

Helen Sähköverkko Oy:n tilaustutkimus  
TUNTIMITTAUS- JA AVOIMEEN DATAAN PERUSTUVA  
KUORMITUSMALLINNUS

Antti Mutanen, Kimmo Lummi ja Pertti Järventausta  
Sähkötekniikan yksikkö, Tampereen yliopisto  
syyskuu 2020

# Sisällysluettelo

1) Johdanto .....	1
2) Verkostolaskennassa käytettävien kuormituskäyrien päivitys (vaihe 1) .....	2
2.1. Tausta ja motivaatio kuormituskäyrien päivitykselle .....	2
2.2. Tavoitteet .....	2
2.3. Tutkimuksessa käytetty klusterointimenetelmä .....	4
2.4. Tutkimusaineisto .....	6
2.5. Klusterien lukumäärän optimointi .....	7
Yksilölliset kuormitusmallit .....	9
Suuret sähkönkuluttajat .....	9
Keskikokoiset sähkönkuluttajat .....	12
Pienet sähkönkuluttajat .....	12
2.6. Kaikkien käyttöpaikkojen klusterointi .....	13
Klustereiden nimeäminen .....	14
Vertailu päivitettyihin Sener-käyriin .....	17
Klusterikuvauksista puuttuvien käyttäjäryhmien luokittelu .....	18
2.7. Luokittelun varmuusluku .....	18
2.8. Liittymien klusterointi .....	22
2.9. Indeksisarjojen muodostus .....	27
2.10. Topografioiden muodostus .....	28
3) Yhteenveto .....	30
Lähteet .....	30

## 1) Johdanto

Helen Sähköverkko Oy:lle (HSV) on kertynyt tuntimittausdataa asiakkaiden sähkönkulutuksesta jo usean vuoden ajan. Lisäksi käytettävissä on avointa dataa, mm. rakennustietoja Helsingin kaupungin rakennusrekisteristä ja säähavaintoja Ilmatieteen laitoksen avoimen datan palvelusta. Näitä tietoja on jo yhdistelty HSV aikaisemmissa sisäisissä projekteissa siten, että rakennustiedot on saatu yhdistettyä sähköliittymiin. Lisäksi maalämpökohteista löytyy erillinen vuonna 2013 diplomityön yhteydessä tehty listaus.

Näitä tietoja halutaan käyttää parantamaan jakeluverkon kuormitusten mallinnusta. HSV:llä ei ole aikaisemmin hyödynnetty etäluettavilta sähkömittareilta kerättyä tuntimittausdataa osana laajamittaista kuormituskäyrien päivitystä ja käytössä on edelleen vanhat SLY:n indeksisarjat. Nämä verkostolaskennassa käytettävät kuormituskäyrät ovat auttamatta vanhentuneita. Lisäksi asiakkaiden ryhmittely eri SLY-käyrille perustuu epävarmaan ja osittain vanhentuneeseen tietoon.

HSV kanssa käytyjen keskusteluiden pohjalta on tunnistettu seuraavat tavoitteet tälle projektille:

- 1) Ensimmäisenä tavoitteena on klusteroida sekä käyttöpaikat että liittymät samalla tavalla käyttäytyviin ryhmiin, jolloin ryhmäkohtaisia kuormitusmalleja voidaan käyttää verkostolaskennan tarkentamiseksi. Tulosten tulee olla vietävissä Trimble NIS -järjestelmään. Myös yksilölliset kuormitusmallit pitää huomioida tarpeen mukaan.
- 2) Toisena tavoitteena on asiakaspalvelun apuvälineenä toimivien vertailukäyrien tuottaminen. Vertailukäyristä pitää nähdä kuinka kunkin asiakkaan sähkönkäyttö eroaa muiden samankaltaisten ja samaan käyttäjäryhmään kuuluvien asiakkaiden keskimääräisestä kulutuksesta. Lisäksi tavoitteena on muodostaa avoimesti jaettavissa olevia tyyppikäyttäjämalleja.
- 3) Kolmantena tavoitteena on tuottaa kuormitusmalleja sähköverkon suunnittelun tarpeisiin. Suunnittelussa käytettävien kuormitusmallien tulee kuvata keskimääräisen sähkönkäytön sijaan nykyaikaisten, tai jopa tulevaisuudessa odotettavissa olevien, sähkönkuluttajien käyttäytymistä. Erityisesti kiinnostuksen kohteena ovat toimistorakennusten sähkönkulutuksen ennustaminen, aurinkosähkökohteet ja sähköisen liikenteen latauspisteet.

Edellä mainitun kolmen päätavoitteen lisäksi keskusteluissa on ollut mukana myös energiavarastolliset asiakkaat sekä tehotariffien aiheuttamat muutokset asiakkaiden kulutuskäyttäytymiseen. Tulevaisuudessa myös sähkönmyyjien tekemällä yö-sähkökuormien ohjauksella voi olla merkittävä vaikutus sähköverkon kuormiin. Näiden asioiden tutkimista ei kuitenkaan ole sisällytetty tämän tutkimuksen tavoitteisiin.

## 2) Verkostolaskennassa käytettävien kuormituskäyrien päivitys (vaihe 1)

### 2.1. Tausta ja motivaatio kuormituskäyrien päivitykselle

Kuormituksen suuruuden ja ajallisen vaihtelun tunteminen on aina ollut tärkeää sähköenergia-alan toimijoille. Tämä tarve ei rajoitu pelkästään kokonaiskuormitukseen, vaan myös kokonaiskuormituksen muodostavien osakuormien tunteminen on tärkeää. Esimerkiksi tariffien suunnittelussa ja jakeluverkon tehonjaon laskennassa tarvitaan tarkkoja kuormitusmalleja. Tyypillisesti kuormitusmalleina käytetään asiakasryhmäkohtaisia kuormituskäyriä. Ryhmittelemällä asiakkaat samalla tavalla käyttäytyviin asiakasryhmiin ja mallintamalla näiden ryhmien keskimääräistä käyttäytymistä, laskennassa käytettävien kuormituskäyrien lukumäärä voidaan pitää kohtuullisena. Asiakasryhmäkohtaista mallinnusta voidaan tarvittaessa täydentää yksittäisten suurkuluttajien erillismalleilla.

Senerin 90-luvulla julkaisemat 46 kuormituskäyrää ovat yksi esimerkki asiakasryhmäkohtaisista kuormituskäyristä. Niitä käytetään erityisesti jakeluverkkojen laskennassa, esimerkiksi verkon suunnittelua ja käytön suunnittelua varten tehtävässä tehonjakolaskennassa sekä operatiivisessa tilaestimoinnissa. Korkean ikänsä ja sähkönkulutuksessa tapahtuneiden muutosten takia Sener-käyrät eivät enää mallinna sähkönkulutusta yhtä hyvin kuin niiden julkaisuvuonna. VTT päivitti Sener-käyriä 2000-luvun alussa (Jalonen 2003), mutta näiden päivitysten rajoitetun saatavuuden takia myös vanhoja Sener-käyriä käytetään yhä.

Korkean iän lisäksi Sener-käyrät kärsivät useista muistakin ongelmista: kuormituskäyrien laskennassa käytetyt otoskoot ovat olleet pieniä, tieto kuormitusten lämpötilariippuvuudesta puuttuu suurelta osin, ja samoilla käyrillä mallinnetaan kuormituksia kaikkialla Suomessa – maantieteellisistä eroista välittämättä. Yksi suurimmista ongelmista on kuitenkin virheet asiakkaiden luokittelussa.

Nykyään asiakkaiden ryhmittely perustuu epävarmaan tietoon asiakkaan tyyppistä ja muista ominaisuuksista esim. lämmitysratkaisusta. Ei ole epätavallista, että merkittävä osa asiakkaista on luokiteltu väärään asiakasryhmään. Tästä puolestaan aiheutuu monia ongelmia: Esimerkiksi kuormituskäyriä laskettaessa heterogeenisten asiakasryhmien kuormituskäyristä tulee tehojen ajallisen risteilyn takia profiililtaan tasaisempia kuin niistä pitäisi ideaalitalanteessa tulla. Tasoittuneessa kuormituskäyrässä huipputeho on pienempi kuin mitä se olisi asiakasryhmän koostuessa vain hyvin homogeenisista kuluttajista ja virheelliset huipputehot vaikuttavat oleellisesti mm. tariffilaskelmiin. Lisäksi ongelmia aiheuttaa se, että osaa asiakkaita mallinnetaan väärällä kuormituskäyrällä.

### 2.2. Tavoitteet

Tutkimuksen ensimmäisen vaiheen tavoitteena on klusteroida kaikki HSV:n 385 000 käyttöpaikkaa ja 35 000 liittymää tuntimittausdatan perusteella samalla tavalla käyttäytyviin asiakasryhmiin. Klusteroinnin jälkeen löydetyille asiakasryhmille lasketaan niiden keskimääräistä kuormitusta mallintavat tyypikuormituskäyrät. Nämä uudet entistä tarkemmat kuormitusmallit mahdollistavat useiden eri prosessien parantamisen. Kuvassa 2.1 on esitetty uusien kuormitusmallien tärkeimmät sovelluskohteet.



Klusteroiduille asiakasryhmille määritetään keskimääräiset kuormituksen odotusarvon ja hajonnan profiilit, sekä lämpötilariippuvuus. Näillä klusterikäyrillä voidaan mallintaa yksittäisiä käyttöpaikkoja tai liittymiä tai monen käyttöpaikan tai liittymän summatehoja. Tuloksena saatavia klusterikäyriä voidaan tarvittaessa käyttää myös verrattain tarkkaan kuormituksen ennustamiseen esim. seuraavan päivän lämpötilaennusteita hyödyntäen, olettaen tietenkin, että verkkoyhtiön käytössä olevat järjestelmät osaavat tätä ominaisuutta hyödyntää.

Tuloksena saatavat klusterikäyrät ilmaistaan Trimblen laskentasovellusten käyttäminä indeksisarjoina. Tuloksen annetaan Excel-tiedostoina, joista tietojen vienti Trimblen tietokantaan on mahdollista.

Käyttöpaikka- ja liittymäkohtaisia klusterikäyriä vertaillaan. Ei kuitenkaan ole odotettavissa, että liittymäkohtaiset tyyppikuormituskäyrät johtaisivat tarkempaan verkostolaskentaan. Liittymäkohtaisten kuormitusmallien mahdolliset edut ovat muualla, mm. siinä, että verkkotietojärjestelmän tarvitsee tehdä vähemmän summauksia tai siinä että suunnittelussa on helpompi käyttää esim. kerrostalon summakulutuksen kuvaavaa liittymämallia kuin koostaa sen kulutus N:stä käyttöpaikasta.

Vaiheen 1 sisältää myös esiselvityksen vaiheiden 2 ja 3 toteutettavuudesta, jotta seuraavien vaiheiden toteutuksesta ja tarkemmasta sisällöstä pystytään päättämään (raportoidaan erikseen).

### 2.3. Tutkimuksessa käytetty klusterointimenetelmä

Klusterointi on datan organisointiin tähtäävä data-analysiteknikka. Klusterointialgoritmit jakavat havainnot aliryhmiin (klustereihin) siten että samaan klusteriin kuuluvat havainnot ovat samankaltaisia ja eri klustereihin kuuluvat erilaisia. Havaintojen samankaltaisuutta ja erikaltaisuutta mitataan yleensä jonkinlaisella etäisyysmitalla, esimerkiksi Euklidisella etäisyydellä. Klusterianalyysin ulostulona on tyyppillisesti jaottelu, jossa kaikki havainnot on luokiteltu johonkin klusteriin. Klusteroinnin tuloksena saadaan tietää mitkä havainnot ovat samankaltaisia ja kaikkia samaan klusteriin kuuluvia havaintoja voidaan mallintaa klusterin keskipisteellä. Tämä mahdollistaa tiedon tiivistämisen, koska useita samankaltaisia ja samaan klusteriin kuuluvia havaintoja voidaan lukuisten yksilöllisten mallien sijaan mallintaa (tiivistetymin) yhdellä klusterin keskipisteellä.

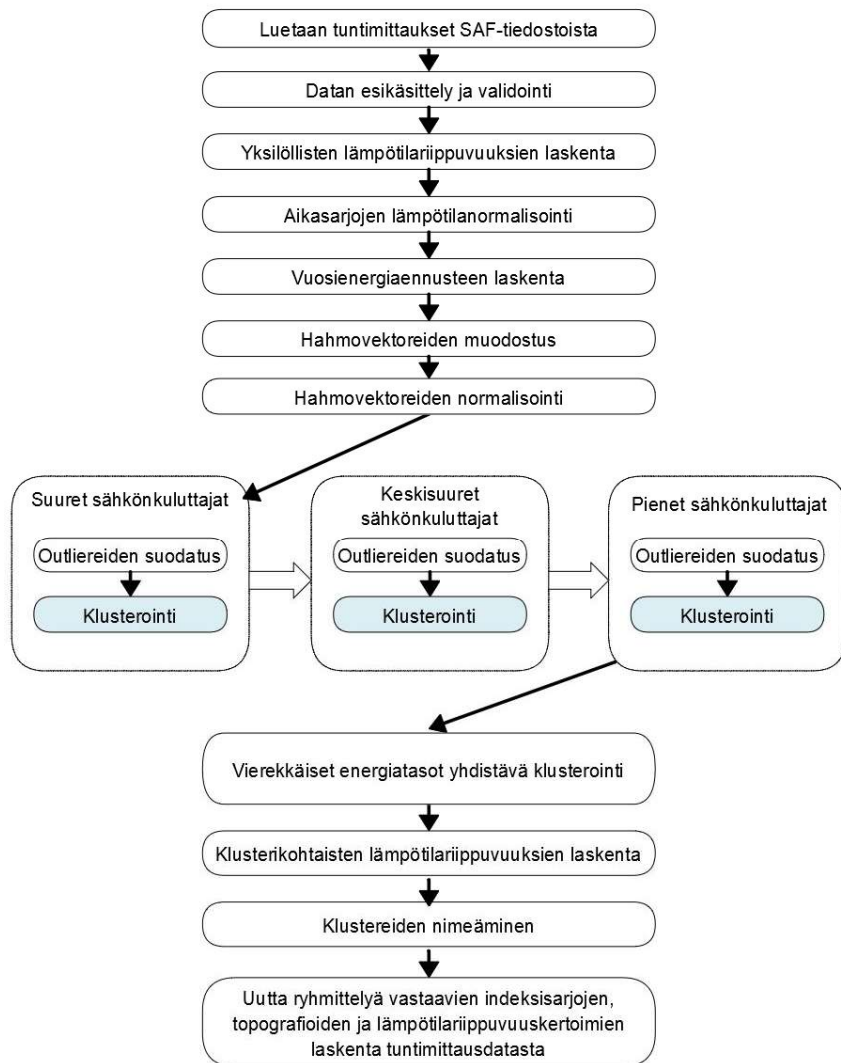
Kirjallisuudessa on esitetty lukuisia klusterointialgoritmeja ja monia niistä on kokeiltu myös sähkönkuluttajien luokitteluun. Klusterointialgoritmin lisäksi eri sähkönkuluttajien klusterointimenetelmät voivat erota siinä minkä suureiden perusteella klusterointi suoritetaan. Monissa ulkomaalaisissa julkaisuissa klusterointi tehdään joko päivittäistä kulutusta kuvaavien vektoreiden (24–96 tunneittaista, puolitunneittaista tai neljännestunneittaista arvoa) tai niistä jotenkin tiivistetyn tiedon perusteella. Datan tiivistykseen voidaan käyttää esimerkiksi pääkomponenttianalyysejä tai muita matemaattisia menetelmiä. Päivittäisten kuormitusvektoreiden klusteroinnissa on se huono puoli, että silloin klusterointi joudutaan toistamaan erikseen jokaiselle tarkasteltavalle tyyppipäivälle ja vuodenaikalle. Tämä monimutkaistaa klusteroinnin tulosten käyttöä ja ymmärrettävyyttä. Sama asiakas voi kuulua moneen eri klusteriin, riippuen päivätyypistä ja vuodenaikasta.

