



TALLINNA TEHNIKAÜLIKOOL

INSENERITEADUSKOND

Elektroenergeetika ja mehhatroonika instituut

EESTI GAASITARBIMISE PÕHILISED MÕJURID NING NENDE PÕHJAL TARBIMISE PÄEV-ETTE PROGNOOSIMINE

MAIN FACTORS AFFECTING ESTONIAN GAS CONSUMPTION AND USING
THEM TO FORECAST GAS DEMAND ON DAY-AHEAD BASIS

MAGISTRITÖÖ

Üliõpilane: Sander Randver

Üliõpilaskood: 163640AAVM

Juhendaja: Hardi Koduvere, Nooremteadur

Tallinn, 2019

AUTORIDEKLARATSIOON

Olen koostanud lõputöö iseseisvalt.

Lõputöö alusel ei ole varem kutse- või teaduskraadi või inseneridiplomit taotletud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

“.....” mai 2019

Autor:

/ allkiri /

Töö vastab bakalaureusetöö/magistritööle esitatud nõuetele

“.....” mai 2019

Juhendaja:

/ allkiri /

Kaitsmisele lubatud

“.....”2019 .

Kaitsmiskomisjoni esimees

/ nimi ja allkiri /

LÕPUTÖÖ LÜHIKOKKUVÕTE

Autor: Sander Randver

Lõputöö liik: Magistritöö

Töö pealkiri: Eesti gaasitarbimise põhilised mõjurid ning nende põhjal tarbimise päev-ette prognoosimine

Kuupäev: 24.05.2019

78 lk (lõputöö lehekülgede arv koos lisadega)

Ülikool: Tallinna Tehnikaülikool

Teaduskond: Inseneriteaduskond

Instituut: Elektroenergeetika ja mehhatroonika instituut

Töö juhendaja(d): Hardi Koduvere

Sisu kirjeldus:

Lõputöö gaasitarbimise põhilistes mõjuritest keskendub gaasitarbimisele olulisimate ilmastikumõjurite leidmisele ning nende arvuliste parameetrite välja selgitamisele. Energia jaeäris on olulisel kohal energiatarbimise prognoosimine lähitulevikuks päev-ette resolutsiooniga. Energiatarbimise prognoosimine on tähtis täpse bilansiplaani koostamiseks ning seeläbi bilansienergia kulude vähendamiseks. Prognoosimine on oluline ka varustuskindluse koha pealt, et tagada alati gaasisüsteemis võimalikult täpne kogus gaasi ja seeläbi optimeerida ka energiasüsteemis esinevaid muid kulusid.

Töö eesmärgiks on välja selgitada erinevate ilmategurite täpne numbriline mõju gaasitarbimisele ning seeläbi saavutada võimalikult täpne väljund gaasitarbimise prognoosimisel lineaarsel regressioonil põhineva prognoosimudeliga rakendustarkvaras R või Python selgitades välja kõige olulisemad mõjurid gaasitarbimisele ning võttes vastavad parameetrid ka prognoosimudelite arvutusprotsessis kasutusele. Mõjuritele arvatud parameetritele antakse hinnang pikaajalise (mitu aastat) keskmise prognoosivea põhjal, mida arvutatakse MAPE põhjal.

Märksõnad: gaasitarbimine, gaasiturg, energiatarbimise prognoosimine, regressioonanalüüs, regressioonimudel.

ABSTRACT

Author: Sander Randver

Type of the work: Master Thesis

Title: Main factors affecting Estonian gas consumption and using them to forecast gas demand on day-ahead basis

Date: 24.05.2019

78 pages (the number of thesis pages including appendices)

University: Tallinn University of Technology

School: School of Engineering

Department: Department of Electrical Power Engineering and Mechatronics

Supervisor(s) of the thesis: Hardi Koduvere

Abstract:

The aim of this thesis work is to determine the impact of the main weather factors to Estonian gas consumption in a numerical way. In energy retail business day-ahead forecasting of energy consumption balance plans are important to compose as accurate as possible and to minimize the balancing cost of providing the energy to the end users. Forecasting the demand is also important from system's security of supply point of view to ensure the most accurate possible volume of gas in the transmission network and therefore to minimize different costs that can occur with inaccurate balance plans.

In addition to finding out the numerical values of impacts of different weather factors, this thesis also focuses on using those numerical values to forecast gas demand in Estonia on day-ahead basis. To achieve this, the practical part of the work uses a forecasting model based on multivariate linear regression and it is composed in R or Python. The numerical values of impact of different weather factors are used in this model as input. Significance and relevance of those factors is assessed by analysing the forecast model's long term MAPE.

Keywords: gas demand, gas market, energy demand forecasting, regression analysis, regression model.

LÕPUTÖÖ ÜLESANNE

Lõputöö teema:	Eesti gaasitarbimise põhilised mõjurid ning nende põhjal tarbimise päev-ette prognoosimine
Lõputöö teema inglise keeles:	Main factors affecting Estonian gas consumption and using them to forecast gas demand on day-ahead basis
Üliõpilane:	Sander Randver
Eriala:	Elektroenergeetika
Lõputöö liik:	magistritöö
Lõputöö juhendaja:	Hardi Koduvere
Lõputöö ülesande kehtivusaeg:	kehtivusaja annab juhendaja
Lõputöö esitamise tähtaeg:	24.05.2019

Üliõpilane (allkiri)

Juhendaja (allkiri)

Õppekava juht (allkiri)

1. Teema põhjendus

Lõputöö gaasitarbimise prognoosimisest keskendub gaasitarbimisele olulisimate ilmastikumõjurite leidmisele ning nende arvuliste parameetrite välja selgitamisele. Energia jaeäris on olulisel kohal energiatarbimise prognoosimine lähitulevikuks päev-ette resolutsiooniga. Energiatarbimise prognoosimine on tähtis täpse bilansiplaani koostamiseks ning seeläbi bilansienergia kulude vähendamiseks. Prognoosimine on oluline ka varustuskindluse kohapealt, et tagada alati gaasisüsteemis võimalikult täpne kogus gaasi ja seeläbi optimeerida ka energiasüsteemis esinevaid muid kulusid.

2. Töö eesmärk

Töö eesmärgiks on välja selgitada erinevate ilmategurite täpne numbriline mõju gaasitarbimisele ning seeläbi saavutada võimalikult täpne väljund gaasitarbimise prognoosimisel lineaarsel regressioonil põhineva prognoosimudeliga rakendustarkvaras R või Python selgitades välja kõige olulisemad mõjurid gaasitarbimisele ning võttes vastavad parameetrid ka prognoosimudelite arvutusprotsessis kasutusele. Mõjuritele arvatatud parameetritele antakse hinnang pikaajalise (mitu aastat) keskmise prognoosivea põhjal, mida arvutatakse MAPE põhjal.

3. Lahendamisele kuuluvate küsimuste loetelu:

Käesolevas lõputöös lahendamisele kuuluvad küsimused on järgnevad:

- Millised gaasitarbimise mõjurid on kõige olulisemad prognoosimisel?
- Millised on täpsed arvulised parameetrid, millega eeltoodud mõjureid gaasitarbimise prognoosimisel arvesse võtta?
- Milline on kõige täpsem prognoosimudel ning selle parameetrid võttes arvesse eelnevalt kindlaks tehtud mõjureid?

4. Lähteandmed

Põhilised andmed on gaasi tarbimisandmed, mis on saadavad Eleringi kodulehelt ning erinevad ilmastikuandmed, mis on saadavad Riigi Ilmateenistuse kodulehel.

5. Uurimismeetodid

Lõputöö raames on plaanis hankida esmalt piisavalt pika ajalooga tarbimis- ja ilmastikuandmed ning uurida erinevate ilmastikutegurite mõju gaasitarbimisele, et saada aru ilmateguri ning tarbimise vahelisest korrelatsioonist. Seejärel on tarvis selgitada välja kõige täpsem lineaarsel regressioonil põhinev prognoosimudel ning selle parameetrid .

6. Graafiline osa

Töö graafiline osa on plaanis siduda töö sisulise osaga ning põhiline graafiline osa peaks olema ka seal. Graafika abil on antud lõputöös kavas kirjeldada gaasitarbimise sõltuvust erinevates mõjuritest ning kasutada graafikuid ka andmaks hinnanguid mudeli parameetrite täpsusele.

7. Töö struktuur

1. Sisukord
2. Lõputöö ülesanne
3. Eessõna
4. Sissejuhatus
5. Statistiline analüüs
 - a. Temperatuuri mõju gaasitarbimisele
 - b. Pilvisuse või päikesepaiste mõju gaasitarbimisele
 - c. Tuuleolude mõju gaasitarbimisele
 - d. Sademete mõju gaasitarbimisele
6. Gaasitarbimise modelleerimine
 - a. Ülevaade automaatsest lineaarsest regressioonmudelist
 - b. Kindlate numbriliste väärtuste omistamine erinevatele mõjuritele

- c. Tulemuste välja selgitamine
- 7. Tulemused ja nende analüüs
- 8. Hinnang tulemustele ning määratud parameetrite täpsusele
- 9. Kokkuvõte
- 10. Kasutatud kirjandus

8. Kasutatud kirjanduse allikad

IEEE Xplore: <https://ieeexplore.ieee.org/Xplore/home.jsp>

Science Direct: <https://www.sciencedirect.com/>

Google Scholar: <https://scholar.google.com/>

TTÜ raamatukogu

9. Lõputöö konsultandid

Vajadusel konsultantide nimed ja töö osad, mille juures abi saadakse.

10. Töö etapid ja ajakava

- 1) Alusandmete kogumine (01.03.2019)
- 2) Alusandmete analüüs (31.03.2019)
- 3) Prognoosimudeli koostamine ja modelleerimine (14.04.2019)
- 4) Järeldused, kokkuvõte ja algne versioon (28.04.2019)
- 5) Töö lõplikult valmis (13.05.2019)

SISUKORD

LÕPUTÖÖ LÜHIKOKKUVÕTE.....	3
ABSTRACT	4
LÕPUTÖÖ ÜLESANNE	5
EESSÕNA.....	10
SISSEJUHATUS	11
1. Eesti gaasisüsteem	13
1.1 Eesti gaasisüsteemi ajalugu	13
1.2 Eesti gaasisüsteemi tulevikutrendid	14
1.3 Gaasiturg Eestis.....	16
2. Lineaarne regressioon	19
2.1 Lineaarse regressiooni kasutusnäited.....	20
2.2 Regressioonanalüüs	21
2.3 Regressioonimudel	26
3. Ilmastikumõjurite statistiline analüüs.....	29
3.1 Ilmastikuolude mõju gaasitarbimisele	30
3.2 Andmestiku regressioonanalüüs.....	38
3.3 Analüüsi tulemused ja järeldused.....	46
4. Gaasitarbimise modelleerimine	49
4.1 Lineaarsel regressioonil põhinev prognoosimudel.....	49
4.2 Modelleerimisel sooritatavad iteratsioonid	51
4.3 Tulemuste analüüs	54
4.4 Tulemuste kokkuvõte ja järeldus	66
KOKKUVÕTE	69
SUMMARY	71
KASUTATUD KIRJANDUS	73

LISAD	76
Lisa 1. Gaasitarbimise prognoosimise mudeli kood.....	76

EESSÕNA

Käesoleva lõputöö teema pakuti välja Elektroenergeetika ja mehhatroonika instituudi nooremteaduri Hardi Koduvere poolt, kes jagas ka lõputöö autorile töö koostamisel ka analüütilisi nõuandeid. Käesolevas töös käsitletakse ilmastikutegurite mõju Eesti gaasitarbimisele ning põhiliseks analüüsimeetodiks on regressioonanalüüs.

Soovin avaldada tänu Elektroenergeetika ja mehhatroonika instituudi nooremteadurile Hardi Koduverele, kelle juhendamisel käesolev lõputöö valmis. Lisaks soovin tänada Eesti Energia prognoosimise tiimijuhti Ingvar Lukast ja Bolti vanem-andmeanalüütikut Cliona Georgia Dalbergi modelleerimisalaste konsultatsioonide eest ning Eesti Energia modelleerijat Meelis Muhki abi eest käesoleva lõputöö alusandmete koostamisel.

Käesolev lõputöö on valminud Tallinnas.

SISSEJUHATUS

Käesolev lõputöö Eesti gaasitarbimise põhilistest mõjuritest keskendub gaasitarbimisele olulisimate ilmastikumõjurite leidmisele ning nende arvuliste parameetrite välja selgitamisele. Energia jaeäris on olulisel kohal energiatarbimise prognoosimine lähitulevikuks päev-ette resolutsiooniga. Energiatarbimise prognoosimine on tähtis täpse bilansiplaani koostamiseks ning seeläbi bilansienergia kulude vähendamiseks. Prognoosimine on oluline ka varustuskindluse kohapealt, et tagada alati gaasisüsteemis võimalikult täpne kogus gaasi ja seeläbi optimeerida ka energiasüsteemis esinevaid muid kulusid. Järjest enam avaneva konkurentsiga gaasiturul on turuosaliste jaoks oluline kõikide muutuvkulude minimeerimine, mis muudab gaasitarbimise prognoosimise aina olulisemaks. Euroopa Liidu Komisjoni määrus number 312/2014 sätestab, et kõik gaasivõrgu kasutajad peavad vastutama oma bilansiportfellide tasakaalus olemise ees, et minimeerida ülekandesüsteemi haldurite vajadust tasakaalustustoimingute järele. Sama määrus näeb ette ka seda, et ülekandesüsteemi haldur peab tasakaalustustoiminguid sooritades omama ka ise prognoosi tasakaalustatava bilansiperioodi kohta. Turuosalistele on gaasiturul ette nähtud ka stiimulid, mis seisnevad gaasi bilansituru toimimises. [1]

Käesolevas lõputöös leitakse vastused järgnevatele küsimustele:

- Millised gaasitarbimise mõjurid on kõige olulisemad prognoosimisel?
- Millised on täpsed arvulised parameetrid, millega eeltoodud mõjureid gaasitarbimise prognoosimisel arvesse võtta?
- Milline on kõige täpsem prognoosimudel ning selle parameetrid võttes arvesse eelnevalt kindlaks tehtud mõjureid?

Töö eesmärgiks on välja selgitada erinevate ilmategurite täpne numbriline mõju gaasitarbimisele ning seeläbi saavutada võimalikult täpne väljund gaasitarbimise prognoosimisel lineaarsel regressioonil põhineva prognoosimudeliga rakendustarkvaras R, et selgitada välja kõige olulisemad mõjurid gaasitarbimisele ning võttes vastavad parameetrid ka prognoosimudelite arvutusprotsessis kasutusele. Mõjuritele arvutatud parameetritele antakse hinnang pikaajalise keskmise prognoosivea põhjal, mida arvutatakse MAPE põhjal.

Lõputöö esimeses peatükis antakse ülevaade Eesti gaasisüsteemist ja selle ajaloost, et selgitada käesoleva lõputöö tausta ja konteksti, milles antud uurimus läbi on viidud. Ühtlasi aitab esile tõsta antud lõputöö olulisust bilansivastutuse seisukohast.

Põhiosa teises peatükis antakse ülevaade käesoleva lõputöö teoreetilisest poolest. Põhiliselt keskendutakse gaasitarbimise analüüsimise käigus regressioonanalüüsile ning mitme muutujaga

regressioonimudeli koostamisele, et anda hinnang eelnenud analüüsi käigus leitud ilmastikutegurite numbrilisele mõjule Eesti gaasitarbimisele. Regressioonanalüüs on aluseks ka käesolevas lõputöös gaasitarbimise prognoosimiseks kasutatavale regressioonimudelile. Kasutades ilmastikutegurite numbriliste parameetrite omistamiseks regressioonanalüüsi on võimalik kasutada leitud väärtusi ka otse regressioonimudelis.

Lõputöö kolmas peatükk keskendub ilmastikumõjurite üldisele ja statistilisele analüüsile ning põhiliseks analüüsimeetodiks on regressioonanalüüs, sest ilmastikuteguritele arvatud numbrilise mõju õigsuse hindamiseks kasutatakse lineaarset regressioonimudelit, mida on eelnenud analüüsi põhjal kõige lihtsam ja adekvaatsem kasutada. Esialgne üldine analüüs ning kogu andmestiku regressioonanalüüs sooritatakse kontoritarkvara Microsoft Excel abil, kuid modelleerimisel kasutatav libiseva treeningreaga regressioonanalüüs sooritatakse rakendustarkvara R abil.

Lõputöö neljas peatükk keskendub esimeses peatükis arvatud parameetrite põhjal gaasitarbimise modelleerimisele. Gaasitarbimise prognoosimisel kasutatakse regressioonimudelit libiseva treeningreaga, mis tähendab, et iga arvatud prognoosi jaoks võetakse eelnevalt defineeritud pikkusega andmestiku osa arvesse. Prognoosimudel on koostatud rakendustarkvaras R, mille baaspaketis sisalduvad juba kõik vajaminevad käsured regressioonanalüüsiks ning selle põhjal regressioonimudeli koostamiseks. Peale prognoosimudeli abil prognooside koostamist antakse tulemustele hinnang MAPE (*mean absolute percentage error*) abil hindamaks kasutatud parameetrite ja nende numbriliste väärtuste olulisust ja täpsust. Mudelite täpsust hinnatakse kontoritarkvara Microsoft Excel abil peale prognoosimudelite väljundite genereerimist prognoosimudeliga, mis on koostatud rakendustarkvaras R.

1. Eesti gaasisüsteem

Käesolevas peatükis antakse lühiülevaade Eesti gaasisüsteemist, mille tarbimise analüüsimisele antud lõputöö keskendub. Gaas on energeetikas, kodumajapidamistes ning tööstuses laialdaselt kasutatav kütus ning tooraine, mille eeliseks on kasutusmugavus, lihtne transporditavus ning laia kauplemist võimaldav maailmaturg. Eestis tarbitav gaas on pärit peamiselt Venemaalt ja Leedust Kaunase gaasiterminalist. Peamiselt kasutatakse Eestis maagaasi kaugkütteks ja tööstuslikel eesmärkidel Eesti gaasiturg on avatud alates 2007. aastast, kuid sinne gaasivõrk on ühendatud vaid Venemaa ja Läti gaasivõrkudega ning AS Eesti Gaas omab valitsevat positsiooni gaasi importimisel Eestisse. 2014. aastani puudus gaasitarnijate vaheline konkurents, kuid seda olukorda aitas parandada Leedus avatud veeldatud gaasi vastuvõtuterminal [2]. Gaasitarnijate vaheline tihedam konkurents muudab olulisemaks ka täpsema gaasitarbimise prognoosimise, sest see aitab vähendada jaeäride bilansikulusid ning sellest tulenevalt pakkuda tarbijatele ka soodsamaid müügmarginaale. Gaasisüsteemi haldur Eestis on Elering AS. Maagasiseadusega on sätestatud bilansivastutuse regulatsioon, mille kohaselt on iga turuosaline vastutav oma gaasibilansi eest ehk bilansihaldur peab tagama oma portfelli tarbimisele võimalikult täpse tarbimisprognoosi ja minimeerima bilansigaasi kogust. Süsteemihaldur vastutab kogu gaasisüsteemi bilansi eest ning gaasiturul tegutseb mitmeid bilansihaldureid. [3]

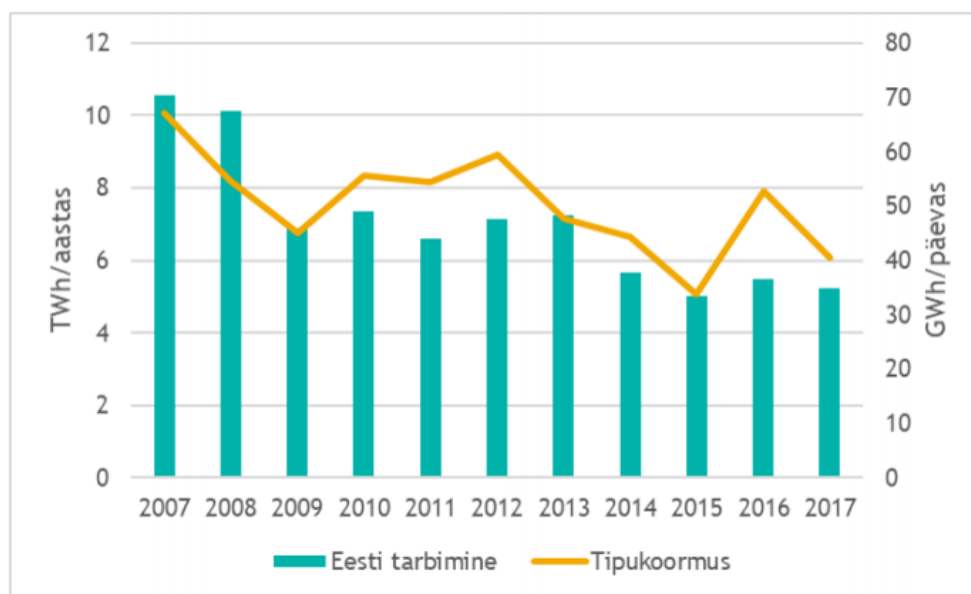
1.1 Eesti gaasisüsteemi ajalugu

Eestis on tarbitud gaasi alates 19. sajandist, kui 1865. aastal valmis esimene gaasivabrik Tallinnas, kus toodeti tehisgaasi Inglismaalt toodud kivisöest. Enamik gaasi tarbiti sellel ajal tänavavalgustuseks, Balti Manufaktuuris, vineerivabrikutes ning Kadrioru lossis ja pargis. Teine gaasivabrik asutati Eestis 1880. aastal Tartus ja 1948. aastal valmis Kohtla-Järve Põlevkivitöötlemise tehas, mille kaudu hakati gaasiga varustama Leningradi. Esimene kõrgsurve gaasiülekanal alustas tööd 1953. aastal Kohtla-Järve ja Tallinna vahel. Sellest tulenevalt kasvas gaasitarbimine 1955. aastaks 65 miljoni kuupmeetrini, mis teisendatuna vatt-tundidesse on umbes 680 gigavatt-tundi. Kuuekümnendate lõpus hakati Leningradist transportima aga maagaasi Eestisse, mis Jõhvis segati põlevkivigaasiga ning seitsmekümnendatel valmisid Tartu-Irboska ja Tartu-Rakvere gaasitorustikud ja seitsmekümnendate lõpuks tarbiti Eestis kokku juba 708 miljonit kuupmeetrit maagaasi ja 273 miljonit kuupmeetrit põlevkivigaasi [4] [5]. 2018. aastal importis ja tarbis Eesti kokku 5,2 teravatt-tundi maagaasi [6]. Tänapäeval koosneb Eesti maagaasi ülekandevõrk 885 km gaasitorustikust, kolmest gaasimõõtejaamast ning 36 gaasijaotusjaamast. Täna sel päeval puudub Eesti gaasisüsteemil oma kompressorjaam gaasitorustikus rõhuerinevuse tekitamiseks ning sellest

tulenevalt tagatakse gaasi ülekandeks vajalik rõhk Venemaa ülekande süsteemis asuvate kompressorjaamadega või Lätis asuvas maa-aluses gaasihoidlas [5].

1.2 Eesti gaasisüsteemi tulevikutrendid

Viimastel aastatel on maagaasi tarbimine liikunud Eestis pigem langevas trendis. Erandiks on 2016. aasta, kui külma talve tõttu gaasitarbimine kasvas ning 2017. aastal toimus gaasitarbimises jällegi 5,1-protsendiline langus. Eesti maagaasi aastast tarbimist ning tipukoormust antud aastal on kujutatud joonisel 1.1. [5]



Joonis 1.1. Eesti maagaasi aastane tarbimine (teravatt-tundi aastas) ja tipukoormus (gigavatt-tundi päevas) aastatel 2007-2017. [5]

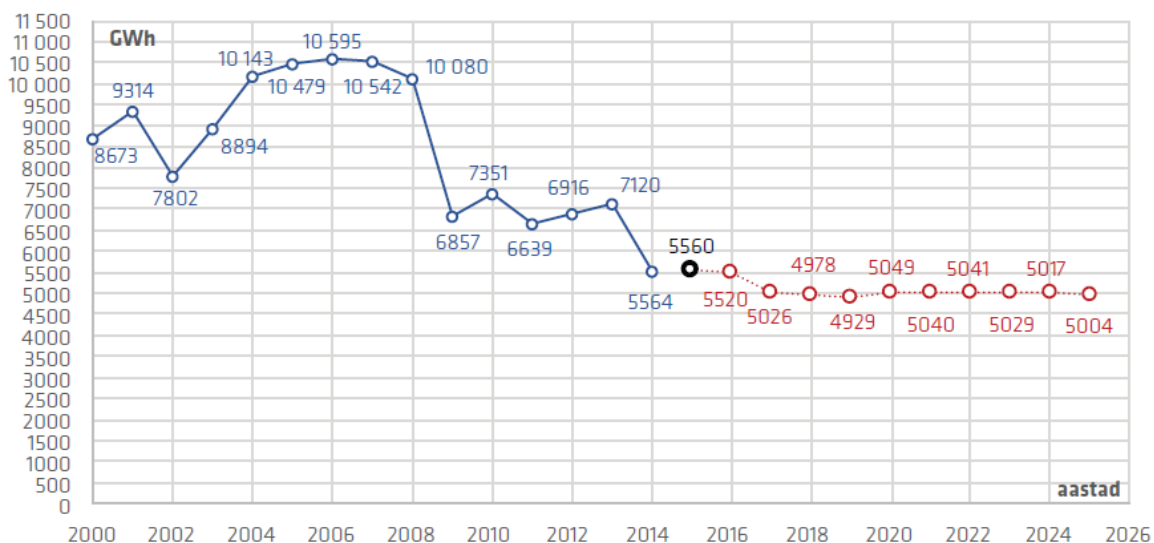
Elering AS-i poolt väljastatud gaasituru käsiraamatus viidatakse tarbimise langustrendi põhjendades sellele, et Eesti maagaasi tarbimisstruktuur on muutumas, sest tööstustarbijad ning elektri- ja soojatootjad on loobunud gaasi kasutamisest kütusena. Lisaks veel on välja toodud energiatarbimise tõhusamaks muutumine ning ka üldine gaasi halb maine tarbijate seas niinimetatud poliitilise kütusena. Gaasi kasutatakse Eestis 2016. aasta andmetele tuginedes järgnevates sektorites vastavate osakaaludega [5]:

- Energiasektor – 38%;
- Tööstus -28%;
- Äri- ja avaliku teeninduse sektor – 19%;
- Kodumajapidamised – 14%;

- Põllumajandus – 2%;
- Transpordisektor – 1%

Suurima muutuse on võrreldes 2008. aastaga läbi teinud tööstussektor, mille osakaal gaasitarbimises on kukkunud kümme protsenti ja energiasektor, mille osakaal on kukkunud viis protsenti. Samas tõusis maagaasi tarbimise osakaal kodumajapidamistes selle ajaga kaheksa protsenti. [5]

