 1)$	-0,5
$Tunti * T(t - 1)^2$	0,1***
$Kuukausi * T(t - 1)$	0,7**
$Kuukausi * T(t - 1)^2$	-0,004
<i>Pyhä</i>	-3,8***

*** p<0.01, ** p<0.05, * p<0.1

Taulukon 10 yhteydessä kullekin regressiokertoimelle on ilmoitettu p-arvo, joka kuvastaa sitä, että eroaako regressiokertoimet tilastollisesti merkitsevästi nolasta. Tilastollinen merkitsevyys kertoo, kuinka todennäköisesti selittävän ja selitettävän muuttujan välinen yhteys havaittaisiin, mikäli muuttujien välillä ei todellisuudessa olisi yhteyttä. Pieni todennäköisyys (p-arvo) kertoo, että vastaavan tuloksen saaminen siinä tapauksessa olisi epätodennäköistä. [50] Tilastollinen merkitsevyys on siis käsite, jolla pyritään kuvaamaan sitä, ettei saadut tulokset ole sattumaa [56]. P-arvot on ilmoitettu kolmella merkitsevyystasolla. Esimerkiksi p<0,01 kertoo, että on 1 prosentin todennäköisyys sille, että nollahypoteesi (regressiokerroin saa arvon 0) hylätään, vaikka todellisuudessa nollahypoteesi olisi pätevä. Taulukosta 10 nähdään, että esimerkiksi muuttujan $Tunti * T(t)^2$ regressiokerroin ei ole tilastollisesti merkitsevä.

Huomionarvoista taulukon 10 tuloksissa on, että pyhäpäivällä on merkittävä vaikutus sähkönkulutukseen. Mikäli muut muuttujat pysyvät samana, vähentää pyhäpäivä sähkönkulutusta 3,79 megawattituntia, kun tammikuun ensimmäisen tunnin kokonaiskulutus oli 86 megawattituntia. Lämpötilan vaikutusta arvioidessa tulee ottaa huomioon kaikkien lämpötilasta riippuvien muuttujien yhteisvaikutus. Mikäli lämpötila nousee yhden asteen, voidaan lämpötilasta riippuvien muuttujien regressiokertoimet laskea yhteen, jolloin saadaan:

$$\beta = -0,7 - 0,02 - 0,5 - 0,04 = -1,26$$

Näin ollen, mikäli lämpötila nousee yhden asteen, pienentää se sähkön kokonaiskulutusta 1,26 megawattituntia, mikäli muut muuttujat pysyvät vakioina. Mikäli puolestaan viimeisen 24 tunnin lämpötilan keskiarvo nousee yhdellä asteella, voidaan vastaavat regressiokertoimet summata, jolloin saadaan:

$$\beta = -2,2 + 0,04 + 0,9 - 0,04 = -1,3$$

Yhden asteen nousu viimeisen 24 tunnin lämpötilan keskiarvossa pienentää sähkön kokonaiskulutusta puolestaan 1,3 megawattituntia. Regressiokerrointen tarkastelu osoittaa, että sähkönkulutus on lämpötilariippuvaista ja pienilläkin lämpötilan muutoksilla on vaikutusta sähkönkulutukseen.

Kun tarkastelu laajennetaan koko vuoteen 2021, lämpötilan vaikutus sähkönkulutukseen näkyy myös muuttujan *Kuukausi* regressiokertoimissa, jotka on esitetty taulukossa 11. *Kuukausi*-muuttujan voidaan ajatella kuvaavan sähkönkulutuksen kausittaista vaihtelua. Regressiokertoimien vertailukohtana toimii tammikuu, jonka regressiokerroin saa siten arvon 0.

Taulukko 11: *Kuukausi*-muuttujan regressiokertoimet.