Tässä tutkimuksessa käytetään Antti Mutasen väitöskirjassa (Mutanen 2018) kehitettyä klusterointimenetelmää. Tämä menetelmä perustuu muokattuun k-means algoritmiin ja hahmovektoreihin, jotka samanaikaisesti tiivistävät dataa ja kuvaavat koko vuoden keskimääräistä kulutusta. Nämä hahmovektorit sisältävät yhden tyyppiviikon jokaiselle kahdelletoista kuukaudelle. Kun mittausaika on yksi tunti, tarkoittaa se, että hahmovektorissa on  $24 \times 7 \times 12 = 2016$  alkia.

Käytetyn klusterointimenetelmän lohkokaavio on esitetty kuvassa 2.2. Kaikki kuvassa esitetyt toimenpiteet suoritetaan Matlab-ohjelmalla. Sähkönkuluttajien klusterointi alkaa mittausdatan

luennalla ja jatkuu datan esikäsittelyllä ja validoinnilla. Tässä tapauksessa mittaukset luetaan SAF-tiedostoista ja tallennetaan matriiseina Matlabin käyttämiin m-tiedostoihin. Tämän jälkeen kaikille sähkökuluttajille lasketaan yksilölliset lämpötilariippuvuuskerroimet ja mittausaikaasarjat normalisoidaan keskimääräisiin pitkän ajan kuukausikeskilämpötiloihin (30 vuoden keskiarvot). Lämpötilanormalisointi tehdään, jotta eri vuosina, ja siten myös eri lämpötiloissa, mitattuja sähkökulutuksia pystytään käsittelemään yhdenvertaisina. Myös kunkin sähkökuluttajan vuosienenergia-arvio lasketaan lämpötilaltaan keskimääräisen vuoden mukaan.

Keskimääräistä sähkökulutusta kuvaavat hahmovektorit lasketaan niin ikään lämpötilanormalisoiduista mittauksista. Seuraavaksi hahmovektorit normalisoidaan, jotta erisuuruisten sähkökuluttajien kuormitusprofileita voidaan verrata keskenään. Normalisointi suoritetaan siten että jokaisen hahmovektorin keskitenhoksi tulee tasan yksi. Tieto sähkökulutuksen suuruudesta säilytetään erillisessä vektorissa, joka sisältää kunkin kuluttajan vuosienenergia-arviot.



**Kuva 2.2.** Käytetyn klusterointimenetelmän vaiheet.

Käytetty klusterointimenetelmä on kaksivaiheinen. Ensimmäistä vaihetta käytetään outliereiden (poikkeavien hahmovektoreiden) suodatukseen. Ne hahmovektorit, jotka eivät sovi yhteenkään

ensimmäisessä klusterointivaiheessa löydettyyn klusteriin, poistetaan tilapäisesti datajoukosta. Outlier-suodatuksen raja-arvo valitaan niin että poikkeavin 10 % hahmovektoreista suodatuu.

Klusteroinnin toisessa vaiheessa klusterointi toistetaan ilman edellä suodatettuja outliereita. Tällä menettelyllä varmistetaan, että harvalukuiset, mutta erittäin poikkeavat hahmovektorit eivät vaikuta klusteroinnin tuloksiin. Klusterointia tasapainottaa myös se, että tässä käytetyssä muokatussa k-means menetelmässä klusterikeskukset lasketaan painotettuina keskiarvoina. Painoarvoina käytetään sähkönkuluttajien vuosienergia-arvioita. Suuret sähkönkuluttajat vaikuttavat siis enemmän klusterikeskusten muodostumiseen kuin pienet sähkönkuluttajat. Klusteroinnin toisen vaiheen jälkeen edellä suodatetut outlierit ja mittausaikasarjojen aukkojen takia vajaiksi jääneet hahmovektorit luokitellaan niitä parhaiten mallintavaan klusteriin.

Lopulliset kuormitusprofiilit lasketaan kuhunkin klusteriin kuuluvien asiakkaiden tuntimittauksista ja kuormitusprofiili esitetään indeksisarjana sekä topografiana. Jokaiselle klusterille lasketaan myös klusterikohtainen lämpötilariippuvuus ja keskihajonta, joka kuvaa kuormitusprofiilin tarkkuutta.

## 2.4. Tutkimusaineisto

Tässä tutkimuksessa on käytössä sähkönkulutuksen tuntimittaukset kolmen vuoden ajalta (2016–2018) kaikkiaan 394 366 kulutusta sisältävältä käyttöpaikalta ja 207 tuotantoa sisältävältä käyttöpaikalta. Tuntimittaukset on toimitettu SAF-tiedostoina, jotka tutkijat ovat lukeneet ja tallentaneet matriiseina Matlabin käyttämiin m-tiedostoihin. Mittausaikasarjat tallennettiin matriiseihin, joissa on 26304 riviä, yhtä monta kuin mittausajanjaksolla on tunteja, ja yhtä monta saraketta kuin on mitattuja käyttöpaikkoja. Tallennusmuoto on hyvin kompakti, sillä rivien aikaleimat voidaan tallentaa yhteen ainoaan infomatriisiin. Tästä huolimatta, tietokoneessa, jossa on 32 Gt keskusmuistia, datan esikäsittely jouduttiin jakamaan neljään osaan. Vasta hahmovektoreiden muodostamisen jälkeen koko asiakasmassaa pystyttiin käsittelemään yhtenä kokonaisuutena.

Tuntimittaustietojen lisäksi HSV:ltä saatiin seuraavat käyttöpaikkakohtaiset tiedot:

- Käyttöpaikkatunnus
- Liittymätunnus
- Vaiheiden lukumäärä ja sulakekoko (A)
- Siirtotuotetunnus
- Forum business group
- Luokittelu verkkotietojärjestelmässä (≈SLY/Sener-ryhmä)

Pientuotantoa sisältävistä liittymistä oli lisäksi tiedossa tuotannon nimellisteho ja käyttöönottopäivä.

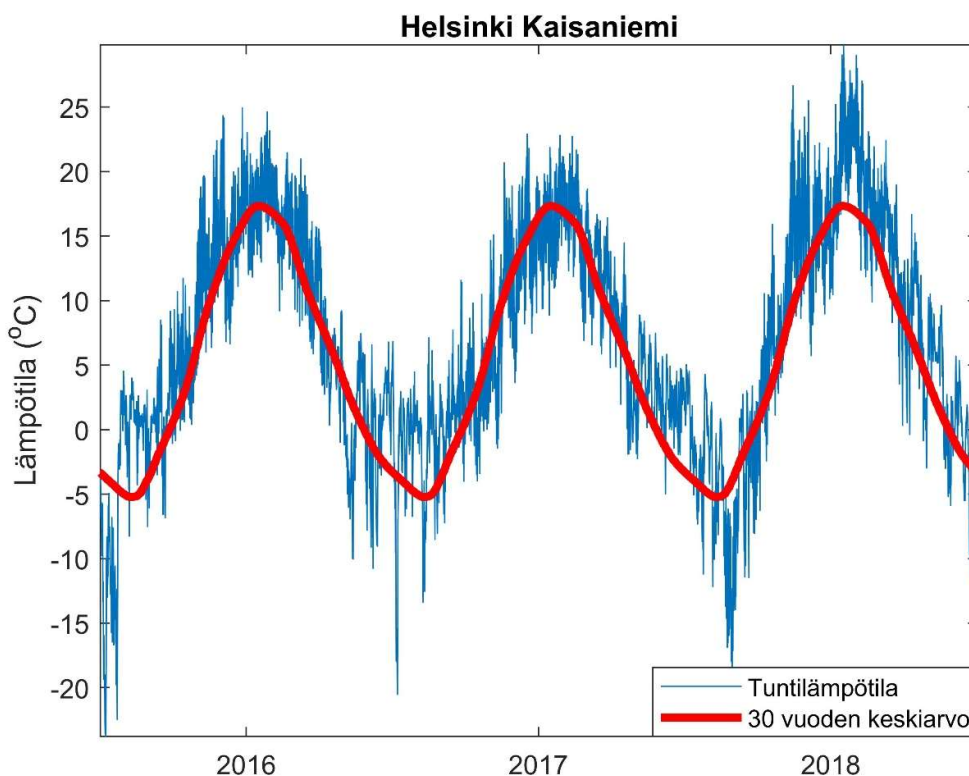
Lisäksi käytettävissä oli liittymiin valmiiksi yhdistetyt kiinteistö ja rakennustiedot, jotka sisälsivät mm.

- Kiinteistötunnukset
- Rakennustunnukset
- Rakennuksen tyyppi
- Rakennuksen käyttötarkoituksen
- Rakennusten pinta-alat
- Kantavan rakenteen materiaalin
- Julkisivumateriaalin
- Lämmitystavan
- Kaupunginosan



HSV:ltä saadun mittaus- ja asiakastiedon lisäksi tässä tutkimuksessa käytettiin Ilmatieteen laitoksen avoimesta tietokannasta haettuja tunneittaisia lämpötilamittauksia. Helsingin alueen pitkän ajan (30 vuotta) kuukausikeskilämpötilat etsittiin niin ikään avoimena data saatavasta keskilämpötilataulukosta. Kuvassa 2.3 on verrattu mitattua lämpötilaa pitkän ajan keskiarvoon.

HSV:n tuntimittausmateriaali todettiin hyvälaatuiseksi. Klustereiden laskennassa käytettiin vain mittausjärjestelmässä statuksella 125 (laskettu) tai 150 (mitattu) olevia mittauksia. Näillä hyväksi luokitelluilla mittauksilla saatiin laskettua täydelliset hahmovektorit 97,9 % käyttöpaikoista. Klustereiden muodostumiseen ja lopullisten kuormitusmallien käyrämuotoihin vaikutti siis vain edellä mainitut 97,9 % käyttöpaikoista. Loppujen 2,1 % käyttöpaikoista hahmovektoreita täydennettiin mittausjärjestelmässä statuksella 25 (epäkelpo), 40 (puuttuva), 100 (arvioitu), 200 (asetettu), tai 255 (tyhjä) olevilla mittauksilla. Vaikka kaikille käyttöpaikoille ei edelleenkään saatu laskettua täydellisiä hahmovektoreita, yhtään täysin tyhjää hahmovektoria ei jäänyt jäljelle ja kaikki käyttöpaikat voitiin luokitella. Kaikki mitatut arvot olivat nolliä 1200 käyttöpaikalla. Nämä käyttöpaikat ryhmiteltiin erilliseen "nollaklusteriin".



**Kuva 2.3.** Helsingin lämpötilamittaus ja pitkän ajan keskiarvo.

## 2.5. Klusterien lukumäärän optimointi

Sähkökuluttajien klusteroinnissa on kaksi oleellista kysymystä:

- Mitkä kuluttajat käyttäytyvät samalla tavalla?
- Kuinka moneen segmenttiin kuluttajat jakautuvat?

K-means klusterointialgoritmissa pitää ensin päättää segmenttien lukumäärä, jonka jälkeen algoritmi laskee, mitkä havainnot ovat samankaltaisia ja kuuluvat siten samaan segmenttiin. On olemassa myös klusterointialgoritmeja, jotka väittävät löytävänsä optimaalisen klusterimäärän automaattisesti.

Nämä algoritmit vaativat kuitenkin usein alkuparametreja, joiden määrittäminen on yhtä vaikeaa kuin itse klusterimäärän optimointi.

Klustereiden optimaalista lukumäärää voidaan arvioida erilaisten tunnuslukujen avulla. Näitä tunnuslukuja on käsitelty Antti Mutasen väitöskirjassa (Mutanen 2018, luku 4.4.1). Käytännössä toimivimmaksi menetelmäksi on havaittu klusteroinnin laskeminen usealla eri klusterimäärällä ja näistä saatujen virheen neliösummien vertailu. Virheen neliösummat lasketaan tässä tapauksessa kaavalla

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} (c_i - x)^2, \quad (1)$$

missä  $k$  on klustereiden lukumäärä  
 $c_i$  on klusterin  $i$  keskus  
 $x$  on klusteriin  $i$  kuuluva havainto.

Virheen neliösummista ei kuitenkaan saada klustereiden optimaalista määrää suoraan, vaan se pitää määrittää käyrästä, jossa virheen neliösumma on piirretty klusterimäärän funktiona. Tähän käyrään muodostuu tyypillisesti ”polvi” tai ”kynnärpää”, joka ilmaisee alueen, jossa optimaalinen lukumäärä sijaitsee. Kun klusterimäärää kasvatetaan, virheen neliösumma pienenee aluksi hyvin nopeasti. Jossain vaiheessa lisäklustereista saatava hyöty kuitenkin pienenee marginaaliseksi. Optimaalisen klusterimäärän voidaan ajatella sijaitsevan taitekohdassa, jossa lisäklustereista saatava hyöty laskee niistä syntyvää haittaa (mallin monimutkaisuutta) pienemmäksi.

