# KAUKOLÄMMÖNTARPEEN MALLINTAMINEN POHJOIS-SAVON ALUEELLA

Lappeenrannan–Lahden teknillinen yliopisto LUT

Energiatekniikan kandidaatintyö

2024

Juha Kapulainen

Tarkastaja: Tutkijatohtori Hannu Karjunen

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUTin energiajärjestelmien tiedekunta

Energiatekniikka

Juha Kapulainen

**Kaukolämmöntarpeen mallintaminen Pohjois-Savon alueella**

Energiatekniikan kandidaatintyö

2024

33 sivua, 9 kuvaa, 9 taulukkoa ja 1 liite

Tarkastaja: Tutkijatohtori Hannu Karjunen

Avainsanat: Kaukolämmöntarve, mallinnus, regressiomalli, Pohjois-Savo

Tässä kandidaatintyössä tutkitaan, voidaanko maantieteellisillä muuttujilla selittää kunnan vuosittaista kaukolämmöntarvetta. Tavoitteena on luoda malli, jolla voisi arvioida kaukolämmöntarvetta Suomen kunnissa. Mallin rakentamisen otoksena käytetään Pohjois-Savon aluetta.

Mallina työssä käytetään moniregressiomallia. Moniregressiomallissa on useampi muuttuja ja muuttujilla parametrit, jotka täytyy ensin ratkaista tilastojen pohjalta. Kaukolämmöntarvetta kuvaaviksi muuttujiksi malliin valittiin kuntien väkiluku ja taajama-aste. Parametrien ratkaisun jälkeen malli on käyttövalmis ja sillä voidaan selvittää kaukolämmöntarvetta.

Mallin antamien tulosten ja tilastojen välisen eron avulla voidaan selvittää mallin tarkkuutta. Pienin ero mallin ja tilastojen välillä oli 0,4 % eräässä kunnassa. Ero mallin ja oikean lämmöntarpeen välillä on parhaimmillaan matala. Suurimmassa osassa kuntia ei päästä kuitenkaan lähelle oikeaa kaukolämmöntarvetta. Mallilla pystytään tällaisenaan saavuttamaan suuntaa antavia tuloksia lämmöntarpeesta kunnassa.

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Energy Systems

Energy Technology

Juha Kapulainen

**Modelling of district heating demand in Northern Savonia**

Bachelor's thesis

2024

33 pages, 9 figures, 9 tables and 1 appendix

Examiners: Post-doctoral researcher Hannu Karjunen

Keywords: District heat demand, modelling, multiple linear regression, North Savonia

The aim of this bachelor's thesis is to research if it is possible to explain municipalities' district heating demand by geographical variables. The goal is to create a model that could be used to estimate municipalities' district heating demand in Finland. Northern Savonia is the sample of Finland used for the model.

Multiple linear regression model is used in this research. Multiple linear regression model has multiple variables and parameters that need to be calculated first according to the sample statistics. Variables used for the model are population and degree of urbanisation. The model is ready for use after the parameters have been calculated.

To evaluate the accuracy of the model, we must calculate the difference between the model and the actual heat demand. The smallest difference between the model and statistics was 0,4 %. The difference between the model and the actual heat demand is at best very low. However, in most municipalities the model is not close to the actual heat demand. The model gives approximate estimates for the municipalities' district heating demand in its current state.

SYMBOLI- JA LYHENNELUETTELO

Roomalaiset

| | |
|---|---|
| $b$ | regressioparametri |
| $Y$ | tulos |
| $y$ | tulos |
| $\bar{y}$ | keskiarvo |
| $x$ | regressio muuttuja |
| $\bar{x}$ | keskiarvo |
| $e$ | satunnaisvirhe |

Kreikkalaiset

| | |
|---|---|
| $\beta$ | regressio parametri |

Dimensiottomat luvut

| | |
|---|---|
| R | korrelaatiokerroin |

Alaindeksit

| | |
|---|---|
| $n$ | lukumäärä |
| $k$ | lukumäärä |
| 0 | vakiotermi |
| 1 | ensimmäinen muuttuja |
| 2 | toinen muuttuja |

Lyhenteet

| | |
|---|---|
| KL | kaukolämpö |

**Sisällysluettelo**

Tiivistelmä

Abstract

Symboli- ja lyhenneluettelo

1 Johdanto ..... 7
2 Tutkimuksen esittely ..... 9
2.1 Pohjois-Savo alueena ..... 9
2.2 Aikaisempaa tutkimusta ..... 13
2.3 Tilastollinen mallintaminen ..... 15
3 Mallin rakennus ..... 17
3.1 Muuttujien valinta ja arviointi ..... 17
3.2 Regressiomalli ..... 21
4 Tulosten tarkastelua ..... 25
4.1 Muuttujien arviointi ..... 25
4.2 Mallin erot tilastoihin verrattuna ..... 26
5 Johtopäätökset ..... 31
Lähteet ..... 32

Liitteet

Liite 1. Työssä käytetyt tilastot Pohjois-Savon kunnista

# 1 Johdanto

Kaukolämpö on tärkeä osa rakennusten ja niiden käyttöveden lämmitystä Suomessa. Kaukolämpöverkosto on keskittynyt kaupunkeihin, sillä kaukolämpöä on järkevää hyödyntää lähellä paikkaa, jossa sitä tuotetaan, koska lämpöhäviöt kasvavat kaukolämpöverkon pidentyessä. Kaukolämpö on kuluttajalle helppo ja huoleton lämmitystapa. Kaukolämpö on järkevä hyödyntää kaupungeissa sillä se mahdollistaa keskitetyn lämmöntuotannon sekä erilaisten lämmöntuotantotapojen hyödyntämisen.

Kaukolämmöntarve riippuu monesta tekijästä. Kaukolämmön tuotantoa suunniteltaessa ja ennen investointeja on tärkeää tietää paljonko sille olisi mahdollisesti tarvetta, että osataan vastata riittävästi kysyntään. Malli, jolla voitaisiin arvioida kaukolämmöntarvetta ottamalla huomioon siihen vaikuttavia tekijöitä, voisi auttaa kaukolämmön kapasiteetin suunnittelussa. Tällaiselle mallille voisi olla käyttöä myös, jos halutaan tutkia energiatehokkuutta parantavien toimien, kuten lämmöneristyksen lisäämisen ja energiankulutuksen muutoksien vaikutuksia kaukolämmöntarpeeseen jollain tietyllä alueella. Mallia voitaisiin myös hyödyntää mahdollisesti pienemmässäkin mittakaavassa esimerkiksi korttelin kaukolämmöntarvetta tarkastellessa.

Tämän kandidaatintyön aiheena on tutkia, pystytäänkö luomaan malli, jolla voidaan arvioida kaukolämmöntarvetta kunnan tasolla tai jopa tarkemmalla tasolla. Mallia muodostaessa tutkitaan, voidaanko käyttää yleisiä muuttujia kuten asukaslukua tai asukastiheyttä kunnan kaukolämmöntarpeen arviointiin. Jos kaukolämmöntarpeen arvioimisen malli toimii tutkittavana olevaan maakunnan eli Pohjois-Savon alueella, voidaan mallia mahdollisesti soveltaa muidenkin maakuntien alueella. Tällaisen mallin luomisesta voi olla hyötyä, kun halutaan tarkastella jonkun uuden kaavoitetun alueen kaukolämmön tuotantoon tarvittavaa kapasiteettia, sillä alueellista kaukolämmöntuotantoa on vaikea arvioida etukäteen. Kaukolämmöntarpeen arviointia tarvitaan myös, kun suunnitellaan uusia tulevaisuuden energian ja sähköntuotantomuotoja, sillä näiden mahdollisesti tuottamaa ylijäämälämpöä voidaan hyödyntää yleensä kaukolämpöverkossa. Esimerkiksi jos Suomessa vedyn tuotanto alkaa kehittyä entisestään ja on saatavilla sopivan hintaista uusiutuvilla tuotettua sähköä, voivat elektrolyyserit olla yksi tapa tuottaa vetyä ja hyödyntää prosessissa syntyvää ylijäämälämpöä kaukolämpöverkossa.

Tässä työssä tutkitaan vuosittaista kaukolämmöntarvetta eli vuodenaikoihin perustuva vaihtelu sisältyy vuosittaiseen tarpeeseen. Työssä tarkastellaan vain yhden vuoden kaukolämmön käyttötilastoja. Tutkimuksen tavoitteena on onnistua luomaan kaukolämmöntarpeelle malli käyttäen Pohjois-Savoa tarkastelualueena. Mallin muodostus toteutetaan niin, että tietyn alueen tunnettuja muuttujia käyttämällä saadaan luotua tilastollinen malli, joka antaa vastaukseksi alueen kaukolämmöntarpeen. Kaukolämmöntarvetta tarkastellaan asuintalojen ja kaikkien rakennusten ja niiden tilojen sekä käyttöveden lämmitykseen tarpeeseen. Ensin tarkastellaan Pohjois-Savoa alueena kaukolämmön näkökulmasta, jonka jälkeen tarkastellaan aiempia tutkimuksia sekä aiheeseen liittyvää kirjallisuutta. Sen jälkeen siirrytään mallin muodostukseen ja tarkastellaan, minkälaisia tuloksia saatiin.

# 2 Tutkimuksen esittely

Tutkimuksessa käytetään mallin rakentamisen pohjana Energiateollisuus ry:n vuoden 2021 kaukolämpötilastoja. Tarkasteltavalla vuodella ei ole kovin suurta merkitystä, sillä jos mallin saa vastaamaan yhden valitun vuoden kaukolämmöntarvetta, saa sen vastaamaan myös toisen vuoden tarvetta vaihtamalla vain muuttujat kyseessä olevaa vuotta vastaaviksi. Tämän pohjadatan avulla pohditaan mahdollisia tekijöitä kaukolämmöntarpeen mallia varten. Tässä kappaleessa esitellään tarkasteltava maakunta, aiempaa tutkimusta sekä tilastollista mallintamista.