Kuukausi	Regressiokerroin
Tammikuu	0
Helmikuu	-1,5***
Maaliskuu	-6,0***
Huhtikuu	-10,4***
Toukokuu	-15,0***
Kesäkuu	-13,1***
Heinäkuu	-17,2***
Elokuu	-25,0***
Syyskuu	-6,8***
Lokakuu	-9,1***
Marraskuu	-3,4***
Joulukuu	-2,2***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Verrattuna tammikuuhun, kaikkien muiden kuukausien regressiokertoimet ovat negatiivisia, jolloin muiden selittävien muuttujien pysyessä vakiona on muilla kuukausilla sähkönkulutusta alentava vaikutus. Sähkönkulutuksen lämpötilariippuvuutta kuvaa hyvin se, että huhtikuusta-elokuuhun regressiokertoimet ovat huomattavan pieniä. Esimerkiksi verrattuna tammikuuhun, elokuussa sähkönkulutus on 25 megawattituntia tunnissa pienempää, mikäli muut muuttujat pysyvät muuttumattomina. Marraskuusta maaliskuuhun *Kuukausi*-muuttujan vaikutus on huomattavasti pienempi, joka kuvastaa puolestaan sitä, että ulkolämpötilan ollessa alhainen on sähkönkulutus korkea. *Kuukausi*-muuttujan regressiokertoimet ovat tilastollisesti merkitseviä, sillä kaikkien muuttujien p -arvo on pienempi kuin 0,01.

Viikonpäivällä ja päivän tunnilla on myös merkittävä vaikutus sähkönkulutukseen. Tarkastelemalla interaktiotermin *Päivä*Tunti* regressiokertoimia, voidaan tehdä johtopäätöksiä siitä, mitkä ajanhetket ovat sähkönkulutuksen kannalta merkityksellisiä. Taulukko 12 esittää kunkin viikonpäivän regressiokertoimet tunneittain.

Taulukko 12: Kokonaiskulutuksen ennustemallin UMR-1 päivän tunnin ja viikonpäivän välisen interaktiotermin regressiokertoimet.

Tunti	Ma	Ti	Ke	To	Pe	La	Su
00:00	0,0	0,8*	0,6	1,1**	1,6***	1,9***	1,1**
01:00	-7,1***	-6,3***	-6,3***	-6,1***	-5,5***	-5,6***	-6,2***
02:00	8,1***	8,6***	8,4***	8,5***	8,3***	-1,6***	-4,8***
03:00	8,1***	8,6***	8,3***	8,4***	8,0***	-0,1	-2,5***
04:00	7,7***	8,0***	7,9***	7,8***	7,4***	-0,0	-1,9***
05:00	6,0***	6,3***	6,2***	6,1***	5,5***	-1,0*	-2,8***
06:00	5,4***	5,6***	5,8***	5,4***	4,8***	-1,3***	-3,3***
07:00	6,0***	6,2***	6,6***	6,1***	5,9***	0,5	-2,0***
08:00	8,5***	8,7***	9,4***	8,6***	9,0***	4,1***	1,2**
09:00	9,2***	9,4***	10,0***	9,5***	10,8***	6,6***	3,5***
10:00	8,6***	9,0***	9,9***	9,3***	11,5***	8,5***	5,3***
11:00	9,1***	9,6***	10,5***	9,8***	11,7***	9,1***	7,0***
12:00	-11,5***	-10,6***	-10,7***	-10,5***	-9,7***	-10,3***	-11,0***
13:00	7,3***	7,6***	8,5***	8,1***	8,9***	6,2***	5,1***
14:00	2,7***	2,8***	3,5***	3,3***	3,6***	1,3**	0,3
15:00	10,1***	9,9***	10,4***	10,6***	10,8***	8,7***	7,7***
16:00	11,4***	10,9***	11,6***	11,9***	12,3***	10,7***	9,8***
17:00	-13,6***	-12,7***	-12,7***	-12,5***	-11,7***	-12,6***	-13,4***
18:00	-13,4***	-12,4***	-12,5***	-12,4***	-11,9***	-13,3***	-14,2***
19:00	-9,7***	-8,9***	-8,9***	-8,7***	-8,3***	-11,9***	-13,1***
20:00	-0,8*	0,3	0,3	0,6	0,5	-8,9***	-10,7***
21:00	3,8***	4,9***	4,7***	5,0***	4,5***	-9,4***	-11,9***
22:00	7,0***	7,9***	7,7***	8,0***	7,7***	-6,1***	-9,5***
23:00	7,5***	8,1***	7,8***	8,0***	7,9***	-3,7***	-7,2***