Polven taitekohta voidaan määrittää kuvasta visuaalisesti, tai sen määrittämiseen voidaan käyttää erilaisia menetelmiä, joita on esitelty ja testattu lähteessä (Mutanen 2018, luku 4.4.2). Parhaiten polven taitekohtaa kuvaa käyrän kaarevuuden maksipiste. Jatkuvalle funktiolle  $f$ , kaarevuus  $K_f(x)$  voidaan määrittää ensimmäisen ja toisen derivaatan avulla (Satopää et al. 2006):

$$K_f(x) = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}} \quad (2)$$

Vaikka jatkuvien funktioiden kaarevuus voidaan määrittää tarkasti, kaarevuutta ei ole helppo laskea diskreetille käyrälle. Ratkaisuna tähän ongelmaan, on mahdollista sovittaa diskreettiin käyrään jokin sitä hyvin kuvaava jatkuva funktio ja laskea kaarevuus kyseiselle sovittelulle. Tässä tutkimuksessa käyrään, joka kuvaa virheen neliösummaa klusterimäärän funktiona, sovitetaan funktio

$$f(x) = ax^b + cx + d. \quad (3)$$

Soveltamalla kaavaa (1), tämän funktion kaarevuudeksi saadaan

$$Kf(x) = \frac{(ab^2 - ab)x^{b-2}}{(1 + (abx^{b-1} + c)^2)^{3/2}}. \quad (4)$$

Sovitteen laskennassa käytettiin Matlab funktiota *fit*. Joissain tapauksissa yksi tai kaksi sovitettavan käyrän ensimmäistä arvoa jouduttiin jättämään pois sovitteen laskennasta. Tämä oli näissä tapauksissa tarpeen, jotta käyrälle saatiin riittävän hyvä sovite. Koska etsittävä taitekohta ei yleensä osu kahden ensimmäisen arvon kohdalle, ei näiden arvojen poisjättäminen vaikuta menetelmän tarkkuuteen merkittävästi.

### Yksilölliset kuormitusmallit

Ennen klusteroinnin aloittamista, klusteroitavien käyttöpaikkojen joukosta poistettiin 10 vuosienergialtaan suurta ja poikkeuksellisesti käyttäytyvää sähkökuluttajaa. Näille suurimmille sähkökuluttajille tullaan muodostamaan yksilölliset kuormitusmallit, jotta näitä tärkeitä ja klustereihin huonosti sopivia kulutuspesteitä pystytään mallintamaan mahdollisimman tarkasti. Vuosienergialtaan suurten ja poikkeuksellisesti käyttäytyvien käyttöpaikkojen erottelu takasi myös, etteivät ne pääse dominoimaan energiapainotteista klusterointia tai aiheuta vain yhden käyttöpaikan sisältävien klustereiden muodostumista. Taulukossa 2.1 on annettu yksilöllisesti mallinnettavien käyttöpaikkojen tiedot.

**Taulukko 2.1.** Yksilöllisesti mallinnettavat sähkökuluttajat.

Data poistettu julkisesta raportista.

Käyttöpaikka	Liittymännumero	Vuosienergia (GWh)	Kuvaus

### Suuret sähkökuluttajat

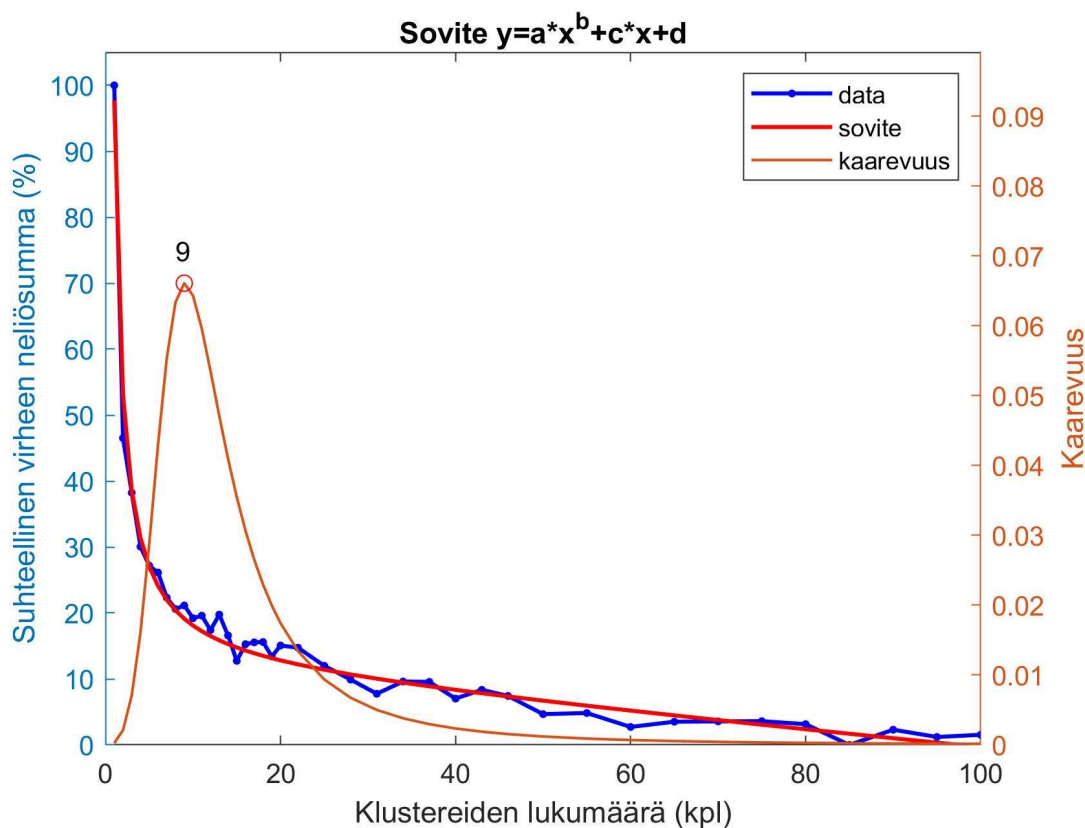
Klustereiden optimaalista lukumäärää lähdettiin selvittämään siten että ensin selvitetään optimaalinen lukumäärä suurille sähkökuluttajille ( $E > 1 \text{ GWh/a}$ ), sitten keskisuurille sähkökuluttajille ( $50 \text{ MWh/a} \leq E \leq 1 \text{ GWh/a}$ ) ja viimeiseksi vähän kuluttaville asiakkaille ( $E < 50 \text{ MWh/a}$ ). Tällä tavalla tätä haastava tehtävä saatiin jaettua pienempiin ja helpommin hallittaviin osiin. Lisäksi tämä lähestymistapa mahdollisti erikokoisten asiakkaiden ominaispiirteiden vertailun. Erikokoisten sähkökuluttajien lukumäärät ja jakautuminen eri siirtotuotteille on esitetty taulukossa 2.2.

**Taulukko 2.2.** Erikokoisten sähkökuluttajien jakautuminen eri siirtotuotteille.

<b>Suuret sähkökuluttajat (E&gt;1 GWh)</b>	
110 kV -teho	5 kpl
KJ-teho	451 kpl
PJ-teho	108 kpl
Siirtotuote ei tiedossa	17 kpl
<b>yht.</b>	<b>581 kpl</b>
<b>Keskikokoiset sähkökuluttajat</b>	
KJ-teho	221 kpl
PJ-teho	5088 kpl
Ohjattu yösiirto	74 kpl
Aikasiirto	377 kpl
Yleissiirto	594 kpl
Siirtotuote ei tiedossa	309 kpl

yht.	6663 kpl
<b>Pienet sähkökuluttajat (E&lt;50 MWh)</b>	
KJ-teho	9 kpl
PJ-teho	3781 kpl
Ohjattu yösiirto	1179 kpl
Aikasiirto	16333 kpl
Yleissiirto	354358 kpl
Siirtotuote ei tiedossa	3029 kpl
<b>yht.</b>	<b>378689 kpl</b>

Suuret sähkökuluttajat klusteroitiin klusterimäärillä 1–100 ja tuloksista piirrettiin virheen neliösummaa kuvaava käyrä. Kuvassa 2.4 on esitetty virheen neliösumma klustereiden lukumäärän funktiona. Silmämääräisen arvion perusteella optimaalinen klusterimäärä on noin 8–12. Käyrään sovitetulle potenssifunktiolle laskettu kaarevuus ja sen maksimiarvo tarkentavat tätä arviota lukumäärään 9. Sovitteen parametrit olivat:  $a=86,3098$ ;  $b=-1,0423$ ;  $c=-0,1256$ ;  $d=11,5269$ .



**Kuva 2.4.** Suurten sähkökuluttajien optimaalisen klusterimäärän selvitys.

Kun optimaalinen klusterimäärä oli saatu selvitettyä, KJ-asiakkaat klusteroitiin yhdeksään segmenttiin käyttäen luvussa 2.3 esiteltyä menetelmää. Tuloksena saaduille klustereille laskettiin asiakasmäärät, keskimääräiset vuosienergiat ja kokonaisenergiat. Lisäksi klustereille annettiin sanallinen kuvaus tarkastelemalla niiden käyrämuotoja ja niihin kuuluvien asiakkaiden käyttötarkoituksia. Nämä tiedot on esitetty taulukoissa 2.3 ja 2.4, joista ensimmäinen sisältää tiedot NIS järjestelmästä poimitusta yleisimmästä asiakastyypistä ja jälkimmäinen kiinteistötiedoista poimituista suurimmista käyttötaparyhmistä energian ja lukumäärän mukaan luokiteltuna.

Kiinteistötiedoista ei ollut apua suurten asiakkaiden ryhmittelyssä. Lähes kaikkien suurten kiinteistöjen rakennusmateriaali on betoni, lämmönjakotapa vesikierto, lämmitysmuoto kaukolämpö ja käyttötapa toimistorakennus.

Yleisesti ottaen suuret asiakkaat käyttävät sähköä samalla tavalla, kulutus on kaikissa klustereissa suurinta päiväaikaan ja viikonloppuna arkea pienempää.

**Taulukko 2.3.** Suurten asiakkaiden klusterit. *Klusterin 7 dataa poistettu julkisesta raportista.*

Klusteri	N (kpl)	E <sub>ave</sub> (GWh/a)	E <sub>tot</sub> (GWh/a)	Yleisin asiakastyyppi	Kuvaus
1	113	1,79	203	Toimisto/palvelu	Toimisto/palvelu, suuri vuorokausivaihtelu, auki arkisin (7-18)
2	74	4,20	311	Toimisto/palvelu	Toimisto/palvelu, pieni vuorokausivaihtelu, auki arkisin (7-18)
3	62	2,63	163	Toimisto/palvelu	Toimisto/palvelu, auki arkisin (7-19), positiivinen lämpötilariippuvuus (ilmastointi)
4	78	2,40	187	Toimisto/palvelu	Toimisto/palvelu, negatiivinen lämpötilariippuvuus, auki arkisin (7-18)
5	77	2,61	201	Tavaratalot ja marketit	Tavaratalot ja marketit, pitkät aukioloajat (7-22), auki viikonloppuisin (8-22 & 9-22)
6	32	3,75	120	Toimisto/palvelu	Tavaratalot ja marketit, lyhyet aukioloajat (8-22), auki viikonloppuisin (9-20 & 10-19)
7				Metro- ja raitiovaunuliikenne	Metro- ja raitiovaunuliikenne
8	57	2,51	143	Virkistys- ja Kulttuuripalvelut	Virkistys- ja kulttuuripalvelut, lämpötilariippuva
9	49	3,51	172	Toimistorakennus	Suuri ja tasainen vuoden- ja vuorokaudenajasta riippumaton kulutus

**Taulukko 2.4.** Suurten asiakkaiden kiinteistötiedoista saadut käyttötapatiedot.

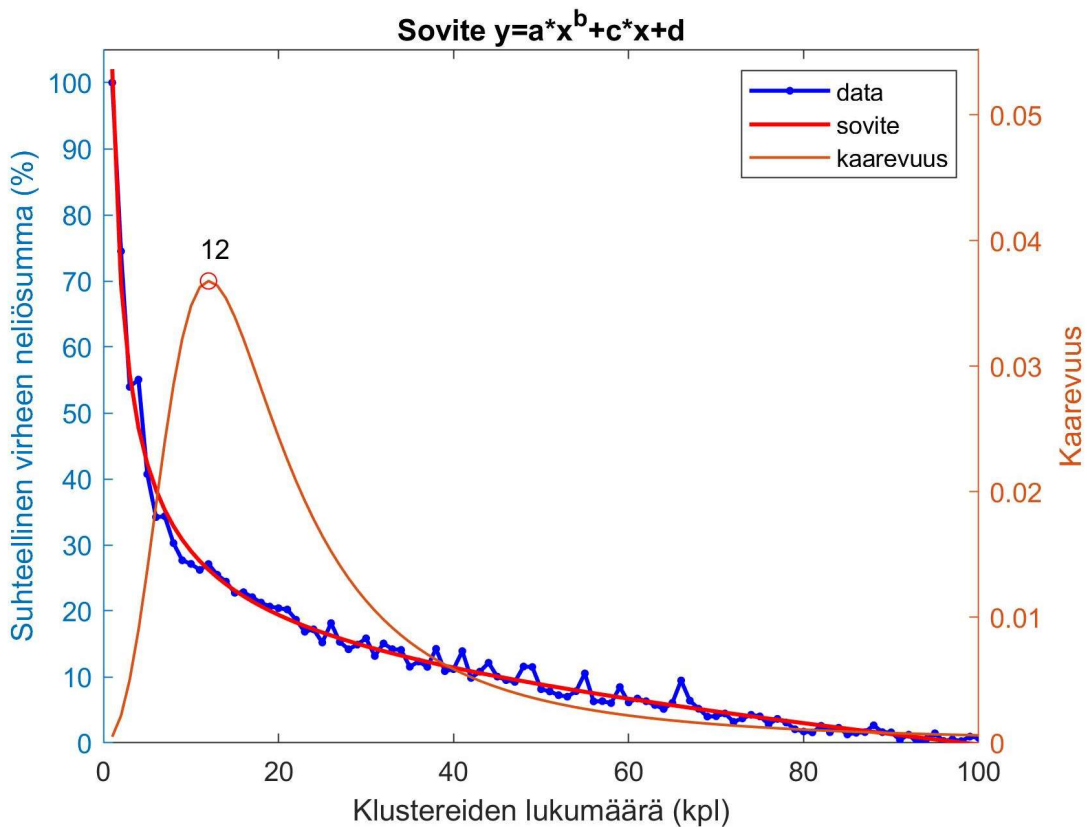
Klusteri	Käyttötapa (energia)	Käyttötapa (lkm)
1	Toimistorakennukset	Toimistorakennukset
2	Toimistorakennukset / kauppa / teollisuus	Toimistorakennukset / kauppa / teollisuus
3	Toimistorakennukset	Toimistorakennukset
4	Toimistorakennukset / teollisuus	Toimistorakennukset
5	Toimistorakennukset	Toimistorakennukset
6	Teollisuus / toimistorakennukset	Toimistorakennukset
7	Toimistorakennukset	Toimistorakennukset
8	Toimistorakennukset / korkeakoulut	Toimistorakennukset
9	Hotellit, motellit, kylpylähotellit	Toimistorakennukset / hotelli / teollisuus

### Keskikokoiset sähkökuluttajat

Keskisuuret sähkökuluttajat klusteroitiin klusterimäärillä 1–100 ja kuvaan 2.5 piirrettiin virheen virheen neliösumma klustereiden lukumäärän funktiona. Käyrään sovitetun potenssifunktion kaarevuus saa maksimiarvonsa klusterimäärällä 12, joka on siis optimaalinen lukumäärä keski suurten asiakkaiden klustereille. Sovitteen parametrit olivat:  $a=97,6475$ ;  $b=-0,5797$ ;  $c=-0,1152$ ;  $d=4,5055$ .