## 2.1 Pohjois-Savo alueena

Pohjois-Savo valikoitui maakunnaksi tähän työhön sen vuoksi, että siellä kaukolämmöllä lämpiävien talojen asukkaiden osuus maakunnan väkiluvusta oli 57 %, joka on lähellä kaikkien Suomen kaukolämmön asukkaiden suhdetta koko Suomen väkilukuun, mikä oli 54 % (Energiateollisuus ry, 2023, Taulukko 8). Pohjois-Savossa KL-asukkaiden osuus on siis hyvin lähellä sitä osuutta mitä KL-asukkaat ovat Suomen väestöstä eli se sopii tältä osin hyvin tarkasteltavaksi maakunnaksi. Pohjois-Savossa on myös useampi kunta, jossa on kaukolämpöä käytössä, joten kaukolämmöntarve jakautuu ympäri maakuntaa. Pohjois-Savossa on väkiluvultaan suuria ja pieniä kuntia, joten se toimii hyvänä otoksena Suomesta. Pohjois-Savossa ei ole lisäksi mitään erityispiirteitä tai pelkkää suurkaupunkialuetta, mitkä saattaisivat heikentää mallin yleispätevyyttä. Maakunta on myös alueena tuttu, joten aluetuntemus auttaa tulosten analysoinnissa.

Pohjois-Savoon on kuulunut vuonna 2021 yhteensä 19 kuntaa, joista 13:sta on KL-asukkaita (taulukko 1). Taajama-aste, joka tarkoittaa taajamissa asuvien osuutta kunnan väkiluvusta, ja asukasluku kunnittain ovat saatavilla avoimesti tilastokeskuksen sivuilta ja ne näkyvät myös taulukossa 1 (Tilastokeskus, 2024). Pohjois-Savossa on yksi kunta, joka erottuu asukasluvultaan muista ja se on Kuopio. Muissa kunnissa asukasluku on noin 20 000 tai alle.

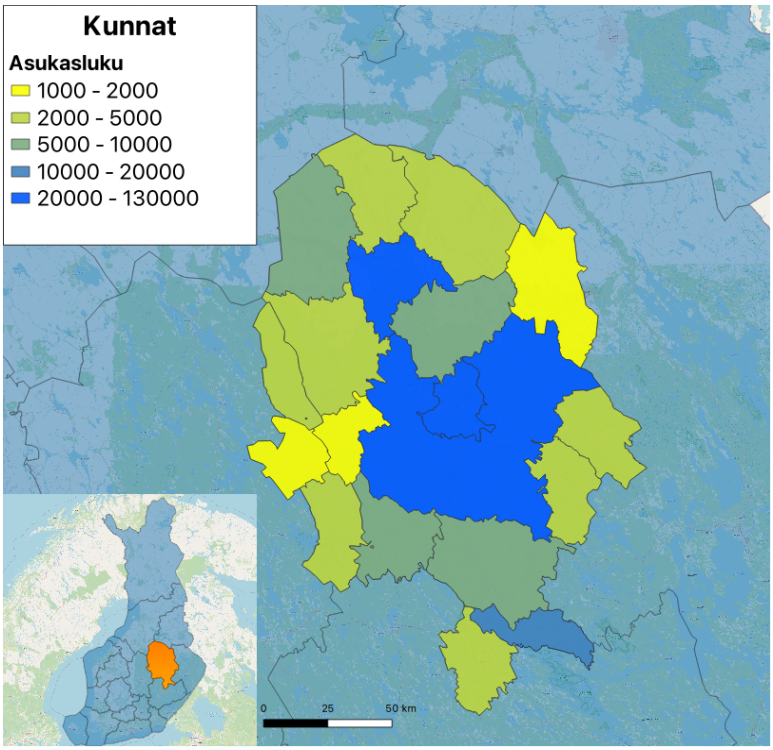
Taulukko 1. Pohjois-Savon kunnat asukasluvun mukaisessa järjestyksessä (Tilastokeskus, 2024)

| Nimi | Asukasluku | Taajama-aste % | KL-asukkaita |
|---|---|---|---|
| Tervo | 1479 | 34,5 | Ei |
| Rautavaara | 1513 | 44,2 | Ei |
| Vesanto | 1941 | 40,1 | Ei |
| Keitele | 2095 | 48,5 | On |
| Tuusniemi | 2420 | 41,5 | Ei |
| Kaavi | 2778 | 46,1 | Ei |
| Rautalampi | 3033 | 48,3 | On |
| Vieremä | 3490 | 39,2 | Ei |
| Sonkajärvi | 3777 | 43,7 | On |
| Pielavesi | 4269 | 42,4 | On |
| Joroinen | 4624 | 58,1 | On |
| Suonenjoki | 6891 | 72,1 | On |
| Kiuruvesi | 7759 | 50,4 | On |
| Lapinlahti | 9247 | 53,8 | On |
| Leppävirta | 9280 | 57,9 | On |
| Varkaus | 19973 | 92,2 | On |
| Iisalmi | 20958 | 76,6 | On |
| Siilinjärvi | 21293 | 81,1 | On |
| Kuopio | 121543 | 87 | On |

Energiateollisuuden taulukoista saatuja kaukolämmön tilastoja sekä tilastokeskuksen sivuilta täydennettyä tietoja yhdyskuntarakenteesta on yhdistelty Excelissä (Tilastokeskus, 2024). Energiateollisuuden taulukoista on ensin suodatettu vain Pohjois-Savoon liittyvät tiedot ja näitä on jalostettu eteenpäin lisäämällä esimerkiksi asukasluku kuntakohtaisesti.

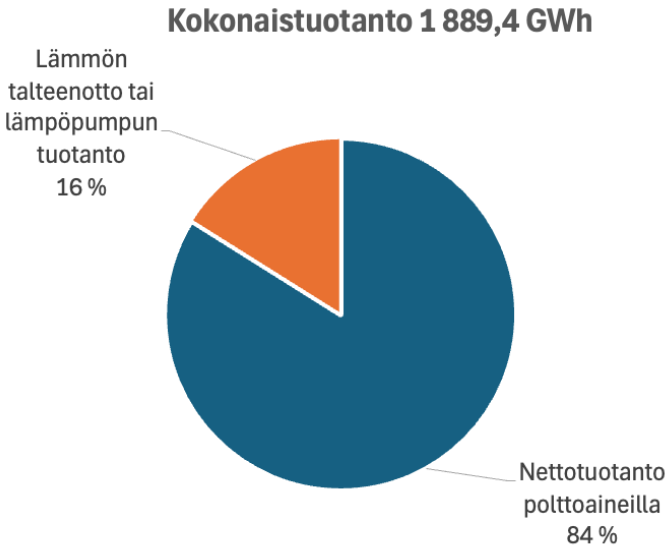
Tässä työssä havainnollistetaan tuloksia ja saatuja tilastoja QGIS-ohjelmalla, joka on avoimen lähdekoodin paikkatieto-ohjelmisto (QGIS Development team, 2009). QGIS-ohjelmaan ladattiin avoimesti saatavilla olevia tasoja maanmittauslaitoksen ja Suomen ympäristökeskuksen avoimista kartta-aineistoista (Syke 2024 & Maanmittauslaitos 2024). Näihin

tasoihin pystytään liittämään myös muualta tietoa esimerkiksi Excelillä tehtyjä taulukoita kuten tässä työssä on tehty. Pohjois-Savoa tarkastellaan tarkemmin kuntatasolla ja kuntien asukaslukua on havainnollistettu kuvassa 1.



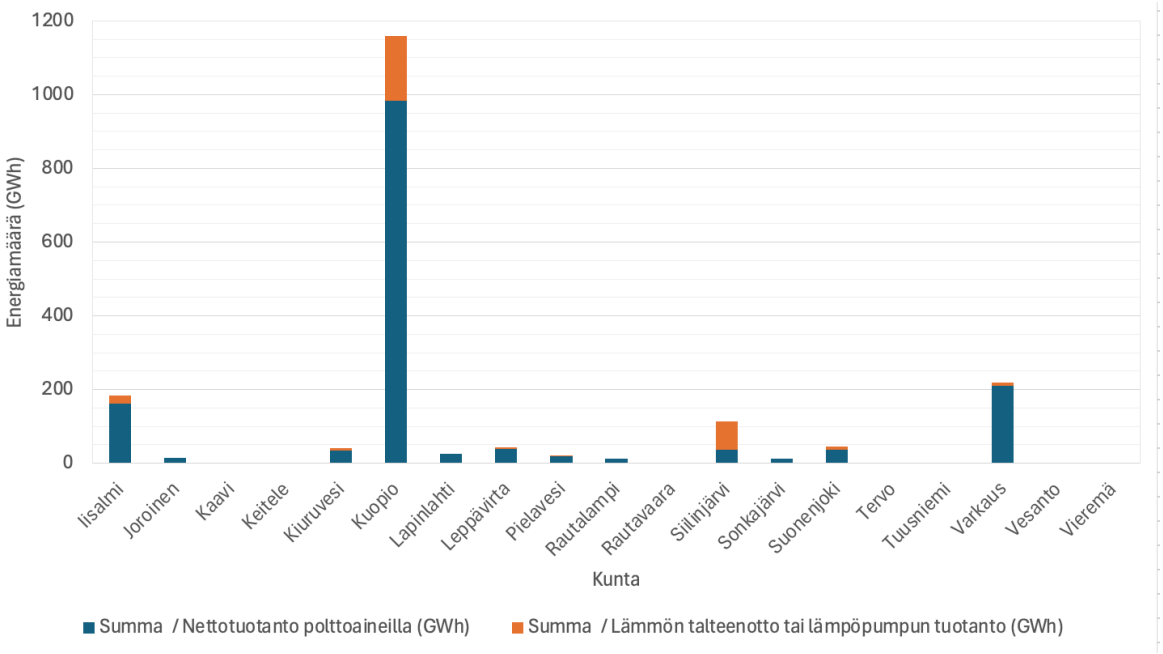

Kuva 1. Pohjois-Savon sijaintia ja kuntien asukaslukua havainnollistettu QGIS-ohjelmalla.