*** p<0.01, ** p<0.05, * p<0.1

Taulukon 12 regressiokertoimien kohdalla vertailu käy maanantain 00:00 alkavaan tuntiin. Kertoimista nähdään, että esimerkiksi arkipäivisin 02:00-11:00 välillä regressiokertoimet ovat positiivisia ja vaihtelevat välillä 4.8 – 11.7. Näin ollen, verrattuna maanantain 00:00 alkavaan tuntiin, arkipäivien aamun tunneilla on sähkönkulutusta selvästi nostava vaikutus. Puolestaan viikonloppuisin 02:00-07:00 alkavilla tunneilla regressiokertoimet ovat pääosin negatiivisia, joka kuvastaa puolestaan sitä, että viikonloppuaamuisin sähkönkulutus on vertailukohtaa pienempää. Huomionarvoista on myös se, että viikonpäivästä riippumatta 17:00-18:00 alkavien tuntien regressiokertoimet ovat hyvin pieniä. Verrattuna maanantain 00:00 alkavaan tuntiin 17:00-18:00 alkavien tuntien aikana sähkönkulutus on

huomattavasti matalampaa, mikäli muut muuttujat pysyvät vakioina. Valtaosa regressiokertoimista on tilastollisesti merkitseviä ($p < 0,01$), mutta esimerkiksi tiistaista perjantaihin 20:00 alkavan tunnin regressiokertoimet eivät ole tilastollisesti merkitseviä, sillä p-arvot nousevat yli 0,1:n.

6 Yhteenveto

Tässä diplomityössä kehitettiin sähköyhtiön lyhyen aikavälin kulutusennustamista. Yhdessä tuotantoennusteen kanssa kulutusennustamisella on iso rooli, kun sähköyhtiö suunnittelee seuraavan päivän sähkömarkkinatoimenpiteitään. Pyrkimällä tarkkaan kulutuksen ennustamiseen, voi sähköyhtiö minimoida altistumista tasesähkön hintariskille, kun tarvittava sähkön hankinta ja myynti voidaan toteuttaa sähköpörssin vuorokausimarkkinalla sen sijaan, että sähköyhtiö joutuisi tasapainottamaan sähkön hankintaa ja myyntiä tasesähkökaupalla.

Aikaisemman kirjallisuuden pohjalta tässä työssä käytetyksi ennustemalliksi valikoitui usean muuttujan lineaarinen regressiomalli. Regressiomallissa selitettävänä muuttujana toimi seuraavan päivän tunnittainen sähkönkulutus ja selittävinä muuttujina lämpötilaan ja aikaan liittyvät muuttujat. Selittävien muuttujien valinta suoritettiin tutkimalla toteutuneen kulutuksen lämpötila- ja aikariippuvuuksia. Näiden riippuvuussuhteiden perusteella luotiin kulutusennustamisen benchmark-malli, jota lähdettiin kehittämään edelleen ottamalla malliin mukaan yksitellen lisää selittäviä muuttujia, kuten tarkasteltavaa tuntia edeltävien tuntien lämpötiloja. Näin luotiin yhteensä yhdeksän ennustemallia, joiden suorituskykyä testattiin tutkimusaineistoon. Työssä käytettiin Vaasan Sähköverkon kulutusdataa vuosilta 2019-2021. Vuodet 2019-2020 toimivat ennustemallien opetusjaksona ja malleja testattiin vuoden 2021 toteutuneita kulutushavaintoja vasten. Ennustemallien suorituskykyä vertailtiin keskimääräisen absoluuttisen prosentuaalisen ennustevirheen avulla.