Keskisuuria sähkökuluttajia klusteroitaessa huomioitiin jo edellä muodostetut suurten sähkökuluttajien klusterit. Mikäli keski suuri kuluttaja oli lähempänä suurten sähkökuluttajien klusteria kuin yhtäkään keski suurten sähkökuluttajien klusteria, luokiteltiin se suurten sähkökuluttajien klusteriin sen kuitenkään vaikuttamatta suurten sähkökuluttajien klusterikeskuksiin.

Keskikokoisten sähkökuluttajien klustereiden nimeäminen ja analysointi on tehty myöhemmin alaluvussa 2.6.



**Kuva 2.5.** Keski suurten sähkökuluttajien optimaalisen klusterimäärän selvitys.

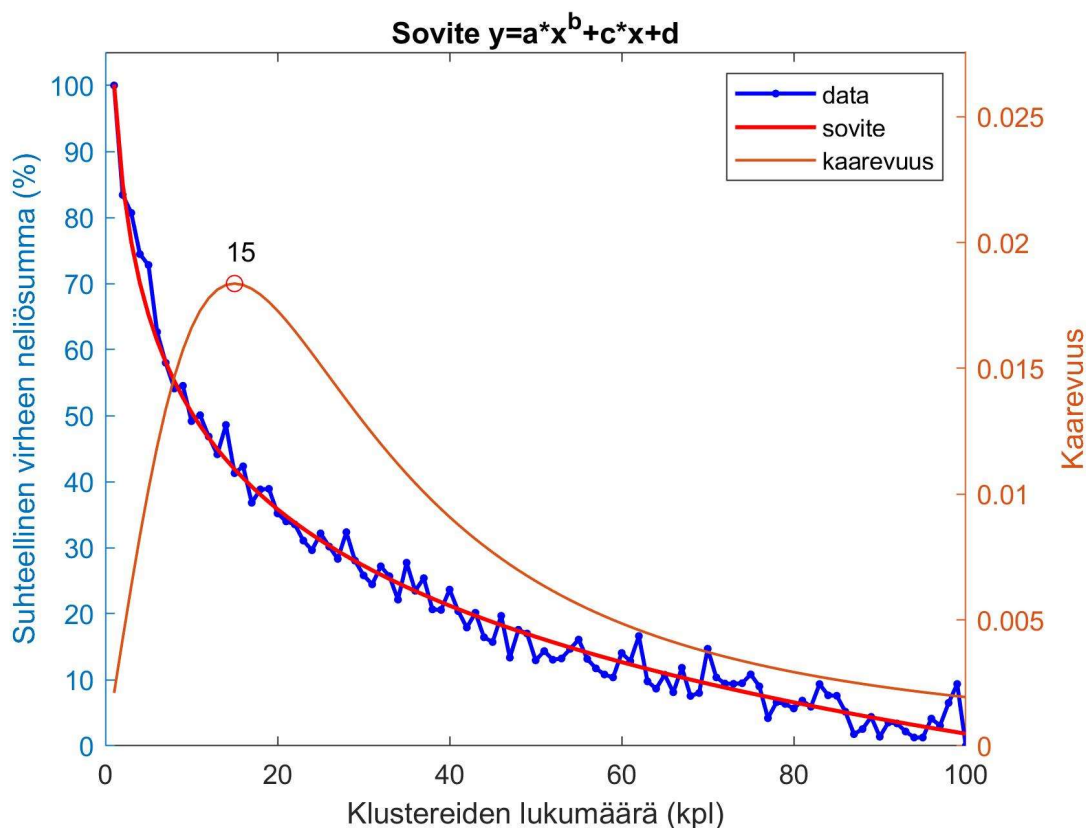
### Pienet sähkökuluttajat

Pienten sähkökuluttajien optimaalista klusterimäärää tutkittiin ottamalla pienistä sähkökuluttajista 10 % satunnaisotos. Asiakkaiden lukumäärän ollessa erittäin suuri, satunnaisotosta klusteroimalla on mahdollista saada koko joukkoa hyvin kuvaavia klustereita. Tässä tutkimuksessa satunnaisotosta on käytetty vain klustereiden optimaalisen lukumäärän selvittämisessä ja lopulliset klusteroinnit on tehty koko asiakasjoukolla. Satunnaisotoksen sähkökuluttajat klusteroitiin klusterimäärillä 1–100 ja kuvaan 2.6 piirrettiin virheen neliösumma klustereiden lukumäärän funktiona. Käyrään sovitetun potenssifunktion kaarevuus saa maksimiarvonsa klusterimäärällä 15, joka on siis optimaalinen

lukumäärä pienten sähkönkuluttajien klustereille. Sovitteen parametrit olivat:  $a=133,52$ ;  $b=-0,0164$ ;  $c=-0,0100$ ;  $d=-123,50$ .

Pieniä sähkönkuluttajia klusteroitaessa huomioitiin jo edellä muodostetut suurten ja keskisuurten sähkönkuluttajien klusterit. Mikäli pieni kuluttaja oli lähempänä suurten tai keskisuurten sähkönkuluttajien klusteria kuin yhtäkään pienten sähkönkuluttajien klusteria, luokiteltiin se suurten tai keskisuurten sähkönkuluttajien klusteriin sen kuitenkään vaikuttamatta suurten tai keskisuurten sähkönkuluttajien klusterikeskuksiin.

Pienten sähkönkuluttajien klustereiden nimeäminen ja analysointi on tehty myöhemmin alaluvussa 2.6.



**Kuva 2.6.** Pienten sähkönkuluttajien optimaalisen klusterimäärän selvitys.

## 2.6. Kaikkien käyttöpaikkojen klusterointi

Edellä tehdyn analyysin perusteella optimaalinen klusterimäärä HSV:n käyttöpaikoille on 36 klusteria ja kun tähän lisätään vielä "nollaklusteri", jolla huomioidaan käyttöpaikat, joissa ei mittauksen perusteella ole lainkaan kulutusta, saadaan lopulliseksi klusterimääräksi 37. Seuraavaksi kaikki käyttöpaikat klusteroidaan näihin 37 klusteriin, käyttäen klusteroinnin alutilana edellä määritettyjä suurten, keskisuurten ja pienten sähkönkuluttajien klusterikeskuksia.

Alaluvussa 2.5 määritettiin erikseen klusterit suurille, keskisuurille ja pienille käyttöpaikoille. Käytännössä erot näiden asiakasryhmien välillä ovat liukuvia, esimerkiksi osa keinoitekoisesti vedetyn vuosienenergiarajan perusteella suureksi sähkönkäyttäjäksi luokitelluista asiakkaita käyttäytyy samalla tavalla kuin osa keskisuuriksi sähkönkäyttäjiksi luokitelluista asiakkaista ja päinvastoin. Tämän vuoksi

kaikkia käyttöpaikkoja klusteroitaessa asiakkaiden sallittiin, tietyin rajoituksin, ylittää edellä mainitut vuosienenergiaan sidotut ryhmäraajat. Rajoitteet olivat:

- 1) Suuret sähkökäyttäjät eivät voi siirtyä pienten sähkökäyttäjien klustereihin
- 2) Pienet sähkökäyttäjät eivät voi siirtyä suurten sähkökäyttäjien klustereihin.

Nämä rajoitteet otettiin käyttöön, jotta klusteroinnin tuloksia käytettäessä ei tulisi tilanteita, joissa joudutaan pohtimaan esim., että mistä syystä pieni kerrostaloasunto on luokiteltu suureksi toimistorakennukseksi. Erityisesti joidenkin pienten sähkökuluttajien kulutus on hyvin satunnaista ja mikään klusteri ei kuvaa näiden käyttäytymistä hyvin. Tässä tilanteessa, ilman rajoituksia, asiakas saatetaan luokitella lähes mihin tahansa klusteriin. Klusteroinnin tuloksiin tulisikin suhtautua aina pienellä varauksella, koska klusterikeskus muodostuu klusteriin kuuluvien asiakkaiden keskimääräisen käyttäytymisen mukaan ja yksittäisten asiakkaiden käyttäytyminen voi poiketa paljonkin klusteriin kuuluvien asiakkaiden keskimääräisestä käyttäytymisestä.

Vajaat NaN-arvoja (Not-a-Number, puuttuva arvo, vastaa tietokannoissa käytettyä tyhjääarvoa NULL) sisältävät hahmovektorit luokiteltiin lähimpään klusteriin käytettävissä olleiden arvojen perusteella ja asiakkaan vuosienenergia estimoitui oletuksella, että hahmovektorin puuttuva osa käyttäytyy kuten lähin klusterikeskus. Tällä tavoin kaikki käyttöpaikat, joista oli vähänkin tuntimittauksia, saatiin luokiteltua.

#### Klustereiden nimeäminen

Siinä missä asiakkaiden luokittelu ja klusterikeskusten laskenta on tietokoneella tehtävää laskentaa, on klustereiden nimeäminen edelleen manuaalista asiantuntijatyötä. Taulukkoon 2.5 on koottu perustiedot kaikista 37 klusterista ja niille annetut sanalliset kuvaukset. Taulukossa kullekin klusterille on annettu klusteriin luokiteltujen käyttöpaikkojen lukumäärä, klusteriin kuuluvien käyttöpaikkojen keskimääräiset ja yhteenlasketut energiat, klusterin yleisin asiakastyyppi sekä klusterin sanallinen kuvaus. Klusterin (lukumääräisesti) yleisin asiakastyyppi poikkeaa joissain tapauksissa klusterille annetusta sanallisesta kuvauksesta. Varsinkin pienten sähkökuluttajien klustereissa, kerrostaloasunnot dominoivat lukumäärällisesti useita klustereita, vaikka niiden osuus klusterin kokonaisenergiasta oli muita käyttäjäryhmiä pienempi. Näissä tapauksissa klusterin nimeämisessä painotettiin energialtaan suurinta käyttäjäryhmää. Näin toimittiin esimerkiksi klusterin 31 kohdalla, joka sisälsi lukumääräisesti eniten kerrostaloasuntoja, mutta jonka energiasta suurin osa tuli omakotitaloista .

Klusterin sanallisten kuvausten muodostuksessa hyödynnettiin lähinnä verkkotietojärjestelmästä saatuja Sener- ja Business-luokkatietoja. Rakennustiedoista tarkasteltiin rakennuksen tyyppiä, käyttötapaa, lämmitystapaa, polttoainetta, rakennusainetta ja julkisivumateriaalia. Näistä ainoastaan käyttötapatieto oli jossain määrin hyödyllinen klustereiden nimeämisessä, ja tämäkin tieto oli usein päällekkäistä Sener- ja Business-luokkatiedoista saadun informaation kanssa. Muista rakennustiedoista oli hyvin vähän hyötyä klustereiden nimeämisessä. Ainoastaan klustereiden 31 ja 32 välistä eroa pohdittaessa, lämmitystapatiedoista nähtiin hyödyllinen tieto, että klusterissa 32 on huomattavasti enemmän sähkölämmitettyjä rakennuksia kuin klusterissa 31.

Taulukossa 2.5 on uudelleennumeroitu ja -järjestetty aikaisemmissa vaiheissa löydetyt klusterit, siten että samanlaisen kuvauksen omaavat klusterit ovat vierekkäin ja taulukon lukeminen on helpompaa. Sama järjestys on käytössä myös tämän tutkimuksen tulostiedostoissa. Varjostuksen väri kertoo mihinkä suuruusluokkaan kaikkien asiakkaiden klusteroinnin pohjana olleet suuruusluokittain jaotellut klusterit kuuluivat. Suurten sähkökuluttajien klusterit on väritetty harmaalla, keskiuurten punaisella ja pienten vihreällä. Jotkin klusterit, esimerkiksi klusteri 24, keräsivät niin paljon asetettujen



luokkarajojen yli siirtyneitä asiakkaita, että klusterin lopullinen kokoluokka vaihtui. Taulukossa 2.6 on esitetty yksilöllisesti mallinnettavat käyttöpaikat, joiden kuormitusmallinumerot on uudelleen numeroitu alkamaan lukuarvosta 1001, jotta ne eivät myöhemmin sekoitu klusterimalleihin.