Eesti maagaasi tarbimise pikaajaline prognoos sõltub mitmetest mõjuritest nagu näiteks energiapoliitika, majanduskasv, ehitiste energiaefektiivsus ja muudest sarnastest tegurites. Hetkel on gaasitarbimine Eestis pigem langustrendis ning gaasi kasutamine sooja- ja elektritootmises ning tööstussektoris. Transpordisektoris aga eeldatakse gaasi tarbimises kasvavat trendi tulenevalt Euroopa Liidu eesmärgist katta kümme protsenti taastuvenergiaga ning seal nähakse suurt potentsiaali biometaani kasutamisel. Gaasi kasutamine soojuse tootmiseks on hakanud langema, sest kasutusele on hakatud võtma energiasäästumeetmeid kaugküttepiirkondades ning samuti on järjest enam hakatud kasutama ka kohalike kütuseid nagu näiteks puiduhake ja turvas. Joonisel 1.2 on toodud Elering AS-i pikaajaline gaasitarbimise prognoos aastani 2026. [7]



Joonis 1.2. Eesti gaasitarbimise statistika ja pikaajaline gaasitarbimise prognoos aastani 2026. [7]

Oluline eeldus Eesti gaasitarbimise pikaajalisel prognoosimisel on ka see, et Nitrofert AS-i väetisetehas ei käivita uuesti oma väetisetehast, mis peatati 2009. aastal [8]. Nitroferti tehase sulgemine on selgelt nähtav ka jooniselt 1.2. Viimane Nitroferti tehase peatati 2013. aastal, mida on samuti näha jooniselt 1.2, kui Eesti gaasitarbimises esines järjekordne järsk langus. [9]

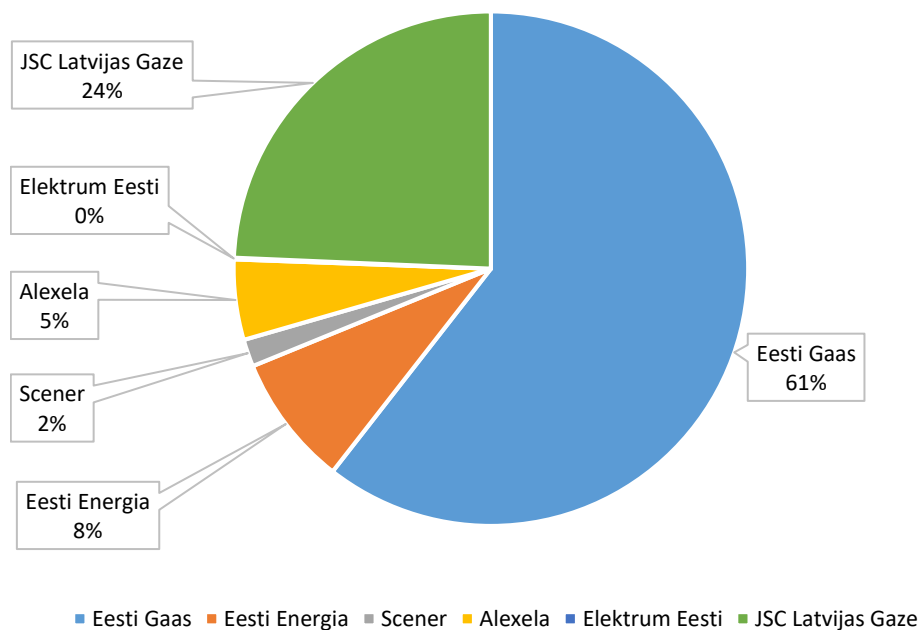
1.3 Gaasiturg Eestis

Eestis, Lätis, Leedus ja Soomes ei toimu maagaasi tootmist, mis tähendab seda, et kogu vajamineva gaasi peab siinsetele turgudele importima Venemaalt ja LNG maailmaturult. Venemaalt imporditakse siinsetele turgudele gaasi läbi kõrgsurve torustike ning LNG maailmaturult imporditav gaas transporditakse veeldatud kujul Klaipeda LNG terminali, mis on ühendatud Baltikumi regiooni maagaasivõrkudega. Imporditavat gaasi võib asendada ka kohalik biometaan, kuid tänasel päeval biometaan Eesti ülekandevõrku ei sisestata, aga selle potentsiaali maagaasi asendamisel hinnatakse kõrgeks. Oluline on siinkohal ka ehitamisjärgus olev Balticconnector, mis ühendab Eesti ja Soome gaasisüsteemid ning annab võimaluse importida Eestisse gaasi teistest allikatest läbi LNG terminali. [5]

Gaasisüsteemi haldur Eestis on Elering AS. Maagasiseadusega on sätestatud bilansivastutuse regulatsioon, mille kohaselt on iga turuosaline vastutav oma gaasibilansi eest. Süsteemihaldur vastutab kogu gaasisüsteemi bilansi eest ning gaasiturul tegutseb mitmeid bilansihaldureid, kelle gaasibilansside selgitamise eest vastutab Elering. Käesoleval ajal on Eesti gaasi jaeturul kuus bilansihaldurit [3] [10]:

- Alexela Energia AS;
- JSC Latvijas Gaze;
- Scener OÜ;
- Eesti Gaas AS;
- Eesti Energia AS;
- Elektrum Eesti OÜ.

2019. aasta märtsikuu seisuga oli Eesti Gaasi turuosa 60,0 protsenti, JSC Latvijas Gaze turuosa 24,1 protsenti, Eesti Energia turuosa 8,2 protsenti, Alexela Energia turuosa 5,9 protsenti, Sceneri turuosa 1,7 protsenti ja Elektrim Eesti turuosa 0,1 protsenti. Bilansihaldurite turuosad 2019. aasta märtsikuu seisuga on graafiliselt kujutatud ka joonisel 1.3. [10]



Joonis 1.3. Eesti gaasi bilansihaldurite jaeturu turuosad 2019. aasta märtsikuu seisuga. [10]

Eesti gaasiturul korraldus sai alguse 1998. aastal koos energiaseaduse jõustumisega, mis liigitas vabatarbijateks kõik gaasitarbijad peale kodutarbijate, mis tähendas, et kõik tarbijad peale kodutarbijate võisid hankida gaasi ükskõik millisel müüjalt. 2003. aastal liigitasid vabatarbijateks kõik tarbijad, kelle aastane gaasitarbimine oli üle 200 tuhande kuupmeetri ning 2007. aastast avanes gaasiturg täielikult. Paraku oli sellel ajal Eesti hulgiturul vaid üks gaasi importija ja monopoolses seisundis müüja, kelleks oli AS Eesti Gaas, kes importis gaasi Venemaalt pikaajalise *take-or-pay* lepingu alusel. Aktsiaseltsile Eesti Gaas kuulusid ka gaasisüsteemihaldur EG Võrguteenus ning suurim gaasi jaotusvõrguettevõtte Eestis AS Gaasivõrgud. Turu tegelik avanemine algas alles 2012. aastal, kui muudeti maagaasiseadust nii, et bilansihaldur ei saanud samal ajal paralleelselt olla ka ülekandesüsteemi omanik, mis päädis sellega, et tegelik konkurents tekkis Eesti gaasiturul alles 2015. aastal, kui toimus süsteemihalduri täielik eraldumine monopoolses seisundis olevast gaasi müüjast ja importijast, mis võimaldas kõigi turuosaliste võrdse ligipääsu Eesti gaasisüsteemile ja –turule. Teine oluline tegur gaasiturul konkurentsi tekkimisel oli Klaipedas 2014. aastal avatud LNG terminaal, mis andis Baltikumile mõnel määral ligipääsu ka LNG maailmaturule. [5]

Iga gaasiturul osalev bilansihaldur on vastutav oma tarbimisportfelli bilansi eest. Bilansihaldurid peavad süsteemihaldurile esitama viis päeva enne arvestuskuu algust järgneva kuu bilansiplaan ning süsteemihaldur võtab selle informatsiooni arvesse tulevikus gaasivoogude planeerimiseks. Kui bilansihaldur soovib oma koguseid bilansiplaanis muuta, siis on seda võimalus teha ka esitades

järgmise päeva koguseid hiljemalt tarnepäevale eelneval päeval kell 17.00. Süsteemihaldur võrdleb iga piiriülese sisend- ja väljundpunkti kohta esitatud määratud tarneid teise süsteemihalduri gaasisüsteemist üle kantavate määratud tarnetega ning virtuaalses kauplemispunktis esitatud määratud tarneid vastaspoole määratud tarnetega. Virtuaalse kauplemispunkti all mõeldakse süsteemihalduri hallatavat füüsiliselt määratlemata kauplemispunkti, kus bilansihaldurid saavad teha bilansiportfellide vahel tehinguid. Juhul kui bilansihaldurite poolt esitatud kogused omavahel ei võrdu, rakendatakse gaasikogustele vähima koguse printsiipi, mis sätestab, et süsteemihaldur võrdleb bilansihalduri poolt esitatud määratud tarneid vastaspoole vastavate määratud tarnetega ning vähendab erinevuse korral suuremat kogust tarnete võrdsustamiseks. Süsteemihaldur müüb või ostab bilansihaldurilt bilansilepingu alusel igal kauplemisperioodil bilansihalduri bilansi tagamiseks vajaliku koguse bilansigaasi. [11]

Bilansigaasi lõpliku hinna määrab süsteemihaldur ning selle arvutamisel lähtutakse selles, et katta bilansigaasi ostuks tehtud põhjendatud kulutused ning bilansigaasi ostu ja müügiga seotud muud põhjendatud kulutused ja tagada ka põhjendatud tulukus. Kui bilansihalduri bilansi selgitamise tulemusena on tema bilans negatiivne ehk tarbimisportfell tarbis oodatust enam gaasi, siis loetakse, et süsteemihaldur on sellel perioodil avatud tarnena müünud bilansihaldurile bilansigaasi koguses, mis on vajalik bilansihalduri bilansi hoidmiseks. Kui bilansihalduri bilansi selgitamise tulemusena on tema bilans positiivne ehk tarbimisportfell tarbis oodatust vähem gaasi, siis loetakse, et süsteemihaldur on sellel perioodil avatud tarnena ostnud bilansihaldurilt bilansigaasi. Üldjuhul on bilansigaasi hind selline, et bilansigaasi peab ostma kallima hinnaga kui päev ette turult gaasi on ostetud ning bilansigaasi peab müüma odavama hinnaga kui päev ette turult gaasi on ostetud. Sellest tulenevalt tekivad ebatäpsetest prognoosidest ka bilansikulud bilansihalduritele. [12]

2. Lineaarne regressioon

Käesoleva lõputöö esimene peatükk keskendub lineaarse regressiooni teooriale. Regressioonimudelid on ühed lihtsamad ja täpsemad moodused energiatarbimise analüüsimiseks ja prognoosimiseks piiratud andmestike puhul [11]. Regressiooni terminit tutvustas esmakordselt briti Francis Galton aastal 1908. Regressioon kirjeldab statistiliselt erinevate muutujate omavahelist sõltuvust.

Käesolevas lõputöös on kasutatud regressioonanalüüsi meetodit eelkõige selle praktilise kasutamise lihtsuse tõttu ning selline meetod aitab lihtsasti leida korrelatsiooni energiatarbimise ning selle põhiliste mõjurite vahel. Vaadeldes gaasitarbimise ja temperatuuri vahelist sõltuvust saab ka täheldada, et gaasitarbimine on võrdlemisi selgelt tükiti lineaarses sõltuvuses temperatuuris ning tulenevalt oma sesoonsusest ja kindlast korduvast profiilist võib eeldada, et kasutatava prognoosimudeli viga allub normaaljaotusele. Mõlema aspekti puhul on tegemist lineaarse regressiooni kasutamiseks oluliste eeldustega. Lisaks lihtsalt implementeerimisele on regressioonimudel võimeline läbi analüüsima suure hulga andmeid võrdlemisi lühikese aja jooksul, mis teeb mudeli kasutamise praktikas lihtsaks ja kasutajasõbralikuks. Regressioonanalüüs võimaldab ühtse meetodika raames hinnata ilmategurite mõju gaasitarbimisele ning seejärel kasutada leitud tulemusi gaasitarbimise prognoosiks. Lisaks annab selline statistiline lähenemine võimaluse hinnata ilmategurite mõju gaasitarbimisele numbriliselt ja objektiivselt. Samas aga uurib regressioonanalüüs erinevate muutujate statistilist suhet läbi ruutvea minimeerimise, mis tähendab, et iga sõltumatu muutuja regressioonfunktsioonis omandab sellise väärtuse, et regressioonfunktsioon omaks kõikide võimalike andmepunktide puhul vähimat summaarset viga, mis tähendab, et käesolev meetodika uurib muutujate statistilist seost mitte põhjuselist seost ning sellest tulenevalt võib tekkida analüüsimise ja modelleerimise käigus suuremaid või väiksemaid vigu. [12]

Regressioonanalüüs ja –mudelid on aegridade analüüsimiseks ja prognoosimiseks kasutatavad klassikalised statistilised meetodid. Lisaks kasutatakse aegridade prognoosimiseks ka erinevaid masinõppel ja närvivõrkudel põhinevaid mudeleid. Regressiooni puhul on tegemist kõige lihtsamini implementeeritava analüüsimeetodiga aegridade analüüsimiseks, uurimiseks ja prognoosimiseks. Regressioonanalüüs annab käesolevas lõputöös võimaluse hinnata erinevate ilmaparameetrite mõju Eesti gaasitarbimisele numbriliselt ning seejärel kasutada neid tulemusi ka regressioonmudelis gaasitarbimise prognoosimiseks. Kuigi on olemas ka erinevaid masinõppel ja närvivõrkudel põhinevaid mudeleid ja meetodikaid aegridade prognoosimiseks, siis käesolevas lõputöös alusandmeteks olevad ilma- ja tarbimisandmed ei oma piisavalt suurt mahtu ega

mitmekesisust, et selliste keerukate mudelite kasutamine õigustatud oleks. Ühtlasi on nimetatud meetodid väga ressursikulukad ning iga iteratsiooni läbimiseks võib selliste mudelite puhul kuluda ka mitmeid tunde.

2.1 Lineaarse regressiooni kasutusnäited

Regressioonanalüüs annab võimaluse uurida ja prognoosida erinevaid suuruseid ning nendele mõju avaldavate muutujate suhet uuritava suurusega, milleks antud lõputöö raames on gaasitarbimise sõltuvus erinevatest ilmastikuteguritest ja muutujatest. Antud meetodit on võimalik väga laialdaselt kasutada erinevates sektorites ja tegevusaladel, kus on vajalik uurida mõne sõltuva muutuja suhet erinevatesse sõltumatutesse muutujatesse. Regressioonanalüüs aitab sellistel juhtudel hinnata esiteks erinevate sõltumatute muutujate olulisust sõltuva muutuja suhtes ning seejärel võimaldab koostada sõltuva muutuja prognoosi teades tulevikus ilmnevat või eeldatavat sõltumatuid muutujaid.

Regressioonanalüüs on energiatarbimise uurimises leidnud ka laialdast kasutust üle maailma. Sloveenias on 2018. aastal läbi viidud uuring erinevatest meetodikatest linnaalade gaasitarbimise prognoosimisel, kus võrreldi lineaarsel regressioonil põhinevat mudelit, kerneli masinõppe mudelit ja erinevaid närvivõrkude mudeleid. Hoolimata oma lihtsusest oli lineaarsel regressioonil põhinev statistiline mudel kõige parema tulemuse tunnipõhises prognoosis 24-tunnise ajahorisondiga. Märkimist väärib ka lineaarsete mudelite vähene ressursikasutus, mis seisneb selles, et lineaarset mudelit kasutades genereerib mudel väljundi umbes kümne sekundi jooksul. Keerulisemate närvivõrkude mudelite puhul võib mudel prognoosi genereerida kuni üheksa tundi, mis näitab, et regressioonmudelite puhul on tegemist ka väga operatiivsete mudelitega. Antud uurimuses saavutati lineaarsel regressioonil põhineva mudeliga prognoosi koostamisel MAPE 8,1 protsenti, aga keerulisemate mudelite puhul oli see 9,3 kuni 42,4 protsenti, mis annab ka käesolevale lõputööle adekvaatse võrdlusaluse. [13]

Rumeenias on läbi viidud uuring kuidas kõige kiiremal moel prognoosida ehitiste energiatarbimist tulevikus, et hinnata nende mõju keskkonnale ning ka sellisel juhul on eelistatud kasutada regressioonil põhinevat mudelit oma kiiruse ja suhteliselt hea täpsuse tõttu. [14] Sarnast mitme muutujaga regressioonanalüüsi kasutatud energiatarbimise prognoosimiseks Hiinas erinevatel aastaagadel. Mudelit on kasutatud seetõttu, et selles saab mitmeid erinevaid mõjureid arvesse võtta hoolimata nende ühikulistest väärtustest ning regressioonanalüüs ning –mudel võimaldab kasutada laia valikut sõltumatuid parameetreid. Lisaks energiatarbimise prognoosimisele võimaldas regressioonanalüüs uurida ka ehitiste erinevate ehituslike parameetrite mõju nende energiakulule. Näiteks võeti analüüsis arvesse ka erinevate ehitismaterjalide soojusisolatsiooni, mis iseloomustab

seda, et kuivõrd erinevaid parameetreid saab antud mudelis kasutada. Lisaks sellele saab regressioonanalüüsi abil prognoosida erinevaid ajahorisonte ning sõltuvaid muutujaid, kui on täidetud eeldused regressioonanalüüsi läbiviimiseks. [15]

2.2 Regressioonanalüüs

Selgitamaks välja gaasitarbimise seoseid ilmastikuoludega viidi käesoleva lõputöö raames läbi regressioonanalüüs erinevate ilmastikuandmete ja Eesti gaasisüsteemi tarbimisandmete vahel. Regressioonanalüüs võimaldab koostada mudeli hindamiseks tunnuste vahelisi seoseid vaadeldes üht tunnust kui sõltuvat ning ülejäänud tunnuseid, mille põhjal oleks võimalik kirjeldada ja prognoosida sõltuva tunnuse väärtusi. Mida tugevam on seos sõltuva ja sõltumatute tunnuste vahel, seda olulisem on kasutada neid sõltumatuid tunnused sõltuva tunnuse prognoosimisel ning seda täpsem on ka prognoosimudel [16]. Käesolevas lõputöös on sõltuvaks tunnuseks gaasitarbimine ning analüüsitakse gaasitarbimise sõltuvust ilmastikuoludest.

Lineaarsel regressioonanalüüsil on käesoleva lõputöö raames kaks olulisemat tüüpi, mis käsitlemisele kuuluvad. Esimest tüüpi regressioonanalüüs on lihtne regressioonanalüüs ühe sõltumatu muutujaga ning teist tüüpi regressioonanalüüs on mitme muutujaga regressioonanalüüs, kus sõltumatuid muutujaid on enam kui üks. Kõigepealt kirjeldatakse lihtsat lineaarset regressiooni ning seejärel antakse ülevaade mitme muutujaga lineaarsest regressioonist.

2.2.1 Lihtne lineaarne regressioon

Käesolevas peatüki punktis antakse ülevaade lihtsa lineaarse regressiooni teooriast ning peatüki järgmises alapunktis antakse ülevaade mitme muutujaga lineaarsest regressioonist.

Eeldused regressioonanalüüsi läbi viimiseks on järgnevad [16]:

- Analüüsitakse arvtunnuseid või binaarseid tunnuseid
- Sõltuv ja sõltumatud tunnused omavad lineaarset seost
- Sõltumatud tunnused on võimalikult tugevalt seotud sõltuva tunnusega, kuid võimalikult vähe seotud omavahel

Regressioonanalüüsi eesmärk on koostada regressioonimudel iseloomustamiseks sõltuva ja sõltumatute tunnuste vahelist suhet, mis matemaatilisel kujul avaldatuna on järgnev [17]:

$$y = f(x_1, x_2, \dots, x_p) + \varepsilon \quad (2.1)$$

Kus

y – sõltuv tunnus;

$f(x_1, x_2, \dots, x_p)$ – regressioonfunktsioon;

ε – juhuslik viga

Regressioonmudel põhineb regressioonfunktsioonil, mille üldkuju on järgnev [16]:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2.2)$$

Kus

y – sõltuv tunnus;

x_n – sõltumatud tunnused;

b_0 – funktsiooni vabaliige;

b_n – regressioonkordaja

Lihtsas lineaarses regressioonis on vaid vabaliige ja üks sõltumatu tunnus ning selle regressioonkordaja.

Analüüsi eesmärgiks on kirjeldada sõltuva muutuja suuruse seost sõltumatute muutujatega ehk regressioonvõrrand ning selle põhjal prognoosida sõltuva tunnuse y väärtuseid erinevate sõltumatute muutujate puhul. Regressioonanalüüsi põhjal saab teha kindlaks ka selle, et millised sõltumatud tunnused on sõltuva tunnuse kujunemisel ja prognoosimisel kõige olulisemad või olulisemad kui teised tunnused.

Sobivate regressioonkordajate leidmiseks kasutatakse vähimruutude meetodit, mis seisneb selles, et tunnuste vahelist seost iseloomustavat andmekogu valitakse kujutlema selline sirge, millest iga andmepunti kauguste ruutude summa on kõige väiksem [18]. Vähimruutude meetodi põhimõte seisneb selles, et leida sellised hinnangud funktsiooni vabaliikmele b_0 ja regressioonkordajale b_n , et summeeritud ruutkaugus tegeliku sõltuva tunnuste y suuruste ja arvutuslike väärtuste $y = b_0 + b_1x_1$ vahel oleks minimaalne üle kogu andmete valimi kõikvõimalike erinevate regressioonkordajate b_0 ja b_1 korral. Matemaatiline valem vähimruudu leidmiseks regressioonanalüüsis on järgnev [17]:

$$(b_0, b_1) = \arg \min_{(b_0, b_1)} \sum_{i=1}^n [y_i - (b_0 + b_1 x_1)]^2 \quad (2.3)$$

Kus

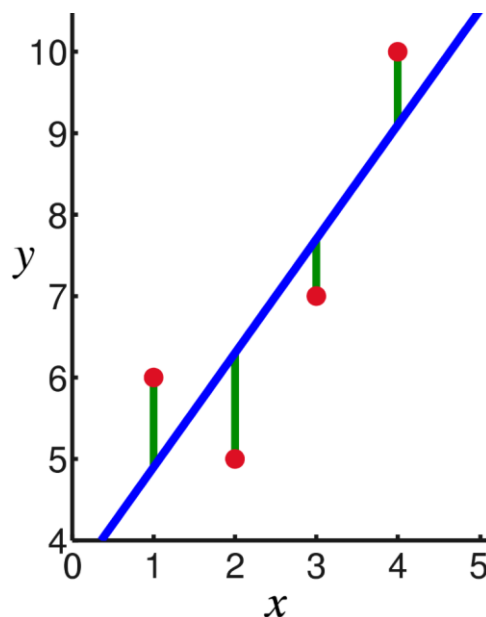
b_0 – regressioonfunktsiooni vabaliige

b_1 – sõltumatu tunnuse regressioonkordaja

x_1 – sõltumatu tunnuse väärtus

y_i – sõltuva tunnuse väärtus

Vähimruutude meetodi eesmärk on leida regressioonkordajate suurused selliselt, et tekkinud regressioonfunktsioon oleks kõige lähemal kõikidele andmepunktidele. Joonisel 2.1 on illustreeritud vähimruutude meetodika ideed ning sinise joonega on märgitud regressioonsirge ja punased on valimis olevad andmepunktid, mille põhjal on koostatud regressioonsirge. Rohelised jooned regressioonsirge ja andmepunktide vahel tähistavad regressioonsirge ja andmepunktide vahelist kaugust, mille ruutude summat minimeeritakse vähimruutude meetodiga. [19]



Joonis 2.1. Vähimruutude summat kirjeldav joonis. [20]

2.2.2 Mitme muutujaga lineaarne regressioon

Käesolevas punktis antakse ülevaade mitme muutujaga lineaarsest regressioonist. Praktilises elus on tihti regressioonanalüüsis käsitletavatel sõltuvatel suurustel mitmeid mõjureid ning seetõttu ei pruugi lihtne lineaarne regressioon anda parimaid tulemusi. Mitme muutujaga lineaarne

regressioonanalüüs uurib ühe sõltuva suuruse ning mitme sõltumatu suuruse lineaarset seost, mis mitmekesistab oluliselt sooritatavate analüüside andmestike ning võimaldab leida seeläbi teoreetiliselt täpsemaid tulemusi. Mõistmaks mitme muutujaga lineaarse regressiooni teoreetilisi aluseid, tuleb vaadelda ka lineaarse ruumi üldisemat olemust ning selles andmestike kuvamise ja analüüsimise võimalusi.

Esmalt antakse lühiülevaade vektorruumist, -alamruumist, kujutamisest ja ruutkeskmisest vormist mitme muutuja kontekstis. Vektorid on geomeetrilised objektid, mis omavad suurust ja suunda alguspunkti A ja lõpp-punkti B vahel märgistusega \overrightarrow{AB} ning need aitavad selgitada punkti B suhtelist asetumist punkti A suhtes. Vektorid alluvad tavalistele algebra seadustele nagu kommutatiivsus, assotsiatiivsus ja jaotatavus. Vektoritega saab sooritada liitmis-, lahutamise-, jagamis- ja korrutustehteid numbritega muutes seeläbi vektorite suunda ja pikkust. Vektorruum on vektorite kogum, mis on piiratud lõpliku arvu vektorite lisamise ja skalaarkorrutistega. Lihtsaim näide Eukleidilisest ruumist, kus iga element on kujutatud reaalarvude jadana on järgnev [17]:

$$x' = (x_1, x_2, \dots, x_n) \quad (2.4)$$

Kus

x' - tähistab vaadeldavat ruumi

$x_1 \dots x_n$ – tähistavad elemente sellest ruumis

F. Hausdorff oli esimene inimene, kes tõestas, et igal vektorruumil on baas, mis muudab algebras võimalikuks lineaarselt kujutada igat vektorruumis olevat vektorit. Näide vektorruumi baasist \mathbb{R}^n on järgnev [17]:

$$\begin{aligned} e'_1 &= (1, 0, 0, \dots, 0)_{1 \times n} \\ e'_2 &= (0, 1, 0, \dots, 0)_{1 \times n} \\ e'_3 &= (0, 0, 1, \dots, 0)_{1 \times n} \\ &\vdots \\ e'_n &= (0, 0, 0, \dots, 1)_{1 \times n} \end{aligned} \quad (2.5)$$

Eltoodud vektorid koosnevad standardsest ortogonaalsest vektorruumi \mathbb{R}^n baasist. Ükskõik milline vektor x' võib olla lineaarne kombinatsioon vektoritest $e'_1, e'_2, e'_3, \dots, e'_n$ järgneval kujul [17]:

$$x = x_1 e_1 + x_2 e_2 + x_3 e_3 + \dots + x_n e_n \quad (2.6)$$

Võttes vektorruumi V ning mittetühja alamruumi W , mis on suletud läbi liitmise ja skalaarkorrutiste, kutsutakse alamruumi W vektorruumi V alamruumiks. Alamruumide, mis sisaldavad vastavaid vektoreid, ristumiskohti kutsutakse selle alamruumi ulatuseks [17].