Aikaisemmassa kirjallisuudessa tutkimukset ovat keskittyneet erilaisten ennustemallien kehittämiseen ja tutkimuksissa ollaan pyritty minimoimaan kulutusennusteen ennustevirhettä. Tyypillisesti kulutusennusteet luodaan laajalla kuluttajajoukolle, kuten koko sähköyhtiön tai sähköverkkoyhtiön asiakkaille aggregoidusti. Tässä diplomityössä kuluttajat jaettiin pienempiin segmentteihin ensin sen perusteella, onko kuluttajalla omaa pientuotantoaan esimerkiksi aurinkopaneelien muodossa. Toinen segmentointi suoritettiin sähkösopimustyyppien mukaan. Segmentoinnilla pyrittiin siihen, että ennustemallit pystyisivät huomioimaan kullekin segmentille tyypillisiä kulutukseen vaikuttavia tekijöitä aggregoitua ennustetta paremmin.

Kullekin segmentille etsittiin pienimmän ennustevirheen tuottanut ennustemalli, jonka jälkeen ennusteiden muodostamaa summaa verrattiin aggregoituu ennusteeseen. Pientuottajien mukaan segmentoimalla saavutettiin 0,04 prosenttiyksikköä pienempi ennustevirhe aggregoituu ennusteeseen nähden, kun taas sopimustyypeittäin segmentoimalla ennustevirhettä saatiin pienennettyä 0,05 prosenttiyksikköä. Näin ollen segmentoimalla saavutettiin vain hieman parempia lopputuloksia kuin aggregoidulla ennusteella. Pientuottajuuden osalta voidaan todeta, että

segmentoinnin hyödyt jäivät vähäiseksi todennäköisesti siitä syystä, että pientuottajien lukumäärä oli tarkasteltavassa aineistossa varsin pieni. Mikäli pientuotannon määrä kasvaa tulevaisuudessa, voitaisiin segmentoinnilla saavuttaa parempia tuloksia. Sopimustyypeittäin segmentoinnin lähtökohtana oli se, että erilaisten sopimustyyppien kulutuskäyttäytyminen eroaisi toisistaan. Vaikka eroja kulutuskäyttäytymisessä havaittiin aikaisemman kulutusdatan perusteella esimerkiksi yritys- ja yksityisasiakkaiden osalta, lienee perusteltua todeta, että sopimustyyppien vaikutus kulutuskäyttäytymiseen on kuitenkin melko vähäinen. Tämä selittää sen, ettei sopimustyypeittäin segmentoidulla ennusteella päästy merkittävästi parempiin lopputuloksiin.

Tulevaisuudessa mielenkiintoinen sopimustyyppisegmentti tulee olemaan pörssisähkösopimukset. Pörssisähkösopimuksissa sähkön hinta seuraa sähköpörssin vuorokausimarkkinan spot-hintaa, jolloin kuluttajilla on taloudellinen kannustin ajoittaa omaa sähkönkulutustaan matalan hinnan hetkiin. Tyypillisesti kiinteähintaisissa sähkösopimuksissa sähkönkulutus on melko epäelastista, eli se reagoi vain vähän sähkön hinnan muutoksiin. Kuitenkin muun muassa sähköautojen lisääntyessä, on kuluttajilla mahdollisuus merkittäviin säästöihin, mikäli sähköauton lataus voidaan ajoittaa matalan hinnan ajanhetkiin. Yhdessä pientuottajuuden lisääntymisen kanssa sähköautot ja niiden lataaminen tuovatkin kulutusennustamiseen täysin uudenlaisia haasteita tulevaisuudessa, kun perinteiset kulutusprofiilit muuttuvat merkittävästi.