Klustereiden kuvauksissa esiintyy usein termi ”vuorokausivaihtelu”, tällä tarkoitetaan kuinka voimakkaasti päivä- ja yöajan kulutukset eroavat toisistaan. Vuorokausivaihtelun suuruus on useissa tapauksissa muuten samankaltaisia klustereita erottavat tekijä. Kuva 2.7 havainnollistaa, kuinka klusterit 2–5 eroavat vuorokausivaihtelun osalta. Kuvasta nähdään, että kaikilla näillä viidellä klusterilla on samankaltainen päivä- ja viikkorytmi, mutta päivä- ja yökulutusten välillä on erisuuruisia eroja.

**Taulukko 2.5.** Kaikkien klustereiden perustiedot. *Klusterin 12 ja 17 dataa poistettu julkisesta raportista.*

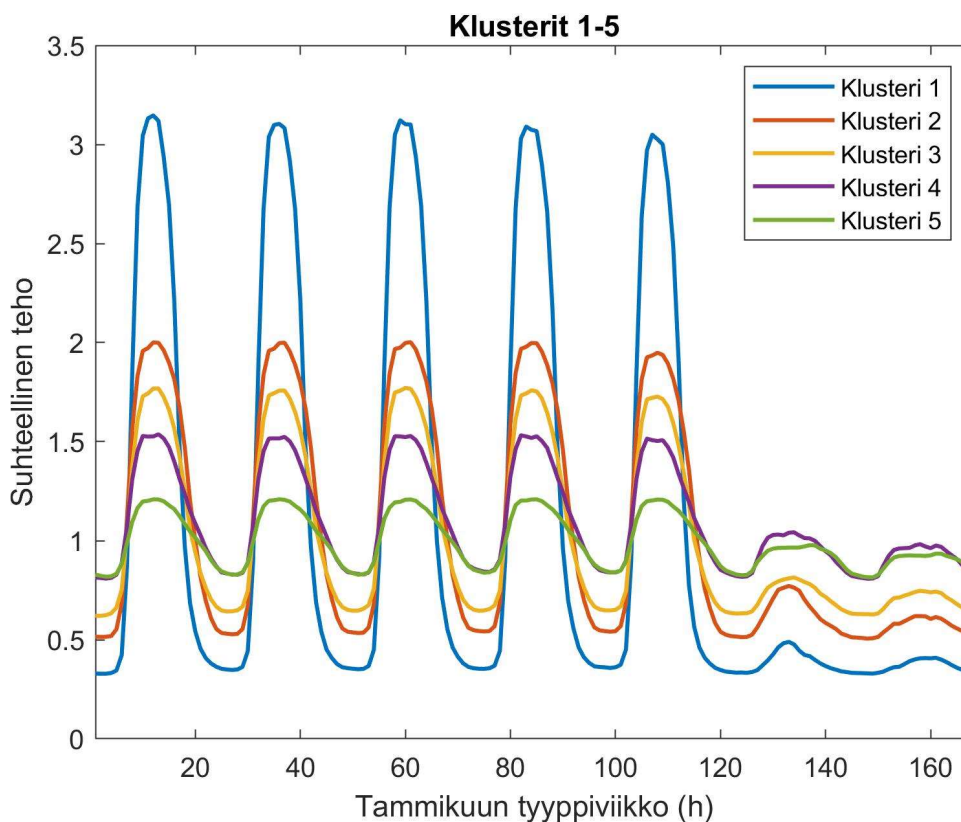
Klusteri	N (kpl)	E <sub>ave</sub> (MWh/a)	E <sub>tot</sub> (GWh/a)	Yleisin asiakastyyppi	Kuvaus
1	3213	10,2	32,9	Toimisto/palvelu	Toimisto/palvelu, pieni vuosienergia, erittäin suuri vuorokausivaihtelu, auki arkisin (8-17)
2	267	579,1	154,6	Toimisto/palvelu	Toimisto/palvelu, suuri vuorokausivaihtelu, auki arkisin (7-18)
3	356	686,7	244,5	Toimisto/palvelu	Toimisto/palvelu, auki arkisin (7-18)
4	298	962,2	286,7	Toimisto/palvelu	Toimisto/palvelu, pieni vuorokausivaihtelu, auki arkisin (7-18)
5	327	1274,0	416,6	Toimisto/palvelu	Toimisto/palvelu, erittäin pieni vuorokausivaihtelu, auki arkisin (7-18)
6	165	1365,2	225,3	Toimisto/palvelu	Toimisto/palvelu, pieni vuorokausivaihtelu, auki arkisin (7-19), positiivinen lämpötilariippuvuus (ilmastointi)
7	260	990,9	257,6	Palvelut	Palvelu ja kauppa, auki arkisin ja viikonloppuisin, ei lämpötilariippuvuutta
8	310	513,3	159,1	Tavaratalot ja marketit	Tavaratalot ja marketit, pitkät aukioloajat (7-22), auki viikonloppuisin (8-22 & 9-22), positiivinen lämpötilariippuvuus (kylmälaitteita)
9	5434	6,1	32,9	Toimisto/palvelu	Vähittäiskauppa, rajoitettu aukiolo viikonloppuisin
10	10699	5,2	55,9	Kerrostaloasunto	Ravintola/kahvila, toimintaa kaikkina viikonpäivinä
11	17480	10,0	174,2	Kerrostaloasunto	Virkistys- ja kulttuuripalvelut, auki arkisin ja viikonloppuisin
12				Metro- ja raitiovaunuliikenne	Metro- ja raitiovaunuliikenne

13	444	340,1	151,0	Kerrostalo	Vesi- ja energiahuolto + muu verrattain tasainen yhteiskunnan vuorokausirytmii seuraava kulutus
14	3261	13,0	42,3	Kerrostaloasunto	Vesi- ja energiahuolto, negatiivinen lämpötilariippuvuus
15	2131	48,4	103,2	Toimisto/palvelu	Koulut ja muu julkinen palvelu, suuri vuorokausivaihtelu
16	2404	44,7	107,5	Toimisto/palvelu	Koulut ja muu julkinen palvelu, pieni vuorokausivaihtelu
17				Ulkovalaistus	Ulkovalaistus
18	3630	17,7	64,2	Kerrostaloasunto	Rakentaminen/työmaa
19	5001	10,0	50,1	Kerrostaloasunto	Työmaa (kerrostalo)
20	9515	15,4	146,6	Kerrostaloasunto	Kerrostalokiinteistö
21	4326	29,1	125,8	Kerrostalokiinteistö	Kerrostalokiinteistö, lauantaista kohti kasvava iltakulutus
22	2405	8,3	20,1	Pien-/talokiinteistö	Kerros-/rivitalokiinteistö, suuret autonlämmityspiikit talviaamuisin
23	33016	2,1	70,3	Kerrostaloasunto	Kerrostaloasunto, suuri vuorokausivaihtelu, huippukuormitus klo 18-19
24	30953	5,3	163,2	Kerrostaloasunto	Kerrostaloasunto, pieni vuorokausivaihtelu, huippukulutus klo 18-19
25	25640	2,4	62,0	Kerrostaloasunto	Kerrostaloasunto, suuri vuorokausivaihtelu, huippukuormitus klo 19-20
26	31391	2,3	70,7	Kerrostaloasunto	Kerrostaloasunto, huippukulutus klo 19-20
27	37934	1,9	71,2	Kerrostaloasunto	Kerrostaloasunto, suuri vuorokausivaihtelu, huippukuormitus klo 20-21
28	34175	3,3	112,8	Kerrostaloasunto	Kerrostaloasunto, huippukulutus klo 20-21
29	45220	2,2	98,1	Kerrostaloasunto	Kerrostaloasunto, asukas ei käy päivätöissä, huippukulutus klo 20-21
30	33991	1,9	64,8	Kerrostaloasunto	Kerrostaloasunto, asukas ei käy päivätöissä, iltavirkku, huippukulutus klo 21-22
31	12452	8,8	109,5	Kerrostaloasunto	Omakotitalo, ei sähkölämmitystä
32	6340	22,2	141,0	OK-/paritalo	Omakotitalo, suora sähkölämmitys
33	7577	16,6	125,7	OK-/paritalo	OK-/pari-/rivitalo, käyttöveden lämmitys yösähköllä
34	1358	20,8	28,2	OK-/paritalo	Omakotitalo, varaava sähkölämmitys, aikasiirto
35	941	29,7	28,0	OK-/paritalo	Omakotitalo, varaava sähkölämmitys, ohjattu yösiirto
36	18237	6,5	119,2	Kerrostaloasunto	Vakiotehokuorma
37	1176	0,0	0,0	Kerrostaloasunto	Ei kulutusta

**Taulukko 2.6.** Yksilöllisesti mallinnettavien käyttöpaikkojen perustiedot.

Data poistettu julkisesta raportista.

Yksilöllinen Klusteri	E (GWh/a)	Asiakastyyppi	Kuvaus

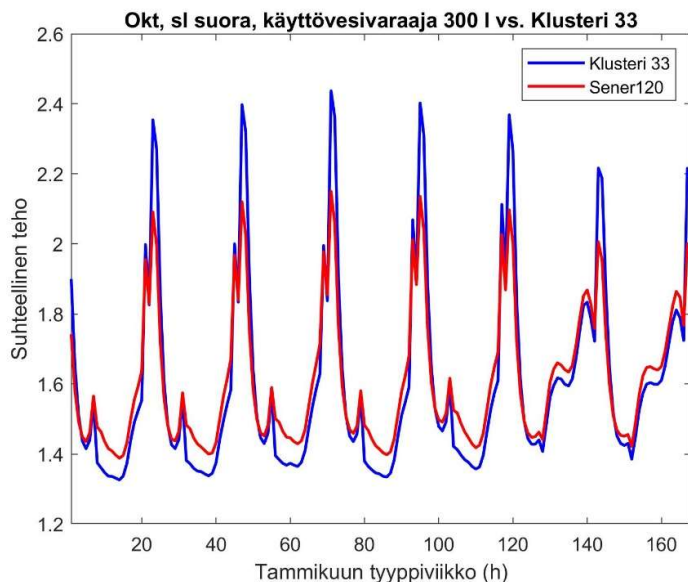


**Kuva 2.7.** Klustereiden 1–5 vuorokausivaihtelun vertailu.

#### Vertailu päivitettyihin Sener-käyriin

Klusteroinnin vaihtoehtona olisi ollut päivittää Sener-käyrät olemassa olevaa luokittelua ja AMR mittauksia hyödyntäen. Tälläkin tavalla olisi saatu käyttökelpoisia kuormituskäyriä, vaikka käyttöpaikkojen luokittelussa Sener-ryhmiin on jonkin verran virheitä. Nämä virheet näkyvät kuormituskäyriissä siten, että samoja AMR mittauksiakin käytettäessä päivitettyjen Sener-käyrien vuorokausivaihtelu jää klusterikäyriä pienemmäksi. Tätä on havainnollistettu kuvassa 2.8. Päivitettyjen sener-käyrien pienemmästä vuorokausivaihtelusta johtuen, myös niiden perusteella

arvioitujen verkon huipputehot jäävät, yhtenäisillä kuormitusalueilla, klusterikäyrien perusteella arvioitua huipputehoja pienemmiksi.



**Kuva 2.8.** Päivitetyin Sener-ryhmän 120 ja klusterin 33 vertailu.

#### Klusterikuvauksista puuttuvien käyttäjäryhmien luokittelu

Uusia asiakkaita verkkotietojärjestelmään lisätessä sähkökuluttajat joudutaan mittaustiedon sijaan luokittelemaan muiden saatavilla olevien tietojen avulla, esimerkiksi kulutuspuisteen käyttötarkoituksen mukaan. Kaikille käyttäjäryhmille ei kuitenkaan löydy omaa klusteria taulukosta 2.5, joten tätä luokittelua helpottamaan taulukkoon 2.7 on koottu muutamia puuttuvia käyttäjäryhmiä (keskimäärin) parhaiten kuvaavat klusterit.

**Taulukko 2.7.** Puuttuvien käyttäjäryhmien lähimmät klusterit.

Käyttäjärühmä	Lähin klusteri
Rivitaloasunto	31-35 lämmitysmuodon mukaan
1-vuoro teollisuus	3
Automyynti ja huoltamatoiminta	3
Sairaalat ja terv.hoito	4
Yleis- ja muu hallinto	4
Julkinen hallinto/järjestyksenpito	4
P-alueet, autotallit	4
Varastot	4
Hotelli- ja majoitustoiminta	5
Posti-/teleliikenne	5
Järjestö-/uskonnollinen toiminta	11
Liikennevalot	13

#### 2.7. Luokittelun varmuusluku

Kuten edellä todettiin, jotkin asiakkaat saattavat sopia hyvin heikosti siihen klusteriin, johon asiakas on lopulta luokiteltu. Toisaalta, osa asiakkaista kuuluu hyvin vahvasti juuri siihen klusteriin, johon

nämä on luokiteltu. Tulosten tulkintaa helpottamaan kehitettiin luokittelun varmuusluku, joka kertoo prosentuaalisena lukuna, kuinka todennäköisesti asiakas on luokiteltu oikeaan klusteriin. Tämän luokittelun varmuusluvun laskenta perustuu fuzzy C-means algoritmin sisältämään jäsenyyden laskentaan. K-means algoritmissa havainnot voivat kuulua vain yhteen klusteriin, mutta fuzzy C-means algoritmissa havainnot kuuluvat jollain jäsenyydellä kaikkiin klustereihin. Jäsenyyksien summa on aina yksi ja asiakas luokitellaan siihen klusteriin, johon sillä on suurin jäsenyys. Jäsenyydet sisältävä matriisi ( $w_{ij}$ ) lasketaan kaavalla

$$w_{ij} = \frac{1}{\sum_{k=1}^C \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (5)$$

missä  $x_i$  on havainto  $i$

$c_j$  on klusterikeskus  $j$

$C$  on klustereiden lukumäärä

$m$  on sumeutuskerroin.

Tässä tapauksessa havainnot ovat asiakkaiden kulutusta kuvaavia hahmovektoreita, klusterikeskukset ovat K-means menetelmällä jo määritetyt klusterikeskukset, joiden lukumäärä on 37 ja sumeutuskerroinena on käytetty arvoa kaksi. Sumeutuskerroin  $m$  määrittää kuinka sumeita tulokset ovat. Mitä suurempi  $m$ , sitä tasaisemmin kertoimet  $w$  ovat jakautuneet, eli sitä sumeampi luokittelun tulos on. Kun  $m$  lähestyy yhtä, luokittelusta tulee selvärajaisempi. Rajalla  $m=1$ , havainnot joko kuuluvat tiettyyn klusteriin (jäsenyys=1) tai eivät kuulu (jäsenyys=0), ja luokittelun tulos on sama kuin  $k$ -means algoritmilla.