Kaukolämmön lähteitä on tarkasteltu Energiateollisuuden (2023) tilastojen perusteella. Saatuja tietoja on havainnollistettu ensin maakuntatasolla kuvassa 2. Kuvan ympyrädiagrammista huomataan, että suurin osa kaukolämmöstä tuotetaan erilaisilla polttoaineilla ja pienempi osa saadaan lämmöntalteenotosta tai lämpöpumpuista.



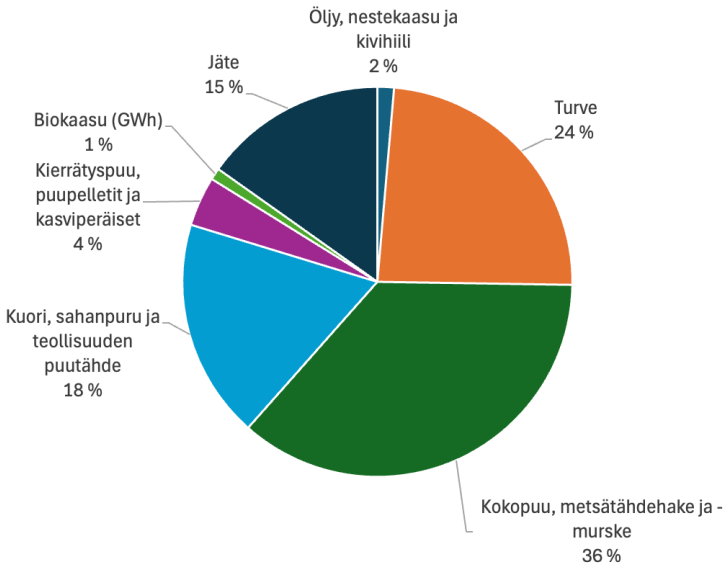

Kuva 2. Ympyrädiagrammi kaukolämmön lähteiden jakautumisesta Pohjois-Savossa.

Esitetään tarkemmin kaukolämmön lähteiden jakautumista kuntatasolla kuvassa 3. Pylväsdiagrammissa näkyvät hyvin kunnat, joissa ei ole KL-asukkaita. Suurimmassa kunnassa Kuopiossa KL-tarve on enemmän kuin muissa kunnissa yhteensä. Eri värit kertovat mistä lähteestä jakelijat saavat kaukolämmön.




Kuva 3. Kaukolämmön lähteiden jakautuminen kunnittain Pohjois-Savossa.

Suurin osa polttoaineilla tuotetusta kaukolämmöstä Pohjois-Savossa tuotetaan yhteistuotannolla eli tuotetaan voimalaitoksessa kaukolämmön lisäksi myös sähköä. Kuvan 4 ympyrädiagrammissa esitetään erillis- ja yhteistuotannossa käytetyt polttoaineet Energiateollisuuden (2023) tilastojen perusteella.




Kuva 4. Kaukolämmön erillis- ja yhteistuotannossa käytetyt polttoaineet Pohjois-Savossa.

Kuvasta 4 havaitaan, että puupohjaisilla biopolttoaineilla tuotetaan suurin osa kaukolämmöstä. Fossiilisten polttoaineiden osuus on pieni kaukolämmön tuotannossa.

## 2.2 Aikaisempaa tutkimusta

Kaukolämmöntarvetta on aiemmin tutkittu esimerkiksi simuloimalla erilaisilla ohjelmilla vallitsevan ulkoilman tilaa malliin vaikuttavana tekijänä. Simulointeja on tehty yksittäisille rakennuksille ja otettu huomioon niiden sisäilman lämmitys, käyttöveden lämmitys, viilennys ja sisäilman lämmitys. Kalamees et al. (2012) ovat tutkineet rakennusten energian kulutukseen vaikuttavia tekijöitä standardisoimalla vallitsevat olosuhteet. Vallitsevia olosuhteita ovat ilman lämpötila, auringon säteily, ilman kosteus ja tuulen nopeus. Tutkimusryhmä selvitti, kuinka paljon muutosta näiden tekijöiden vaihtelusta aiheutuu omakotitalon ja

toimistorakennuksen energiankulutukselle. Ilmastolliset parametrit vaikuttavat energiankulutukseen standardin EN ISO 15927-4 mukaan. (Kalamees et al., 2012.)

Hietaharju tutkimusryhmineen (2021) tutkivat kuinka rakennusten lämmöntarve tulee muuttumaan tulevaisuudessa ilmastonmuutoksen myötä. Mallin luomisessa rakennuksen lämmöntarpeelle käytettiin hyödyksi aiempaa tilastotietoa rakennuksesta. Mallissa otettiin huomioon rakennusten purkaminen ja uudelleenrakentaminen ja rakennusten ihmisten lukumääriä. Näiden tekijöiden avulla luotiin dynaaminen malli, joka antoi suuntaa sille, kasvaako vai vähentyykö rakennusten lämmöntarve. (Hietaharju et al., 2021.)

Jylhä tutkimusryhmineen (2015) tutkivat omakotitalon lämmityksen ja jäähdytyksen tarvetta ja niiden hinnan nousua ilmastonmuutoksen myötä. He käyttivät tähän rakennuksen energiankulutusta simuloivaa ohjelmaa ja ilmastonmuutoksen vaikutuksia tutkittiin erilaisten ilmastomallien avulla. He tutkivat tätä Suomessa, koska täällä ulkoilman lämpötila ja auringon säteily kesällä vaikuttavat rakennusten lämmön ja jäähdytyksen tarpeeseen. (Jylhä et al., 2015.)

Fang ja Lahdelma (2016) tutkivat lämmöntarvetta kaupungin kokoiselle alueelle. Lämmöntarpeen arvioinnissa käytettiin regressioanalyysia käyttäen parametreina tuntikohtaista ulkolämpötilaa ja tuulen nopeutta. Lämmöntarvetta arvioitiin myös SARIMA mallin avulla ottaen huomioon sään tekijöitä ja aikaisempaa kulutusta. He käyttivät reaaliaikaista dataa Espoosta 20 viikon ajalta malleihin käytettäväksi. Regressioanalyysissa otettiin huomioon myös sosiaalinen komponentti, kun huomioitiin viikoittainen rytmi lämmön käytössä. (Fang ja Lahdelma, 2016.)

Ma tutkimusryhmineen (2014) tutkivat puolestaan energian kulutusta rakennuksissa energiankulutusmallien (ECP Energy Consumption Pattern) kautta. ECP:n avulla Ma loi matemaattisen mallin, jossa käytettiin aikaa ja rakennuksen tyyppiä tekijöinä ja näiden muuttujien suhdetta arvoitiin Gaussin sekoitus mallilla (GMM Gaussian mixture model). Datana Ma käytti oikean kaukolämpöjärjestelmän toiminta-arvoja. Ma käytti lämmöntarpeen määrittämiseen Gaussin jakaumaa, sillä mittaustulokset jakautuvat yleensä Gaussin käyrälle. (Ma et al., 2014.)

## 2.3 Tilastollinen mallintaminen

Tilastollisen mallintamisen tarkoituksena on luoda olemassa olevan tilaston kuvaamiseen tai analysointiin malli. Tilastotieteiden tarkoituksena on oppia tilastoista. Tilastotieteissä kerätään ja analysoidaan saatua tilastotietoa ja hyödynnetään saatua tietoa insinööritieteissä tai esimerkiksi lääketieteessä vahvistamaan teoriaa tai päätöksenteon tukena. (Davison, 2003, 1.)

Muuttujien välisen yhteyden selvittämiseen löytyy erilaisia tapoja. Muuttujien välisestä yhteydestä esimerkkinä on kuinka hyvin joku muuttuja kuvaa jotain tulosta tai kuinka paljon joku muuttuja vaikuttaa toiseen muuttujaan. Malleja, joita käytetään tilastollisessa analyysissa, ovat esimerkiksi regressiomallit, jotka ovat hyviä selvittämään riippuvuussuhteita (Ross 2004, 351).

Yksi tapa tarkastella muuttujien yhteyttä on korrelaatiokerroin (R). Korrelaatiokerroin kuvaa muuttujien välisen lineaarisen suhteen vahvuutta. Jos korrelaatiokerroin on suurempaa kuin nolla, on muuttujien välinen suhde positiivinen eli jos toisen muuttujan arvo kasvaa, kasvaa myös toisenkin. Jos korrelaatiokerroin on alle nolla, on muuttujien välinen suhde negatiivinen eli tällöin toisen muuttujan arvon kasvaessa toisen arvo pienenee. Korrelaatiokerroin voi saada arvoja väliltä -1,0–1,0. Arvo -1,0 kuvaa täydellistä negatiivista korrelaatiota ja 1,0 vastaavasti täydellistä positiivista korrelaatiota. Korrelaatiokertoimen arvo 0 kuvaa sitä, että muuttujien välillä ei ole minkäänlaista lineaarista yhteyttä. (Nickolas 2023.)

Korrelaatiokertoimen yhtälö on muotoa

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

jossa $\bar{x}$ ja $\bar{y}$ ovat muuttujien keskiarvoja (Ross 2004, 36)

Yhtälön 1 osoittajasta huomataan, että jos toinen muuttuja alkaa pienemään alle keskiarvon, niin korrelaatiokertoimen arvo on silloin negatiivista.