7 Lähteet

- [1] J. Partanen, S. Viljakainen, J. Lassila, S. Honkapuro, K. Salovaara, S. Annala ja M. Makkonen, ”Sähkömarkkinat - opetusmoniste,” LUT-yliopisto, 2014.
- [2] Tuulivoimayhdistys ry, ”Tuulivoimalla katettiin noin 10% Suomen sähkönkulutuksesta vuonna 2020,” <https://tuulivoimayhdistys.fi/ajankohtaista/tiedotteet/tuulivoimalla-katettiin-noin-10-suomen-sahkonkulutuksesta-vuonna-2020>, (Viitattu 17.8.2022).
- [3] LUT-yliopisto, ”Aurinkoenergia ja aurinkosähkö Suomessa,” <https://www.lut.fi/fi/artikkelit/aurinkoenergia-ja-aurinkosahko-suomessa> (Viitattu 15.8.2022).
- [4] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang ja H. Zareipour, ”Energy Forecasting: A Review and Outlook,” *IEEE Open Access Journal of Power and Energy*, vol. 7, s. 376-388, 2020.
- [5] Fingrid, ”Johdanto sähkömarkkinoihin”, <https://www.fingrid.fi/sahkomarkkinat/markkinoiden-yhtenaisyyss/johdanto-sahkomarkkinoihin/#saatosahko--ja-reservimarkkinat->, (Viitattu 20.07.2022).
- [6] Nord Pool, ”Annual Review 2020,” Nord Pool 2020. [Verkkoaineisto]. Saatavissa: <https://www.nordpoolgroup.com/49eea7/globalassets/download-center/annual-report/annual-review-2020.pdf>, 2020
- [7] M. Kolehmainen, ”Palvelutuottajan Elbas-kaupankäynnin kehittäminen tasekustannusten minimoimiseksi perustuen säätösähkön hinnan odotusarvoon,” Diplomityö, Tampereen teknillinen yliopisto, 2016.
- [8] Nasdaq Commodities AB, ”Trading Appendix 2 / Clearing Appendix 2: Contract Specifications: Commodity Derivatives”, 2022. [Verkkoaineisto]. Saatavissa: <https://www.nasdaq.com/docs/2021/12/30/2022-01-01-Joint-Appendix-2-Contract-specifications-clean.pdf>
- [9] Fingrid, ”Säätösähkö- ja säätökapasiteettimarkkinat,” <https://www.fingrid.fi/sahkomarkkinat/reservit-ja-saatosahko/saatosahko--ja-saatokapasiteettimarkkinat/#saatosahkon-hinnoittelu>, (Viitattu 18.6.2022).
- [10] Fingrid, ”Reservimarkkinat,” <https://www.fingrid.fi/sahkomarkkinat/reservit-ja-saatosahko/>, (Viitattu 7.7.2022).

- [11] Fingrid, "Tasepalvelut" <https://www.fingrid.fi/sahkomarkkinat/tasepalvelut/>, (Viitattu 13.7.2022).
- [12] Fingrid, "Tasemallin kuvaus," Saatavissa: <https://www.fingrid.fi/sahkomarkkinat/tasepalvelut/tasemallin-kuvaus/>, (Viitattu 3.7.2022).
- [13] Fingrid, "Yksihintajärjestelmä," Saatavissa: <https://www.fingrid.fi/sahkomarkkinat/tasepalvelut/tasemallin-kuvaus/Yksihinta-jarjestelma/>, (Viitattu 17.7.2022).
- [14] Fingrid, "Säätösähkön hintarajat nousevat 1.11," Saatavissa: <https://www.fingrid.fi/ajankohtaista/tiedotteet/2022/saatosahkon-hintarajat-nousevat-1.11.2022-menessa/> (Viitattu 13.9.2022).
- [15] Nordpool AS, "Day-ahead Market Regulations, Nordic/Baltic Market and CE Market," 2022. [Verkkoaineisto]. Saatavissa: https://www.nordpoolgroup.com/49d91b/globalassets/download-center/rules-and-regulations/day-ahead-market-regulations_nordic-baltic-ce_valid-from-05.04.21.pdf
- [16] Fingrid, "Tasevastuuta ja taseselvitystä koskeva käsikirja. Osa 1: Fingrid Oyj:n yleiset taseselvityksen ehdot," Fingrid, 2021. [Verkkoaineisto]. Saatavissa: <https://www.fingrid.fi/globalassets/dokumentit/fi/palvelut/tasesahkokauppa-ja-taseselvitys/liite-1-osa-1-fingrid-oyjn-yleiset-tasehallinnan-ehdot-.pdf>
- [17] T. Hong ja S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, nro 3, s. 914-938, 2016.
- [18] A. Yang, W. Li ja Y. Xuan, "Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines.," *Knowledge-Based Systems*, vol 163, s. 159-173, 2019.
- [19] J. Nowotarski, B. Liu, R. Weron ja T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol 98, s. 40-49, 2016.
- [20] Y. Wang, Q. Chen, T. Hong ja C. Kang, "Review of Smart Meter Data Analytics: Application, Methodologies and Challenges," *IEEE Transactions on Smart Grid*, vol. 10, nro. 3, s. 3125-3148, 2019.
- [21] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli ja R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, nro. 3, s. 896-913, 2016.