Kaavalla (5) lasketuista jäsenyyksistä, lasketaan kunkin havainnon  $i$  jäsenyyksille keskiarvo sekä keskihajonta. Olettaen normaalijakauma, näiden perusteella lasketaan sitten mitä kumulatiivista kertymää ( $y$ ) kukin jäsenyyssarvo vastaa ja näiden kertymien perusteella lasketaan varmuusluku  $VL$  kaavalla

$$VL_j = \frac{1}{1 - y_j} \times 100 \% \quad (6)$$

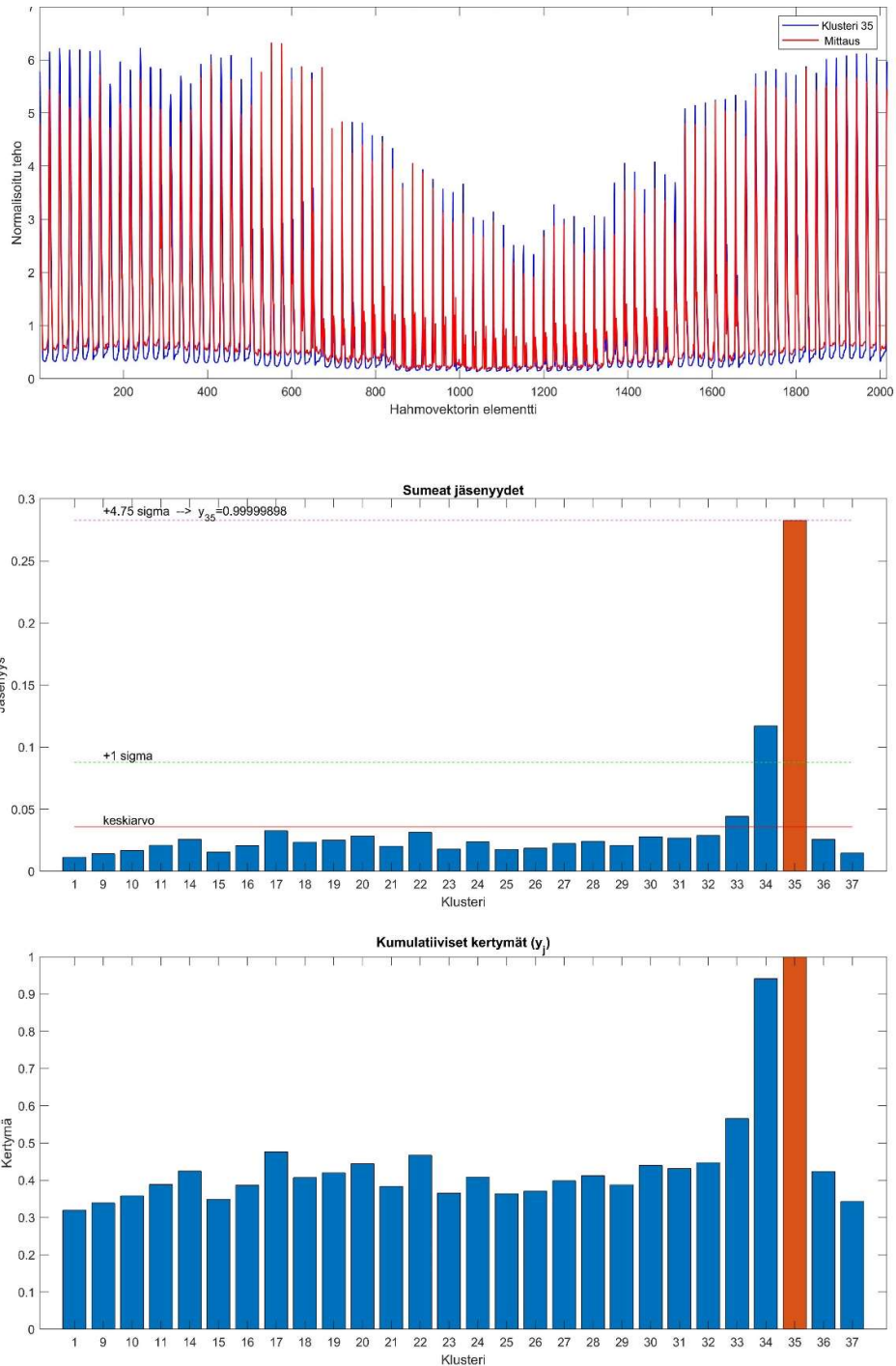
missä  $y_j$  on klusterin  $j$  kertymä

$C$  on klustereiden lukumäärä.

Kuvassa 2.9 on esitetty esimerkki varmuuslukujen laskennan vaiheista. Ylimmässä kuvassa on hahmovektori ja sitä lähinnä oleva klusterikeskus. Nähdään että hahmovektori on hyvin lähellä klusterikeskusta. Keskimmaisessä kuvassa näkyy hahmovektorin jäsenyydet eri klustereihin, ja alimmaisessa kuvassa kumulatiiviset kertymät, jotka kaavaan (6) syöttämällä saadaan lopullinen luokittelun varmuusluku 99,994 %. Tässä tapauksessa käyttöpaikka on luokiteltu oikeaan klusteriin erittäin suurella todennäköisyydellä. Kuvassa 2.10 on esitetty epäselvempi tapaus, jonka luokittelun varmuusluku on vain 7,1 %. Tämän käyttöpaikan hahmovektori on varsin kaukana kaikista klusterikeskuksista, eikä lähinkään klusterikeskus mallinna sitä hyvin. Pienikin poikkeama mitatussa kulutuksessa olisi saattanut johtaa tämän asiakkaan luokitteluun toisin.

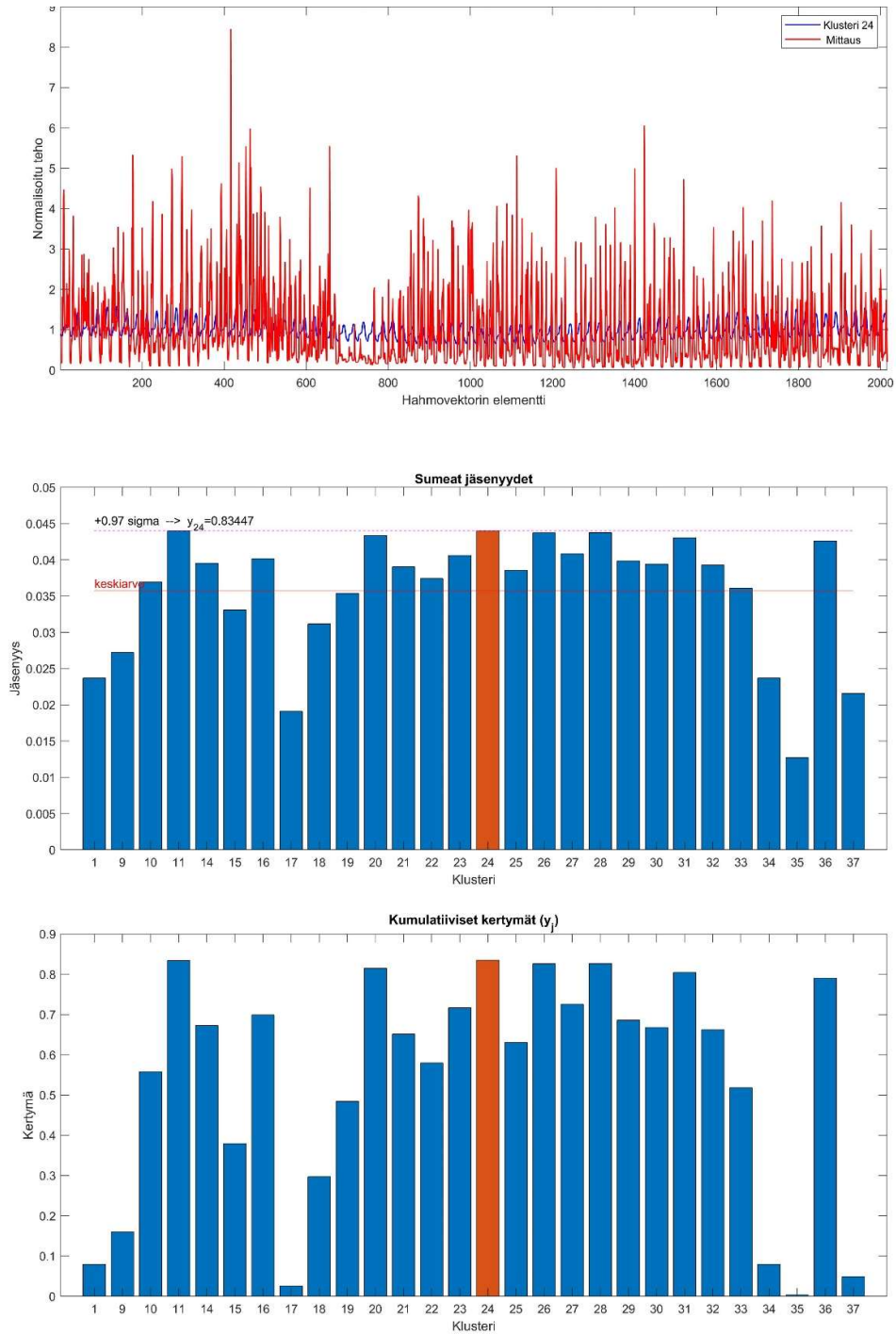
Yleisesti ottaen, kaikki yli 50 % varmuusluvut tarkoittavat, että asiakas on mitä todennäköisimmin luokiteltu oikeaan klusteriin. Yli 50 % varmuuslukuun ylsi yli puolet asiakkaista. Pienempiä

varmuuslukuja voidaan tulkita esimerkiksi siten, että 25 % varmuusluvulla asiakkaalla on ollut neljä suurinpiirtein yhtä sopivaa klusteria ja varmuusluvulla 10 % kymmenen suunnilleen yhtä sopivaa klusteria. Suurin osa asiakkaista pystytään luokittelemaan varsin hyvällä varmuudella. Kuvassa 2.11 on kaikkien luokiteltujen asiakkaiden varmuusluvut suuruusjärjestyksessä. kuvasta 2.12 nähdään kuinka varmuuslukujen keskiarvot riippuvat asiakkaiden vuosienergioista. Tästä kuvasta nähdään, että vuosikulutukseltaan suuret asiakkaat pystytään luokittelemaan varmemmin.

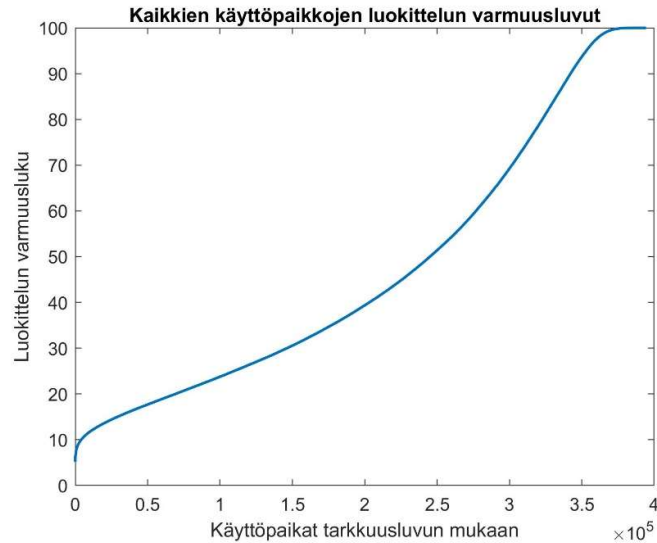


**Kuva 2.9.** Esimerkki suurella todennäköisyydellä oikein luokitellusta käyttöpäikasta "Omakotitalo, varaava sähkölämmitys".

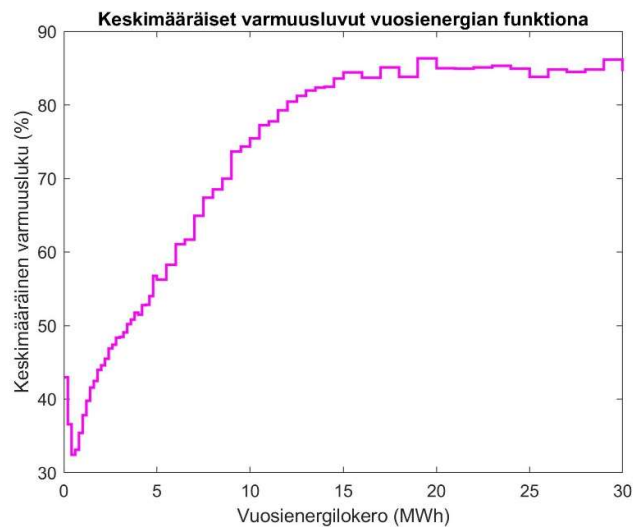
On huomioitava, että tässä esitetty varmuusluku ei ole oikea tilastollinen todennäköisyys sille, onko asiakasluokiteltu oikein vaan ko. arvo on päätöksenteon apuvälineeksi tarkoitettu pseudotodennäköisyys.



**Kuva 2.10.** Esimerkki pienellä todennäköisyydellä oikein luokitellusta käyttöpaikasta "Kerrostaloasunto".



**Kuva 2.11.** Kaikki lasketut luokittelun varmuusluvut suuruusjärjestyksessä.



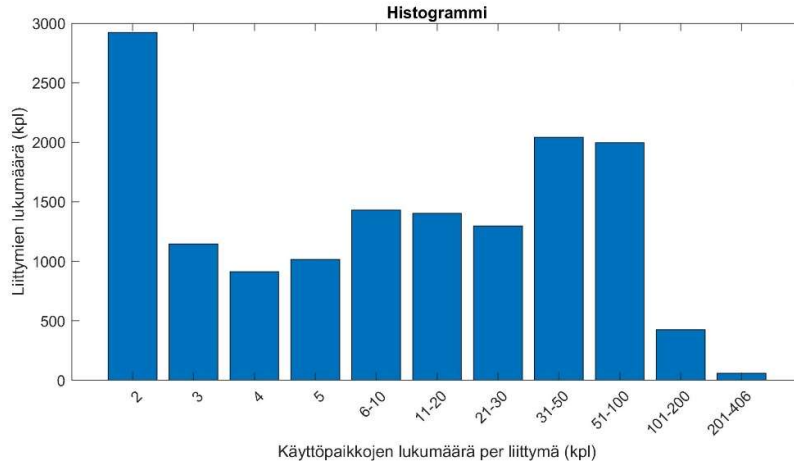
**Kuva 2.12.** Vuosien energian vaikutus keskimääräiseen luokittelun varmuuslukuun.