Eräs tapa tarkastella muuttujien pätevyyttä kuvaamaan jotain haluttua tulosta on nollahypoteesi. Nollahypoteesi voidaan valita niin, että todetaan että jollain muuttujalla ei ole vaikutusta toiseen muuttujaan tai että joku muuttuja on nolla. Sen jälkeen tarkoituksena on yleensä päätellä, voidaanko nollahypoteesi kumota, niin että on jotain todistetta siitä, että muuttujilla

onkin jonkinlaista yhteyttä tai että muuttuja on jotain muuta kuin nolla. (MacInnes 2022, 12.)

Nollahypoteesin oikeellisuus selvitetään niin, että ratkaistaan ylittääkö nollahypoteesin todennäköisyys valitun merkitsevyyden arvon. Merkitsevyys valitaan tapauskohtaisesti mutta yleensä sen valitaan olevan 0,05. Jos testitulokseksi saatu p-arvo ylittää 0,05 eli valitun merkitsevyyden, nollahypoteesi hyväksytään ja vastaavasti alittaessa hylätään ja tällöin kumotaan nollahypoteesi. Tiivistettynä p-arvo kuvaa hypoteesin uskottavuutta. (Ross 2004, 296.)

# 3 Mallin rakennus

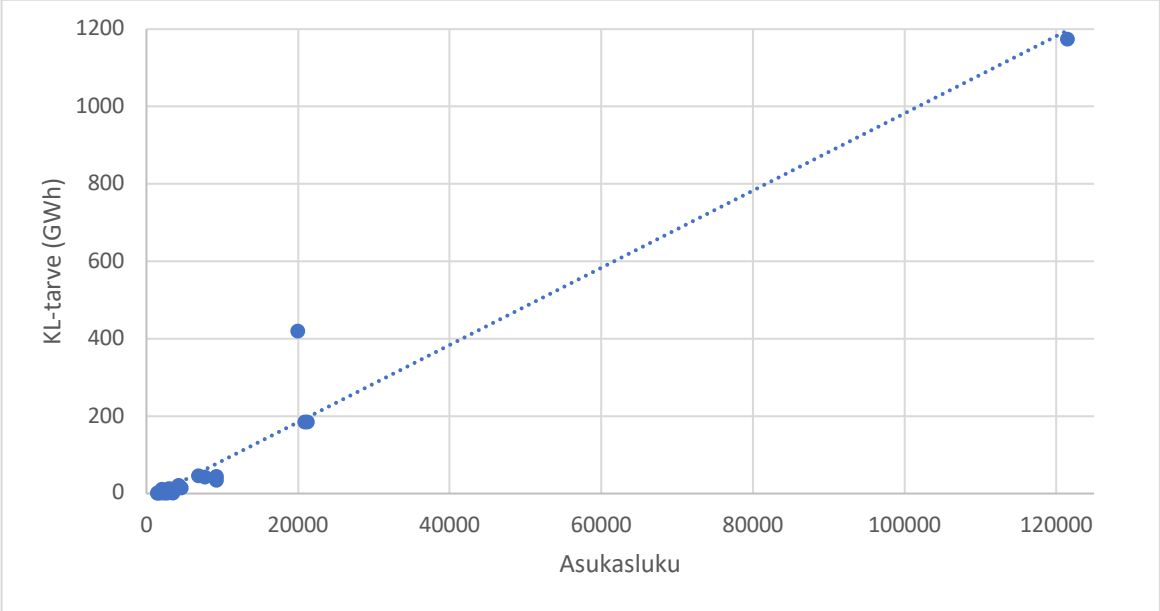
Tässä työssä luotavan mallin tavoitteena on löytää KL-tarvetta selittäviä tekijöitä. Moniregressiomalli sopii työhön hyvin, koska se on hyvä muuttujien välisen suhteen määrittämiseen. Tässä kappaleessa tarkastellaan ensiksi mallin tekijöiden valintaa ja sen jälkeen esitellään tehty moniregressiomalli.

## 3.1 Muuttujien valinta ja arviointi

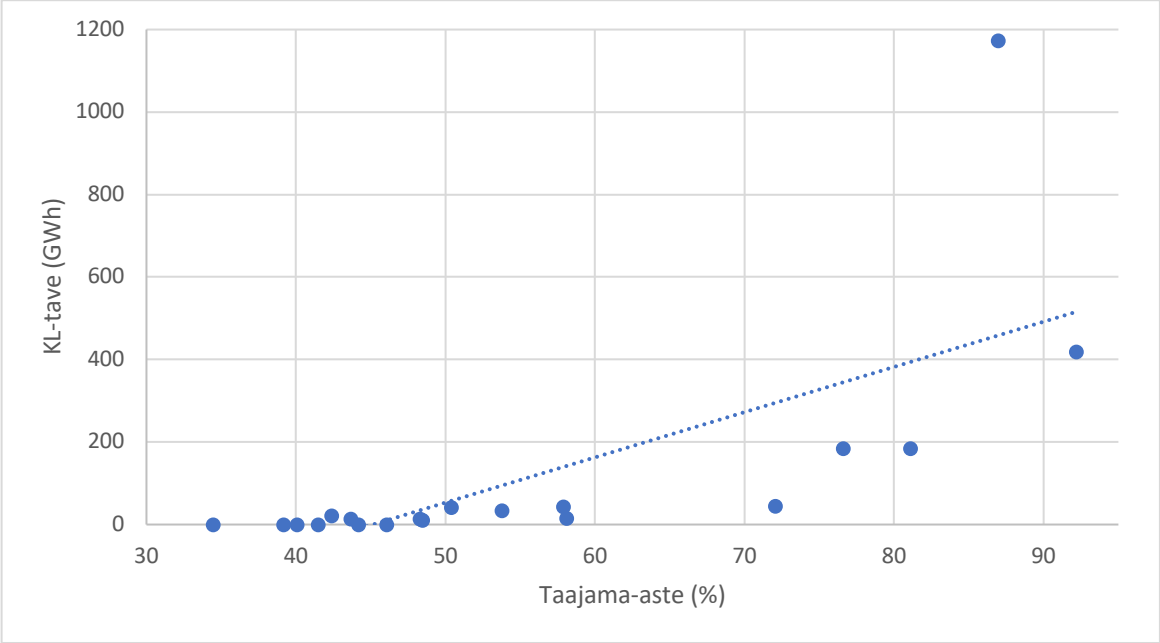
Tekijöiksi malliin otetaan maantieteellisiä parametreja, sillä näiden käytöstä KL-tarpeen arvioinnissa ei löytynyt kovin paljoa tutkimustietoa. Mallin tavoitteena olisi kuvata alueen vuosittaista KL-tarvetta. Muuttujien valinnan arvioinnissa voidaan käyttää erilaisia menetelmiä. Tässä työssä muuttujien sopivuuden valinnassa käytetään korrelaatiokerrointa R ja p-arvoa. Korrelaatiokertoimen avulla selvitetään, onko muuttujan ja halutun tuloksen eli KL-tarpeen välillä lineaarista yhteyttä. Muuttujien keskinäistä suhdetta tarkastellaan myös, sillä olisi parempi, jos niiden välillä ei olisi kovin suurta lineaarista korrelaatiota, koska suuri korrelaatio vaikuttaa regressiomallin kertoimiin (Kiernan 2014, kappale 8).

Energiateollisuuden KL-tilastoista saadaan jokaiselle Pohjois-Savon kunnalle KL-tarve. Tilastokeskuksen sivuilta saadaan kuntien avainluvut vuodelle 2021 ja sieltä saadaan kunnittain taajama-aste, väkiluku ja pinta-ala (Liite 1). Selvitetään näiden kolmen muuttujan sopivuutta regressiomalliin. Moniregressiomalliin sopii useampikin muuttuja, joten selvitetään seuraavaksi muuttujien välistä lineaarista yhteyttä toisiinsa ja KL-tarpeeseen.

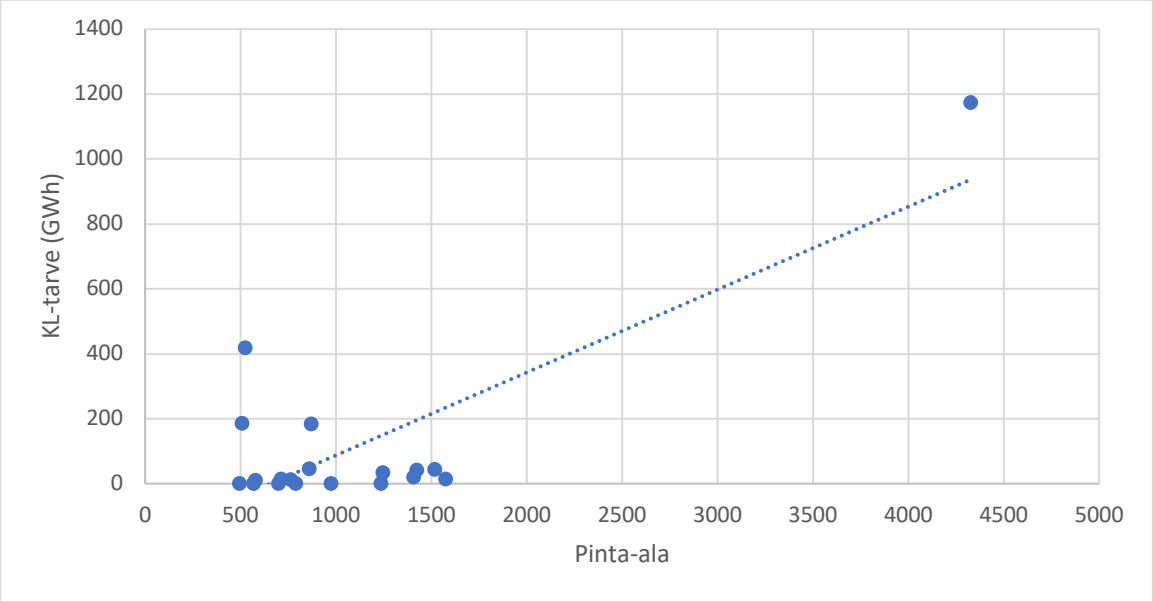
Aluksi voidaan tarkastella muuttujien välisiä suhteita toisiinsa ja haluttuun vastausmuuttujaan eli KL-tarpeeseen pistekuvaajilla. Kuvasta 5 nähdään, että asukasluvun ja KL-tarpeen välillä on selvästi havaittavissa positiivista korrelaatiota, sillä pisteiden välille pystyy hyvin sijoittamaan suoran. Kuvasta 6 ja 7 nähdään, että taajama-asteen ja KL-tarpeen sekä pinta-alan ja KL-tarpeen välillä ei ole yhtä selkeää lineaarista korrelaatiota, sillä pistejoukkoihin ei voi kovin hyvin sijoittaa suoraa.