Kui vektorruumis V esineb lõplik hulk vektoreid v_1, v_2, \dots, v_n ning nullist erinevad skalaarid a_1, a_2, \dots, a_n sellisel kujul, et:

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n = 0 \quad (2.7)$$

Siis öeldakse, et vektorid v_1, v_2, \dots, v_n on lineaarses sõltuvuses ning nende seas on vähemalt üks vektor, mida saab avaldada ülejäänud vektorite kombinatsiooniga. [17]

Kui on kaks vektorit $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ja $\mathbf{y} = (y_1, y_2, \dots, y_n)$ Eukleidilises ruumis \mathbb{R}^n , siis nende kahe vektori korrutist avaldatakse järgnevalt:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n \quad (2.8)$$

Kui nende kahe vektori korrutis ei ole null, siis öeldakse, et need vektorid on ortogonaalsed ning asetsevad omavahel risti. Kui θ on nurk nende kahe vektori vahel, siis kahe vektori vahelise nurga koosinus avaldub järgnevalt [17]:

$$\cos(\theta) \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}||\mathbf{y}|} = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \sqrt{y_1^2 + y_2^2 + \dots + y_n^2}} \quad (2.9)$$

Sellist korrutist saab kasutada ka projektsiooniks. Vektori \mathbf{y} projekteerimine vektorile \mathbf{x} moodustab uue vektori, mille suund on sama, mis vektoril \mathbf{x} ning vektori pikkus avaldub kujul $|\mathbf{y}| \cos(\theta)$, milles $|\mathbf{y}|$ tähistab vektori \mathbf{y} pikkust ning θ tähistab nende kahe vektori vahelist nurka. Selline projektsioon $P_{\mathbf{x}} \mathbf{y}$ avaldub järgmiselt:

$$P_{\mathbf{x}} \mathbf{y} = |\mathbf{y}| \cos(\theta) \frac{\mathbf{x}}{|\mathbf{x}|} = |\mathbf{y}| \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}||\mathbf{y}|} \frac{\mathbf{x}}{|\mathbf{x}|} = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{x_1^2 + x_2^2 + \dots + x_n^2} \mathbf{x} = \lambda \mathbf{x} \quad (2.10)$$

Kus λ on skalaar ja avaldub järgnevalt:

$$\lambda = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{x_1^2 + x_2^2 + \dots + x_n^2} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{x} \cdot \mathbf{x}} \quad (2.11)$$

Sellest tulenevalt saab järeldada, et vektori \mathbf{y} projektsioon vektoril \mathbf{x} on vektori \mathbf{x} korrutamine skalaariga λ , mis tähistab vektorite vahelise nurga koosinust $\cos(\theta)$. Kui vektorid \mathbf{x} ja \mathbf{y} asetsevad

ruumis \mathbb{R}^n , siis saab uurida nende vahelist vahevektorit e . Vektor e on risti vektori x suhtes juhul kui:

$$\lambda = \frac{x \cdot y}{xx} \quad (2.12)$$

See tähendab, et vektor e on kõige lühem vektori, mis algab vektori y lõpp-punktist ning lõppeb ükskõik millisel punktil vektoril x . Kui vektorruum S on p -dimensionaalne vektorruumi V alamruum, siis on võimalik projekteerida vektoreid vektorruumis V alamruumi S . Kui alamruumil S on ortonormiline baas (w_1, w_2, \dots, w_n) , siis ükskõik milline vektor y ruumis V on võimalik projekteerida alamruumi S järgnevalt:

$$P_S y = \sum_{i=1}^p (y \cdot w_i) w_i \quad (2.13)$$

Vektori projektsioon lineaarses ruumis S on tegelikult vektori lineaarne teisendus, mida saab avaldada ka maatrikskorrutisena vektorist. Projektsioonmaatriks P on $n \times n$ suurusega ruutmaatriks, mis annab projektsiooni ruumist \mathbb{R}^n alamruumi S . Projektsioonmaatriksi tulpades on baasvektorite projektsioonid. Projektsioonmaatrikseid saab kasutada parima lineaarse sõltumatu hinnangu saamiseks regressioonimudelil. [17]

2.3 Regressioonimudel

Käesolevas lõputöös kasutatakse mitme sõltumatu muutujaga regressioonimudelit, mille eesmärgiks on uurida ühe sõltuva ja mitme sõltumatu tunnuse vahelist sõltuvust. Mitme muutujaga regressioon võimaldab korraga uurida mitme sõltumatu mõju ühele sõltuvale tunnusele nagu on ka käesoleva lõputöö eesmärk – uurida erinevate ilmastikutegurite mõju gaasitarbimisele. Regressioon aitab määrata erinevatele sõltumatutele muutujatele, antud juhul siis ilmastikuteguritele, kindlad numbrilised väärtused ning anda hinnang iga sõltumatu muutuja olulisusele käsitletavas regressioonimudelil. [21]

Kui uurimise all on mõne sõltuva muutuja y suhe mitme erineva sõltumatu muutujaga $x_1, x_2 \dots x_n$, siis regressioonfunktsioonil põhineva regressioonimudeli üldkuju on järgnev [17]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2.14)$$

Kus

Y – sõltuv muutuja või prognoositav väärtus;

β_0 – regressioonimudeli vabaliige;

β_n - sõltumatu muutuja regressioonkordaja;

x_n - sõltumatu muutuja väärtus;

ε - regressioonimudeli juhuslik viga.

Sõltumatud muutujad hoomavad endas regressoreid ja nende parameetreid ning klassikalises regressioonimudeli käsitluses eeldatakse, et juhuslik viga allub normaaljaotusele, mille keskmine suurus võrdub nulliga. Lineaarset mudelit nimetatakse lineaarseks siis, kui selle mudeli parameetrid on defineeritavad lineaarsetena ehk tema regressioonkordajad kirjeldavad sõltumatute muutujate suhet lineaarsena. [17]

Mitme muutujaga regressioonimudeli parameetrite olulisuse ja täpsuse hindamiseks on mitu võimalust. Ühed olulisemad ja enim kasutatud näidikud mudeli headuse hindamiseks on parameetrite P-väärtused, T-statistikud ning R-ruut.

P-väärtus määrab statistilistes mudelites igale parameetrile olulisuse selle mõjust sõltuval muutujale. P-väärtuse abil lükatakse ümber või kinnitatakse nullhüpootees parameetri olulisusest ning sellest tulenevalt on võimalik otsustada parameetrite olulisuse üle mudelis. Selle abil saab kindlaks teha, et kas andmete valimi keskmine erinevus ja varieeruvus kahe andmegrupi vahel on suurem või võrdne kui tegelikes mõõdetud andmetes [22]. P väärtus võimaldab parameetrite olulisusele käsitletavas valimis anda numbrilise mõõtme hindamaks nende olulisust. Mida madalam on P väärtus, seda olulisem ja suurem on vaadeldava sõltumatu muutuja mõju sõltuva muutuja suhtes. P väärtusel on mitu lävendit, mille järgi otsustatakse, et kas vaadeldav muutuja on oluline käsitletavas kontekstis või mitte. Põhilised praktikas kasutatavad lävendid on 0,05, 0,01, 0,005 või 0,001. Kui P väärtus on madalam kasutatavast lävendist, siis saab öelda, et vaadeldav parameeter on statistiliselt oluline ning kui P väärtus ületab lävendi, siis saab öelda, et vaadeldav parameeter ei ole statistiliselt oluline mõjur vaadeldava sõltumatu suuruse jaoks. [23]

Lisaks P-väärtusele saab mudelis kasutatavate parameetrite täpsust ja olulisust hinnata ka T-statistikute abil, mille näol on tegemist parameetriga, mis kujutab endast eeldatavat prognoositud parameetri väärtuse ja hüpooteetilise standardhälbe erinevust. T-statistikut leitakse järgneva valemi abil [18]:

$$t_{\beta_n} = \frac{\beta_n - \beta_0}{\sigma_{\beta_n}} \quad (2.15)$$

Kus

β_0 – on mittejhuslik konstant, mis võib või ei või võrduda tegeliku teadaoleva parameetriga β ;

β_n - on prediktor ehk sisuliselt käesolevas kontekstis regressioonkordaja;

σ_{β_n} – konstandi β_0 standardhälve tegeliku väärtuse β suhtes.

T statistik võimaldab testida nullhüpoteesi, et kas mittejhuslik konstant β_0 võrdub teadaoleva parameetriga β . [18]

Lisaks eelnevale aitab mudeli parameetrite olulisust ja relevantsust hinnata ka andmestiku R-ruut näidik. Tegemist on statistilise näitajaga, mida kasutatakse statistilistes mudelites, mille eesmärk on prognoosida mingite parameetrite väärtusi tulevikus teiste teadaolevate parameetrite väärtuste põhjal. R-ruut annab üldise hinnangu mudelis kasutatavatele parameetrite olulisusele ning sisuliselt annab hinnangu sellele, et kui täpselt on võimaline vaadeldava koosseisuga mudel prognoosima tulevase andmepunkte. Mida suurem on R-ruudu väärtus, seda täpsema mudeliga on tegu. Üldjuhul jääb R-ruudu väärtus 0 ja 1 vahele. Kui R-ruudu väärtus on 1, siis saab öelda, et tegemist on ideaalse mudeliga, mis on alati täielikult täpne oma prognoosides. [24]

Lisaks eelnevale võimaldab parameetrite olulisust mudelis mingil määral hinnata ka regressioonkordaja numbriline suurus. Samas aga peab regressioonkordaja põhjal olulisust hinnates arvesse võtma ka sõltumatute muutujate võimalike väärtusi, mille kordajateks need regressioonkordajad on, sest väga väikse regressioonkordaja puhul võib tegemist olla parameetriga, mis võib varieeruda sadades ühikutes ning kui mõni parameeter omab väga suurt regressioonkordajat, siis võib tema väärtus varieeruda väga väikesel määral, mis võib siiski tulemust moonutada.

3. Ilmastikumõjurite statistiline analüüs

Käesolev peatükk keskendub ilmastikunähtuste mõju analüüsile gaasitarbimisel Eestis. Ilmastikuandmed on pärit Riigi Ilmateenistuse ilmavaatluse andmestikust 7 ilmajaama kohta üle Eesti ning ilmajaamade valikul arvesse võetud andmete kättesaadavust ja kvaliteeti ning asukohtade rahvastiku arvu. Uurimuses kasutatud ilmajaamad on järgnevad [22]:

- Tallinn-Harku
- Tartu-Tõravere
- Pärnu-Sauga
- Viljandi
- Võru
- Väike-Maarja
- Narva

Uurimiseks kasutatavad parameetrid valiti selle põhjal, et andmed kirjeldaks valitsevaid ilmastikuolusid võimalikult objektiivselt ning numbriliselt. Ühtlasi võeti parameetrite valikul arvesse ka seda, et andmete kvaliteet oleks stabiilne ning vigaseid andmepunkte esineks minimaalselt või üldse mitte. Uurimuses kasutatud ilmastikuandmed on järgnevad [22]:

- Temperatuur
- Sademed
- Tuule kiirus
- Tuule suund
- Nähtavus kilomeetrites

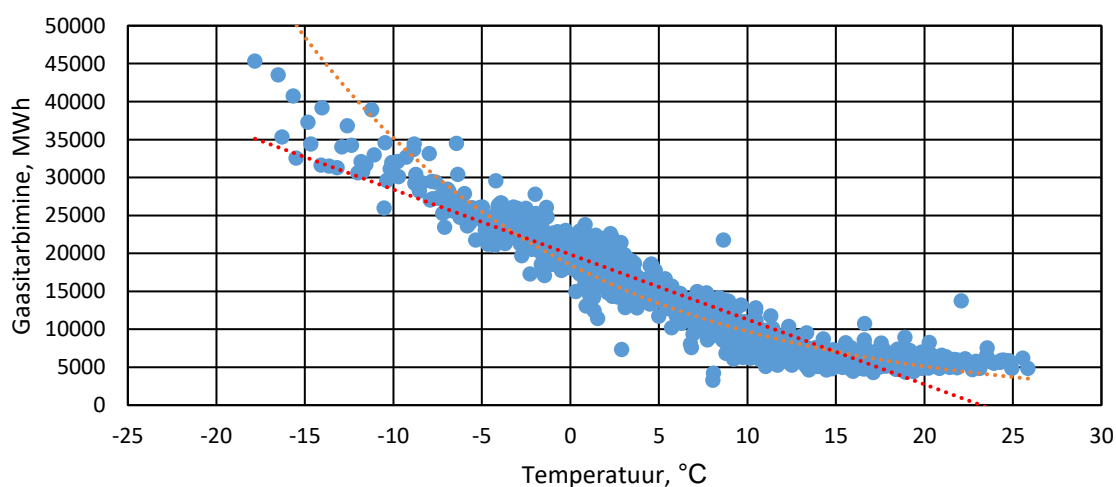
Ilmastikuandmete olulisus selgitatakse regressioonanalüüsi tulemusel ning põhilised olulisuse hindamiskriteeriumid on ilmastikuandmete P-väärtus ja regressioonkordaja absoluutväärtus. Regressioonanalüüsiks on kasutatud ka Eesti gaasitarbimise andmeid Eleringi andmestikust ning regressioonanalüüs viidi läbi andmete kohta ajavahemikul 2016. aasta märtsikuu kuni 2019. aasta veebruarikuu [23]. Esimese sammuna andmete regressioonanalüüsis võetakse aluseks kõik andmed ning uuritakse nende mõju gaasitarbimisele. Erinevate ilmajaamade andmed on töödeldud ja agregeeritud uurimuse käigus nii, et igale kuupäevale vastaks üks kokkuvõttev number. Temperatuurid, tuule kiirus, tuule suund ja nähtavus on ilmajaamade ja tundide lõikes keskmistatud ning sademed on summeeritud igas ilmajaamas nii, et tekiks asukoha- ja päevapõhine ülevaade andmetest. Seejärel on andmed agregeeritud bilansiperioodi põhisteks andmepunktideks keskmistamise teel. Üks bilansiperiood algab Eesti gaasisüsteemis kell seitse hommikul ning kestab 24 tundi.

3.1 Ilmastikuolude mõju gaasitarbimisele

Uurides ilmastikuolude mõju gaasitarbimisele eeldatakse käesoleva lõputöö raames seda, et mida halvem on ilm üldiselt, seda kõrgem on ka gaasitarbimine. Siinkohal on oluline defineerida ka see, et mida loetakse halvaks ilmaks. Halva ilma all käsitletakse käesolevas lõputöös jahedat, sademeterohket, tuulist ja vähese nähtavusega ilma. Hea ilma all käsitletakse sooja, väheste sademetega, tuulevaiksed ja suure nähtavusega ilma. Enne analüüside sooritamist on eeldatud ka seda, et kõik ilmastikutegurid mõjutavad gaasitarbimist eelnevalt mainitud suhte alusel. Juhul kui selline eeldus ei pea otseselt ja kindlalt paika, siis uuritakse analüütiliselt, et mis selliste seoste põhjus võib olla ning otsitakse ka keerulisemaid meetmeid ilmastikutegurite mõju analüüsiks gaasitarbimisele.

3.1.1 Bilansiperioodi keskmise temperatuuri mõju gaasitarbimisele

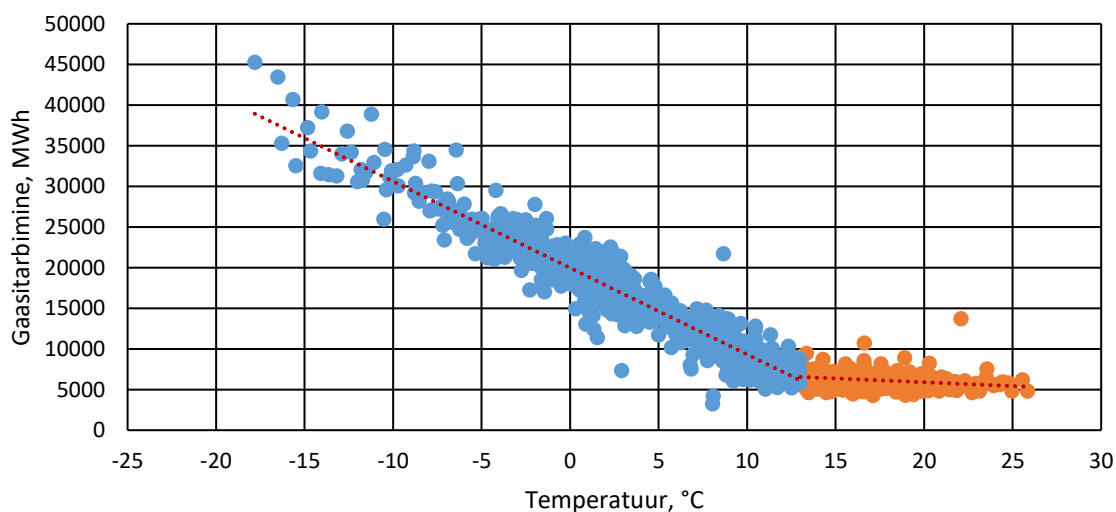
Ilmastikuolude mõju uurimise esimese sammuna vaadeldakse gaasitarbimise üldist sõltuvust ilmastikuoludest. Esialgu uuritakse iga ilmastikuparameetri ja tarbimise vahelist suhet eraldi, et selgitada vajadust keerulisemate analüüside järgi. Uurimise alguses on eeldatud, et üks olulisimaid gaasitarbimise mõjureid on temperatuur, mida illustreerib ka joonis 3.1, kus on kujutatud terve bilansiperioodi jooksul tarbitud gaasi kogus megavatt-tundides vastava bilansiperioodi keskmise temperatuuri juures. Joonisele 3.1 on lisatud ka lineaarne suunajoon, mis joonisel on kujutatud punase punktiirjoonega ja eksponentsiaalne suunajoon, mis joonisel on kujutatud oranži punktiiriga. Vaadeldes aga nimetatud joonist, siis selgub, et gaasitarbimine pole temperatuurist täielikult lineaarselt sõltuv, vaid tegemist on pigem tükiti lineaarse suhtega. Gaasitarbimise sõltuvust temperatuurist lineaarsena vaadeldes võib täheldada, et lineaarne suunajoon eksib just äärmuslikes oludes kõige rohkem ehk siis kõrge ja madala keskmise temperatuuriga päevadel.



Joonis 3.1. Gaasitarbimise lineaarne (punane punktiir) ja eksponentsiaalne (oranž punktiir) sõltuvus temperatuurist.

Lisaks eeltoodud lineaarsele sõltuvusele on joonisel 3.1 illustreeritud ka gaasitarbimise sõltuvus temperatuurist eksponentsiaalse suunajoone abil. Visuaalse vaatluse põhjal saab öelda, et eksponentsiaalne suunajoon järgib üldist gaasitarbimise sõltuvuse kuju temperatuurist mõnevõrra täpsemini. Samas on ka eksponentsiaalse sõltuvuse korral näha, et väga jahedate ilmade korral ei järgi suunajoon vaadeldavat andmekogu kõige täpsemini. Siiski on mõlemat liiki sõltuvuse korral näha, et sarnase temperatuuriga ilmade korral võib kogu Eesti gaasisüsteemi tarbimine varieeruda kuni 10 gigavatt-tundi ühes bilansiperioodis, mis on tõestuseks sellele, et lisaks temperatuurile esineb ka teisi olulisi mõjureid, mis kujundavad gaasitarbimist.

Vaadeldes joonist 3.1 on näha, et gaasitarbimise sõltuvus temperatuurist on pigem tükiti lineaarne, sest mõlemal joonisel esineb piirkondi, kus suunajoon on selgelt ühele poole valimit kallutatud ega kajasta tegeliku olukorda. Gaasitarbimise tükiti lineaarne sõltuvus temperatuurist annab ka aluse kasutada andmestiku analüüsimisel ja selle põhjal prognoosides vaid kindlat osa andmetest, mis on vaadeldaval ajahetkel analüüsi ja prognoosi koostamiseks aktuaalne ja relevantne. Sellisel juhul on ka prognoosimisel oluline kasutada kindlalt defineeritud treeningrea pikkust, mille põhjal prognoose koostatakse. Joonis 3.2 kirjeldab gaasitarbimise tükiti lineaarset sõltuvust temperatuurist.



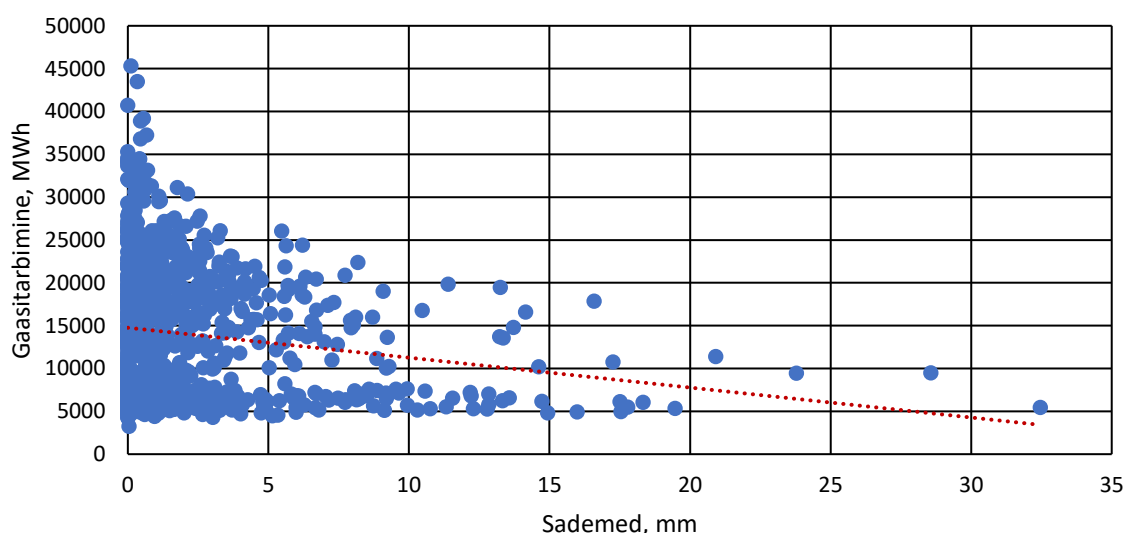
Joonis 3.2. Gaasitarbimise tükiti lineaarne sõltuvus temperatuurist.

Joonisel 3.2 on jaotatud vaadeldavad andmepunktid temperatuuri põhjal kahte ossa. Oranži värvi andmepunktid kirjeldavad gaasitarbimise andmepunkte päevadel, mil bilansiperioodi jooksul keskmine temperatuur on kõrgem kui 13 kraadi ning sinised andmepunktid kirjeldavad gaasitarbimise andmepunkte päevadel, mil bilansiperioodi keskmine temperatuur on alla 13 kraadi. Omistades sellise jaotuse korral mõlemale variandile eraldi suunajoone võib näha, et üleval- ja allpool suunajoont on andmepunktide jaotus oluliselt võrdsem ja järjekindlam, mis kinnitab väidet,

et gaasitarbimine on temperatuurist tükiti lineaarselt sõltuvuses ning erinevatel perioodidel avaldab temperatuur kogu Eesti gaasitarbimisele erineval määral mõju.

3.1.2 Sademete, tuule kiiruse ja nähtavuse mõju gaasitarbimisele

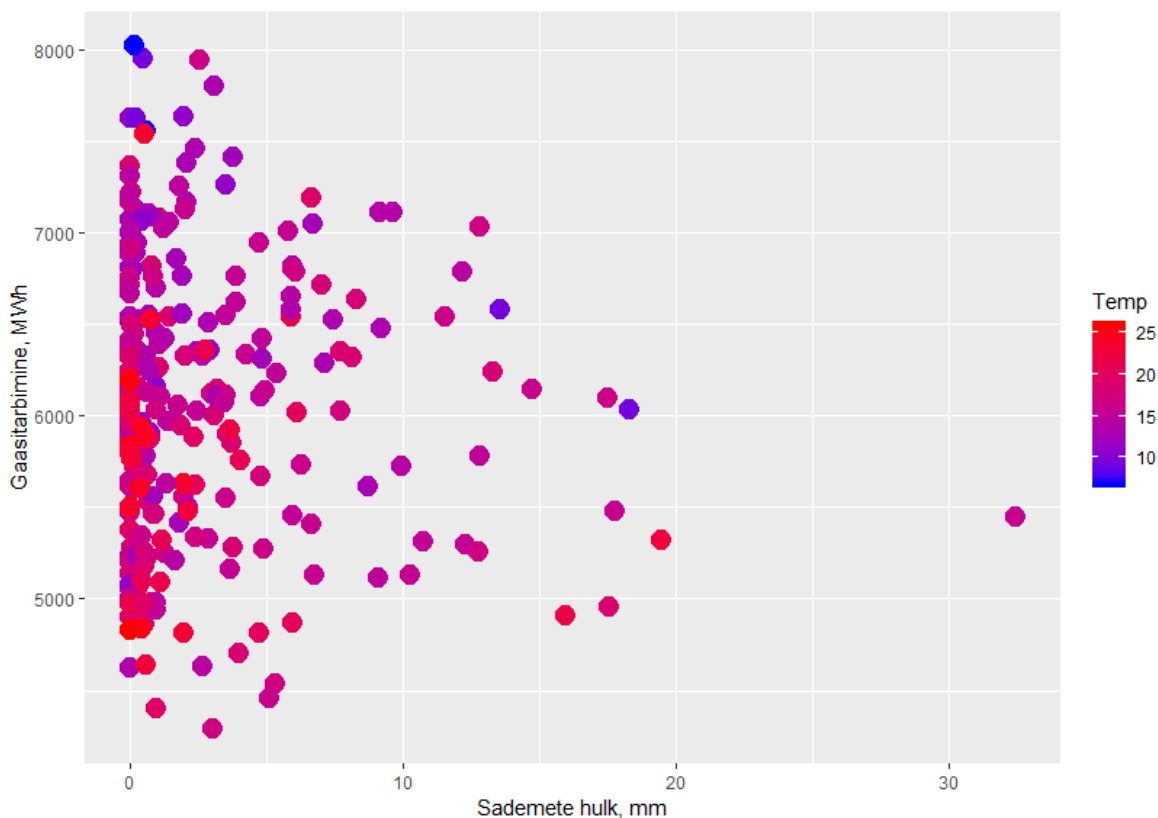
Lisaks temperatuurile vaadeldakse käesolevas lõputöös ka teiste ilmastikutegurite mõju gaasitarbimisele. Esmalt vaadeldakse, et kas ilmastikutegurid omavad eraldiseisvalt kindlat mõju gaasitarbimisele ning kas on võimalik tuvastada kindlat korrelatsiooni ilmastikutegurite ja gaasitarbimise vahel neid eraldi vaadeldes. Peale temperatuuri vaadeldakse esmalt gaasitarbimise sõltuvust sademetest. Enne analüüside sooritamist eeldati antud lõputöö raames, et suurem hulk sademeid põhjustab tõusu gaasitarbimises. Joonis 3.4 aga lükkab selle eelduse ümber, sest sellelt selgub, et just suurema hulga sademete korral on gaasitarbimine Eestis madalam ning kõrgema gaasitarbimisega päevadel on sademete hulk madalam. Joonisel 3.3 on graafiliselt kujutatud sademete hulk ühes bilansiperioodis käesoleva lõputöö raames kasutatud ilmajaamades.



Joonis 3.3. Gaasitarbimise sõltuvus sademetest.