## 2.8. Liittymien klusterointi

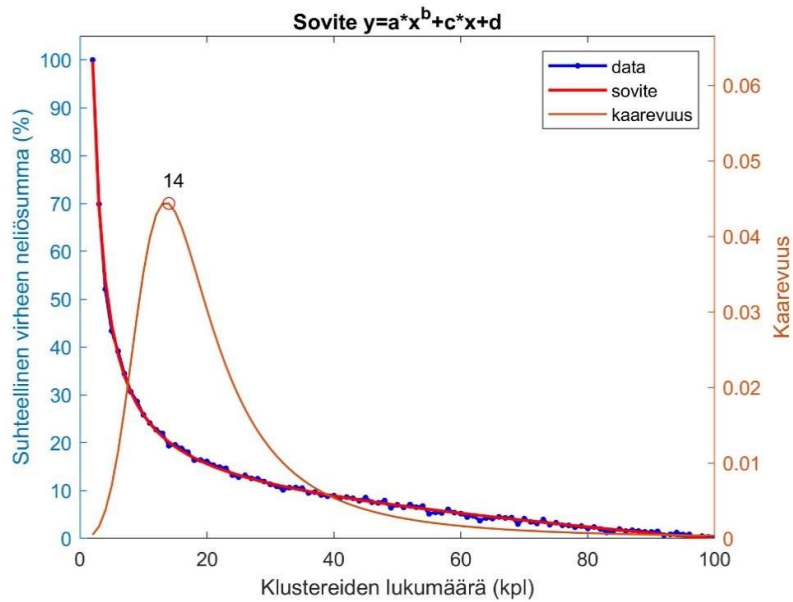
Tässä luvussa klusteroidaan yhtenä massana kaikki liittymät, jotka sisältävät useamman kuin yhden käyttöpaikan. Tällaisia liittymiä on 14649 ja kuvan 2.13 histogrammi havainnollistaa kuinka liittymien käyttöpaikkamäärät jakautuvat tässä joukossa. Yhden käyttöpaikan liittymiä on 18319.

Kuten edellisissä luvuissa, liittymiä klusteroitaessa selvitetään ensin klustereiden optimaalinen lukumäärä. Kuvaan 2.14 piirrettyjen tulosten perusteella optimaalinen lukumäärä on 14. Tähän lisätään vielä yksi "nollaklusteri", johon laitetaan liittymät joiden kaikki mittaukset ovat nolliä. Lopullinen klusterimäärä on siis 15. Ennen klusterointia, taulukossa 2.6 esitetyt yksilöllisesti mallinnettavat käyttöpaikat poistettiin liittymistä.





**Kuva 2.13.** Histogrammi liittymien käyttöpaikkamääristä.



**Kuva 2.14.** Liittymien optimaalisen klusterimäärän selvitys.

Liittymien klusteroinnin tulokset on esitetty taulukossa 2.8. Klustereiden nimeämisessä on jälleen käytetty apuna verkkotietojärjestelmästä saatuja Sener- ja Business-luokkatietoja, taustatietoina saatuja kiinteistö- ja rakennustietoja, sekä kokemuseräistä asiantuntemusta kuormituksista.

**Taulukko 2.8.** Liittymäklustereiden perustiedot.

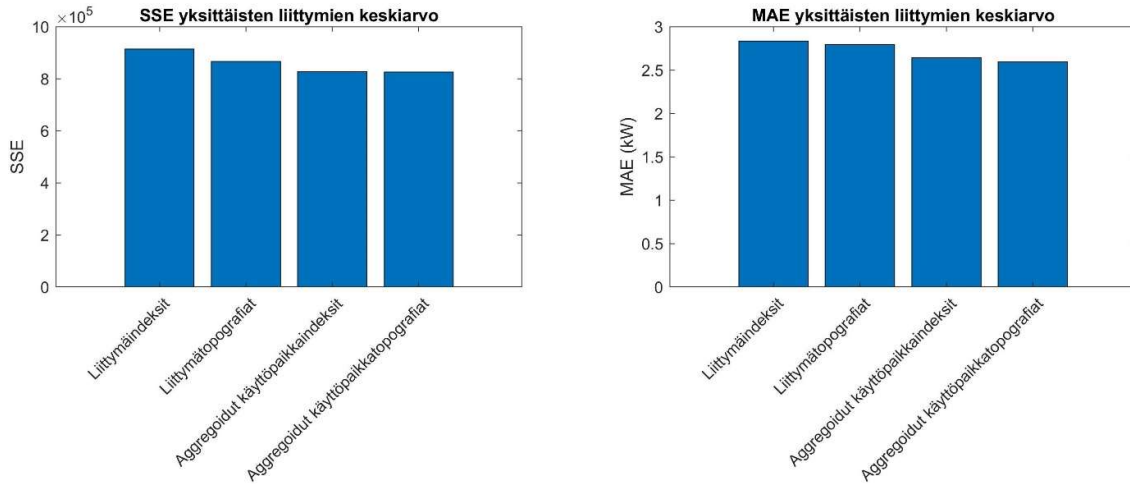
Klusteri	N (kpl)	$E_{ave}$ (MWh/a)	$E_{tot}$ (GWh/a)	Yleisin asiakastyppi	Kuvaus
1	207	299,6	62,0	Toimisto/palvelu	Toimisto/palvelu, suuri vuorokausivaihtelu, auki arkisin (7-18)
2	237	568,0	134,6	Toimisto/palvelu	Toimisto/palvelu, auki arkisin (7-18)
3	94	1138,1	107,0	Teollisuus/toimisto	Teollisuus, 1-vuoro, (pieni vuorokausivaihtelu)

4	263	301,7	79,4	Teollisuus/toimisto	Teollisuus, 1-vuoro, (pieni vuorokausivaihtelu), (lähellä KP klusteria 4), lämpötilariippuva
5	200	726,4	145,3	Vähittäiskauppa	Kaupat ja marketit, auki arkisin ja viikonloppuisin, ei lämpötilariippuvuutta
6	2271	76,1	172,7	Kerrostalo	Kerrostalo, suuri vuorokausivaihtelu jokaisena viikonpäivänä, kulutushuippu klo 19-20
7	1748	155,8	272,3	Kerrostalo	Kerrostalo, suuri vuorokausivaihtelu, lauantaita kohti nouseva kulutus, kulutushuippu klo 18-19
8	1569	138,0	216,5	Kerrostalo	Kerrostalo, pieni vuorokausivaihtelu, kulutushuippu klo 19-20
9	2070	87,8	181,6	Kerrostalo	Kerrostalo, kulutushuippu klo 19-20, osittainen lämpötilariippuvuus
10	1276	84,3	107,5	Kerrostalo	Kerrostalo, pieni vuorokausivaihtelu, kulutushuippu klo 19-20, lämpötilariippuva
11	796	203,5	162,0	Kerrostalo	Kerrostalo, pieni vuorokausivaihtelu, suuri päiväkulutus (asukas ei käy päivätöissä), kulutushuippu klo 17-18
12	1735	57,1	99,0	Rivi-/paritalo	Rivi-/paritalo, suora sähkölämmitys
13	1900	66,7	126,7	Rivi-/paritalo	Rivi-/paritalo, käyttöveden lämmitys yösaikolla, (identtinen käyttöpaikkaklusterin 33 kanssa), huippukulutus klo 22-23(4)
14	278	1257,4	349,6	Kerrostalo	Kerrostalo, suhteellisen tasainen päivän- ja vuodenajasta riippumaton kulutus
15	5	0,0	0,0	Kerrostalo	Ei kulutusta

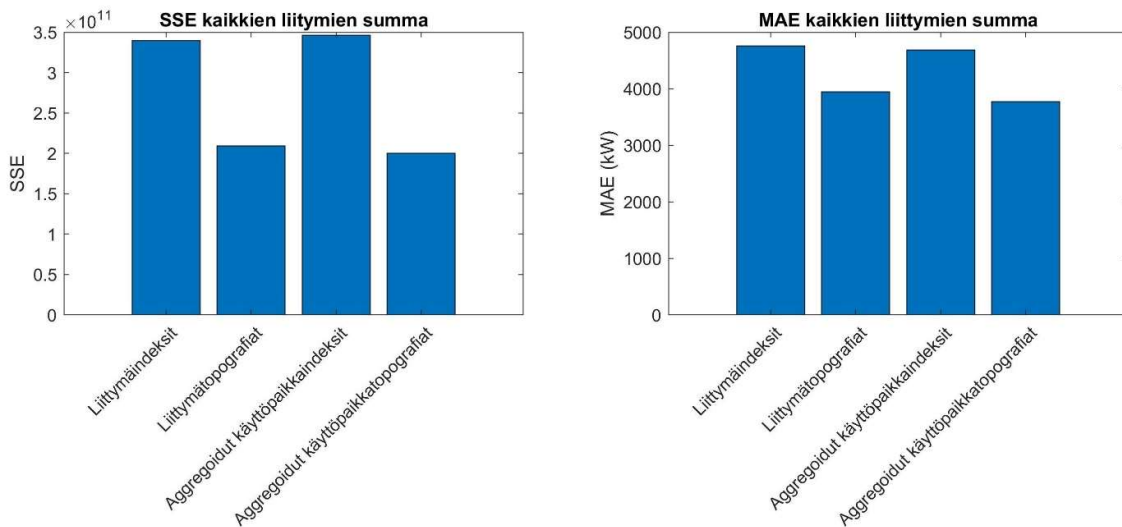
Liittymäklustereiden muodostaminen nostaa esiin mielenkiintoisen kysymyksen, onko liittymien mallinnus tarkempaa liittymäklustereilla vai aggregoiduilla käyttöpaikkaklustereilla? Kuvassa 2.15 on esitetty keskimääräiset tarkkuudet yksittäisten liittymien kuormituksille eri tavoilla mallinnettuna: lähimmän liittymäklusterin indeksisarjalla mallinnettuna, lähimmän liittymäklusterin topografiolla mallinnettuna, aggregoiduilla lähimpien käyttöpaikkaklustereiden indeksisarjoilla mallinnettuna ja aggregoiduilla lähimpien käyttöpaikkaklustereiden topografiolla mallinnettuna. Tulokset kertovat, että liittymiä voidaan mallintaa liittymäklustereita tarkemmin käyttämällä aggregoituja käyttöpaikkaklustereita. Tätä saattaa selittää se, että käyttöpaikkaklustereita muodostettiin tässä tutkimuksessa huomattavasti enemmän (37 kpl) kuin liittymäklustereita (15 kpl). Topografioiden indeksisarjoja parempi mallinnustarkkuus selittyy osittain topografioiden yhteydessä käytetyllä tarkemmalla lämpötilariippuvuusmallilla. Käyttöpaikkoja mallinnettaessa käytettiin kuukausikohtaisia lämpötilariippuvuuskerroimia, kun taas indeksisarjojen yhteydessä käytettiin vain yhtä koko vuoden

kattavaa lämpötilariippuvuuskerrointa. Kuvasta 2.16 on esitetty samat tulokset liittymien summakuormituksen mallinnukselle. Tässä kuvassa indeksisarjojen topografioita heikompi suorituskyky johtuu edellä mainitusta yksinkertaisemmasta lämpötilariippuvuusmallista, erikoispäivien huonommasta mallinnuksesta ja myöhemmin alaluvussa 2.9 kuvatusta lauantaiaamujen mallinnusvirheestä.

Kuvien 2.15 ja 2.16 taustalla olevissa laskelmissa käytettiin 7837 liittymän vuodelle 2018 muodostettuja kuormitusmalleja. Tarkkuuslaskelmissa huomioitiin vain ne liittymät, joiden kaikista käyttöpaikoista oli täydelliset aukottomat mittaukset. Valittujen liittymien yhteenlaskettu vuosienenergia oli 1,2 TWh.



**Kuva 2.15.** Liittymäkohtaisten mallien tarkkuusvertailu (SSE=Sum of Squared Errors , MAE=Mean Absolute Error).



**Kuva 2.16.** Liittymien summakulutuksen mallien tarkkuusvertailu.

Taulukossa 2.9 on vertailtu liittymäklustereita lähimpiin käyttöpaikkaklustereihin. Useimmille liittymäklustereille löytyy selkeä vastapari käyttöliittymäklustereiden joukosta. Poikkeuksena liittymäklusterit 4, 7 ja 11. Liittymäklusterin 4 lämpötilariippuvuus on huomattavasti suurempi kuin lähimmän käyttöpaikkaklusterin lämpötilariippuvuus. Liittymäklusterille 7 ei löytynyt yhtä selkeää

vastinetta käyttöpaikkojen joukosta, koska sen etäisyys käyttöpaikkaklustereista 24 ja 29 oli samaa suuruusluokkaa. Liittymäklusterille 11 ei löytynyt profiiltaan vastaavaa klusteria käyttöpaikkaklustereiden joukosta, vaikka se luokiteltiin hyvin suurella varmuudella käyttöpaikkaklusteriin 11. Tämä kertoo, että vaikka liittymäklusteri oli kaukana käyttöpaikkaklusterista 11, oli sen etäisyys muista käyttöpaikkaklustereista vielä huomattavasti suurempi. Luokittelun varmuusluku 100 ei siis taulukossa 2.9 tarkoita, että klustereiden kuormitusprofiilit olisivat täysin identtiset. Sen sijaan suuri luokittelun varmuusluku kertoo ettei luokittelua ole tehty vahingossa, esimerkiksi kuormituksen suuren hajonnan takia, väärin. Tässä tapauksessa luokittelun varmuusluvut ovat suurempia kuin yksittäisten liittymien tai käyttöpaikkojen tapauksissa, koska liittymäklusterit on laskettu keskiarvona suuresta liittymäjoukosta jolloin hajonta on pientä ja siitä johtuen myös luokitteluvirheiden todennäköisyys on pienempi.