Kuva 5. KL-tarpeen ja asukasluvun suhde Pohjois-Savon kunnissa.




Kuva 6. KL-tarpeen ja taajama-asteen suhde Pohjois-Savon kunnissa.




Kuva 7. KL-tarpeen ja pinta-alan suhde Pohjois-Savon kunnissa.

Korrelaatiokertoimella ja p-arvolla voidaan selventää muuttujien ja KL-tarpeen välistä yhteyttä toisiinsa. Määritetään R ja p-arvo syöttämällä jokaisen Pohjois-Savon kunnan muuttujat ja KL-tarve MATLAB-ohjelmaan (MathWorks 2022). MATLAB:n avulla saadaan laskettua R-arvo kunkin muuttujan ja KL-tarpeen välille. Samalla saadaan laskettua kullekin muuttuja-KL-tarve parille p-arvo. P-arvo on laskettu niin, että nollahypoteesi on, ettei muuttujan ja KL-tarpeen välillä ole suhdetta. Merkitsevyyden rajana on 0,05 tarkoittaen sitä, että tätä pienemmät p-arvot tarkoittavat, että nollahypoteesi kumotaan ja muuttujan ja KL-tarpeen välillä on suhde. P-arvon ollessa yli 0,05 nollahypoteesi on voimassa eli muuttujan ja KL-tarpeen välillä ei ole suhdetta.

Muuttujien korrelaatiokertoimet ja p-arvot verrattuna KL-tarpeeseen on esitetty taulukossa 2. Kaikkien kolmen muuttujan ja KL-tarpeen välillä näyttäisi olevan jonkinasteista positiivista korrelaatiota. Väkiluvun lineaarinen korrelaatio näyttää kaikkein vahvimmalta kuten se jo aikaisemmin havaittiin kuvasta 5. Taulukon 2 korrelaatioiden p-arvot ovat myös alle 0,05 osoittaen, että korrelaatiot ovat merkittäviä.

Taulukko 2. KL-tarpeen ja kunkin muuttujan R-luku ja p-arvo.

| | R | p-arvo |
|---|---|---|
| Taajama-aste—KL-tarve | 0,6954 | 0,009 |
| Väkiluku—KL-tarve | 0,9777 | < 0,0000 |
| Pinta-ala —KL-tarve | 0,7942 | < 0,0000 |

Muuttujia verrataan myös toisiinsa, jotta saadaan selville, onko niillä keskenään lineaarista korrelaatiota. Taulukossa 3 on esitetty muuttujien keskinäiset R-luvut ja p-arvot. Taulukosta 3 huomataan, että vahvin korrelaatio on pinta-alan ja väkiluvun välillä ja tämän p-arvo todistaa myös. Taajama-asteen ja väkiluvun välillä on kohtalainen korrelaatio, jonka p-arvo vahvistaa myös. Pinta-alan ja taajama-asteen välillä on puolestaan hyvin heikko korrelaatio ja p-arvo osoittaa myös tämän sillä se on yli 0,05 vahvistaen nollahypoteesin oikeaksi.

Taulukko 3. Muuttujien keskinäiset R-luvut ja p-arvot

| | R | p-arvo |
|---|---|---|
| Pinta-ala—Taajama-aste | 0,2991 | 0,2135 |
| Pinta-ala —Väkiluku | 0,8647 | < 0,0000 |
| Taajama-aste—Väkiluku | 0,6170 | 0,0049 |

Taulukon 2 perusteella nähdään, että kaikki muuttujat kuvaavat kohtalaisen hyvin KL-tarvetta mutta taulukosta kolme nähdään, että pinta-alan ja väkiluvun välillä on sen verran korkea korrelaatio, että niitä molempia ei olisi hyvä ottaa regressiomalliin mukaan. Jos muuttujien välinen korrelaatio on suurempaa kuin 0,7, voi käydä niin, että regressiomallin kertoimet eivät ole tarpeeksi merkittäviä, koska ne eivät eroa juurikaan nollasta (Kiernan, 2014, kappale 8). Tämän vuoksi valitaan muuttujaksi mieluummin väkiluku, sillä sen korrelaatio on paras KL-tarpeen suhteen. Toiseksi muuttujaksi jää siis taajama-aste ja se sopii, koska sillä ei ole liian suurta suhdetta väkilukuun ja lisäksi sen suhde KL-tarpeeseen on kohtalainen. Kaksi muuttujaa on tähän työhön sopiva määrä ottaen huomioon työn laajuus. Kahden

muuttujan esittäminen suhteessa haluttuun muuttujaan eli KL-tarpeeseen vaatii jo 3D-kuvaajaa, joten jos tähän lisäisi vielä kolmannen muuttujan, menisi lopullisten tulosten esittäminen hankalaksi. Lähdetään rakentamaan mallia asukasluvun ja taajama-asteen avulla.

## 3.2 Regressiomalli

Moniregressiomalli tarkoittaa regressiomallia, jossa on useampi kuin yksi tekijä. Moniregressiomalli on yhtälön 2 mukainen

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + e \tag{2}$$

jossa $Y$ on tulos, $\beta$ on regressiomallin parametri, jotka täytyy selvittää tilastojen pohjalta, $x$ on malliin valittu muuttuja, $k$ kertoo monesko tekijä on kyseessä ja $e$ on satunnaisvirhe (Ross 2004, 394.)

Yhtälöstä 2 on tavoitteena selvittää $\beta$-arvot, jotta sitä voidaan käyttää arvioimaan KL-tarvetta muualla. Nyt syötetään Pohjois-Savon jokaisen kunnan halutut muuttujien arvot sekä KL-tarve yhtälöön. Ratkaistaan $\beta$-arvot muodostamalla matriisit seuraavasti

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

jossa $\boldsymbol{Y}$ on $n \times 1$ matriisi, $\boldsymbol{X}$ on $n \times p$, jossa $p = k + 1$, $\boldsymbol{\beta}$ on $p \times 1$ ja $\boldsymbol{e}$ on $n \times 1$ (Ross 2004, 396)

ja nyt voidaan esittää moniregressiomalli muodossa

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{3}$$

Syötetään matriisi muodossa Pohjois-Savon kuntien KL-tarve $\boldsymbol{Y}$-matriisiin, $\boldsymbol{X}$-matriisiin syötetään asukasluku ja taajama-aste. Selvitetään regressiomallin parametrit $\beta$ ja $e$ tilastojen

perusteella. Valitaan ensimmäiseksi muuttujaksi väkiluku ja toiseksi muuttujaksi taajama-aste. Regressiomallin parametrit ovat taulukossa 4.

Taulukko 4. Moniregressiomallin parametrit

| Parametri | Arvo |
|---|---|
| $\beta_0$ | -23,1534 |
| $\beta_1$ | -0,0292 |
| $\beta_2$ | 1,5496 |
| $\beta_{12}$ | 0,0004369 |

Moniregressiomalli saa nyt muodon MATLABin ’regress’ funktiota käyttämällä

$$Y = -23{,}1534 - 0{,}0292x_1 + 1{,}5496x_2 + 0{,}0004369 \cdot x_1x_2 \qquad (4)$$

Verrataan yhtälön 4 mukaisia mallin tuloksia muutamaan yksittäiseen kuntaan Pohjois-Savossa taulukossa 5. Huomataan, että saadaan negatiivisia arvoja Kiuruvedelle KL-tarpeeksi, mikä ei kuulosta järkevältä. Erot ovat paikoitellen pieniä sekä suuria tilastotiedon ja mallin välillä.

Taulukko 5. Mallin antamia tuloksia ’regress’-funktiota käyttämällä

| KL-tarve | Malli (GWh) | Tilasto (GWh) | Ero (GWh) | Ero (%) |
|---|---|---|---|---|
| Kiuruvesi | -0,77 | 41,60 | 42,37 | 102 |
| Iisalmi | 185 | 183,6 | 1,36 | 0,7 |
| Joroinen | 49,23 | 14,2 | 35 | 246 |
| Kuopio | 1182,5 | 1172,7 | 9,8 | 0,8 |
| Rautalampi | 27,13 | 12,5 | 14,6 | 117 |
| Lapinlahti | 53,8 | 32,9 | 25,3 | 77 |
| Varkaus | 341 | 418,6 | 77,5 | 19 |

Lähestytään mallin rakentamista toisenlaisella funktiolla. Käytetään MATLABin ’fitlm’ funktiota, joka sovittaa regressiomallin valittuihin muuttujiin ja ennustettavaan muuttujaan. Funktio ei ota regressiomalliin mukaan arvoja, joilla ei ole numeroa eli tässä tapauksessa se tarkoittaa kuntia ilman KL-tarvetta. Funktion laskemat parametrit moniregressiomallille ovat taulukossa 6.

Taulukko 6. Moniregressiomallin parametrit ’fitlm’-funktiolla

| Parametri | Arvo |
|---|---|
| $b_0$ | -189,8 |
| $b_1$ | 0,0089105 |
| $b_2$ | 3,1609 |

$$y = b_0 + b_1 x_1 + b_2 x_2 \tag{5}$$

$$y = -189{,}8 + 0{,}0089105 x_1 + 3{,}1609 x_2$$

Regressiomalli saa yhtälön 5 mukaisen muodon. Verrattaan mallia muutamaan yksittäiseen kuntaan Pohjois-Savossa taulukossa 7. Huomataan että negatiivisia tuloksia tulee Rautalammilla, jolla on pieni KL-tarve.