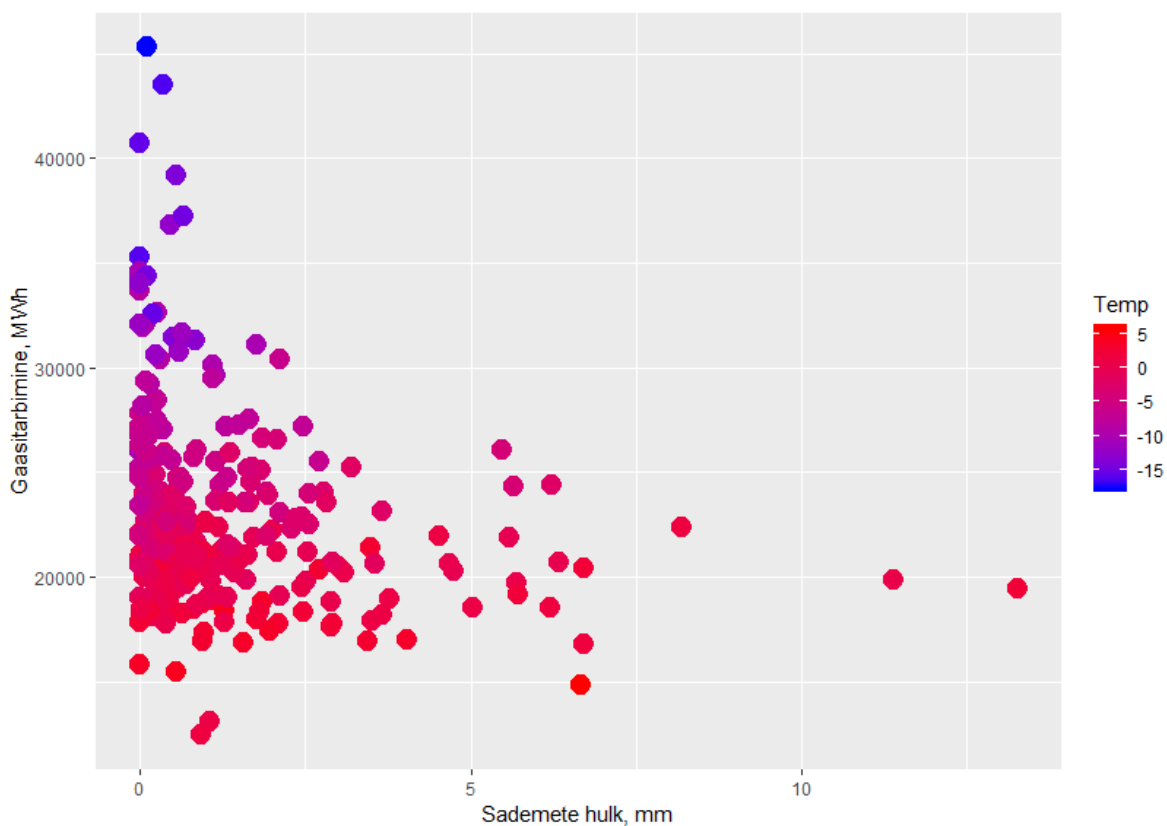
Joonise 3.3 põhjal saab öelda, et sademete hulk ei korreleeru piisavalt selgelt vaadeldava perioodi jooksul, et nendele omistada niivõrd lihtsal meetodil mingit kindlat arvulist väärtust, et kui suurel määral sademete hulk gaasitarbimist mõjutab. Samuti viitab ka lineaarne suunajoon pigem negatiivsele korrelatsioonile, kuigi sademete puhul on eeldatud positiivset korrelatsiooni sademete hulga ja gaasitarbimise vahel. Antud olukord võib tekkida näiteks juhul, kui suurema sademete hulgaga päevad on esinenud Eestis soojemal perioodil, kui gaasitarbimine on juba vaikimisi madalam ning ilmastikutegurid üldisemas pildis mõjutavad gaasitarbimist vähemalt määral. Joonised 3.4, 3.5 ja 3.6 illustreerivad gaasitarbimise sõltuvust sademetest suve-, talve- ja

sügiskuudel. Lisaks on joonistel määratud igale andmepunktile värvus sõltuvalt temperatuurist nii, et soojemad temperatuurid on punast värvi ning jahedamad temperatuurid on sinist värvi. Esimesel joonisel 3.4, mis illustreerib gaasitarbimise sõltuvust sademetest suvekuudel, ei ole märgata eriti kindlat mustrit sademete mõjust gaasitarbimisele.



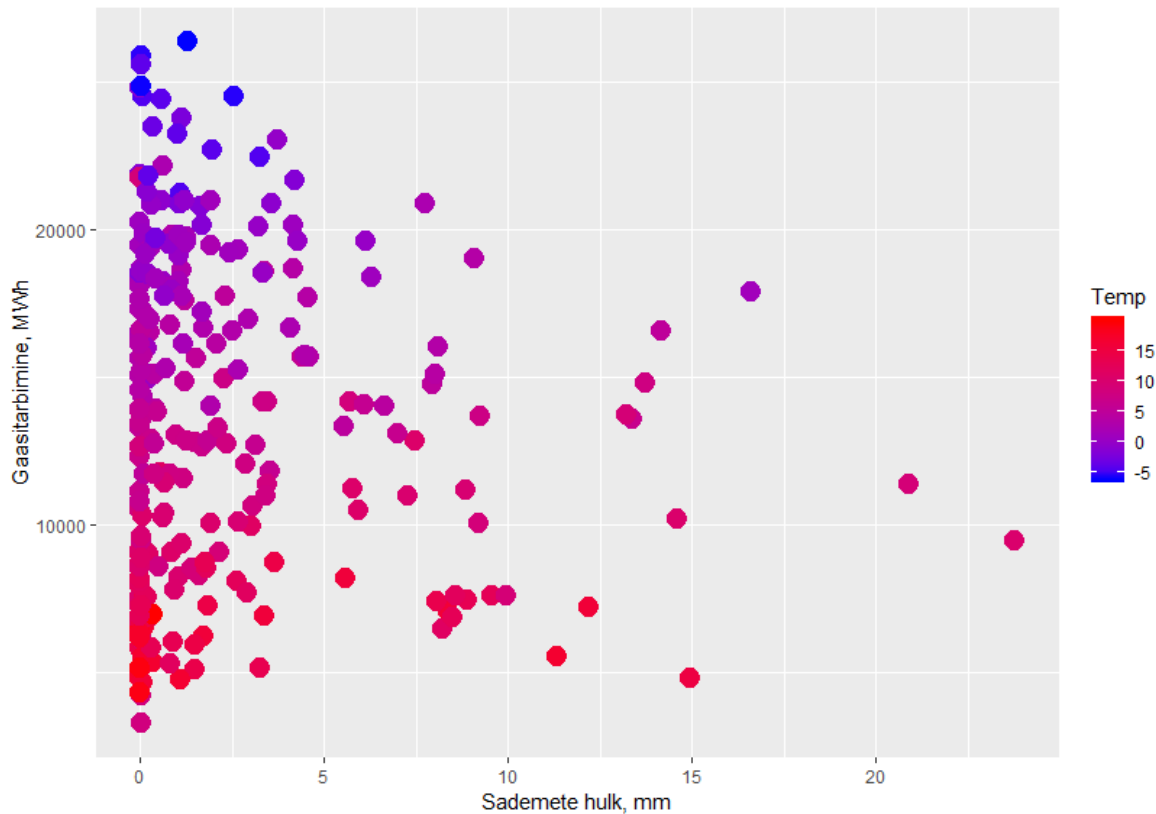
Joonis 3.4. Gaasitarbimise sõltuvus sademetest suvekuudel. Kuvatõmmis rakendustarkvarast R.

Joonisel 3.5 on graafiliselt kujutatud gaasitarbimise sõltuvust sademetest talvekuudel. Erinevalt joonisest 3.4 on joonisel 3.5 juba selgemalt näha korrelatsioone tarbimise ja temperatuuri ning sademete vahel



Joonis 3.5. Gaasitarbimise sõltuvus sademetest talvekuudel. Kuvatõmmis rakendustarkvarast R.

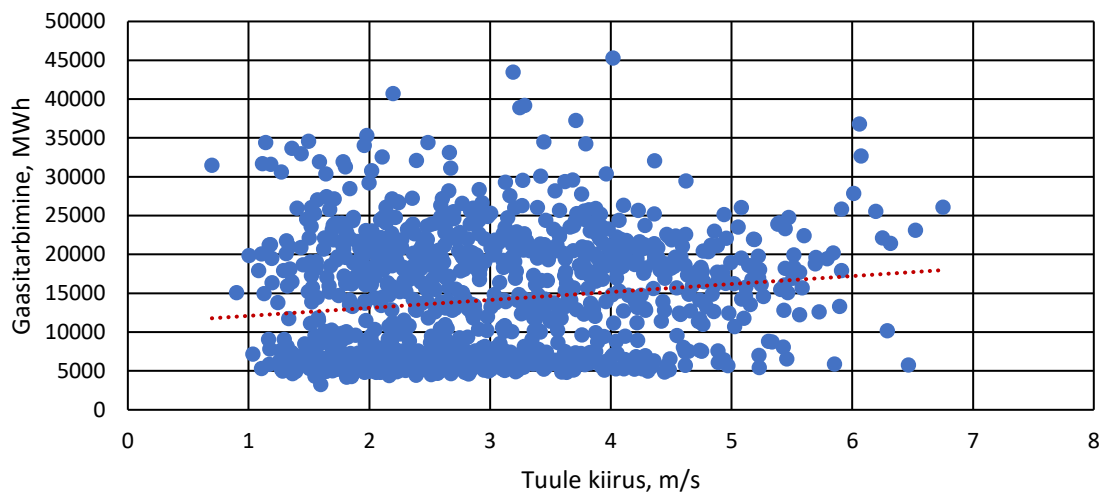
Joonisel 3.5 on näha väga kindlat suhet ka temperatuuri ja gaasitarbimise vahel, mis võib mõjutada gaasitarbimist piisavalt tugevalt, et lihtsate meetoditega ei ole võimalik kindlaks teha sademete mõju gaasitarbimisele. All oleval joonisel 3.6 on illustreeritud gaasitarbimise sõltuvus sademetest sügiskuudel. Sügiskuudel on samuti märgata, et soojemate ilmadega päevad koonduvad endiselt madalamasse tarbimisgruppi graafikul.



Joonis 3.6. Gaasitarbimise sõltuvus sademetest sügiskuudel. Kuvatõmmis rakendustarkvarast R.

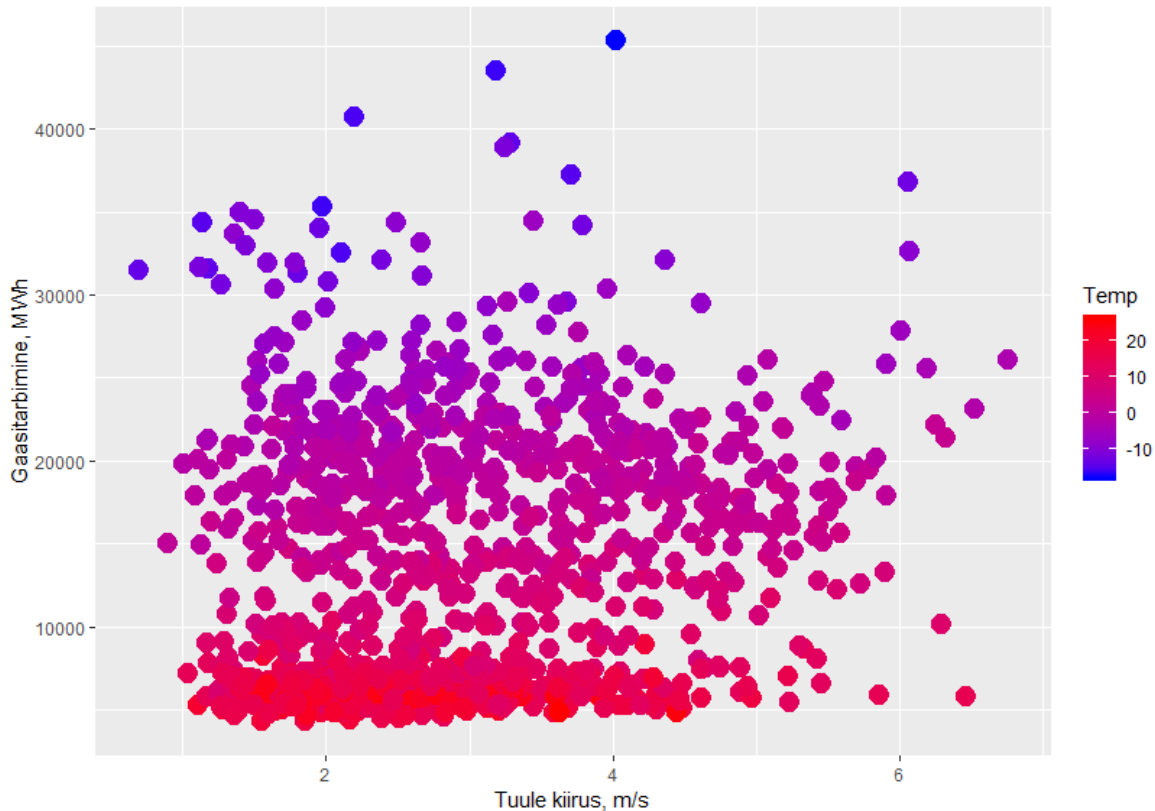
Kokkuvõtvalt saab sademete mõju kohta öelda tulenevalt asjaolust, et soojem temperatuur põhjustab madalamat tarbimist ning joonistel 3.4 kuni 3.6 kujutatul põhjal on võimalik öelda, et suurem hulk sademeid esineb üldjuhul soojematel päevadel, mistõttu kujutades vaid gaasitarbimist ja sademete hulka võib statistiliselt jõuda järelduseni, et sademete hulk korreleerub gaasi tarbimisega negatiivselt.

Järgmisena on vaadeldud tuule kiiruse ja gaasitarbimise vahelist sõltuvust joonisel 3.7. Vaatlemisele eelnevalt on eeldatud, et tuule kiirus ja gaasitarbimine on omavahel positiivselt korreleeruvad, sest eelduste kohaselt mõjub tuul jahutava tegurina ning seeläbi suurendab ka vajadust kasutada gaasi sooja saamiseks erinevates ehitistes. Joonisel 3.4 võib lineaarse suunajoone abil märgata väikest positiivset korrelatsiooni. Kuid vaadates antud joonist võib näha suurt andmepunktide hajuvust, mis tähendab, et andmepunktide kaugus suunajoonest on kuni 30 gigavatt-tundi ööpäeva jooksul, mis ühtlasi viitab ka sellele, et võrreldes temperatuuri mõjuga on tuule kiiruse mõju oluliselt väiksem gaasitarbimisele.



Joonis 3.7. Gaasitarbimise sõltuvus tuule kiirusest.

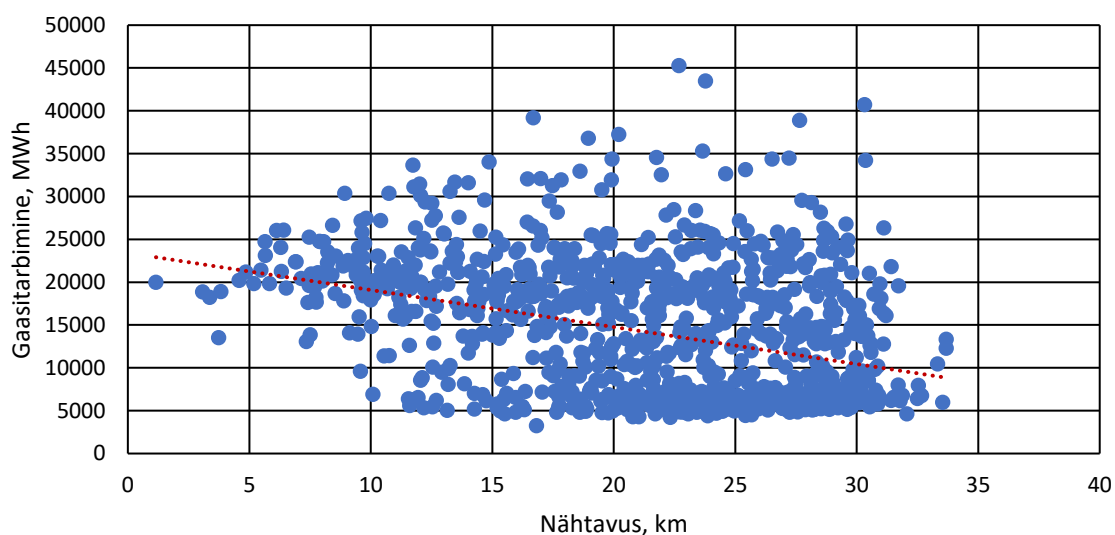
Sarnaselt sademete mõju uurimisele koos temperatuuriga on joonisel 3.8 kujutatud gaasitarbimise sõltuvust tuule kiirusest nii, et iga päeva keskmine temperatuur on graafikul kujutatud erineva värviga nii, et soojemate bilansiperioodi keskmiste temperatuuride korral on andmepunktid punased ning jahedamate bilansiperioodide keskmiste temperatuuride korral on andmepunktid sinised.



Joonis 3.8. Gaasitarbimise sõltuvus tuule kiirusest koos temperatuuriga. Kuvatõmmis rakendustarkvarast R.

Sarnaselt sademete puhul uuritule saab ka tuule kiirust uurides öelda, et tulenevalt temperatuuri väga suurest mõjust gaasitarbimisele ei saa tuule kiiruse mõju üksinda uurides teha kindlaid järeldusi tuule kiiruse mõjust gaasi tarbimisele.

Viimane ilmastikutegur, mida vaadeldakse, on nähtavus. Nähtavus on välja toodud antud uurimuses kilomeetrites ehk kui kaugele on võimalik maapinnal näha. Nähtavust on kasutatud antud uurimuses, sest see ilmastikutegur aitab koondada mitmeid teisi ilmastikutegureid enda alla nagu näiteks udu ja pilvisus. Halva nähtavuse korral ei avalda ka päike niivõrd suurt soojendavat mõju maapinnale ja ehitistele, mis annab alust eeldada, et gaasitarbimise ja nähtavuse vahel on negatiivne seos. Joonisel 3.5 on kujutatud gaasitarbimise andmepunkte erineva nähtavuse korral ning suunajoon viitab samuti sellele, et mingil määral negatiivne seos on gaasitarbimise ja nähtavuse vahel olemas. Samas on sellelt jooniselt ka näha jällegi võrdlemisi suurt hajuvust, mis tähendab, et vaid nähtavuse põhjal on raske teha järeldusi gaasitarbimise kohta.



Joonis 3.5. Gaasitarbimise sõltuvus nähtavusest.

Tulenevalt sellest, et lihtsama analüüsi käigus oli võimatu tuvastada üksikute parameetrite kindlat mõju gaasitarbimisele on vajalik kasutada keerulisemat ja ülevaatlikumat meetodikat ilmastikutegurite mõju analüüsimiseks gaasitarbimisele. Põhiliseks probleemiks gaasitarbimise mõjurite uurimisel oli asjaolu, et temperatuur avaldab gaasitarbimisele niivõrd suurt mõju ning seetõttu ei ole võimalik teha kindlaid järeldusi erinevate tegurite mõjust gaasitarbimisele.

3.2 Andmestiku regressioonanalüüs

Tulenevalt sellest, et ilmastikutegurite mõju uurimine eraldi tegurite haaval ei anna piisavalt head ja objektiivset ülevaadet tegurite mõjust gaasitarbimisele on käesolevas lõputöös kohaldatud tervele olemasolevale andmestikule regressioonanalüüsi meetodit, mis võimaldab uurida ilmastikutegurite koosmõju gaasitarbimisele. Regressioonanalüüs võimaldab koostada mudeli hindamiseks tunnuste vahelisi seoseid vaadeldes üht tunnust kui sõltuvat ning ülejäänud tunnuseid, mille põhjal oleks võimalik kirjeldada ja prognoosida sõltuva tunnuse väärtusi. Mida tugevam on seos sõltuva ja sõltumatute tunnuste vahel, seda olulisem on kasutada neid sõltumatuid tunnused sõltuva tunnuse prognoosimisel ning seda täpsem on ka prognoosimudel [16]. Käesolevas lõputöös on sõltuvaks tunnuseks gaasitarbimine ning analüüsitakse gaasitarbimise sõltuvust ilmastikuoludest. Regressioonanalüüs aitab selgitada välja ka gaasitarbimist kõige enam mõjutavad ilmastikutegurid ning võimaldab selgitada välja ka nende mõju olulisuse gaasitarbimisele läbi P-väärtuse. Käesolevas uurimuses on kogu andmestiku regressioonanalüüs läbi viidud viies iteratsioonis, kus igas iteratsioonis eemaldatakse andmestikust eelnevas iteratsioonis kõige vähemolulisem ilmastikutegur ning seejärel on leitud ülejäänud ilmastikuteguritele uued regressioonkordajad.

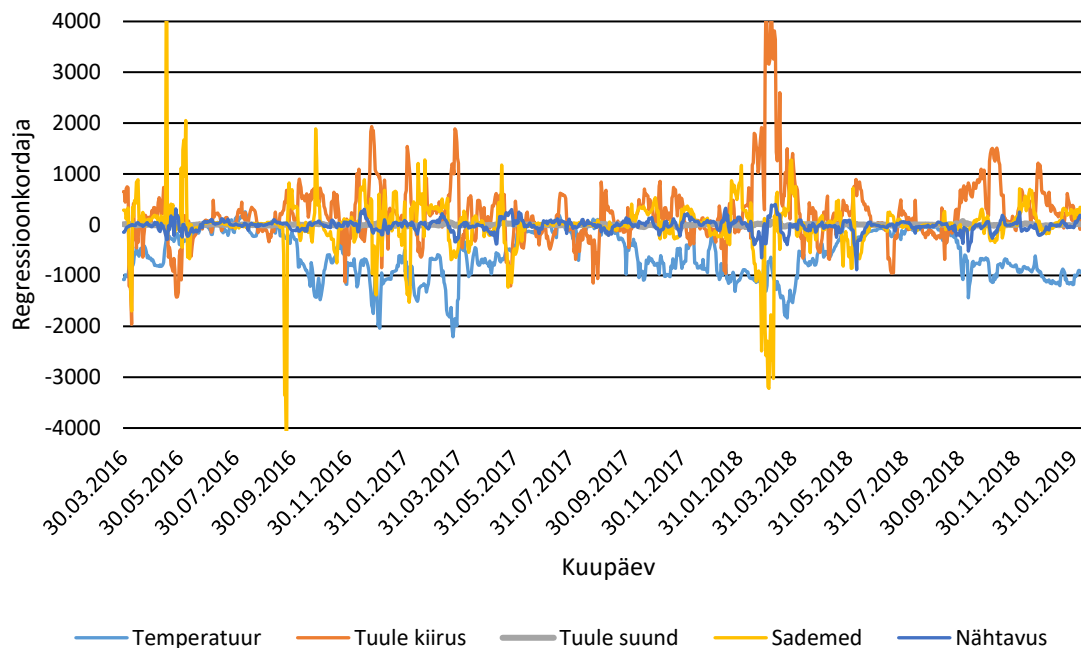
3.2.1 Terve andmestiku regressioonanalüüs

Esimese sammuna andmete regressioonanalüüsis vaadeldakse kõiki ilmastiku parameetreid ning uuritakse nende seost Eesti gaasitarbimisega. Lisaks annab esimene iteratsioon võrdlusaluse järgnevateks iteratsioonideks. Esimese tervikanalüüsi tulemusel selgitatakse välja, et millised ilmastiku parameetrid on gaasitarbimise seisukohalt kõige olulisemad ja omavad suurimat seost tarbimise kujunemisel. All toodud tabelis 3.1 on terve andmestiku regressioonanalüüsi tulemused, kus koefitsiendid määravad sõltumatute tegurite regressioonkordajad ning statistilised näitajad regressioonmudelil.

Tabel 3.1. Terve andmestiku regressioonanalüüsi tulemused.

	Koefitsiendid	Standardhälve	t-stat	P-väärtus	Alumine 95%	Ülemine 95%
Vabaliige	22219,46	356,36	62,4	0	21520,21	22918,72
Temperatuur	-825,17	8,95	-92,2	0	-842,73	-807,61
Tuule suund	-7,97	1,15	-6,9	$6,93 \cdot 10^{-12}$	-10,23	-5,72
Tuule kiirus	267,22	64,21	4,2	$3,41 \cdot 10^{-5}$	141,24	393,21
Sademed	-68,76	22,76	-3,0	$2,60 \cdot 10^{-3}$	-113,42	-24,10
Nähtavus	-79,98	11,88	-6,7	$2,71 \cdot 10^{-11}$	-103,28	-56,67

Tabeli 3.1 põhjal saab järeldada, et kõige suurema olulisuse ja mõjuga gaasitarbimisele on temperatuur tulenevalt oma madalast P-väärtusest ning kõrgest regressioonkordajast. Edasiste iteratsioonide käigus eemaldatakse igal korral kõike väiksema olulisusega ilmaparameeter analüüsist ning sooritatakse uus analüüs seni kuni järgi jääb ainult kõige olulisem parameeter. Regressioonanalüüsi esimeses iteratsioonis oli kõige vähemoluline ilmastikutegur tuule suund tulenevalt oma väga väiksest regressioonkordajast. Tulenevalt gaasitarbimise ja selle parameetrite mõjust on oluline vaadelda ka erinevate ilmaparameetrite regressioonkordajaid ajas muutuvatena. Joonisel 3.6 on kõikide esimeses iteratsioonis analüüsitud ilmaparameetrite regressioonkordajad, mis aitavad iseloomustada erinevate ilmaparameetrite mõju gaasitarbimisele numbriliselt. Antud joonisel on sooritatud iga päeva kohta regressioonanalüüs eelmise 15 päeva andmete põhjal.



Joonis 3.6. Regressioonkordajate muutumine ajas regressioonanalüüsi esimeses iteratsioonis.

Joonisel 3.6 on võimalik selgelt näha eelkõige temperatuuri mõju sesoonsust gaasitarbimisele, mille regressioonkordaja võib omada külmematel kuudel väärtust kuni 2000 ning soojematel kuudel omab jällegi väga väikest väärtust. Regressioonkordaja sätestab sisuliselt selle, et mitme megavatt-tunnist mõju omab vastava parameetri muutus ühe ühiku võrra gaasitarbimisele. Samuti on kõrgemaid tippe näha ka teiste parameetrite nagu näiteks sademete ja tuule kiiruse regressioonkordajates just külmematel kuudel. Samas on näha ka mitmeid äkilisi varieeruvusi näiteks sademete ja tuule kiiruse regressioonkordajates, mis kohati omavad lühikese perioodi jooksul tavapärasest väga palju erinevat regressioonkordajat. Sellised olukorrad tekivad siis, kui analüüsi aluseks oleva perioodi jooksul omab mõni ilmaparameeter tavapärasest väga palju väiksemat väärtust võrreldes oma tavapärase väärtusega. Samalt jooniselt on näha ka seda, et tuule suund omab võrreldes teiste parameetritega väga väikest mõju gaasitarbimisele, mis toetab hinnangut, et tegemist on kõige vähemolulisema parameetriga andmete valimis. Regressioonkordajate sesoonsust põhjustab ka gaasitarbimise sesoonsus, sest saavutamaks regressioonfunktsiooni abil arvutades suuremat gaasitarbimise numbrit peavad ka ilmastikuparameetrite regressioonkordajad omama suuremat väärtust.