- [22] T. Hong, P. Pinson ja S. Fan, "Global Energy Forecasting Competition 2012," *International Journal of Forecasting*, vol. 30, nro. 2, s. 357-363, 2014.
- [23] R. J. Hyndman ja S. Fan, "Density Forecasting for Long-Term Peak Electricity Demand," *IEEE Transactions on Power Systems*, vol. 25, nro. 2, s. 1142-1153, 2009.
- [24] N. Elamin ja M. Fukushige, "Modeling and forecasting hourly electricity demand by SARIMAX with interactions," *Energy*, Vol. 165, Part B, pp. 257-268, 2018.
- [25] T. Hong, "Short Term Electric Load Forecasting", väitöskirja, Operations Research and Electrical Engineering, University of North Carolina, Raleigh, 2010.
- [26] R. J. Hyndman ja F. Shu, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Transactions on Power Systems*, vol. 27, nro. 1, s. 134-141, 2012.
- [27] J. W. Taylor ja R. Buizza, "Neural Network Load Forecasting With Weather Ensemble Predictions," *IEEE Transactions on Power Systems*, vol. 17, nro 3, s. 626-632, 2002.
- [28] A. D. Papalexopoulos ja T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Transactions on Power Systems*, vol. 5, nro 4, s. 1535-1547, 1990.
- [29] R. Ramanathan, R. Engle, C. W. Granger, F. Vahid-Araghi ja C. Brace, "Short-run forecasts of electricity loads and peaks," *International Journal of Forecasting*, vol. 13, nro. 2, s. 161-174, 1997.
- [30] P. Wang, B. Liu ja T. Hong, "Electric load forecasting with recency effect: A big data approach," *International Journal of Forecasting*, vol. 32, nro. 3, s. 585-597, 2016.
- [31] J. Xie, Y. Chen, T. Hong ja T. D. Laing, "Relative Humidity for Load Forecasting Models," *IEEE Transactions on Smart Grid*, vol. 9, nro. 1, s. 191-198, 2018.
- [32] R. J. Hyndman ja G. Athanasopoulos, *Forecasting: Principles and Practice, 2nd edition*, Melbourne, Australia: OTexts, 2018.
- [33] S.-J. Huang ja K.-R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Transactions on Power Systems*, vol. 18, nro. 2, s. 679-679, 2003.
- [34] T. Hastie ja R. Tibshirani, "Generalized Additive Models," *Statistical Science*, vol. 1, nro. 3, s. 297-318, 1986.
- [35] R. Nedellec, J. Cugliari ja Y. Goude, "GEFCom2012: Electric load forecasting and backcasting with semi-parametric models,"