Taulukko 7. Malli sijoitettuna yksittäisiin Pohjois-Savon kuntiin

| KL-tarve | Malli (GWh) | Tilasto (GWh) | Ero (GWh) | Ero (%) |
|---|---|---|---|---|
| Kiuruvesi | 38,6 | 41,60 | 3 | 7 |
| Iisalmi | 239,1 | 183,6 | 55,5 | 30 |
| Joroinen | 35,1 | 14,2 | 20,9 | 147 |
| Kuopio | 1168,2 | 1172,7 | 4,5 | 0,4 |
| Rautalampi | -10,1 | 12,5 | 22,6 | 180 |
| Lapinlahti | 62,7 | 32,9 | 29,8 | 91 |
| Varkaus | 279,6 | 418,6 | 139 | 33 |

Jälkimmäisellä funktiolla (yhtälö 5) saadaan lähempänä tilastoa olevia tuloksia. Yhtälön 4 mukainen malli ei kuvaa missään satunnaisesti valitussa Pohjois-Savon kunnassa KL-tarvetta paremmin kuin yhtälöllä 5 toteutettu malli. Kuitenkin ”regress”-funktiolla tehty malli antaa Iisalmessa ja Kuopiossa arvion KL-tarpeelle hyvin lähelle oikeaa, kun taas yhtälön 5 mukainen malli pääsee lähelle vain Kuopiossa. Yhtälön 4 mukaisella mallilla voisi saada useampia tarkkoja lukuja. Yhtälön 4 mukainen malli sisältää muuttujien keskinäisen tulon termin, kun taas yhtälön 5 mukainen malli ei sisällä ristikerronta termiä. Yhtälön 5 mukaisella mallilla voidaan ehkä helpommin tarkastella muuttujien vaikutusta KL-tarpeeseen ja sen käytöstä on helpompi tulkita, kun siinä ei oteta huomioon muuttujien keskinäisen tulon vaikutusta. Koska tässä tutkimuksessa ei lähdetä kehittämään mallia sen pidemmälle ja halutaan tarkastella nyt tätä mallia, niin jälkimmäinen ’fitlm’-funktio on tähän tarkoitukseen sopivampi (MathWorks, 2024).

# 4 Tulosten tarkastelua

Tässä kappaleessa tarkastellaan aluksi muuttujien keskinäisiä suhteita ja niiden sopivuutta mallissa. Tämän jälkeen pohditaan saatuja tuloksia ja mallin onnistumista tilastotietoon eli todellisuuteen nähden.

## 4.1 Muuttujien arviointi

Mallin sopivuutta voidaan arvioida erilaisilla parametreillä. Yksi tähän sopiva parametri on $R^2$-luku. Se kertoo kuinka suuren osan muuttujat selittävät arvioitavasta arvosta ja se voi saada arvoja väliltä 0–1. $R^2$-luku saattaa nousta korkeammaksi mitä enemmän muuttujia malliin lisää. Tähän ongelmaan puuttuu tarkennettu $R^2$-luku, joka antaa paremman kuvauksen muuttujien sopivuudesta. (Kiernan 2014, kappale 8.)

$R^2$-luku on tämän mallin tapauksessa 0,971 ja tarkennettu $R^2$-luku on 0,965 eli ne ovat hyvin suuria. Tämä tarkoittaa sitä, että ainakin toinen muuttujista kuvaa erittäin hyvin KL-tarvetta.

Jos halutaan tarkemmin tarkastella vielä yksittäisen muuttujan vaikutusta arvioitavaan arvoon, voidaan se tehdä t-testin avulla ja testin merkittävyyttä arvioidaan p-arvon avulla. Mitä suurempi t-arvo on, sitä enemmän muuttuja vaikuttaa lopulliseen arvioitavaan muuttujaan. (Kiernan 2014, kappale 8.)

T-testin arvot ja p-arvot ovat esitetty taulukossa 8. Taulukosta huomataan, että väkiluvun t-arvo on suuri ja p-arvon perusteella merkittävä (<0,05). Väkiluvun merkitys KL-tarpeeseen huomattiin jo muuttujia valittaessa, kun väkiluvun korrelaatio KL-tarpeeseen oli merkittävät. Taajama-asteen t-arvo ei ole aivan yhtä suuri ja p-arvo alkaa lähestyä myös merkitsevyyden rajaa (0,05) eli tästä voi päätellä, että taajama-asteen merkitys KL-tarpeen suhteen on kohtalainen tai hieman heikko. Taajama-asteen korrelaatio KL-tarpeeseen havaittiin jo muuttujien arviointi vaiheessa olevan hieman rajoilla merkittävyyden suhteen.

Taulukko 8. Kahden muuttujan t-arvot ja p-arvot

| | *t*-arvo | *p*-arvo |
|---|---|---|
| Väkiluku | 13,101 | 0,0000 |
| Taajama-aste | 2,505 | 0,0311 |

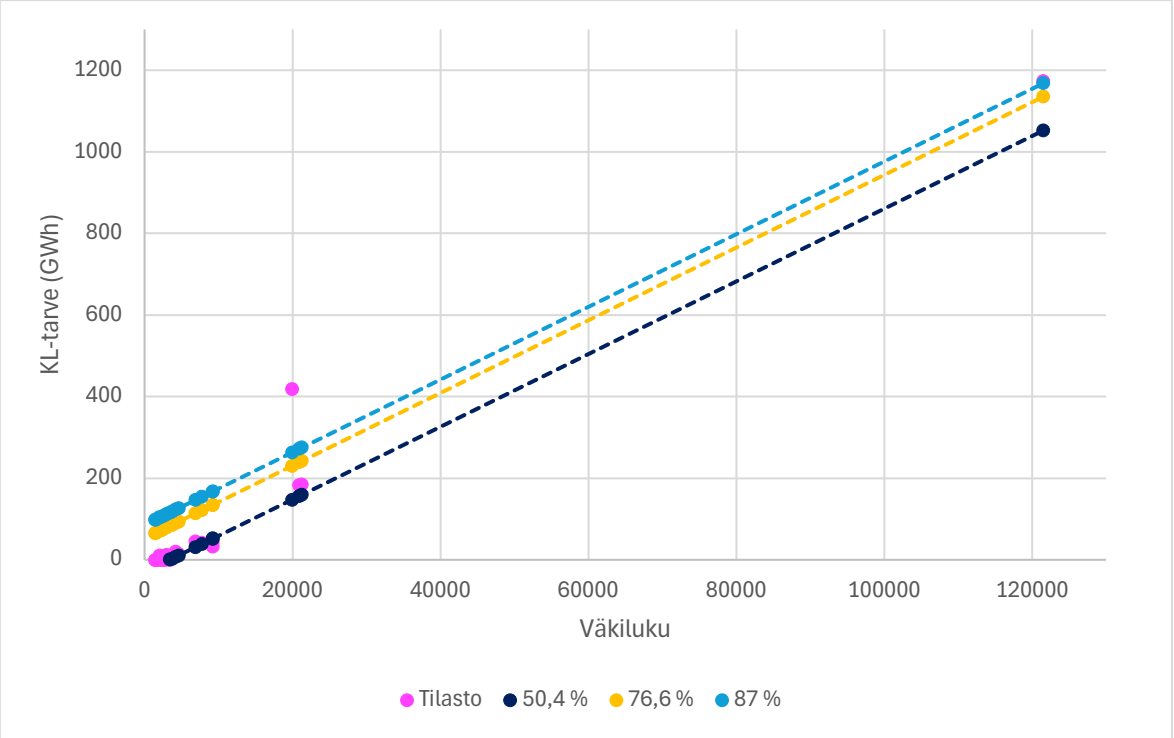
Mallin arviointivaiheessa myös huomattiin (taulukko 3), että muuttujilla on keskinäistä lineaarista suhdetta suunnilleen saman verran kuin taajama-asteella on KL-tarpeen kanssa (taulukko 2). Eli muuttujilla on keskenään hienoista suhdetta toisiinsa ja on mahdollista, että tämä on voinut vaikuttaa moniregressiomallin kertoimiin. Mutta toisaalta jouduttiin tekemään kompromissi sen suhteen, koska pinta-ala olisi ollut vielä vahvemmin yhteydessä väkilukuun, joka on paras selittävä muuttuja KL-tarpeelle tässä tutkimuksessa.

Muuttujien valinnassa malliin ajateltiin myös sitä, että tilastot mallia varten olisivat helposti saatavilla. Käytettiin siis sellaisia tilastoja, jotka ovat yleisesti saatavilla. Tämä toisaalta tarkoittaa sitä, että kovin spesifejä muuttujia mallille ei saada. Vielä tarkempien lämmönkulutukseen liittyvien muuttujien kerääminen tarkoittaisi haastavampaa tiedonkeräystä ja se ei tämän työn laajuus huomioon ottaen ollut mahdollista.