3.2.2 Teise iteratsiooni regressioonanalüüs

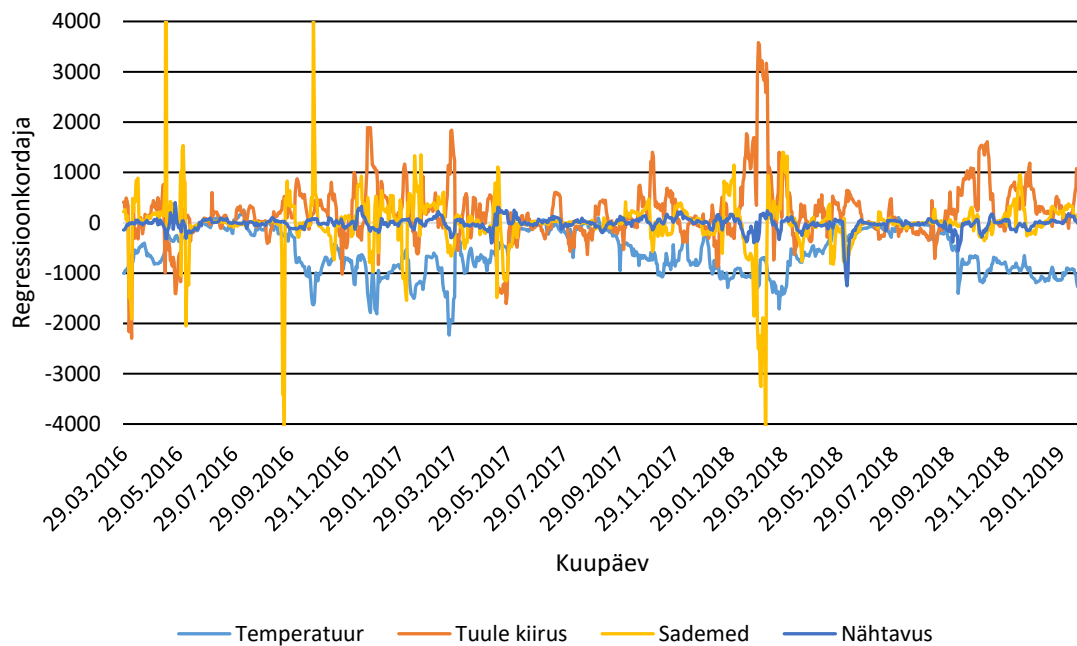
Regressioonanalüüsi teises iteratsioonis eemaldati analüüsist tuule suund, sest tuule suund omas kõige väiksemat olulisust vastavalt tabelis 3.1 toodud andmetele. Tabelis 3.2 on toodud regressioonanalüüsi teise iteratsiooni tulemused.

Tabel 3.2. Teise iteratsiooni regressioonanalüüsi tulemused.

	Koefitsiendid	Standardhälve	t-stat	P-väärtus	Alumine 95%	Ülemine 95%
Vabaliige	20983,35	315,33	66,54	0	20364,62	21602,09
Temperatuur	-825,99	9,14	-90,35	0	-843,93	-808,05
Tuule kiirus	243,44	65,51	3,72	$2,13 \cdot 10^{-4}$	114,90	371,99
Sademed	-78,11	23,21	-3,36	$7,93 \cdot 10^{-4}$	-123,66	-32,56
Nähtavus	-88,35	12,07	-7,32	$4,96 \cdot 10^{-13}$	-112,04	-64,66

Tabeli 3.2 põhjal saab järeldada, et sel korral on sademed kõige vähem oluline parameeter gaasitarbimisele tulenevalt kõrgeimast P-väärtusest ja madalaimast regressioonkordajast. Sellest tulenevalt eemaldatakse regressioonanalüüsi kolmandaks iteratsiooniks sademete andmed, et selgitada välja edasised tähtsaimad ja vähim olulisemad parameetrid. Samuti on märkimist väärt asjaolu, et sademete ja nähtavuse parameetrid omandasid tuule suuna eemaldamisel mõnevõrra suurema osakaalu regressioonkordaja poolest ning tuule kiiruse osakaal regressioonkordaja poolest muutus madalamaks. P väärtuse põhjal hinnates kaotasid ka tuule kiirus ja sademete hulk oma olulisust, kuid nähtavuse parameeter muutus tuule suuna eemaldamisel P väärtuse põhjal olulisemaks.

Joonisel 3.7 on toodud regressioonkordajate muutumine ajas regressioonanalüüsi teises iteratsioonis. Võrreldes esimese iteratsiooniga ei toimunud suuri muutusi ning endiselt on antud jooniselt näha, et temperatuuri, tuule kiiruse ja sademete mõju gaasitarbimisele omavad sarnast sesoonsust nagu on näha ka joonisel 3.6. Samas on endiselt märgata sademete ja tuule kiiruse regressioonkordaja järske muutusi, mis viitavad sellele, et vaadeldaval perioodil on esinenud ilmastikus püsivalt tuulevaikseid ja ka sademeteta perioode.



Joonis 3.7. Regressioonkordajate muutumine ajas regressioonanalüüsi teises iteratsioonis.

Joonisel 3.7 on näha ka seda, et sademete puhul omab mitmel juhul sademete regressioonkordaja tavapärasest väga suurel määral erinevat regressioonkordajat, mille järgi võib eeldada ka seda, et sademete hulga puhul on tegemist võrdlemisi volatiilse muutujaga, mis võib samas ka periooditi omada pikema aja jooksul väga väikest väärtust. Sellest tulenevalt saab kinnitada, et statistiliselt on keeruline võtta sademete hulka gaasitarbimise prognoosimisel arvesse ning lisaks sellele, et tegu on P väärtuse järgi kõige vähem olulisema ilmastikuparameetriga gaasitarbimise suhtes, saab ka öelda, et prognoosimisel nende andmete kasutamine võib põhjustada suuri prognoosivigasid, sest omades niivõrd suurt regressioonkordajat võib sademetevaesele perioodile järgnev sademetega ilm tekitada reaalsusest suurel määral erineva prognoosi.

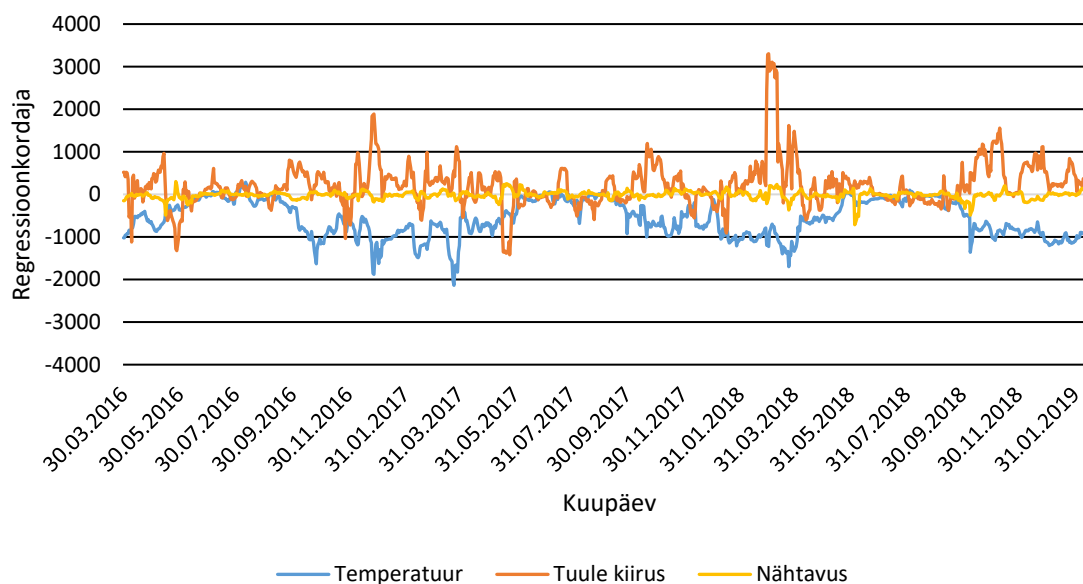
3.2.3 Kolmanda iteratsiooni regressioonanalüüs

Regressioonanalüüsi kolmandas iteratsioonis eemaldati andmete valimist sademete hulk tänu suurenenud P väärtusele ning kõige madalamale regressioonkordajale. Tabelis 3.3 on toodud regressioonanalüüsi kolmanda iteratsiooni tulemused. Regressioonanalüüsi kolmandas iteratsioonis langes oluliselt tuule kiiruse regressioonkordaja ning tõusis suurel määral selle P väärtus, mille põhjal otsustati järgnevas iteratsioonis eemaldada andmete valimist tuule kiirus. Nähtavusel on küll madalam regressioonkordaja absoluutväärtus, kuid tulenevalt väga madalast P väärtusest otsustati see andmestiku jätta ka järgmiseks iteratsiooniks.

Tabel 3.3. Kolmanda iteratsiooni regressioonanalüüsi tulemused.

	Koefitsiendid	Standardhälve	t-stat	P-väärtus	Alumine 95%	Ülemine 95%
Vabaliige	20822,12	313,16	66,49	0	20207,64	21436,60
Temperatuur	-834,04	8,87	-94,08	0	-851,44	-816,65
Tuule kiirus	188,28	63,73	2,95	$3,20 \cdot 10^{-3}$	63,23	313,33
Nähtavus	-77,58	11,70	-6,63	$5,25 \cdot 10^{-11}$	-100,53	-54,62

Joonisel 3.8 on toodud ilmaparameetrite regressioonkordajate muutumine ajas regressioonanalüüsi kolmandas iteratsioonis. Endiselt on näha, et regressioonkordajate sesoonsuses ei ole märgata olulisi muutusi võrreldes eelmiste regressioonanalüüsi iteratsioonidega. Samas on näha, et tuule kiiruse sesoonsus ei ole niivõrd korrapärane kui näiteks temperatuuri regressioonkordaja sesoonsus. Sellise sesoonsuse puudumist iseloomustab ka tuule kiiruse kõige suurem P väärtus. Tulenevalt gaasitarbimise sesoonsusest on alust eeldada seda, et ilmaparameetrite regressioonkordaja absoluutväärtus suureneks sarnase sesoonsusega ning niivõrd selge sesoonsuse puudumist väljendabki tuule kiiruse suhteliselt suur P väärtus võrdluses nähtavuse P väärtusega.



Joonis 3.8. Regressioonkordajate muutumine ajas regressioonanalüüsi kolmandas iteratsioonis.

Tulenevalt nõrgemast sesoonsusest ja kohati tavapärasest väga suurel määral erinevast regressioonkordajast saab öelda, et tuule kiiruse eemaldamine sisendandmete valimist peale kolmandat iteratsiooni on õigustatud.

3.2.4 Neljanda iteratsiooni regressioonanalüüs

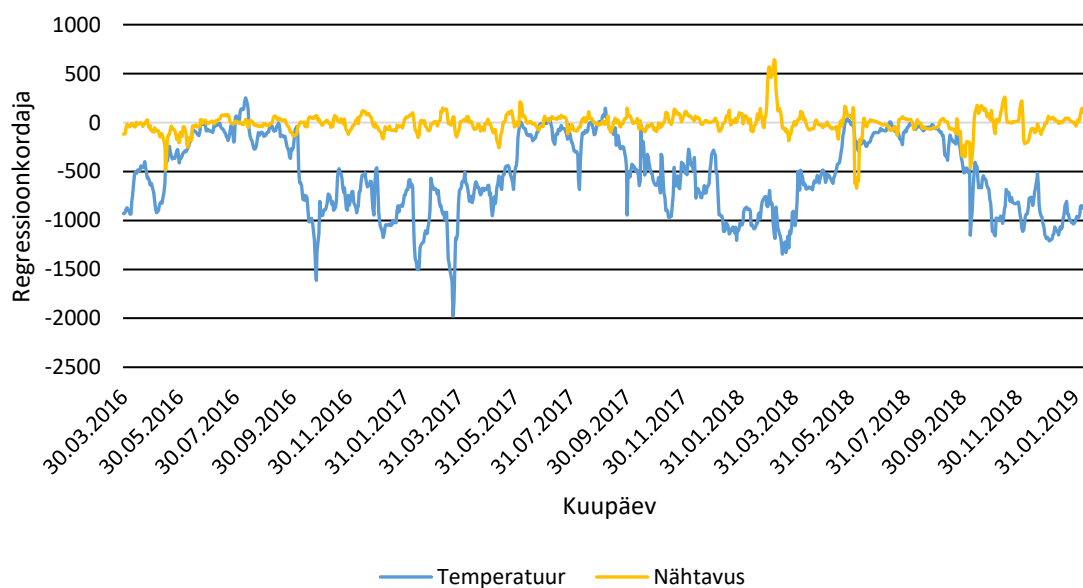
Regressioonanalüüsi neljandas iteratsioonis eemaldati andmestikust tuule kiiruse parameeter tulenevalt eelnevalt toodud põhjustest. Neljandas iteratsioonis oli sõltumatutest muutujatest andmestiku jäetud veel vaid temperatuur ja nähtavus kilomeetrites. Tabelis 3.4 on toodud regressioonanalüüsi neljanda iteratsiooni tulemused.

Tabel 3.4. Neljanda iteratsiooni regressioonanalüüsi tulemused.

	Koefitsiendid	Standardhälve	t-stat	P-väärtus	Alumine 95%	Ülemine 95%
Vabaliige	21379,68	250,80	85,24	0	20887,56	21871,80
Temperatuur	-837,55	8,82	-94,99	0	-854,85	-820,25
Nähtavus	-76,49	11,73	-6,52	$1,09 \cdot 10^{-10}$	-99,51	-53,46

Tuule kiiruse eemaldamisel andmete valimist ei muutunud märkimisväärselt ülejäänud muutujate regressioonkordajad ning suurimaks muutuseks oli nähtavuse P väärtuse suurenemine. Tulenevalt kõrgest P väärtusest otsustati viimaseks regressioonanalüüsi iteratsiooniks eemaldada valimist nähtavuse parameeter ning sooritada viies iteratsioon vaid temperatuuri parameetriga.

Joonisel 3.9 on toodud regressioonkordajate muutumine ajas regressioonanalüüsi kolmandas iteratsioonis.



Joonis 3.9. Regressioonkordajate muutumine ajas regressioonanalüüsi neljandas iteratsioonis.

Endiselt on säilinud sarnane sesoonsus võrreldes eelnevate iteratsioonidega. Samas on märgata, et järsud muutused mõnes teises regressioonkordajas on mõnel määral üle kandunud nähtavuse ja temperatuuri regressioonkordajatele. Parimaks näiteks on 2018. aasta märtsikuus tekkinud järsk lühiajaline tõus. See võib tuleneda ka sellest, et märtsikuus oli järsk tõus ka gaasitarbimises, mistõttu proovitakse leida regressioonanalüüsis selliseid regressioonkordajaid, mis vastaks ka sellisele kõrgele gaasitarbimisele ning sellest tulenevalt võibki mõne parameetri regressioonkordaja lühiajaliselt väga kõrge numbrilise väärtuse omandada.

3.2.5 Viienda iteratsiooni regressioonanalüüs

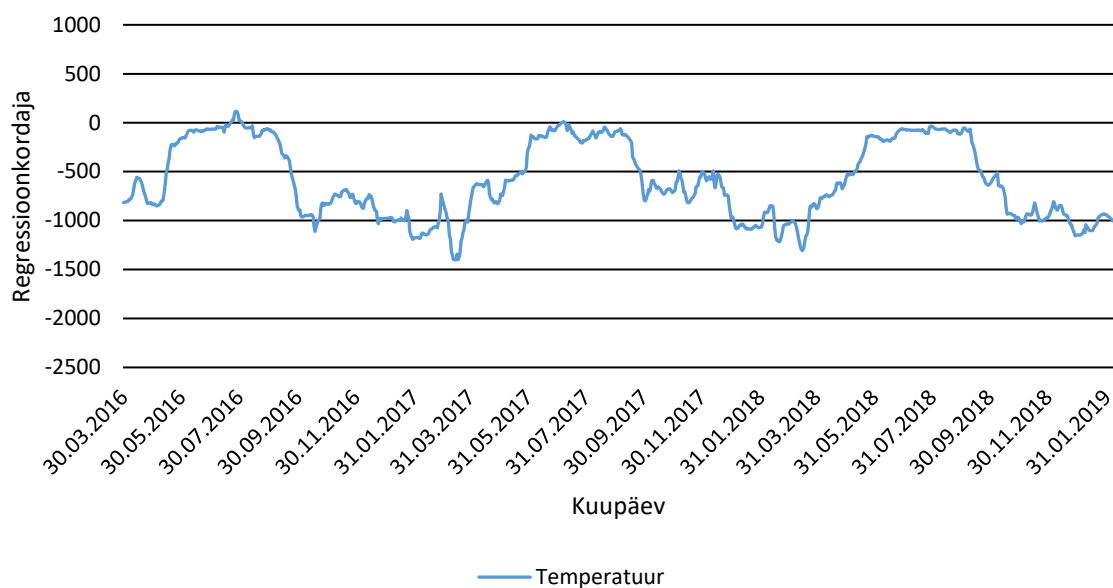
Regressioonanalüüsi viiendas iteratsioonis vaadeldi vaid temperatuuri ja gaasitarbimise vahelist sõltuvust ning analüüsimisel saadud tulemused on toodud tabelis 3.5. Nähtavuse eemaldamisel valimist tõusis mõnel määral temperatuuri regressioonkordaja absoluutväärtus ning vähenes vabaliikme suurus.

Tabel 3.5. Neljanda iteratsiooni regressioonanalüüsi tulemused.

	Koefitsiendid	Standardhälve	t-stat	P-väärtus	Alumine 95%	Ülemine 95%
Vabaliige	19856,74	92,93	213,66	0	19674,38	20039,09
Temperatuur	-855,92	8,51	-100,52	0	-872,63	-839,21

Sooritatud regressioonanalüüsi iteratsioonid on aluseks ka neljandas peatükis sooritatavatele modelleerimise iteratsioonidele, kus igas iteratsioonis kasutatakse vastavas regressioonanalüüsi iteratsioonis kasutatud parameetreid.

Joonisel 3.10 on toodud regressioonkordajate muutumine ajas regressioonanalüüsi viiendas iteratsioonis. Viiendas iteratsioonis on sisendandmetena ainsa sõltumatu muutujana kasutatud temperatuuri andmeid, mistõttu on antud joonisel välja toodud vaid temperatuuri regressioonkordaja muutumine ajas. Joonisel on näha, et temperatuuri regressioonkordaja on säilitanud väga selge sesoonsuse ning sellest tulenevalt on tegemist ka statistiliselt kõige olulisema parameetriga, mis mõjutab gaasitarbimist.



Joonis 3.10. Regressioonkordajate muutumine ajas regressioonanalüüsi viiendas iteratsioonis.

Kokkuvõtvalt saab öelda, et temperatuur avaldab gaasitarbimisele statistiliselt kõige suuremat mõju ning kogu andmestiku raames analüüsidest jääb temperatuuri regressioonkordaja absoluutväärtus vahemikku 825 kuni 856, mis tähendab, et üks kraad temperatuuri muutust mõjutab gaasitarbimist 824 kuni 855 megavatt-tundi päevas.

3.3 Analüüsi tulemused ja järeldused

Käesolevas peatükis vaadeldi erinevate ilmastikutegurite mõju kogu Eesti gaasitarbimisele vahemikus 2016. aasta märtsikuu kuni 2019. aasta veebruarikuu. Esmases analüüsi põhjal jõuti järeldusele, et temperatuur mõjutab gaasitarbimist väga suurel määral, mistõttu ei olnud võimalik lihtsa analüüsi tulemusel selgitada erinevate parameetrite numbrilist mõju gaasitarbimisele. Sellest tulenevalt sooritati saadavaloleva andmestiku põhjal regressioonanalüüs viie erineva iteratsiooni põhjal ning igas iteratsioonis eemaldati andmestikust ilmastikuparameeter, mis oli eelmises iteratsioonis statistiliselt kõige vähem oluline. Regressioonanalüüsi tulemusel omistatakse regressioonfunktsioonis igale sõltumatule muutujale regressioonkordaja, mis sisuliselt iseloomustab seda, et mitu megavatt-tundi päevas muutub gaasitarbimine sõltumatu kordaja väärtuse muutumisel ühe ühiku võrra. Numbriliselt saab väljendada erinevate ilmastikuparameetrite hinnanguliselt keskmist mõju kogu vaadeldava perioodi jooksul, ning need andmed on toodud tabelis 3.6.

Tabel 3.6. Ilmastikuparameetrite suurimad ja vähimad regressioonkordajate absoluutväärtused erinevate iteratsioonide jooksul.

Ilmastikuparameeter	Suurim regressioonkordaja	Vähim regressioonkordaja
Temperatuur	855,92	825,17
Nähtavus	88,35	76,49
Tuule kiirus	267,22	188,28
Sademed	78,11	68,76
Tuule suund	7,97	7,97

Tabeli 3.6 põhjal saab öelda, et kõige enam mõjutab gaasitarbimist temperatuur ning kogu vaadeldava perioodi jooksul tähendas 1 kraad temperatuuri muutmust 825,17 kuni 855,92 megavatt-tunni suurust muutust kogu bilansiperioodi gaasitarbimises. Samas vaadeldes regressioonkordajaid ajas muutuvatena võis märgata, et temperatuuri muutus ühe kraadi võrra võib sesoonselt omada mõju gaasitarbimisele ka kuni 2000 megavatt-tundi päevas. Nähtavus mõjutas kogu Eesti gaasitarbimist keskmiselt kogu vaadeldava perioodi vältel 76,49 kuni 88,35 megavatt-tundi päevas ühe kilomeetri nähtavuse kohta. Nähtavuse parameeter oli ka ajas muutudes vähese sesoonsusega ning selle mõju oli kogu vaadeldava perioodi jooksul võrdlemisi stabiilne. Tuule kiiruse muutus 1 meeter sekundis omas kogu vaadeldava perioodi jooksul keskmiselt mõju 188,28 kuni 267,22 megavatt-tundi päevas. Sademed omasid kogu vaadeldava perioodi jooksul mõju gaasitarbimisele 68,76 kuni 78,11 megavatt-tundi päevas 1 millimeetri sademete hulga kohta. Tuule suund omas kõige väiksemat mõju gaasitarbimisele ning 1 kraad tuule suuna muutuses mõjutas kogu Eesti gaasitarbimist keskmiselt 7,97 megavatt-tundi päevas kogu vaadeldava perioodi jooksul.

Üldiselt võis regressioonanalüüsi käigus näha seda, et mida kindlama sesoonsusega muutus ilmastikuparameetrite regressioonkordaja ajas, seda statistiliselt olulisema parameetriga gaasitarbimise suhtes on tegu ning seda kinnitas ka P väärtus. Tulenevalt ilmastikuparameetrite erinevast varieerumisest ja numbrilistest väärtustest võis varieeruda ka regressioonkordaja. Näiteks tuule suunda arvestati analüüsis erinevusena põhjasuunast kraadides, mis tähendas, et tuule suuna parameetri väärtus võis varieeruda teoreetiliselt vahemikus 0 kuni 180 kraadi. Samas aga näiteks kõige suurem keskmine tuule kiirus bilansiperioodi jooksul oli 6,75 meetrit sekundis ning kõige väiksem 1,16 meetrit sekundis. Keskmine temperatuur ühes bilansiperioodis varieerus

andmestikus -17,82 kuni 25,85 kraadi, mis tähendab, et hoolimata vähemast varieeruvusest võrreldes näiteks tuule suuna parameetriga, omab see endiselt väga suurt mõju gaasitarbimise suhtes.

Regressioonanalüüsi põhjal võib eeldada ka seda, et prognoosimudeli kaks esimest iteratsiooni on võrreldes järgmiste iteratsioonidega ebatäpsemad, sest tuule suuna ja sademete regressioonkordaja ei omanud ajas muutuvana kindlat sesoonsust või andmestiku analüüsimisel esinesid olukorrad kus vastavate parameetrite regressioonkordajad omasid lühikese perioodi jooksul tavapärasest väga erinevaid väärtusi, mis tähendab, et kasutades neid sellistel hetkedel prognoosimudeli sisendiks, kus need parameetrid omavad tavapärasest väga palju suuremat regressioonkordajat võrreldes tavapärasega, võib nende tegelik mõju prognoositavas perioodis olla väga palju erinev statistilisest hinnangus. Sellest tulenevalt tekib prognoosimudelis eeldatavasti ka suur prognoosiviga sellistes olukordades.

4. Gaasitarbimise modelleerimine

Käesoleva lõputöö neljandas peatükis keskendutakse gaasitarbimise modelleerimisele ja prognoosimisele esimeses peatükis leitud parameetrite põhjal, et anda objektiivne hinnang leitud parameetrite täpsusele. Modelleerimiseks kasutatakse lineaarse regressiooni mudelit, mis on koostatud rakendustarkvaras R. Kui ilmastiku parameetrite olulisust hinnati kogu andmestiku põhjal, siis modelleerimisel kasutatakse libisevat treeningrida, mis tähendab seda, et iga prognoosi koostamiseks võetakse ajaloolistest andmetest aluseks kindla pikkusega andmerida, et tõsta esile ilmastiku parameetrite olulisuse sesoonsust ning kasutada igal ajahetkel vastava aktuaalsuse, osakaalu ja väärtusega erinevaid parameetreid.

Gaasitarbimise modelleerimisel on prognoosimiseks kasutatud tegelike ilmaandmeid, et eemaldada mudelist ilmaprognooside määramatus ja viga ning keskenduda vaid mudeli ja parameetrite väärtuse täpsusele.