Liittymäklustereiden sanalliset kuvaukset poikkeavat joiltain osin käyttöpaikkaklustereiden kuvauksista. Teollisuusasiakkaat nousivat pienemmässä liittymien tausta-aineistossa selkeämmin esille kuin käyttöpaikkojen tapauksessa, jossa teollisuus jäi lukumääräisesti suurempien toimisto- ja palveluryhmien varjoon. Kuormituksen vuorokausivaihtelua kuvatessa otettiin huomioon vain kulloinkin käsiteltävässä klustereiden joukossa esiintyvä suhteellinen vaihtelu, eikä vuorokausivaihtelun sanalliset kuvaukset siten ole täysin vertailukelpoisia liittymä- ja käyttöpaikkaklustereiden kesken. Muutamissa tapauksissa kuormitusten huipputunneissa on eroja, koska pienikin muutos kuormitusprofiilissa saattaa keikauttaa keskimääräisen kulutushuipun pyöristyksen yhdeltä tunnilta toiselle.

**Taulukko 2.9.** Liittymäklustereita lähinnä olevat käyttöpaikkaklusterit

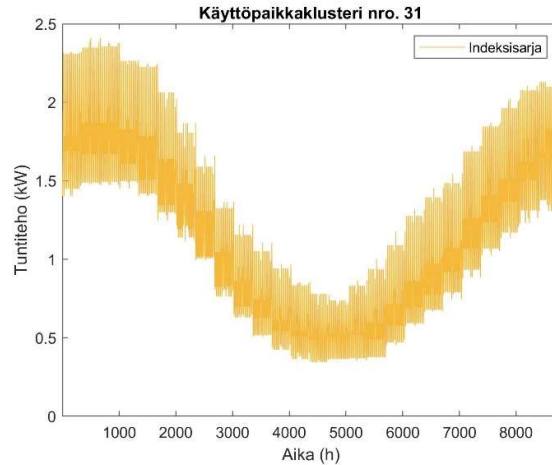
Liittymä- Klusteri	Liittymäklusterin kuvaus	käyttö- paikka- klusteri	Lähimmän käyttöpaikkaklusterin kuvaus	Luokittelun varmuus- luku
1	Toimisto/palvelu, suuri vuorokausivaihtelu, auki arkisin (7-18)	2	Toimisto/palvelu, suuri vuorokausivaihtelu, auki arkisin (7-18)	100,0
2	Toimisto/palvelu, auki arkisin (7- 18)	3	Toimisto/palvelu, auki arkisin (7-18)	100,0
3	Teollisuus, 1-vuoro, (pieni vuorokausivaihtelu)	6	Toimisto/palvelu, pieni vuorokausivaihtelu, auki arkisin (7-19), positiivinen lämpötilariippuvuus (ilmastointi)	97,2
4	Teollisuus, 1-vuoro, (pieni vuorokausivaihtelu), Lämpötilariippuva	4	Toimisto/palvelu, pieni vuorokausivaihtelu, auki arkisin (7-18)	96,7
5	Kaupat ja marketit, auki arkisin ja viikonloppuisin, ei lämpötilariippuvuutta	7	Palvelu ja kauppa, auki arkisin ja viikonloppuisin, ei lämpötilariippuvuutta	100,0
6	Kerrostalo, suuri vuorokausivaihtelu jokaisena viikonpäivänä, kulutushuippu klo 19-20	28	Kerrostaloasunto, huippukulutus klo 20-21	99,5
7	Kerrostalo, suuri vuorokausivaihtelu, lauantaita kohti nouseva kulutus, kulutushuippu klo 18-19	29	Kerrostaloasunto, asukas normaalin työelämän ulkopuolella, huippukulutus klo 20-21	64,1

8	Kerrostalo, pieni vuorokausivaihtelu, kulutushuippu klo 19-20	24	Kerrostaloasunto, pieni vuorokausivaihtelu, huippukulutus klo 18-19	100,0
9	Kerrostalo, kulutushuippu klo 19-20, osittainen lämpötilariippuvuus	28	Kerrostaloasunto, huippukulutus klo 20-21	100,0
10	Kerrostalo, pieni vuorokausivaihtelu, kulutushuippu klo 19-20, lämpötilariippuva	20	Kerrostalokiinteistö	100,0
11	Kerrostalo, pieni vuorokausivaihtelu, suuri päiväkulutus (asukkaat eivät työelämässä), kulutushuippu klo 17-18	11	Virkistys- ja kulttuuripalvelut, auki arkisin ja viikonloppuisin	100,0
12	Rivi-/paritalo, suora sähkölämmitys	32	Omakotitalo, suora sähkölämmitys	100,0
13	Rivi-/paritalo, käyttöveden lämmitys yösähköllä, (identtinen käyttöpaikkaklusterin 33 kanssa), huippukulutus klo 22-23	33	OK-/pari-/rivitalo, käyttöveden lämmitys yösähköllä	100,0
14	Kerrostalo, suhteellisen tasainen päivän- ja vuodenaajasta riippumaton kulutus	36	Vakiotehokuorma	100,0
15	Ei kulutusta	37	Ei kulutusta	100,0

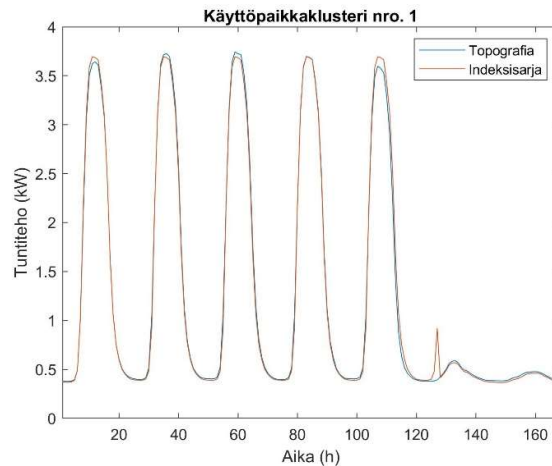
## 2.9. Indeksisarjojen muodostus

Tässä tutkimuksessa käyttöpaikka- ja liittymäklustereiden indeksisarjat muodostettiin nykyisen käytössä olevan Trimblen tietokantataulujen rakenteen mukaisesti. Vuosi jaettiin 26 kaksiviikkojaksoon ja kullekin klusterille laskettiin 26 ulkoista indeksiä, jotka kuvaavat suhteellista vuodenaikavaihtelua ja  $26 \times 3 \times 24 = 1872$  sisäistä indeksiä, jotka kuvaavat 2-viikkojaksojen sisäistä vaihtelua, ja niitä vastaavat 1872 prosentuaalista keskihajontaa. Kaikkien arkipäivien oletetaan käyttäytyvän samalla tavalla ja lauantai sekä sunnuntai mallinnetaan erikseen. Myöhemmin indeksisarjoja käytettäessä, erikoispäiviä (aatot ja arkipyhät) mallinnetaan lauantai- ja sunnuntaipäivien indekseillä. Kuvassa 2.17 on esitetty miltä indeksisarjasta laskettu koko vuoden mittainen aikasarja näyttää. Kuvan esimerkki on esitetty pitkän aikavälin (30 vuotta) Helsingin keskimääräisessä 2-viikkolämpötilassa.

Päivittäiset indeksisarjat määritettiin alkamaan tunnista 07.00-08.00 ja päättymään tuntiin 06.00–07.00. Tästä ja siitä että kaikkien arkipäivien oletetaan olevan samanlaisia, seuraa yksi erikoisuus, joka näkyy virheellisinä lauantaiamujen mallinnuksena niissä klustereissa joissa lauantaiamujen kulutus poikkeaa merkittävästi arkiamujen kulutuksesta. Kuva 2.18, jossa indeksisarjoilla mallinnettua viikkoa verrataan topografioilla mallinnettuun viikkoon havainnollistaa tätä virhettä. Kuvassa indeksisarjalla mallinnetussa lauantaiamussa näkyy virheellinen kuormituspiikki. Virhe johtuu siitä, että indeksisarjamallissa lauantaiamua mallinnetaan arkipäivän mallilla klo 07.00 asti.



**Kuva 2.17.** Käyttöpaikkaklusterin nro. 31 (omakotitalo, ei sähkölämmitystä) indeksisarjan perusteella vuodelle 2018 muodostettu aikasarja, oletuksena 10000 kWh vuosienergia.



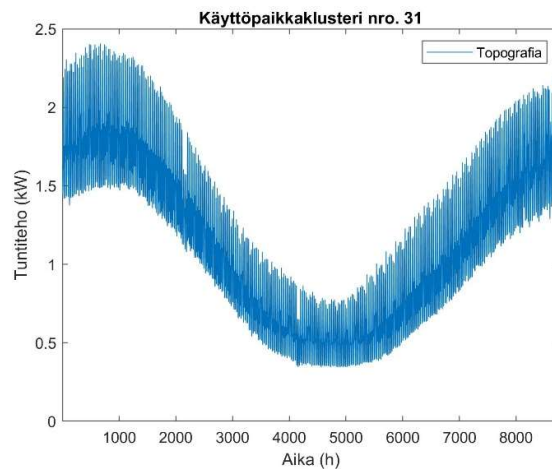
**Kuva 2.18.** Käyttöpaikkaklusterin nro. 1 (toimisto/palvelu) topografian ja indeksisarjan perusteella muodostetut tammikuun 2018 toisen viikon aikasarjat, oletuksena 10000 kWh vuosienergiat.

## 2.10. Topografioiden muodostus

Tässä tutkimuksessa käyttöpaikka- ja liittymäklustereiden topografiat muodostettiin eri periaatteita käyttäen kuin indeksisarjat. Indeksisarjoihin verrattuna, topografioihin tehtiin seuraavat mallinnus-tarkkuutta parantavat muutokset:

- 1) Viikonpäivät mallinnettiin erikseen, jolloin myös eri arkipäivien väliset erot tulevat mallinnetuiksi ja siirtymät viikonpäivien välillä ovat sujuvia (vrt. indeksisarjat kuvassa 2.18)
- 2) Erikoispäivät (arkipyhät ja aatot) mallinnettiin erikseen. Erikoispäiviksi luokiteltiin seuraavat päivät:
  - Uudenvuodenpäivä
  - Loppiainen
  - Pitkäperjantai
  - Lauantai Pitkäperjantain ja Pääsiäisen välissä
  - Pääsiäispäivä
  - Pääsiäispäivä

- Vappu
  - Helatorstai
  - Helluntai
  - Juhannusaatto
  - Juhannuspäivä
  - Juhannussunnuntai
  - Pyhäinpäivä
  - Itsenäisyyspäivä
  - Jouluaatto
  - Joulupäivä
  - Tapaninpäivä
- 3) Topografiat laskettiin 2-viikkojaksojen sijaan 30 vrk liukuvalla ikkunalla. Tämän valinnan vaikutus näkyy kuvassa 2.19, jossa ei esiinny vastaavassa indeksisarjakuvassa 2.17 esiintyviä porrasmaisia siirtymiä.
- 4) Topografiat normalisoitiin päivittäisiin keskilämpötiloihin, 2-viikkokeskilämpötilojen sijaan.
- 5) Topografioiden vuosienenergiat normalisoitiin arvoon 10000 kWh/vuosi ja vuoden jokaiselle tunnille laskettiin myös keskihajonta. Tulostiedoissa annetut topografiat on annettu vuoden 2018 kalenterissa. Topografioita käytettäessä on huomioita, että useimpien erikoispäivien ajallinen sijainti topografiassa vaihtelee vuosittain.



**Kuva 2.19.** Käyttöpaikkaklusterin nro. 31 (omakotitalo, ei sähkölämmitystä) topografian perusteella vuodelle 2018 muodostettu aikasarja, oletuksena 10000 kWh vuosienenergia.

### 3) Yhteenveto

Tutkimuksen ensimmäisessä vaiheessa HSV:n käyttöpaikat ja liittymät klusteroitiin samalla tavalla käyttäytyviin ryhmiin tuntimittausten kertoman kulutuskäyttäytymisen mukaan. Tavoitteena oli mallintaa HSV:n asiakkaiden kulutuskäyttäytymistä optimaalisella määrällä klusterikohtaisia kuormitusmalleja. Käyttöpaikkojen optimaaliseksi klusterimääräksi määritettiin 37 ja liittymien 15. Klustereiden nimeämisessä hyödynnettiin käyttöpaikkoihin ja liittymiin yhdistettyjä rakennustietoja.

Klusterikäyrien ensisijainen käyttökohde on jakeluverkon verkostolaskenta ja käyttöönoton helpottamiseksi niille laskettiin täsmälleen samanmuotoiset indeksisarjat mitä HSV:n järjestelmissä nykyäänkin käytetään. Myös koko vuotta tunneittain mallintavat topografiat laskettiin käyttöpaikka- ja liittymäklustereille. Tutkimuksessa tehtyjen tarkkuusvertailujen mukaan käyttöpaikkoja mallintavien klusterikäyrien aggregointi on tarkempi tapa mallintaa liittymien kulutusta kuin liittymiä mallintavien klusterikäyrien käyttö. Liittymäklustereiden käytölle tulee siis löytää muita perusteluja kuin tarkkuus, esimerkiksi niiden käytön helppous uutta verkkoa suunniteltaessa.

Tarkkuusvertailut paljastivat myös, että topografiat mallintavat kulutusta indeksisarjoja paremmin. On siis suositeltavaa, että indeksisarjoista siirryttäisiin tulevaisuudessa topografioihin. Lisäksi yhden koko vuodelle saman lämpötilariippuvuuskertoimen käytöstä tulisi siirtyä tarkemmin kuormituksen vuodenaikakohtaista lämpötilariippuvuutta mallintaviin käytäntöihin, esimerkiksi kuukausikohtaisiin lämpötilariippuvuuskertoimiin. Tämän tutkimuksen tulostiedostot sisältävät myös kaikkien laskettujen klustereiden topografiat ja kuukausittaiset lämpötilariippuvuuskertoimet.

### Lähteet

Mutanen Antti (2018) *'Improving Electricity Distribution System State Estimation with AMR-Based Load Profiles'*, Väitöskirja, Tampereen teknillinen yliopisto.

Satopää Ville, Albrecht Jeannie, Irwin David, ja Raghavan Barath (2006) *'Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior'*, International Conference on Distributed Computing Systems Workshops (ICDCSW), Minneapolis, MN, USA.