## 4.2 Mallin erot tilastoihin verrattuna

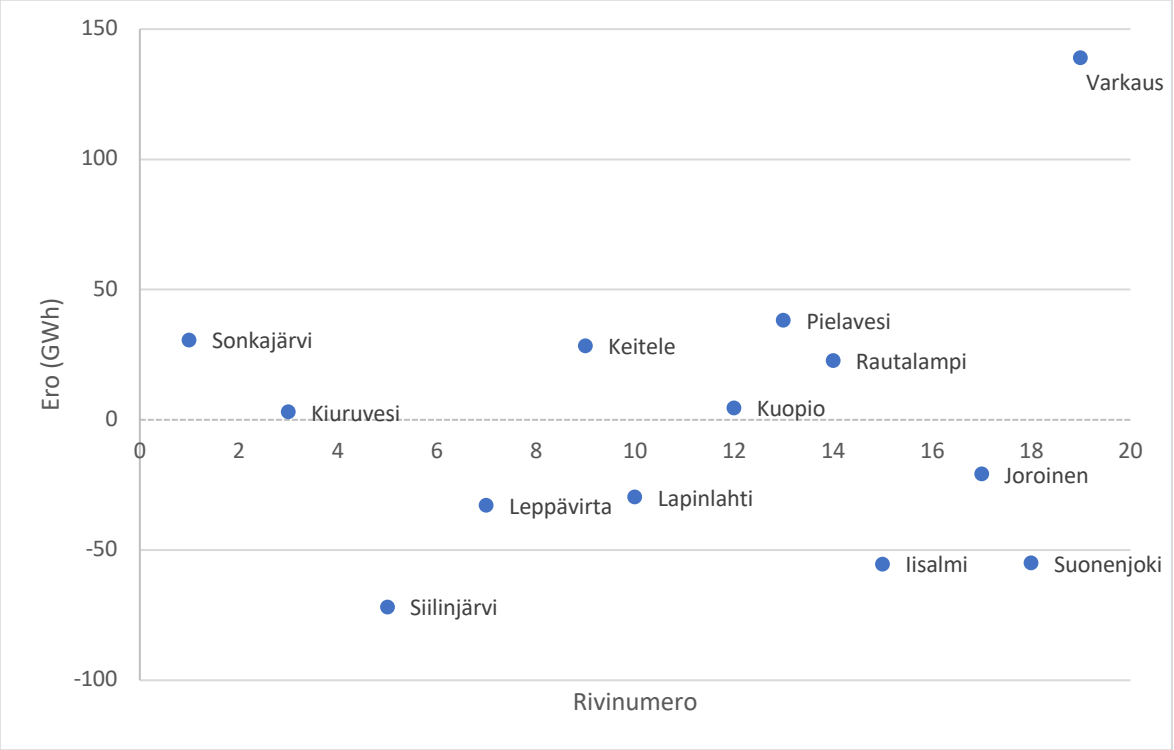
Verrataan moniregressio mallilla saatuja arvoja tilastoihin kuvan avulla. Kahdella muuttujalla yhtälöstä tulee kolmiulotteinen, mutta mallin ja tilastojen eroja on helpompi havainnollistaa kahdella dimensiolla. Asetetaan taajama-aste vakioksi, sillä asukasluku havainnollistaa paremmin KL-tarvetta. Jätetään toinen muuttuja vakioksi niin, että yhtälöön 5 sijoitetaan muuttujan $x_2$ paikalle yksi taajama-asteen arvo ja lasketaan KL-tarve yhtälöllä. Toistetaan tämä lisäksi kahdella muulla taajama-asteen arvolla, jolloin saadaan kolme eri pistejonoa ja näihin sovitettua suorat. Sijoitetaan oikeat arvot tilastoista myös kuvaan ja vertaillaan saatuja käyriä (kuva 8). Kuvassa 8 on esitetty muutamalla keskeisimmällä taajama-asteen arvolla mallin antamia tuloksia. Parhaiten malli vastaa tilastoa ainakin pienemmissä kunnissa taajama-asteen arvon ollessa 50,4 %. Toisaalta yhden muuttujan vakioksi jättäminen ei anna tarkkaa kuvaa mallin ja todellisuuden suhteesta, mutta se antaa siihen suuntaan olevia

tuloksia. Kuvasta 8 voi myös huomata, että 20 000 asukkaan kohdalla oleva piste (Varkaus) erottuu tilastopisteiden joukosta.




Kuva 8. Mallin ja tilaston väliset erot kolmella eri taajama-asteen arvolla

Tarkastellaan mallin ja tilaston välistä absoluuttista eroa KL-tarpeelle. Kuvassa 9 esitetään kuinka paljon mallin antamat arvot kaukolämmöntarpeelle eroavat tilastoidusta arvosta. Kuvassa keskiviiva nollan kohdalla tarkoittaa tilastoitua arvoa. Positiivinen ero tarkoittaa sitä, että malli on arvioinut KL-tarpeen kunnassa liian matalaksi tilastoon verraten ja negatiivinen ero puolestaan sitä, että mallista saatu KL-tarve kunnassa on tilastoa suurempi. Kuvassa on esitetty vain ne Pohjois-Savon kunnat, joissa on KL-tarvetta. Kuvasta nähdään, että malli vastaa parhaiten tilastoitua KL-tarvetta Kiuruvedellä ja Kuopiossa. Eniten heittoa mallin ja oikean KL-tarpeen välillä on Varkaudessa ja Siilinjärvellä. Useimpien kuntien ero jää alle 50 GWh:n.




Kuva 9. Mallin ja tilaston ero, jossa kunnat liitteen 1 mukaisessa järjestyksessä.

Varkaudessa absoluuttinen ero on korkea, mutta toisaalta kuten taulukosta 7 nähdään, suhteellinen ero mallin ja tilaston välillä on vain noin 30 %. Taulukosta 7 nähdään myös, että suhteellinen ero pienemmissä kunnissa kuten Joroisissa tai Rautalammilla on suuri, vaikka kuvasta 9 nämä kunnat eivät erotu. Toisaalta ei voi myöskään todeta, että malli sopisi huonosti pienempiin kuntiin, koska lähimpänä tilastoa oli mallin antama arvio Kiuruvedelle, joka on pienempi kunta.

Tarkastellaan Varkauden tapausta hieman tarkemmin, sillä suuri ero mallin ja tilastoidun välillä kertoo, ettei se sovi lineaariseen malliin kuten kuvasta 8 nähdään. Asiaa voi selittää se, että Varkaus on ainoa Pohjois-Savon kunta, jossa teollisuuden KL-tarve on suurempaa kuin asukaskäyttö (Energiateollisuus, 2023). Mallin muuttujat ovat sellaiset, että ne eivät huomioi teollisuuden KL-tarvetta, sillä kaupungin koko ei vaikuta suoranaisesti teollisuuden määrään. Tämän voi huomata Pohjois-Savon kohdalla siinä, ettei Varkaus ole suurimmasta tai pienimmästä päästä mutta siellä teollisuuden osuus on silti suuri. Asukasluku ja taajama-aste ottavat enemmän huomioon asumiseen liittyvän KL-tarpeen.

Taulukossa 9 on vertailtu saatua monilineaariregressio mallia Pohjois-Savon ulkopuolisiin suomalaisiin kuntiin. Taulukosta 9 huomataan, että parhaimmillaan malli heittää 30 GWh tilastoidusta eli todellisesta arvosta. Tyypillinen ero on noin 80 GWh ja huonoimmillaan lähestytään yli 200 GWh eroa todelliseen KL-tarpeeseen. Toisaalta Pohjois-Savoon kunnissa ero todellisen ja mallin antaman välillä oli parhaimmillaan 3 GWh mutta muuten tyypillisesti noin 30 GWh (kuva 9).

Taulukko 9. Mallin erot Pohjois-Savon ulkopuolella olevissa kaupungeissa.

| KL-tarve | Malli (GWh) | Tilasto (GWh) | Ero (GWh) | Ero (%) |
|---|---|---|---|---|
| Porvoo | 520,4 | 341,2 | 179,2 | 53 |
| Tampere | 2280,6 | 2356,5 | 75,9 | 3 |
| Lahti | 1188,2 | 1353,2 | 165 | 14 |
| Joensuu | 745,2 | 639 | 106,2 | 17 |
| Oulu | 1976,1 | 2003,1 | 27 | 1 |
| Rauma | 451,9 | 316,4 | 135,5 | 43 |

Omakotitalon KL-tarpeeksi voidaan arvioida noin 0,02 GWh vuodessa (Motiva et al., 2024). Näin ollen ero mallinnetun ja tilastoiden KL-tarpeen välillä parhaimmillaan Kiuruvedellä (3 GWh) vastaa 150:n omakotitalon tarvetta. Kiuruvedellä on asuintaloasiakkaita noin 140, joten, vaikka ero on pieni, niin suhteutettuna kunnan kokoon se on suuri (Energiateollisuus, 2023). Toinen esimerkki parhaiten osuvasta on Kuopio, jossa ero mallin ja tilaston välillä on noin 4,5 GWh. Tämä ero tarkoittaa noin 225:tä omakotitaloa ja Kuopiossa kaukolämmön käyttäjinä on noin 5500 asuintaloa, joten ero ei suhteessa ole kovin suuri (Energiateollisuus, 2023). Prosentuaalisesti ero Kuopiossa oli mallin ja tilaston välillä paras, sillä se oli vain 0,4 %. Suurimmassa osassa Pohjois-Savon kunnissa malli ei ole kovin lähellä oikeaa KL-tarvetta. Taulukosta 9 havaitaan, että malli soveltuu huonommin Pohjois-Savon ulkopuolisiin kuntiin, sillä niissä mallin ja tilaston välinen absoluuttinen ero on suurempaa, vaikka prosentuaalisesti ero olisi pientä.

Asukasluku on hyvä muuttuja kuvaamaan KL-tarvetta, kun taas taajama-aste ei ole yhtä hyvä muuttuja kuvaamaan lineaarista suhdetta KL-tarpeelle. Malliin olisi voinut valita muuttujaksi mahdollisesti toisen muuttujan, joka kuvaisi paremmin esimerkiksi teollisuuden KL-

tarvetta, sillä nykyiset muuttujat eivät siitä kerro mitään. Toisaalta hankala sanoa voiko tilastojen perusteella määritetyt parametrit saada mallista niin tarkkaa, että ero mallin ja tilaston välillä olisi vain 1 GWh eli noin 50 omakotitaloa, joka voisi olla tarpeeksi tarkka, jos mallia haluaisi käyttää avuksi uuden alueen suunnittelussa. Mallin tarkkuutta voisi myös parantaa se, että käyttäisi kaikkien Suomen kuntien tilastoja. Tämä toisaalta lisäisi analysoinnin määrää, kun tarkasteltavien kuntien määrä olisi niin suuri. KL-tarpeen vuosittainen vaihtelu voitaisiin ottaa myös huomioon käyttämällä mallin rakentamisessa useamman vuoden tilastoja. Nykyisessä tilanteessa mallilla saadaan parhaassakin tapauksessa vain suuntaa antava KL-tarve.