4.1 Lineaarsel regressioonil põhinev prognoosimudel

Gaasitarbimise lineaarsel regressioonil põhinev prognoosimudel on koostatud rakendustarkvaras R ning kasutatava prognoosimudeli üldkuju on järgnev [24]:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in} \quad (4.1)$$

Kus

Y_i – sõltuv tunnus e funktsioontunnus;

β_0 – prognoosimudeli vabaliige;

$\beta_{1...n}$ – regressioonkordaja;

X_i – Sõltumatu tunnus ehk prediktor

Gaasitarbimise prognoosimiseks regressioonmudeliga kasutatakse rakendustarkvara R ning lõplik koostatud mudel on leitav käesoleva lõputöö lisades. Käesoleva lõputöö raames koostati mudel kahes etapis. Esimeses etapis sooritatakse kindlaksmääratud treeningrea põhjal regressioonanalüüs ning seejärel koostatakse selle analüüsi põhjal regressioonmudel ning iga prognoosi koostamise järel liiguvad treeningrea indeksid ühe päeva võrra edasi ning sooritatakse uus regressioonanalüüs ning koostatakse selle põhjal ka uus regressioonmudel, et imiteerida

libisevat treeningrida võimalikult lihtsalt ja kiirelt. R-i baaspaketis oleva regressioonanalüüsi süntaks on järgnev [25]:

$$lm(Y \sim mudel)$$

Kus

Y – prognoositav väärtus või sõltuv muutuja;

$mudel$ – matemaatilise mudeli avaldis;

$lm()$ – käsurida arvutamaks lineaarse regressiooni koefitsente;

Rakendustarkvara R baaspaketis sisalduv funktsioon $lm()$ võimaldab sooritada määratud treeningreale regressioonanalüüsi ning funktsiooni tehnilisem üldkuju ja olulisemate argumentide vaikimisi väärtused on järgnevad [26]:

```
lm(formula, data, subset, weights, na.action,...)
```

Kus

$formula$ – määrab mudeli sõltuva ja sõltumatud muutujad, mille raames sooritatakse andmestikule regressioonanalüüs;

$data$ – määrab andmestiku, mille põhjal sooritatakse eelnevalt defineeritud mudeli raames regressioonanalüüs;

$subset$ – võimaldab määrata kindla vahemiku andmestikus, mida hakatakse uurima;

$weights$ – võimaldab määrata sõltumatutele muutujatele erinevad kaalud analüüsi sooritamiseks;

$na.action$ – määrab kuidas arvestab mudel andmestikus olevaid „NA“ väärtuseid.

Eelnevalt kirjeldatud funktsiooni abil koostatud regressioonanalüüsi põhjal koostati rakendustarkvaras R funktsiooni $predict()$ abil lineaarsel regressioonil põhinev prognoosimudel. Nimetatud funktsioon võimaldab käesolevas uurimuses koostada iga uue päeva gaasiproгноosi koostamiseks ja funktsiooni üldkuju ning olulisemad argumendid on järgnevad [27]:

```
predict(object, newdata,  
        interval = c("none", "confidence", "prediction"),  
        level = 0.95, ...)
```

Kus

$object$ – funktsiooni $lm()$, mille põhjal koostatakse prognoos ;

`newdata` – prognoositava perioodi kohta teadaolevad andmed, mis on tarvis vastavusse viia regressioonanalüüsi parameetritega;

`interval` – määrab prognoosimisel kasutatava intervalli tüübi;

`level` – prognoosimisel kasutatav usaldusvahemik.

Prognoosimisel saadud tulemuste täpsust mõõdetakse MAPE (*mean absolute percent error*) abil, sest tegemist on kõige lihtsama meetodiga võrdlemaks sõltuva ja sõltumatute tunnuste vahelist täpsust. MAPE arvutusvalem on järgnev [24]:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - Y}{Y_i} \right| \cdot 100\% \quad (4.2)$$

Kus

Y_i – tegelik väärtus, mida prognoositakse

Y – prognoositud väärtus, mille täpsust hinnatakse

MAPE kujutab endast keskmist absoluutvea protsenti. Mida madalam on MAPE, seda väiksem on prognoosi absoluutviga ehk koostatud prognoos on täpsem. Kuigi MAPE on asümmeetriline näidik, mis teoreetiliselt lähenedes kujutab tegelikust väärtusest suuremalt prognoositud väärtusi kõrgema veaprotsendiga ning prognoositavast väärtusest väiksemaid väiksema veaprotsendiga, siis käesolevas töös ei hinnatud seda oluliseks, sest modelleerimisel saavutatav keskmine veaprotsent on eeldatavasti ligikaudu 10 protsenti. MAPE annab universaalse näidikuna ka võimaluse võrrelda mudeli erinevate iteratsioonide väljundit omavahel ning oma laia kasutuse tõttu välismaailmas on võimalik tuua paralleele ka teiste sooritatud modelleerimistega. Lisaks aitab MAPE hinnata ka mudeli täpsuse trendi liikumist modelleerimisel sooritatud iteratsioonides.

4.2 Modelleerimisel sooritatavad iteratsioonid

Käesolevas alampeatükis antakse ülevaade prognoosimudeliga läbi viidud iteratsioonidest. Prognoosimudeli iteratsioonid põhinevad kolmandas peatükis läbi viidud analüüside iteratsioonidel, kus igas iteratsioonis eemaldati andmestikust kõige vähem oluline ilmastikuparameeter ning sooritati uus analüüs väiksema mahuga andmestikuga. Regressioonmudel võtab arvesse ka kolme erinevat päevatüüpi, milleks on tööpäevad esmaspäevast reedeni, laupäevad ning pühapäevad ja riigipühad. Päevatüüpi arvestatakse regressioonmudelis faktortunnusena, mis omistab igale erinevale tüübile oma kindla regressioonkordaja igaks prognoosiks kasutusel oleva treeningrea alusel.

Käesoleva töö raames on eeldatud, et mudeli iga iteratsioon on täpsem kui sellele eelnenud iteratsioon. Igal iteratsioonil kasutati kolme erineva pikkusega treeningrida, milleks olid 15, 30 ja 45 päeva, et hinnata ka erinevate parameetrite sesoonsusest tulenevat mõju. Lisaks kasutati igas iteratsioonis ka eelnevatele treeningrea variantidele treeningrea lisa prognoositavale päevale eelneva aasta perioodist 15 või 30 päeva. Näiteks aastaegade vahetudes esinevad gaasitarbimises suured muutused, mis on tingitud peamiselt küttevajaduse järsust tõusust või langusest olenevalt aastaajast. Sellistel juhtudel ei pruugi pikk treeningrida tagada parimat mudeli täpsust olenemata sellest, et analüüsitavate andmete hulk on suurem. Sellises olukorras on oht, et pikk ja stabiilne ajalooline gaasitarbimine omab prognoosimudelis liiga suurt osakaalu ning mudel ei reageeri ilmastikumuutustele seetõttu piisavalt kiirelt ning prognoosib statistika seisukohalt kõige tõenäolisemat tarbimismahtu ka järgneva perioodiks. Sellest tulenevalt on igas iteratsioonis lisatud ka stsenaariumid, kus mudeli treeningrea mingi osa on prognoositavale päevale eelnenud aastast võetud. Peale igat prognoosimudeli iteratsiooni sooritatakse prognoosimudeli väljundite põhjal ka tulemuste analüüs. Põhiliseks mõõdikuks prognoosimudeli tulemuste hindamisel on MAPE. Prognoosimudeli täpsust hinnatakse kõikide prognoositud päevade keskmise MAPE-ga ning lisaks võetakse arvesse ka suurimat ning vähimat viga, mis erinevate iteratsioonide prognoosimisel esinesid.

Prognoosimudeli esimese iteratsiooni andmestik sisaldab endas kõiki käesolevas töös kasutatud andmeid ning esimese iteratsiooni näol on tegemist ka võrdlusaluse ja suunisega järgnevate iteratsioonide jaoks. Esimeses prognoosimudeli iteratsioonis kasutatakse prognoosimudelis 15, 30 ja 45 päeva pikkuseid treeningridu. See tähendab, et prognoositava päeva kohta prognoosi kasutades võetakse arvesse vastava arvu päevade ilma- ja tarbimisandmed. Lisaks esialgsele treeningreale lisatakse mudeli sisendandmetesse ka 15 ja 30 päeva pikkune lisatreeningrida prognoositavale päevale eelnenud aastatagusest perioodist. Lisatreeningrea alguspunkt on täpselt 365 päeva prognoositavast päevast tagasi ning vastavalt lisatreeningrea pikkusele on lisatreeningrea lõpp-punkt 15 või 30 päeva edasi. Esimeses iteratsioonis kasutatavad ilmaandmed on järgnevad:

- Temperatuur
- Sademed
- Tuule kiirus
- Tuule suund
- Nähtavus kilomeetrites

Peale prognoosimudeli esimese iteratsiooni sooritamist sooritatakse ka tulemuste analüüs, mis on välja toodud tulemuste analüüsi alampeatükis. Prognoosimudelite iteratsioone võrreldakse

põhiliselt iga prognoositud päeva kohta arvatud keskmise MAPE-ga. Samuti tuuakse välja iga iteratsiooni suurimad ja väikseimad vead.

Prognoosimudeli teises iteratsioonis eemaldatakse mudeli sisendandmetest tuule suuna andmed ning modelleeritakse prognoosimist 15, 30 ja 45 päeva pikkuse treeningreaga. Sarnaselt esimesele iteratsioonile kasutatakse iga treeningrea korral ka kolme erinevat lisatreeningrea varianti, mis on 0, 15 ja 30 päeva pikkused, et kompenseerida ilmastikuparameetrite hooajalist mõju muutumist gaasi tarbimisele. Prognoosimudeli teise iteratsiooni sisendandmete koosseis on järgnev:

- Temperatuur
- Sademed
- Tuule kiirus
- Nähtavus kilomeetrites

Prognoosimudeli kolmandas iteratsioonis eemaldatakse mudelist järjekordselt kõige vähem oluline parameeter, milleks antud juhul on sademete hulk ööpäeva jooksul. Kolmandas prognoosimudeli iteratsioonis kasutatakse mudelis 15, 30 ja 45 päeva pikkust treeningrida ning lisaks ka kolme erinevat lisatreeningrea varianti, milleks on 0, 15 ja 30 päeva pikkused lisatreeningread. Kolmanda prognoosimudeli iteratsiooni sisendandmete koosseis on järgnev:

- Temperatuur
- Tuule kiirus
- Nähtavus kilomeetrites

Prognoosimudeli neljandas iteratsioonis eemaldatakse eelnenud regressioonanalüüsi tulemusel järjekordne kõige vähem oluline ilmastiku parameeter, milleks oli tuule kiirus. Neljandas iteratsioonis kasutatakse prognoosimudeli 15, 30 ja 45 päeva pikkust treeningrida ning 0, 15 ja 30 päeva pikkuseid lisatreeningrea variante. Neljanda prognoosimudeli iteratsiooni sisendandmete koosseis lisaks ajaloolisele gaasitarbimisele oli temperatuur ja nähtavus kilomeetrites. Temperatuuri ja nähtavuse näol on tegemist kahe statistiliselt kõige olulisema parameetriga, mis mõjutavad gaasitarbimist.

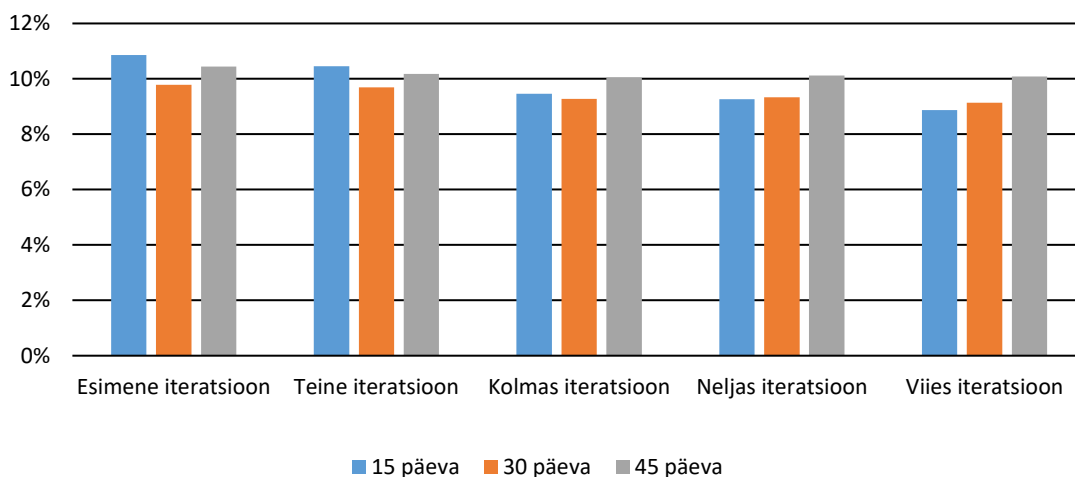
Regressioonanalüüsi tulemusel selgus, et statistiliselt kõige olulisem ilmastiku parameeter gaasitarbimise suhtes on temperatuur. Sellest tulenevalt on prognoosimudeli viiendas iteratsioonis võetud arvesse vaid ööpäeva keskmise temperatuuri mõju gaasitarbimisele. Kasutades vaid ühte parameetrit on võimalik luua esimesele iteratsioonile, milles kasutati kõiki saadavalolevaid ilmaparameetreid, vastanduv iteratsioon. Selle põhjal on võimalik hinnata lisaks optimaalsele treeningrea pikkusele ka kõige optimaalsemat prognoosimudeli koosseisu.

4.3 Tulemuste analüüs

Käesolevas alampeatükis antakse ülevaade prognoosimudeli tulemustest eelnevalt leitud ilmastikuolude parameetrite ja prognoosimudeliga sooritatud iteratsioonide põhjal. Prognoosimudeliga viidi läbi antud uurimuse käigus viis erinevat iteratsiooni erinevate ilmastikuparameetrite hulgaga. Iteratsioonide erinevus seisneb selles, et peale igat iteratsiooni eemaldati mudelist kõige vähem olulisem ilmastikuparameeter, et uurida kas mudeli täpsuse seisukohalt on keerulisem mudeli sisend pigem hea või halb. Igal iteratsioonil kasutati kolme erineva pikkusega treeningrida ja lisaks ka eelnevatele treeningrea variantidele lisatreeningreaga, et prognoosida täpsemalt aastaegade vahetudes esinevaid muutuseid gaasitarbimises, mis on tingitud peamiselt küttevajaduse järsust tõusust või langusest olenevalt aastaajast ning sellistel perioodidel ei pruugi pikk treeningrida tagada parimat mudeli täpsust olenemata sellest, et analüüsitava andmete hulk on suurem. Sellises olukorras on oht, et pikk ja stabiilne ajalooline gaasitarbimine omab prognoosimudelis liiga suurt osakaalu ning mudel ei reageeri ilmastikumuutustele seetõttu piisavalt kiirelt ning prognoosib statistika seisukohalt kõige tõenäolisemat tarbimismahtu ka järgnevat perioodiks. Sellest tulenevalt on igas iteratsioonis lisatud ka stsenaariumid, kus mudeli treeningrea mingi osa on prognoositavale päevale eelnenud aastast võetud. Gaasitarbimist prognoositi vahemikus 2016. aasta märtsikuu kuni 2019. aasta veebruarikuu tulenevalt sellest, et antud ajavahemikul olid kõik uuritavad andmed täielikult olemas ning avalikult ligipääsetavad.

4.3.1 Esialgsed prognoosimudeli tulemused

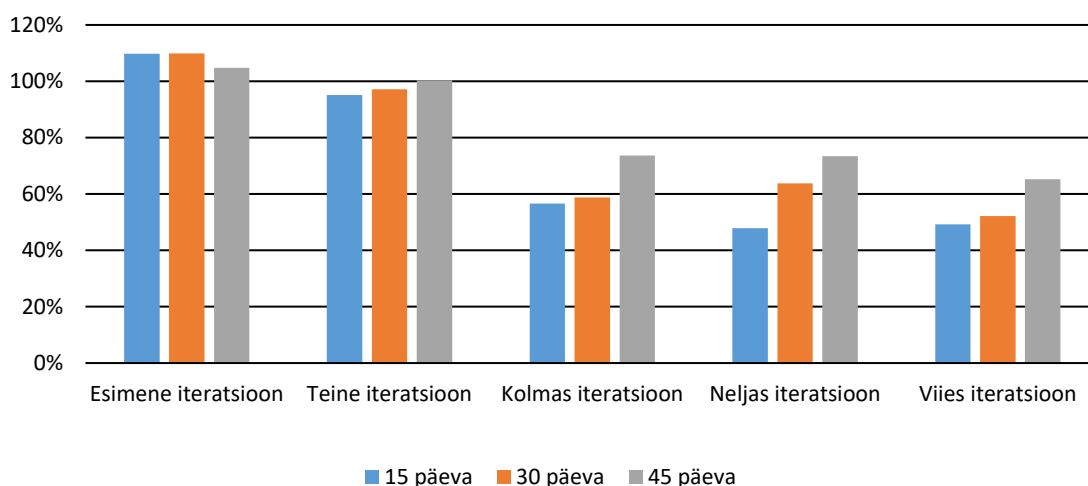
Esmalt prognoositi gaasitarbimist kasutades vaid prognoositavale päevale vahetult eelnenud andmeid vastavalt kasutatava treeningrea pikkusele. Joonisel 4.1 on toodud prognoosimudeli iteratsioonide MAPE-d erinevate iteratsioonide ja treeningrea pikkuste korral. All toodud jooniselt võib näha, et prognoosimudelite täpsus liikus eeldatud suunas ehk iga iteratsiooniga paranes pidevalt prognoosimudeli täpsus ning kõige parem tulemus saavutati viiendas iteratsioonis 15 päeva pikkuse treeningreaga kasutades sisendandmetena vaid ajaloolist gaasitarbimist ning ööpäeva keskmist temperatuuri.



Joonis 4.1. Prognoosimudeli iteratsioonide keskmised veaprotsendid.

Kõige enam paranes prognoosimudeli täpsus iteratsioonide lõikes 15 päeva pikkuse treeningreaga mudelil, mille esimese iteratsiooni täpsus oli 10,86 protsenti ning viienda iteratsiooni täpsus oli 8,87 protsenti ehk sõltuvalt sisendandmete hulgast oli sellise treeningrea puhul täpsuse paranemine kogu modelleerimise käigus 1,99 protsenti. Gaasi mahtudes väljendatuna tähendaks see 2018. aasta kogu Eesti gaasitarbimise põhjal vastavalt esimese iteratsiooni eabilansi mahuks keskmiselt umbes 564,7 gigavatt-tundi aastas ning viienda iteratsiooni eabilansi mahuks keskmiselt umbes 461,2 gigavatt-tundi aastas [6]. Kõige vähem paranes aga 45 päeva pikkuse treeningreaga mudeli täpsus, mille esimese iteratsiooni keskmine absoluutvea protsent oli 10,44 ning parim täpsus esines sellise pikkusega treeningrea puhul kolmandas iteratsioonis, kui prognoosimudel saavutas MAPE 10,04 protsenti ning mudeli täpsus paranes kõige parema ja halvema täpsuse võrdluses 0,40 protsenti. 30 päeva pikkuse treeningreaga mudel oli kolme esimese iteratsiooni jooksul kõige täpsem ning alates neljandast iteratsioonist muutus 15 päeva pikkuse treeningreaga mudel täpsemaks. 30 päeva pikkuse treeningreaga mudeli täpsus oli esimeses iteratsioonis 9,79 protsenti ning parima täpsuse saavutas see mudel viiendas iteratsioonis ning selleks oli 9,14 protsenti ehk mudeli täpsus paranes iteratsioonide käigus 0,65 protsenti. Prognoosimudeli tulemusi vaadeldes võib täheldada, et pikema treeningreaga mudel paraneb vähem ja aeglasemini ning lühema treeningreaga mudeli täpsus paraneb iteratsioonide lõikes oluliselt kiiremini.

Joonisel 4.2 on toodud välja prognoosimudelite suurimad vead ühel prognoositaval päeval iteratsioonide lõikes. Joonisel 4.2 võib märgata mõnevõrra sarnaseid trende nagu joonisel 4.1 ehk üldiselt muutub prognoosimudeli suurim viga aina väiksemaks.

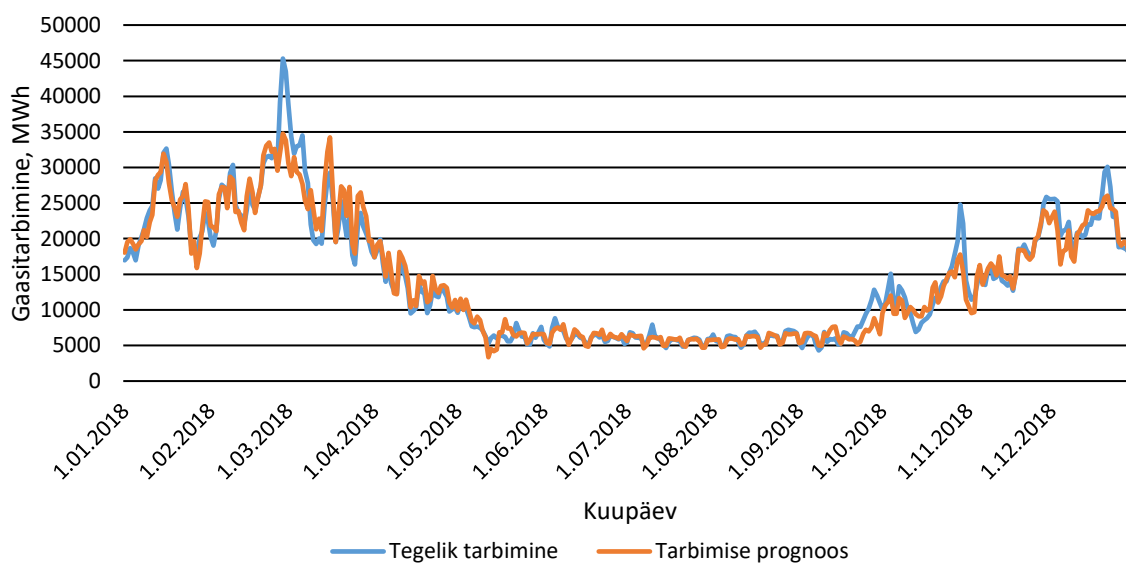


Joonis 4.2. Prognosimudeli suurimad veaprotsendid ühel päeval iteratsioonide lõikes.

Suurimat veaprotsenti vaadates võib näha olulist muutust peale teist iteratsiooni, millest võib järeldada, et tuule suunda ja sademete hulka arvesse võttes tekkivad mudelis kohati väga suured prognoosivead. Esimese kahe iteratsioonis suurimad vead on tekkinud üldjuhul soojema temperatuuriga perioodidel, kus üks parameetrist oma kogu treeningrea ulatuses tavapärasest väga palju väiksemat väärtust ning seejärel on prognoositaval päeval selle parameetri väärtus järsult kasvanud. Seega leiab kinnitust ka regressioonanalüüsi käigus tehtud eeldus, et prognoosimudeli esimesed kaks iteratsiooni on võrreldes ülejäänud iteratsioonidega ebatäpsemad tulenevalt sademete ja tuule suuna regressioonkordajate järskudest kõikumistest. Näiteks esimeses iteratsioonis tekkis suurim viga 15 päeva pikkuse treeningrea puhul 24. septembril aastal 2016. Sademete keskmine hulk ööpäevas eelnenud treeningrea jooksul oli 0,04 millimeetrit ning prognoositaval päeval oli sademete hulk 1,59 millimeetrit. Selles olukorras oli mudel omistanud sademete hulga parameetrile väga suure regressioonkordaja, sest regressioonanalüüsi käigus minimeeritakse andmepunktide vahel vigasid ja järelkult oli väikseima veaga regressioonfunktsioon selline, mille põhjal eeldati, et sademed kindlasti mõjutavad tarbimist, aga oma väikese arväärtuse tõttu pidi regressioonkordaja olema väga suur, et seda mõju esile tuua. Sellest tulenevalt tekkis ka sademete hulga järsu suhtelise muutuse korral suur prognoosiviga mudeli väljundis, millest saab järeldada ka seda, et regressioon mudelit kasutades peab treeningrea jooksul olevate andmete arväärtused varieeruma võimalikult palju. Sarnane olukord põhjustas ka 30 päeva pikkuse mudeli suurima vea, kui 7. juunil 2017. aastal oli treeningrea jooksul keskmine sademete hulk ööpäevas 0,76 millimeetrit ning prognoositaval päeval 7,05 millimeetrit. Samuti oli tuule suund mõlemal juhtumil pöördunud, mis tähendab, et ka sellele parameetrile omistati

tavapärasest suurel määral erinev regressioonkordaja. See õigustab ka mõlema parameetri eemaldamist treeningrea andmestikus esimeses kahes iteratsioonis, sest tegemist ei ole kasutatavate treeningridade jooksul järjepidevaid arväärtusi omavate parameetritega. Käsitledes suurimat veaprotsenti üldisemalt käesoleva analüüsi kontekstis saab öelda, et regressioonimudeli nõrgaks kohaks on üldjuhul väga stabiilsed parameetrid, mis prognoositaval perioodil omavad tavapärasest suurel määral erinevaid väärtuseid.

Joonisel 4.3 on toodud prognoosimudeli väljundi ja tegeliku gaasitarbimise võrdlus 2018. aasta jooksul. Joonisel on kujutatud esimese modelleerimise jooksul kõige täpsema väljundi saavutanud iteratsiooniga, milleks oli viies iteratsioon, mis kasutas sisendandmete ulatusena 15 päeva pikkust treeningrida. Antud iteratsioonis kasutati sisendandmetena vaid ööpäeva keskmist temperatuuri ja ajaloolist gaasitarbimist ning täpsuseks kogu modelleeritava perioodi jooksul saavutati MAPE 8,87 protsenti.

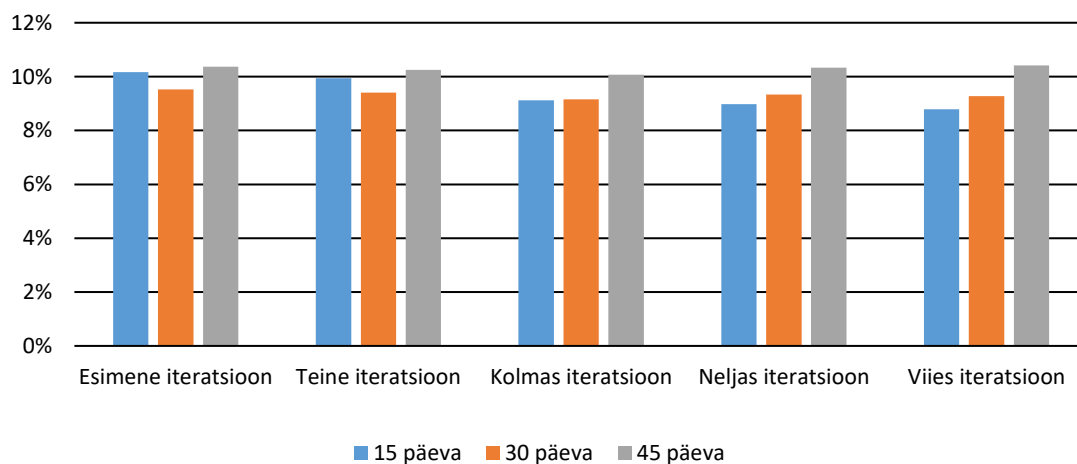


Joonis 4.3. Tegeliku gaasitarbimise ja prognoositud gaasitarbimise võrdlus 2018. aasta jooksul.