- International Journal of Forecasting*, vol. 30, nro. 2, s. 375-381, 2014.
- [36] Y. Goude, R. Nedellec ja N. Kong, "Local Short and Middle Term Electricity Load Forecasting With Semi-Parametric Additive Models," *IEEE Transactions on Smart Grid*, vol. 5, nro. 1, s. 440-446, 2013.
- [37] J. W. Taylor ja P. E. McSharry, "Short-Term Load Forecasting Methods: An evaluation Based on European Data," *IEEE Transactions on Power Systems*, vol. 22, nro. 4, s. 2213-2219, 2007.
- [38] H. S. Hippert, C. E. Pedreira ja R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, nro. 1, s. 44-55, 2001.
- [39] C. Kuster, Y. Rezgui ja M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustainable Cities and Society*, vol. 35, s. 257-270, 2017.
- [40] A. Khotanzad, R. Afkhami-Rohani ja D. Maratukulam, "ANNSTLF - Artificial Neural Network Short-Term Load Forecaster - Generation Three," *IEEE Transactions on Power Systems*, vol. 13, nro. 4, s. 1413-1422, 1998.
- [41] T. Hong ja P. Wang, "Fuzzy interaction regression for short term load forecasting," *Fuzzy Optimization and Decision Making*, vol. 13, nro. 1, s. 91-103, 2014.
- [42] A. Al-Kandari, S. Soliman ja M. El-Hawary, "Fuzzy short-term electric load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 26, nro. 2, s. 111-122, 2004.
- [43] N. I. Sapankevych ja R. Sankar, "Time Series Prediction Using Support Vector Machines: A Survey," *IEEE Computational Intelligence Magazine*, vol. 3, nro. 2, s. 24-38, 2009.
- [44] P.-F. Pai ja W.-C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electric Power Systems Research*, vol. 74, nro. 3, s. 417-425, 2005.
- [45] C.-C. Hsu ja C.-Y. Chen, "Regional load forecasting in Taiwan - applications of artificial neural networks," *Energy Conversion and Management*, vol. 44, nro. 12, s. 1941-1949, 2003.
- [46] B.-J. Chen, M.-W. Chang ja C.-J. Lin, "Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001," *IEEE Transactions on Power Systems*, vol. 19, nro. 4, s. 1821-1830, 2004.

- [47] T. Hastie, R. Tibshirani ja J. Friedman, *The Elements of Statistical Learning: Data Mining, inference and prediction*, Stanford: Springer, 2009.
- [48] S. B. Taieb ja R. J. Hyndman, "A gradient boosting approach to the Kaggle load forecasting competition," *International Journal of Forecasting*, vol. 30, nro. 2, s. 383-394, 2014.
- [49] J. R. Lloyd, "GEFCom2012 hierarchial load forecasting: Gradient boosting machines and Gaussian processes," *International Journal of Forecasting*, vol. 30, nro. 2, s. 369-374, 2014.
- [50] M. Kaakinen ja N. Ellonen, Regressioanalyysi, teoksessa "Kvantitatiivisen tutkimuksen verkkokäsikirja," Tampere: Yhteiskuntatieteellinen tietoaarkisto. [Verkkoaineisto]. (Viitattu 4.8.2022). Saatavissa: <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/regressio/analyysi/>
- [51] K.-B. Song, Y.-S. Baek, D. H. Hong ja G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Transactions on Power Systems*, vol. 20, nro. 1, s. 96-101, 2005.
- [52] Motiva, "Aurinkosähkön perusteet," https://www.motiva.fi/ratkaisut/uusiutuva_energia/aurinkosahko/aurinkosahkon_perusteet (Viitattu 3.8.2022).
- [53] Ilmatieteen laitos, "Ilmatieteen laitoksen avoin data ja lähdekoodi", <https://www.ilmatieteenlaitos.fi/avoin-data> (Viitattu 3.8.2022).
- [54] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," *International Journal of Forecasting*, vol. 16, nro. 4, s. 437-450, 2000.
- [55] StataCorp, *Stata Statistical Software: Release 17*, College Station, TX: StataCorp LLC, 2021.
- [56] Tilastokeskus, "Tilastollinen merkitsevyys," https://www.stat.fi/meta/kas/til_merkitsevy.html (Viitattu 23.8.2022).