Vaikka aiemmissa tutkimuksissa on huomattu säähän liittyvillä muuttujilla saatavan hyviä tuloksia KL-tarpeen arvioinnissa, voi tämä yleisempiä tilastoja hyödyntävä regressiomalli, tai toisenlainen tilastollinen malli, olla toimiva tapa selvittää alueen suuntaa antavaa KL-tarvetta. Toisaalta yleisempiä tilastoja hyödyntävän mallin muuttujien valintaan ja testaukseen täytyy käyttää enemmän aikaa, koska muuttujat eivät ole räätälöity sopivaksi mallia varten.

# 5 Johtopäätökset

Tässä työssä selvitettiin, voidaanko valmiina olevien kaukolämmöntarpeen tilastojen pohjalta luoda malli KL-tarpeelle kunnassa. Tavoitteena oli luoda malli KL-tarpeelle käyttäen otoksena Pohjois-Savon kuntien tilastoja, niin että mallia voisi käyttää arvioimaan KL-tarvetta myös muissa kunnissa Suomessa.

Malliksi valittiin useamman muuttujan regressiomalli. Mallin tavoitteena on tutkia yhdyskuntarakenteeseen liittyvien muuttujien vaikutusta KL-tarpeeseen ja tutkia pystytäänkö niiden avulla arvioimaan tätä. Muuttujiksi valittiin asukasluku sekä taajama-aste ja havaittiin, että näillä on kohtalaista lineaarista yhteyttä KL-tarpeeseen. Sen jälkeen laskettiin lineaarisen moniregressiomallin parametrit käyttämällä Pohjois-Savon kuntien KL-tarvetta, asukaslukuja ja taajama-asteita. Tämän jälkeen testattiin kuinka suuri ero regressiomallin antamien tulosten ja tilastojen välillä oli. Havaittiin että eroa oli pienimmillään noin 0,4 % mallin ja oikean KL-tarpeen välillä. Mallilla päästään joissain yksittäisissä kunnissa lähelle oikeaa KL-tarvetta mutta suurimmassa osassa kuntia ero jää suureksi.

Jatkoa tälle tutkimukselle voisi olla se, että tutkisi regressiomallilla KL-tarvetta vaihtamalla taajama-asteen toiseen muuttujaan. Uusi malli voisi ottaa paremmin huomioon alueen erityispiirteitä kuten suurempaa teollisuuden osuutta tai muita paljon kaukolämpöä tarvitsevia kohteita. Voisi myös tarkastella useamman vuoden aikaisia kaukolämmön tilastoja, jolloin malli ottaisi huomioon eri vuosien väliset erot lämmöntarpeessa. Tutkimusta voisi tehdä myös niin, että tarkastelisi kuntien ja kaupunkien KL-tarvetta ilman, että rajaisi sitä tiettyyn maakuntaan. Laajemmalla tilastoaineistolla voisi mahdollisesti optimoida mallin parametrit paremmiksi, niin että mallia voisi soveltaa useampaan kaupunkiin.

On mahdollista, että yhdyskuntarakenteeseen liittyvillä tekijöillä on mahdollista saada tämän tutkimuksen mukaisia parhaassa tapauksessa suuntaa antavia arvoja. Jos haluttaisiin tarkemmin tarkastella KL-tarvetta, niin siihen täytyisi tarkastella ulkolämpötilaa ja sääoloja, joiden on havaittu aikaisemmissa tutkimuksissa vaikuttavan merkittävästi kaukolämmöntarpeeseen.

# Lähteet

Davison, A. C. 2003. Statistical Models. Cambridge, Cambridge University Press, [E-kirja]. [Viitattu 16.7.2024] DOI: 10.1017/CBO9780511815850

Energiateollisuus ry. 2023 Kaukolämpötilasto 2021, Taulukot. [Verkkoaineisto]. [Viitattu 30.10. 2023] Saatavissa: https://energia.fi/tilastot/kaukolampotilasto/.

Fang, T. ja Lahdelma, R. 2016. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Applied Energy 179 s. 544–552.

Hietaharju, P., Pulkkinen, J., Ruusunen, M. ja Louis, J. 2021. A stochastic dynamic building stock model for determining long-term district heating demand under future climate change. Applied Energy 295 s. 116962.

Jylhä, K., Jokisalo, J., Ruosteenoja, K., Pilli-Sihvola, K., Kalamees, T., Seitola, T., Mäkelä, H. M., Hyvönen, R., Laapas, M. ja Drebs, A. 2015. Energy demand for the heating and cooling of residential houses in Finland in a changing climate. Energy and Buildings 99 s. 104–116.

Kalamees, T., Jylhä, K., Tietäväinen, H., Jokisalo, J., Ilomets, S., Hyvönen, R. ja Saku, S. 2012. Development of weighting factors for climate variables for selecting the energy reference year according to the EN ISO 15927-4 standard. Energy and Buildings 47 s. 53–60.

Kiernan, D. 2014. Natural Resources Biometrics. State University of New York at Geneseo: Open SUNY Textbooks, [E-kirja]. [Viitattu 12.8.2024]. Saatavissa: https://milnepublishing.geneseo.edu/natural-resources-biometrics/.

Ma, Z., Li, H., Sun, Q., Wang, C., Yan, A. ja Starfelt, F. 2014. Statistical analysis of energy con-sumption patterns on the heat demand of buildings in district heating systems. Energy and Buildings 85 s. 464–472.

Maanmittauslaitos. 2024. Kuntajakopohjaiset hallinnolliset aluejaot. [Verkkoaineisto]. [Viitattu 10.3.2024] Saatavissa: https://kartat.kapsi.fi.

MacInnes, J. 2022. The Challenge and Promise of Inference. Teoksessa: MacInnes, J. (toim.) MStatistical Inference and Probability. London: SAGE Publications Ltd, 12, [E-kirja]. [Viitattu 15.8.2024] DOI: 10.4135/9781529682748

MathWorks. 2022. MATLAB. R2022b

MathWorks. 2024. Regress documentation. [Verkkosivu]. [Viitattu 23.8.2024]. Saatavissa: https://se.mathworks.com/help/stats/regress.html.

Motiva, Kesto ja JAMK. 2024. Pientalon Lämmitystapojen Vertailulaskuri. [Viitattu 12.8.2024]. Saatavissa: https://www.motiva.fi/koti_ja_asuminen/energiatehokas_pientalo/lammitysjarjestelman_valinta/vertaile_lammitysjarjestelmia/pientalon_lammitystapojen_vertailulaskuri.

Nickolas, S. 2023 Correlation Coefficients: Positive, Negative and Zero. [Verkkosivu]. [Viitattu 12.8. 2024]. Saatavissa: https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp.

QGIS Development team. 2023. Qgis. 3.32.3-Lima.

Ross, S. M. 2004. Introduction to Probability and Statistics for Engineers and Scientists. San Diego: Elsevier Science & Technology, [E-kirja]. [Viitattu 27.3.2024]

Suomen ympäristökeskus. 2024. Yhdyskuntarakenteen aluejaot 2022. [Verkkoaineisto]. [Viitattu 10.3.2024] Saatavissa: https://www.syke.fi/fi-FI/Avoin_tieto/Paikkatietoaineistot/Ladattavat_paikkatietoaineistot.

Tilastokeskus. 2024. Kuntien Avainluvut 2021 Aluejaolla. [Verkkoaineisto].[Viitattu 26.3. 2024]. Saatavissa: https://pxdata.stat.fi/PxWeb/pxweb/fi/Kuntien_avainluvut/Kuntien_avainluvut__2021/.

Liite 1. Työssä käytetyt tilastot Pohjois-Savon kunnista

| Nimi | Pinta-ala | Väkiluku | Taajama-aste % | KL-tarve (GWh) |
|---|---|---|---|---|
| 1. Sonkajärvi | 1576,79 | 3777 | 43,7 | 12,6 |
| 2. Tuusniemi | 699,42 | 2420 | 41,5 | 0 |
| 3. Kiuruvesi | 1422,9 | 7759 | 50,4 | 41,60 |
| 4. Rautavaara | 1235,25 | 1513 | 44,2 | 0 |
| 5. Siilinjärvi | 507,81 | 21293 | 81,1 | 184,40 |
| 6. Tervo | 494,3 | 1479 | 34,5 | 0 |
| 7. Leppävirta | 1519,64 | 9280 | 57,9 | 42,9 |
| 8. Vieremä | 973,38 | 3490 | 39,2 | 0 |
| 9. Keitele | 578,3 | 2095 | 48,5 | 10,5 |
| 10. Lapinlahti | 1245,16 | 9247 | 53,8 | 32,9 |
| 11. Vesanto | 569,81 | 1941 | 40,1 | 0 |
| 12. Kuopio | 4326,35 | 121543 | 87 | 1172,7 |
| 13. Pielavesi | 1406,51 | 4269 | 42,4 | 20,3 |
| 14. Rautalampi | 761,98 | 3033 | 48,3 | 12,5 |
| 15. Iisalmi | 872,17 | 20958 | 76,6 | 183,6 |
| 16. Kaavi | 789,59 | 2778 | 46,1 | 0 |
| 17. Joroinen | 711,75 | 4624 | 58,1 | 14,2 |
| 18. Suonenjoki | 862,34 | 6891 | 72,1 | 44,5 |
| 19. Varkaus | 524,48 | 19973 | 92,2 | 418,6 |