4.3.2 Prognoosimudeli tulemused 15 päeva pikkuse lisatreeningreaga

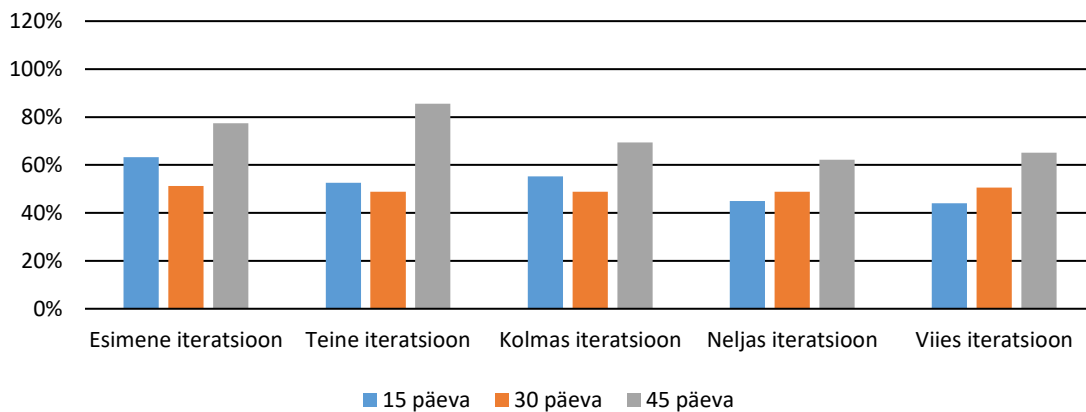
Modelleerimise käigus täheldati, et kevadel ja sügisel, kui algab või lõppeb kütteperiood, muutub prognoosimudel oluliselt ebatäpsemaks tulenevalt tarbimise baastaseme kiirest muutumisest. Sellest tulenevalt eeldati, et kui sisestada mudelile lisaks prognoositavale päevale vahetult eelnevale treeningreale ka lisatreeningrida prognoositavale päevale eelnenud aastatagusest perioodis, on võimalik gaasitarbimise üldise taseme muutust täpsemini prognoosida.

Lisatreeningrida võeti eelmisest aastast sellise loogikaga, et lisatreeningrea alguspunkt oli 365 päeva prognoositavast päevast tagasi ning lõpp-punkt olenes lisatreeningrea pikkusest, mis oli kas 15 või 30 päeva. Joonisel 4.4 on välja toodud prognoosimudeli iteratsioonide MAPE-d erinevate iteratsioonide ja treeningrea pikkuste korral. Nimetatud jooniselt võib näha veaprotsendi muutuse trendis mõnevõrra erinevaid tulemusi eelnenud metoodikast. Näiteks 45 päeva pikkuse treeningreaga mudel muutus esimesest viienda iteratsioonini kokkuvõttes vähem täpsemaks. 45 päeva pikkuse treeningreaga mudeli parim saavutatud täpsus oli kolmandas iteratsioonis 10,07 protsenti ning halvim täpsus oli viiendas iteratsioonis 10,43 protsenti. 30 päeva pikkuse treeningreaga mudeli parima täpsusega iteratsioon oli kolmas iteratsioon veaprotsendiga 9,16 protsenti. Ainsana näitas pidevat paranemist täpsuses 15 päeva pikkuse treeningreaga mudel, mille kõige vähem täpsem tulemus saavutati esimeses iteratsioonis 10,17 protsenti ning parim tulemus oli 8,79 protsenti ehk mudeli täpsus paranes iteratsioonide lõikes 1,38 protsenti. Võrreldes esialgsete iteratsioonidega paranes 15 päeva pikkuse mudeli täpsus 0,08 protsenti, kui võrrelda mõlema variandi parimat tulemust ehk viiendat iteratsiooni. Kõige ebatäpsema mudeli eabilansi maht kogu Eesti 2018. aasta gaasitarbimise alusel oleks keskmiselt 541,8 gigavatt-tundi aastas ning kõige täpsema mudeli puhul oleks eabilansi maht 457,1 gigavatt-tundi aastas [6].



Joonis 4.4. 15 päeva pikkuse lisatreeningreaga prognoosimudeli iteratsioonide keskmised veaprotsendid.

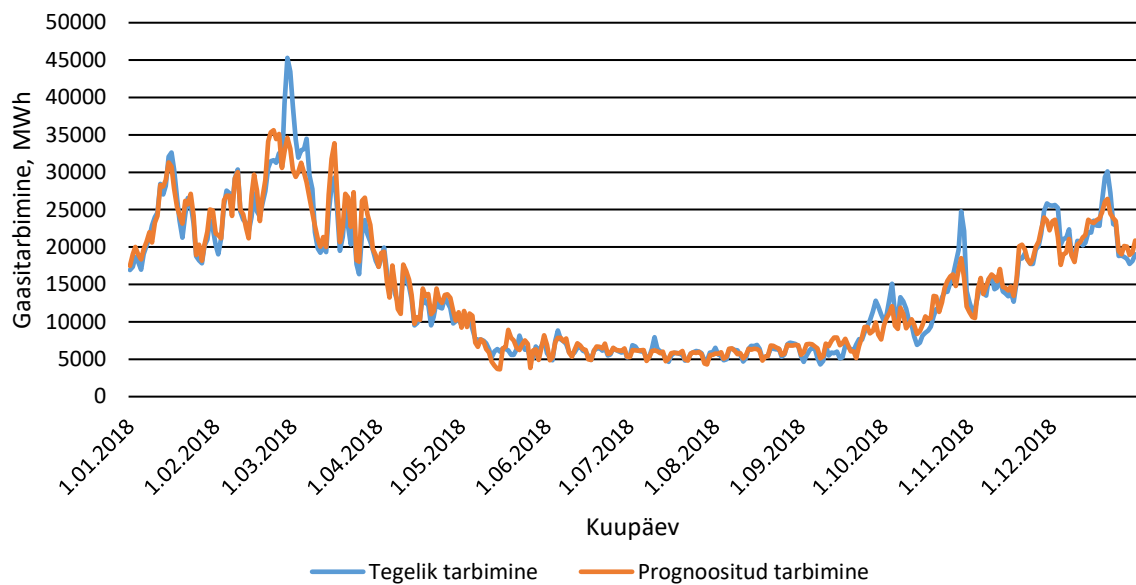
Lisatreeningrea kasutamisel saadud tulemustest võib järeldada, et mudeli reageerimiskiirus muutlikutele ilmastikuoludele paranes mõnevõrra, kuid uurima peab ka mudeli suurimaid vigu, et saada aru, kas lisatreeningrea kasutamine mitmekesistas sisendandmeid. Lisatreeningreaga mudelite iteratsioonide suurimad veaprotsendid on toodud joonisel 4.5. Kõikides modelleeritud iteratsioonides ja lisatreeningrea variantides oli vähim viga 0,00 protsenti.



Joonis 4.5. 15 päeva pikkuse lisatreeningreaga prognoosimudeli suurimad veaprotsendid.

Sarnaselt tavalise treeningreaga mudeli iteratsioonidel esinesid ka siin 45 päeva pikkuse treeningreaga mudeli puhul esimeses kahes iteratsioonis suured vead, mis tähendab seda, et järelkult sattus sellise treeningreaga puhul valimisse endiselt perioode, kus kõikumised tuule suuna ja sademete hulga regressioonkordajas põhjustasid üksikutel päevadel suuri prognoosivigaid. Kõige enam paranes esimese kahe iteratsiooni suurim veaprotsent, millel ilma lisatreeningreata juhtudel esinesid prognoosivead üle 100 protsendi. Esimeses iteratsioonis paranes suurim veaprotsent 27,45 kuni 58,60 protsenti ning teises iteratsioonis paranes suurim veaprotsent 48,39 kuni 14,86 protsenti võrreldes esialgse meetodikaga. Kolmanda kuni viienda iteratsiooni puhul oli veaprotsendi paranemine oluliselt väiksem ning üldjuhul jäi täpsuse paranemine alla kümne protsendi. Siiski on märgata, et lisatreeningreaga sobitamine mudeli sisendisse aitas oluliselt parandada tuule suuna ja sademete hulga järskudest muutustest põhjustatud prognoosivigu, sest lisatreeningreaga mitmekesistab oluliselt sisendandmete valimit ning oluliselt väiksema tõenäosusega tekib olukord, kus terve treeningreaga vältel on mõnel ilmaparameetril püsivalt oluliselt erinev arväärtus võrreldes prognoositava päevaga.

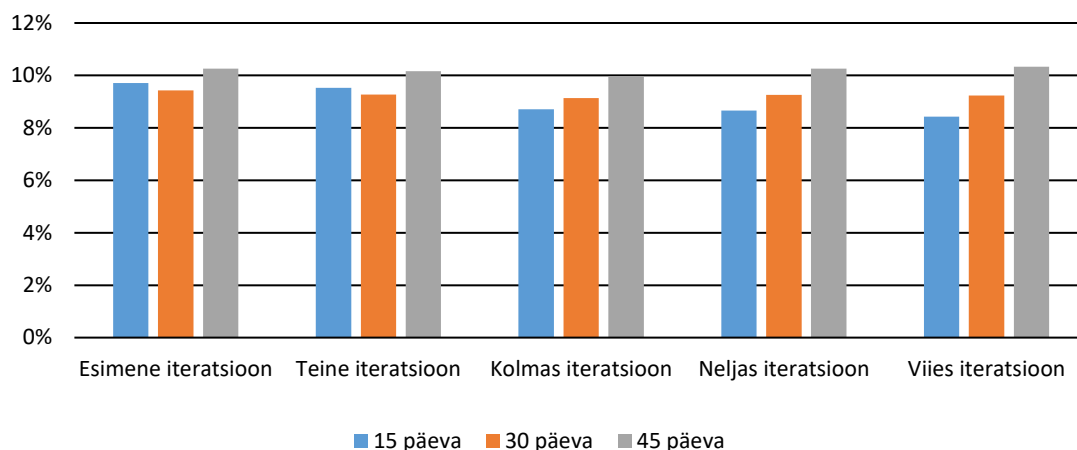
Joonisel 4.6 on toodud 15 päeva pikkuse lisatreeningreaga prognoosimudeli väljundi ja tegeliku gaasitarbimise võrdlus 2018. aasta jooksul. Joonisel on kujutatud 15 päeva pikkuse lisatreeningreaga modelleerimise jooksul kõige täpsema väljundi saavutanud iteratsiooniga, milleks oli viies iteratsioon. Antud iteratsioonis kasutati sisendandmetena vaid ööpäeva keskmist temperatuuri ja ajaloolist gaasitarbimist ning veaprotsendiks saavutati kogu prognoositud perioodi jooksul 8,79 protsenti.



Joonis 4.6. Tegeliku gaasitarbimise ja 15 päeva pikkuse lisatreeningreaga prognoosimudeliga prognoositud gaasitarbimise võrdlus 2018. aasta jooksul.

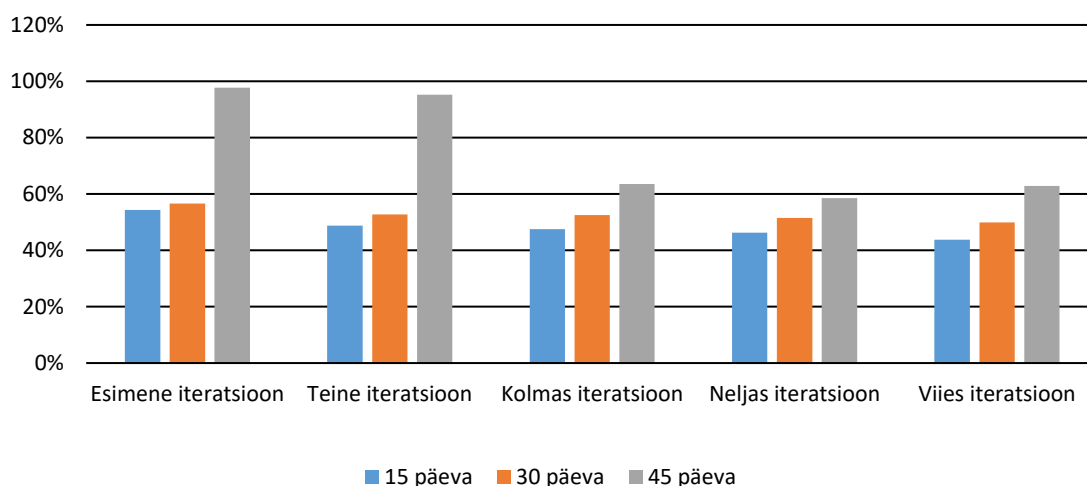
4.3.3 Prognoosimudeli tulemused 30 päeva pikkuse lisatreeningreaga

Prognoosimudelisse sobitati veel ühe variandina ka 30 päeva pikkune lisatreeningrida, mis töötab samal põhimõttel nagu eelnev lisatreeningrida, kuid lisatreeningrea pikkus on sel korral 30 päeva. Joonisel 4.7 on välja toodud prognoosimudeli iteratsioonide MAPE-d erinevate iteratsioonide ja treeningrea pikkuste korral 30 päeva pikkust lisatreeningrida kasutades. 30 päeva pikkust lisatreeningrida kasutades selgus ka kogu andmestiku peale kõige täpsem prognoosimudel, mis oli 15 päeva pikkuse treeningreaga ning 30 päeva pikkuse lisatreeningreaga prognoosimudeli viies iteratsioon, milles sisendandmetena kasutati vaid temperatuuri ja ajaloolist gaasitarbimist. Mudeli saavutatud täpsus kogu andmestiku jooksul oli 8,43 protsenti. Kõige ebatäpsema mudeli eabilansi maht kogu Eesti 2018. aasta gaasitarbimise alusel oleks keskmiselt 537,2 gigavatt-tundi aastas ning kõige täpsema mudeli puhul oleks eabilansi maht 438,4 gigavatt-tundi aastas [6].



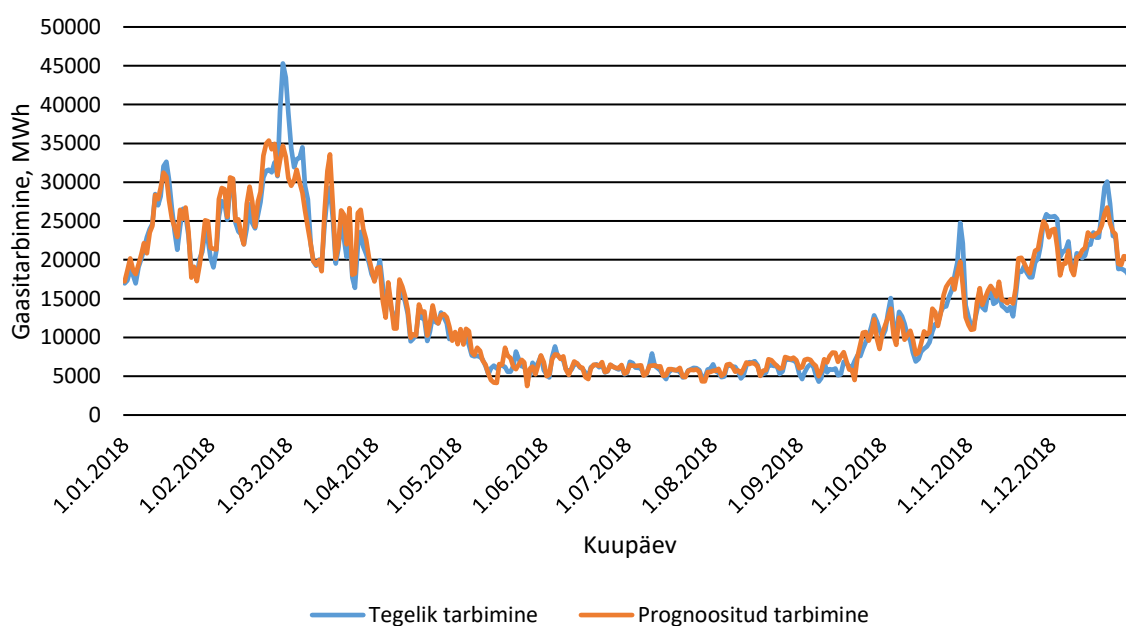
Joonis 4.7. 30 päeva pikkuse lisatreeningreaga prognoosimudeli iteratsioonide keskmised veaprotsendid.

30 päeva pikkuse lisatreeningreaga kasutamise korral esinesid suurimad muutused 15 päeva pikkuse treeningreaga mudeli variandi korral. Prognoosiviga paranes sellisel mudelil esimese ja viienda iteratsiooni jooksul 9,71 protsendilt 8,43 protsendini, mis tähendab, et mudeli veaprotsent paranes iteratsioonide käigus 1,28 protsenti. 30 päeva pikkuse treeningreaga mudel saavutas parima täpsuse kolmandal iteratsioonil ning veaprotsent oli selles iteratsioonis 9,13 protsenti, mis paranes võrreldes esimese iteratsiooniga 0,30 protsenti. Samuti saavutas 45 päeva pikkuse treeningreaga mudel parima täpsuse kolmandas iteratsioonis, kui prognoosimudel saavutas veaprotsendiks 9,97 protsenti. Joonisel 4.8 on toodud 30 päeva pikkuse lisatreeningreaga mudelite suurimad veaprotsendid. Võrreldes 15 päeva pikkuse treeningreaga ei ole sel korral suurimates vigades väga suuri muutusi, kui välja arvata 45 päeva pikkuse treeningreaga mudeli kaks esimest iteratsiooni. Põhiliseks prognoosivea põhjustajaks oli ka sel korral sademete hulga järsk muutus 21. juunil 2018. aastal, kui pika treeningreaga jooksul oli sademete keskmine hulk ööpäeva jooksul 0,39 millimeetrit ja prognoositaval päeval oli sademete hulk 14,73 millimeetrit. Mida pikem on treeningrida, seda enam on oluline, et igal parameetril oleks võimalikult varieeruv arvuline väärtus ning käesolev juhtum illustreerib veelkord seda, et pikk treeningrida annab küll palju põhjalikuma analüütilise võimekuse mudelile, kuid liiga stabiilsete parameetrite arväärtuste korral ei ole mudel võimeline reageerima äkilistele muutustele ilmastikus, mis põhjustab suuri prognoosivigu.



Joonis 4.8. 30 päeva pikkuse lisatreeningreaga prognoosimudeli suurimad veaprotsendid.

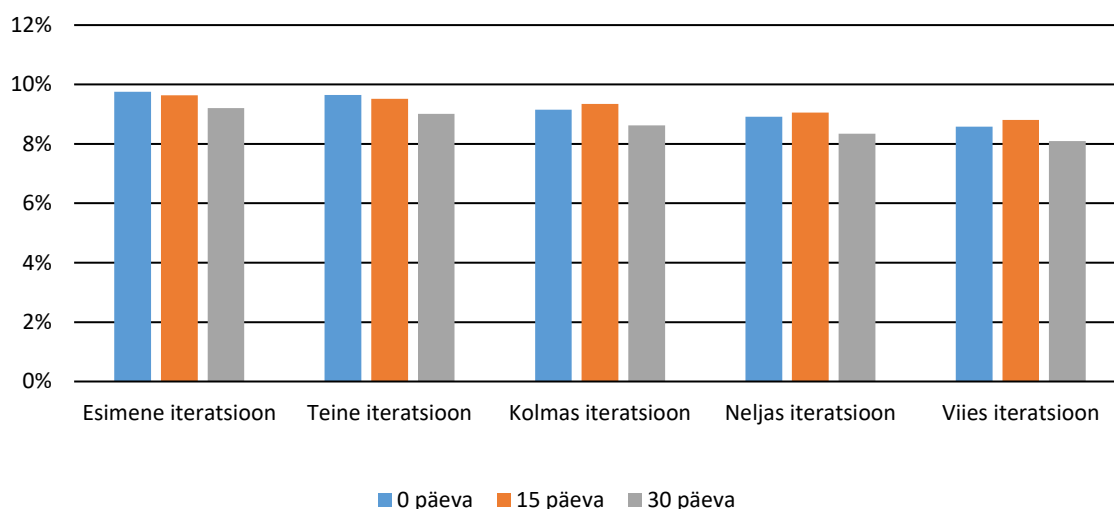
Joonisel 4.9 on toodud 30 päeva pikkuse lisatreeningreaga prognoosimudeli väljundi ja tegeliku gaasitarbimise võrdlus 2018. aasta jooksul. Joonisel on kujutatud 30 päeva pikkuse lisatreeningreaga modelleerimise jooksul kõige täpsema väljundi saavutanud iteratsiooniga, milleks oli viies iteratsioon, mis kasutas 15 päeva pikkust treeningrida. Antud iteratsioonis kasutati sisendandmetena vaid ööpäeva keskmist temperatuuri ja ajaloolist gaasitarbimist ning veaprotsendiks saavutati kogu prognoositud perioodi jooksul 8,43 protsenti.



Joonis 4.9. Tegelik gaasitarbimise ja 30 päeva pikkuse lisatreeningreaga prognoosimudeliga prognoositud gaasitarbimise võrdlus 2018. aasta jooksul.

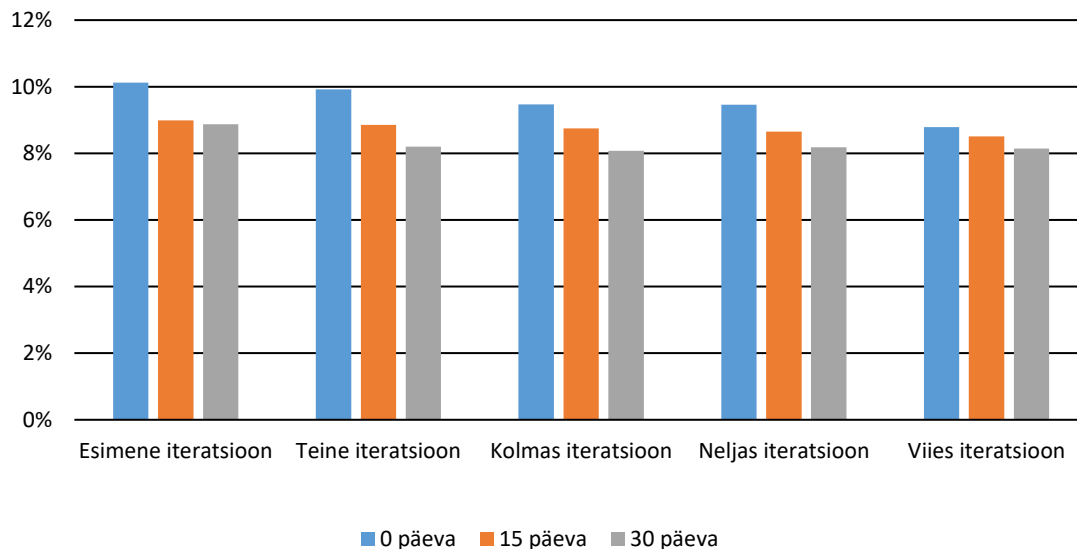
4.3.4 Lisatreeningrea mõju prognoosimudeli tulemustele

Lisatreeningrea kasutamisel on oluline aspekt see, et mudeli esimese aasta prognoosid ei saanud võtta arvesse eelnenud aastat, ning sellest tulenevalt on vajalik vaadelda ka aastate põhiskeskmi veaprotsenti erinevate mudeli variantide puhul, kus vaadeldakse ühe täisaasta jooksul esinenud keskmi prognoosiviga. Joonis 4.9 illustreerib lisatreeningreaga parima mudeli prognoosi täpsust 2017. aasta jooksul ning joonis 4.10 illustreerib parimat lisatreeningreaga mudeli prognoosi täpsust 2018. aastal. Iga iteratsiooni puhul on välja toodud parimad aasta keskmised veaprotsendid vastavalt lisatreeningrea pikkusele. Esialgne modelleerimine, kus lisatreeningrida ei kasutatud, on järgnevatel joonistel märgitud 0 päeva pikkuse lisatreeningreaga.



Joonis 4.9. 2017. aasta prognoosimudelite täpsused iteratsioonide ja lisatreeningrea pikkuste kaupa.

Jooniselt 4.9 selgub, et 2017. aastal oli kõige täpsem prognoosimudel selline, mis kasutas 30 päeva pikkust lisatreeningrida 15 päeva pikkust treeningrida ning sisendandmetena kasutati vaid ööpäeva keskmist temperatuuri ning ajaloolist gaasitarbimist. Nimetatud mudel saavutas keskmiseks veaprotsendiks 8,09 protsenti. Erinevate lisatreeninggridadega mudelite täpsus oli viiendas iteratsioonis keskmiselt 8,49 protsenti, mis märgib keskmiselt 1,12 protsendilist paranemist võrreldes esimese iteratsiooni keskmise prognoosiveaga, milleks oli 9,21 protsenti. Keskmised veaprotsendid paranesid iga iteratsiooniga. Joonisel 4.10 on toodud välja kõige paremad prognoosimudeli veaprotsendid iteratsioonide ja lisatreeningrea pikkuste kaupa 2018. aasta kohta.

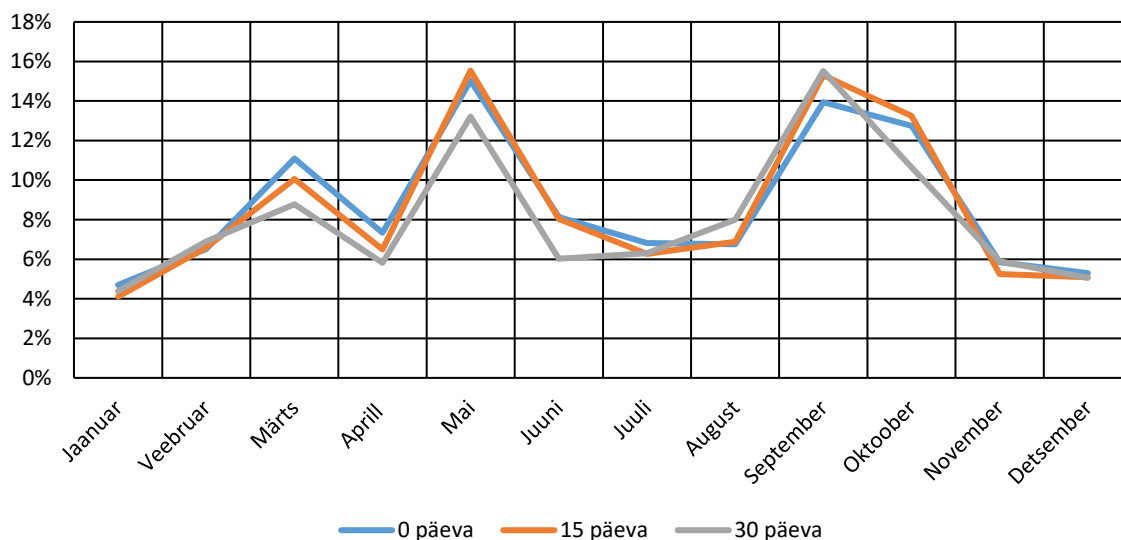


Joonis 4.10. 2018. aasta prognoosimudelite täpsused iteratsioonide ja lisatreeningrea pikkuste kaupa.

Joonisel 4.10 on märgata sarnaseid trende nagu joonisel 4.9. Vaadeldes 2018. aasta keskmiseid veaprotsente on lisatreeningrea kasutamine parandanud prognoosiviga märgatavalt. Esimesel iteratsioonil paranes prognoosiviga 1,25 protsenti, teisel iteratsioonil paranes prognoosiviga 1,72 protsenti, kolmandal iteratsioonil paranes prognoosiviga 1,40 protsenti, neljandal iteratsioonil paranes prognoosiviga 1,28 protsenti ning viiendal iteratsioonil paranes prognoosiviga 0,65 protsenti. 2018. aastal saavutas parima täpsuse prognoosimudel kolmandas iteratsioonis ning prognoosiviga oli 8,08 protsenti. Tegemist oli 15 päeva pikkuse treeningreaga ning 30 päeva pikkuse lisatreeningreaga mudeliga, mis kasutas sisendandmetena tuule kiirust, nähtavust, temperatuuri ja ajaloolist gaasitarbimist.

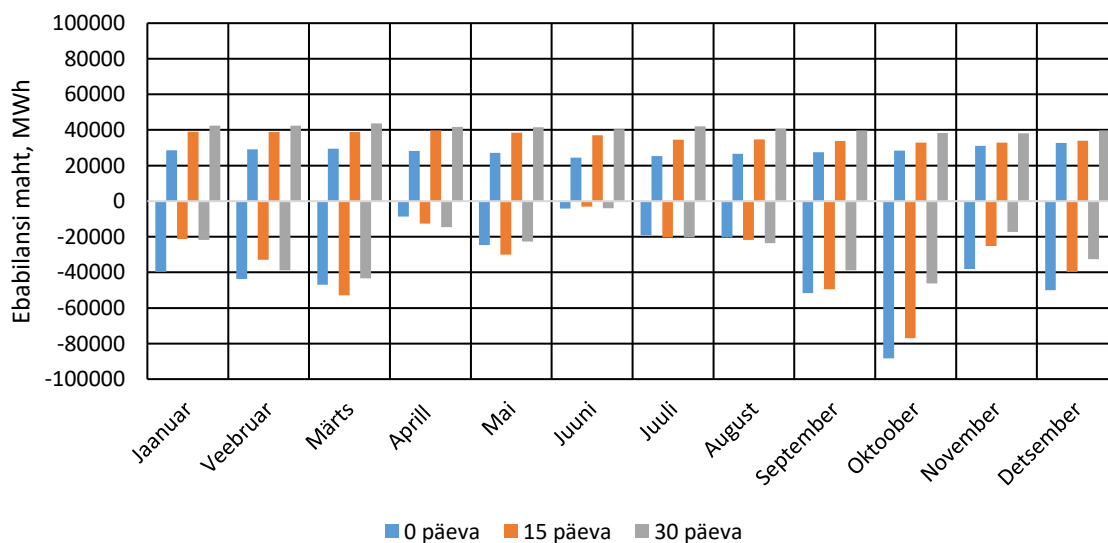
Lisatreeningrida võeti kasutusele tulenevalt sellest, et modelleerimise käigus täheldati, et kevadel ja sügisel, kui algab või lõppeb kütteperiood, muutub prognoosimudel oluliselt ebatäpsemaks tulenevalt tarbimise baastaseme kiirest muutumisest. Sellest tulenevalt eeldati, et kui sisestada mudelile lisaks prognoositavale päevale vahetult eelnevale treeningreale ka lisatreeningrida prognoositavale päevale eelnenud aastatagusest perioodis, on võimalik gaasitarbimise üldise taseme muutust täpsemini prognoosida. Joonisel 4.11 on kujutatud prognoosimudeli iga kalendrikuu jooksul saavutatud keskmine viga kogu prognoositud perioodi vältel erinevate lisatreeningridade korral. Nimetatud joonis näitab, et mudeli veaprotsent on märgatavalt suurem mai- ja septembrikuus ehk perioodidel, kui tarbimine kasvab või langeb kiirelt. Samas suvisel ja talvisel perioodil stabiilse tarbimise korral saavutab mudel võrdlemisi hea täpsuse. Jooniselt on näha, et suurem osa täpsuse paranemisest lisatreeningrea lisamisel tuleneb üldjuhul esimesest

poolaastast. Samas on aga näha ka seda, et sügise tarbimise tõusuga on lisatreeningrea lisamine mudelisse mudeli väljundi hoopiski ebatäpsemaks muutnud.



Joonis 4.11. Prognoosimudeli kalendrikuu keskmised vead erinevate lisatreeningrea variantide korral.

Prognoosimudeli ebatäpsemaks muutumine võib viidata sellele, et lisatreeningrida kasutades on prognoosimudeli kitsaskohaks erinevatel hetkedel saabuval jahedama ilmal. Joonisel 4.12 on toodud ka iga kuu summeeritud eabilansi mahud kuude lõikes kogu prognoositud perioodi jooksul erinevate lisatreeningrea variantide kaupa. Positiivne suurus sellel joonisel näitab, et prognoositud on kõrgemat tarbimist, kui oli tegelik tarbimine ning negatiivne suurus tähendab, et prognoositud tarbimine oli väiksem kui tegelik tarbimine.



Joonis 4.12. Prognoosimudeli vigade mahud kogu prognoositud perioodi jooksul kuude lõikes.

Jooniselt 4.12 on üpriski selgelt näha, et prognoosimudeli viga liigub lisatreeningrea pikenemise korral pigem selliselt, et prognoosimudeli väljund kipub pigem olema suurem tegelikust tarbimisest. See võib tuleneda sellest, et võttes arvesse lisatreeningrida kasutades prognoositavale päevale eelnenud aasta andmeid tõuseb prognoositav tarbimine, sest eelnenud aasta tarbimine oli kõrgem prognoositava perioodi tarbimisest. Arvesse tuleb võtta ka seda, et Eleringi hinnangul liigub gaasi tarbimine aastast aastasse langevas trendis tulenevalt tarbimisstruktuuri muutusest Eestis [5]. Kui gaasitarbimine liigub langustrendis, siis võttes prognoosimudelis arvesse andmeid prognoositavale perioodile eelnevast aastast võib mudeli sisendandmetes olla arvestatud ka prognoositavast perioodist kõrgemat tarbimist. Lisatreeningrida mitmekesisus mudeli sisendit selliselt, et see arvestab prognoositavale päevale järgnevat perioodi eelnevast aastast, mis mingil määral kompenseerib ilmategurite numbrilise mõju muutumist ajas, kuid see eeldab ka seda, et kütteperioodi lõpp ja algus kattuksid omavahel võimalikult täpselt aastast aastasse. Paraku aga ei pruugi ilmad muutuda jahedamaks või soojemaks igal aastal samal ajal, mis tähendab, et ilmastikutegurite mõju muutumisele ajas on võrdlemisi raske reageerida mudeliga, mis prognoosib tarbimist kogu prognoositava perioodi vältel täpselt samasuguse loogikaga. Hoolimata mõne üksiku kuu täpsuse halvenemisest on pikaajaline keskmine veaprotsent lisatreeningrida kasutades siiski madalam kui ilma lisatreeningreata mudelil.

4.4 Tulemuste kokkuvõte ja järeldus

Kokkuvõttes saab öelda, et parim pikaajaline keskmine täpsus saavutati prognoosimudeli viiendas iteratsioonis 15 päeva pikkuse treeningreaga ning 30 päeva pikkuse lisatreeningreaga. Antud mudel kasutas sisendandmetena vaid ööpäeva keskmist temperatuuri ning ajaloolist gaasitarbimist. Hoolimata keerukatest analüüsides ja sisendandmete mitmekesisusest saavutas siiski kõige täpsema prognoosi kõige lihtsam prognoosimudel. Nimetatud mudel oli üle kõikide iteratsioonide ja lisatreeningrea variantidega pikaajaliselt kõige täpsem prognoosimudel. Mitmekesine andmestik peaks teoreetiliselt olema väga täpne, sest seeläbi peaks olema võimalik statistiliselt määratleda temperatuuri ja gaasitarbimise sõltuvuses määramatus, mis Eestis võib sarnaste ööpäevaste keskmiste temperatuuride korral varieeruda kuni 10 gigavatt-tundi nagu selgub jooniselt 3.1 ja 3.2. Samas aga eeldab täpne regressioonanalüüs ja -mudel, et analüüsimise käigus arvesse võetavad parameetrid oleks võimalikult varieeruvad ning gaasitarbimine järgiks ideaalilähedaselt eeldatavaid korrelatsioone ilmaparameetrite ja gaasitarbimise vahel. Reaalselt koostatud analüüs ja mudel aga näitab, et praktikas on selliseid sisendandmeid paraku võimatu tekitada, sest ilmastikuolud võivad mudelis kasutatava treeningrea ja prognoositava päeva vahel oluliselt erineda ning see põhjustab prognoosimisel ebatäpseid prognoose. Modelleerimisel selgus ka see, et kasutades pikemat

treeningrida ehk võimalikult suurt hulka sisendandmeid, omistab mudel erinevatele parameetritele sellised regressioonkordajad, mis muutuvad ajas aeglaselt, sest sisendandmete hulk on väga suur ning iga päev ühe päeva jagu lisandunud sisendandmed omavad regressioonkordajate suhtes väikest mõju. Sellest tulenevalt kaotab mudel võime ilmastikum muutustele kiirelt reageerida. Liiga lühikese treeningrea korral reageerib aga mudel mõnevõrra väiksematele muutustele ilmastikuparameetrites prognoositaval päeval liiga kiirelt ning seeläbi saavutatakse jällegi ebatäpne prognoosimudeli väljund. Sellepärast ongi tarvilik välja selgitada prognoosimudeli treeningrea optimaalne pikkus ning sisendandmete hulk, et tagada pikaajaliselt keskmiselt täpsem prognoosimudeli väljund.

Kasutades modelleerimise esimestes iteratsioonides kõiki võimalike parameetreid prognoosimudeli sisendandmeteks saavutati üldjuhul parim tulemus keskmise pikkusega treeningreaga mudeliga. Samas aga edasistes iteratsioonides paranes oluliselt enam kõige lühema treeningreaga mudel ning 30 ja 45 päeva pikkuste treeningridadega mudelid näitasid mõnevõrra väiksemaid paranemisi täpsuses järgnevatel iteratsioonides. Sellised trendid erinevate mudelivariantide prognoositäpsuse paranemises andsid alust järeldusele, et kõige parema täpsusega prognoosimudel on selline, mis kasutab lühikest treeningrida ning väheseid võimalike sisendandmeid. Lisatreeningrea implementeerimisel mudelisse võis samuti täheldada prognoosimudeli täpsuse paranemist, mis tõestab seda, et ilmaparameetrite mõju gaasitarbimisele on sesoonselt erinev ning kasutades 15 päeva pikkust prognoositavale päevale vahetult eelnevat treeningrida ning 30 päeva pikkust lisatreeningrida prognoositavale päevale eelnenud aastast, on võimalik koostada veelgi täpsema väljundiga prognoosimudel. Ilmaparameetrite mõju sesoonsust gaasitarbimisele saab tõestada nii sellepärast, et prognoosimudelis lisatreeningrea implementeerimisel tekib andmete valim eelmise aasta sellisest perioodist, mis järgneb prognoositavast päevast täpselt aastatagusele päevale ning sellest tulenevalt on võimalik mudelis analüüsida andmeid eelseisva perioodi kohta ajaloolises vaates. Samas aga eeldab selline lähenemine seda, et igal aastal peab temperatuur jahedamaks või soojemaks minema võimalikult samal kuupäeval ning kui mõni aasta langeb temperatuur sügisel oluliselt hiljem või vastupidi, siis võib tekkida endiselt sarnane või isegi halvem veaprotsent kui lisatreeningrida mitte kasutades.

Kasutades lineaarset regressiooni ilmastikuandmete analüüsiks ning selle põhjal prognoosimudeli koostamiseks on võimalik saavutada prognoosimudeli pikaajaliseks täpsuseks 8,4 kuni 10,9 protsenti sõltuvalt prognoosimudelis kasutatavast sisendandmete hulgast ja mitmekesisusest. Lühemate perioodide lõikes nagu näiteks üks aasta võib mudeli täpsuseks saavutada ka 8,1 protsenti sõltuvalt kasutatavatest sisendandmete hulgast ja mitmekesisusest. Võrdluseks saab

siinkohal tuua Sloveenias 2018. aastal sooritatud uurimustöö, mille käigus katsetati erinevaid meetodikaid gaasitarbimise prognoosimiseks ning selles uurimuses saavutati lineaarsel regressioonil põhineva prognoosimudeliga prognoositäpsuseks 8,1 protsenti [13]. Selle põhjal saab väita, et ka käesolevas lõputöös sooritatud modelleerimine õnnestus, sest saavutati võrdväärne mudeli täpsus ka väljaspool käesolevat uurimust koostatud prognoosimudeliga. Antud prognoosimudeli täpsust oleks võimalik arendada ka kasutades tunnipõhist prognoosimist, sest tunnipõhiselt on perioodi jooksul tarbitud gaasi hulk numbriliselt vähem varieeruv ning sellest tulenevalt võib eeldada ka seda, et prognoosimudeliga on kergem prognoosida arvväärtusi, mis on ajas vähemal määral muutuvad. Samas keskenduti käesolevas lõputöös modelleerimise osas prognoosi koostamisele ühes bilansiperioodiks, milleks Eesti gaasiturul on 24 tundi.

KOKKUVÕTE

Käesoleva lõputöö eesmärgiks oli uurida erinevate ilmastikutegurite mõju Eesti gaasitarbimisele ning selgitada välja nende mõjude täpsed numbrilised väärtused regressioonanalüüsi abil ja selle põhjal koostada prognoosimudel, et hinnata uuritavate mõjurite olulisust ja kasutamise optimaalsust praktikas. Prognoosimudeli sisendandmestiku mitmekesisuse ning kasutatavate sisendandmete mahu optimaalsusele ja praktilisusele anti käesolevas töös hinnang prognoosimudeli täpsuse põhjal.

Ilmastikuparameetrite mõju hindamine gaasitarbimisele on oluline gaasitarbimise prognoosimise seisukohalt, mis tuleneb sellest, et gaasiturul omavad bilansihaldurid vastutust oma tarbimisportfellide bilansi suhtes. Eabilanss tarbimisportfellis tekitab eelkõige bilansikulusid ning läbi bilansikulude vähendamise on gaasi jaemüüjatel võimalik oma muutuvkulusid vähendada ja seeläbi omada ka paremat konkurentsivõimet aina tiheneva konkurentsiga gaasiturul.

Käesolevas lõputöös otsiti vastused järgnevatele küsimustele:

- Millised gaasitarbimise ilmastikumõjurid on kõige olulisemad prognoosimisel?
- Millised on täpsed arvulised parameetrid, millega eeltoodud mõjureid gaasitarbimise prognoosimisel arvesse võtta?
- Milline on kõige täpsem prognoosimudel ning selle parameetrid võttes arvesse eelnevalt kindlaks tehtud mõjureid?

Ilmastikutegurite mõju gaasitarbimisele uuriti regressioonanalüüsi abil ning uurimise alla kuulusid sellised ilmastikutegurid nagu temperatuur, sademete hulk, tuule kiirus ja suund ning nähtavus kilomeetrites. Kõikide andmestiku analüüsi iteratsioonides oli näha, et temperatuur omab kõikidest ilmastikuteguritest vaieldamatult kõige suuremat mõju gaasitarbimisele. Mõju suuruse ja olulisuse poolest oli ülejäänud tegurite mõju järjestus alates kõige olulisemast tegurist järgnev: nähtavus kilomeetrites, tuule kiirus, sademete hulk ja tuule suund. Kuigi regressioonanalüüs aitab leida erinevate tegurite täpsed numbrilised mõjud gaasitarbimisele siis tulenevalt nende väärtuste sesoonsusest ja pidevast muutumisest ajas ei ole praktikas prognoosimiseks mõistlik kasutada kõiki võimalike andmeid, sest treeningreana kasutatud sisendandmete jooksul võib erinevate tegurite mõju gaasitarbimisele olla märgatavalt erinev nende mõjust gaasitarbimisele prognoositaval perioodil.

Regressioonanalüüs oli aluseks käesoleva lõputöö praktilise poole teisele osale, milles kasutati praktilise poole esimese osa ehk regressioonanalüüsi tulemusi regressioonmudeli ja selle iteratsioonide koostamiseks. Modelleerimise käigus leiti, et tulenevalt prognoosimiseks

kasutatavate sisendandmete hulgast ja mitmekesisusest võib prognoosimudeli täpsus varieeruda umbes kolm protsenti, mis tähendab, et prognoosivea maht väheneks Eesti gaasisüsteemi 2018. aasta tarbimise põhjal umbes 156 gigavatt-tundi. Parima pikaajalise täpsuse saavutas prognoosimudel, mis kasutas 15 päeva pikkust prognoositavale päevale vahetult eelnevat treeningrida ning 30 päeva pikkust lisatreeningrida prognoositavale päevale eelnenud aastast. Sisendandmetena kasutati selles mudelis vaid temperatuuri ja ajaloolist gaasitarbimist. Nimetatud mudeliga saavutati keskmiseks pikaajaliseks veaprotsendiks 8,43 protsenti. Kuigi üldiselt võib eeldada, et mida rohkem ja põhjalikumaid sisendandmeid prognoosimiseks kasutada, seda täpsemaks muutub ka prognoosimudeli väljund, siis käesoleva lõputöö modelleerimise osas jõuti järeldusele, et hoopis võimalikult vähe sisendandmeid lühikese treeningreaga on võimalik koostada täpsemaid prognoose. See asjaolu tuleneb sellest, et ilmastikuparameetrite mõju gaasitarbimisele on ajas pidevalt muutuv ning parameetrid omavad prognoositaval perioodil erinevat mõju gaasitarbimisele võrreldes prognoositavale päevale eelnenud ajale. Mida lühem on treeningrida, seda kiiremini on prognoosimudel võimeline reageerima pidevatele muutustele parameetrite mõjus gaasitarbimisele. Sellest tulenevalt oli märgata prognoosimudeli veaprotsendi paranemist siis kui mudelile implementeeriti lisatreeningrida, mis arvestas andmeid prognoositava perioodi kohta eelnenud aastast. Sellisel juhul on mudeli sisendandmetes kajastatud ka sesoonselt muutuvad ilmastikuparameetrite mõjud gaasitarbimisele.

Käesolevas lõputöös selgitati välja erinevate ilmastikuparameetrite numbrilised mõjud kogu Eesti gaasitarbimisele ning selgitati välja nende olulisus prognoosimisel läbi prognoosivea minimeerimise erinevate iteratsioonide kaupa, et selgitada välja kõige optimaalsema koosseisuga prognoosimudel. Modelleerimise tulemustest selgus, et prognoosimudeli väljund muutub üldiselt igas iteratsioonis täpsemaks. Tulenevalt sellest, et lineaarne regressioon keskendub vaadeldavas perioodis keskmise ruutvea minimeerimisele, siis saab järeldada, et ka sellel põhinev regressioonimudel hindab parameetrite keskmist mõju vaadeldava perioodi jooksul, mis ei pruugi alati kokku langeda parameetrite tegeliku mõjuga prognoositaval perioodil.

Töö autori hinnangul on käesolevas lõputöös tehtud järeldused rahuldavad ning edaspidistes uuringutes oleks vajalik uurida detailsemalt prognoosimisel kasutatavate sisendandmete kasutamisevõimalusi erinevate treeninggridade näol. Ühtlasi oleks teoreetiliselt võimalik parandada ka käesolevas töös kasutatud prognoosimudeli täpsust, kui kasutada 24-tunnise bilansiperioodi prognoosimise asemel tunnipõhist prognoosimist, sest võib eeldada, et ühes tunnis varieerub gaasi tarbimine numbriliselt vähem, mis võib olla täpsema statistilise mudeli aluseks. Koostatud prognoosimudeli täpsus on töö autori hinnangul rahuldav ning sellest tulenevalt võib käesoleva uurimuse ja lineaarsel regressioonil põhineva prognoosimudeli koostamise õnnestunuks lugeda.

SUMMARY

The aim of this thesis was to determine the impact of different weather factors on Estonian gas demand in a numerical way by means of regression analysis and to compose a regression model for forecasting Estonian gas demand on day-ahead basis to assess the importance of those factors and the optimality of using such factors for forecasting in practice. The assesment for importance and optimality of diversity and volume of input data was carried out by means of accuracy of the forecasting model.

Determining the impact of different weather factors on Estonian gas demand is important from gas demand forecasting point of view due to the fact that balance responsible parties are held responsible for the balance of their consumption portfolio. Imbalance in consumption portfolio causes balance costs and through minimizing the balancing costs utility retailers are able to decrease their variable cost and therefore increase their copetitiveness on the open gas market.

The aim of this thesis was to answer the following questions:

- Which weather factors are the most important for gas demand forecasting?
- What are the numerical values of impact that those weather factors have on gas forecasting?
- Which is the most accurate forecasting model based on previously found numerical impacts of weather factors and its parameters?

The impact of different weather factors was analysed with regression analysis and under investigation were parameters such as: temperature, percipitation, wind speed and direction and visibility in kilometers. All the iterations of regression analysis showed that the most important factor with the biggest impact on gas demand is temperature. In order of decreasing impact the rest of the factors were visibility, wind speed, percipitation and wind direction. Although regression analysis helps to find out the exact numerical values of the impact that different weather factors have on gas demand, then due to seasonality of the impact and continuous changes in the values that those factors have it is not practical to use so diverse and wide range of input data as the factors might have severely different impact on gas demand on the day that is forecasted than during the training of the model.

Regression analysis was the basis of the second part of this thesis where the result of regression analysis was used in a regression model to forecast gas demand. The model had exactly the same iterations that regression analysis had. During the modeling process it was fount that depending on the range and diversity of input data the accuracy of the forecasting model can vary up to 3 per

cent, which means that based on the year 2018 Estonian gas consumption the volume of imbalance gas could vary about 156 gigawatt-hours. The best long term accuracy was achieved with a model that only used 15 days of training and additional 30 days of training from the previous year of the forecasted day. The input data for this model consisted only of temperature and historical gas consumption. Said model achieved long term accuracy of 8,43 per cent throughout the whole forecasted period. Although it is possible to assume that diverse and wide range of input data could produce more accurate results in forecasting then in the modelling part of current thesis it was found that the most accurate results can be achieved with using only one independent variable and short training in the model. This could be caused by the fact that the impact of weather factors varies in time and those factors have different impact on gas demand during forecasted day than during the training of the model. The shorter the training the more agile the model is to cope with the changes in the impact. Therefore it could be seen that the accuracy of forecasting models was significantly improved when the additional training was implemented in forecasting model that took into account the data from the previous year of the forecasted day. In this case the previous year of the forecasted day represents the coming estimated changes in the impact of weather factors which can then be taken into account in the forecasting model.

In the current thesis the exact numerical values of impact on Estonian gas demand were found for different weather factors and also the importance of using them in the forecasting model through minimizing the forecast error through different iterations of the model to find out the most optimal composition for said model. The results of the forecasting model indicated that the model's accuracy increases throughout iterations. As linear regression focuses on minimizing the square error throughout the model's training then it can be assumed that the model assesses the average impact of weather factors during the training which might not match the impact of those factors during the forecasted period.

The author of this thesis finds that the conclusions of current thesis are satisfying and the future research the different training periods with different input data compositions should be investigated in more detail. Also it would be theoretically possible to improve the forecasting model composed in current thesis by using hourly resolution during forecasting instead of 24-hour balance period because the consumption in one hour should vary numerically less compared to previous periods which could be the basis of more accurate statistical forecasting model. The accuracy of composed forecasting model is satisfying and therefore the current research can be said to be accomplished.

KASUTATUD KIRJANDUS

- [1] Euroopa Liidu Komisjon, „Komisjoni Määrus (EL) nr 312/2014,“ Euroopa Liidu Teataja, Brüssel, 2014.
- [2] Majandus- ja Kommunikatsiooniministeerium, „Gaasiturg | Majandus- ja Kommunikatsiooniministeerium,“ [Võrgumaterjal]. Available: <https://www.mkm.ee/et/tegevused-eesmargid/energeetika/gaasiturg>. [Kasutatud 31 03 2019].
- [3] Konkurentsiamet, „Aruanne elektri- ja gaasiturust Eestis 2016,“ Konkurentsiamet, Tallinn, 2017.
- [4] Eesti Gaas AS, „Ajalugu - Eesti Gaas,“ [Võrgumaterjal]. Available: <https://www.gaas.ee/ettevottest/ajalugu/>. [Kasutatud 22 03 2019].
- [5] Elering AS, Gaasituru käsiraamat, Tallinn: Elering AS.
- [6] Elering AS, „Maagaasi tarbimine püsis mullu 2017. aasta tasemel | Elering,“ 30 01 2019. [Võrgumaterjal]. Available: <https://elering.ee/maagaasi-tarbimine-pusis-mullu-2017-aasta-tasemel>. [Kasutatud 11 04 2019].
- [7] Elering AS, „Eesti pikaajaline gaasitarbimise prognoos,“ Elering AS, Tallinn, 2016.
- [8] Tööstusuudised.ee, „Nitrofert peatab ajutiselt tootmise,“ 11 02 2009. [Võrgumaterjal]. Available: <https://www.toostusuudised.ee/uudised/2009/02/11/nitrofert-peatab-ajutiselt-tootmise>. [Kasutatud 18 04 2019].
- [9] Postimees, „Aastaid tagasi tootmise lõpetanud Nitrofert teenis mullu kasumit 6 miljonit eurot,“ 24 11 2017. [Võrgumaterjal]. Available: <https://majandus24.postimees.ee/4321833/aastaid-tagasi-tootmise-lopetanud-nitrofert-teenis-mullu-kasumit-6-miljonit-eurot>. [Kasutatud 26 04 2019].
- [10] Elering AS, „Märtsis langes maagaasi tarbimine veerandi võrra,“ 25 04 2019. [Võrgumaterjal]. Available: <https://elering.ee/martsis-langes-maagaasi-tarbimine-veerandi-vorra>. [Kasutatud 01 05 2019].

- [11] N. J. Johannesen, M. Kolhe ja M. Goodwin, „Relative evaluation of regression tools for urban area electrical energy demand forecasting,” *Journal of Cleaner Production*, kd. 218, pp. 555-564, 2019.
- [12] N. Fumo, „A review on the basics of building energy estimation,” *Renewable and Sustainable Energy Reviews*, nr 31, pp. 53-60, 2014.
- [13] R. Hribar, P. Potocnik, J. Šilc ja G. Papa, „A comparison of models for forecasting the residential natural gas demand of an urban area,” *Energy*, kd. 167, pp. 511-522, 2019.
- [14] T. Catalina, V. Iordache ja B. Caracaleanu, „Multiple regression model for fast prediction of the heating energy demand,” *Energy and Buildings*, kd. 57, pp. 302-312, 2013.
- [15] S. Zhou ja N. Zhu, „Multiple regression models for energy consumption of office buildings in different climates in China,” *Energy*, kd. 7, nr 1, pp. 103-110, 2013.
- [16] K. Niglas, „Regressioonanalüüs,” 2013. [Võrgumaterjal]. Available: <http://www.cs.tlu.ee/~katrin/wp/wp-content/uploads/2013/11/regressioon.pdf>.
- [17] X. Yan ja X. Gang Su, *Linear Regression Analysis: Theory and Computing*, Singapore: World Scientific Publishing CO. Pte. Ltd., 2009.
- [18] A. Sen ja M. Srivastava, *Regression Analysis: Theory, Methods and Applications*, New York: Springer-Verlag New York, Inc., 1990.
- [19] PennState Eberly College of Science, „1.2 - What is the "best fitting line"? | STAT 501,” [Võrgumaterjal]. Available: <https://newonlinecourses.science.psu.edu/stat501/node/252/>. [Kasutatud 29 03 2019].
- [20] O. Alexandrov, *Illustration of linear least squares*, Wikimedia Commons, 2008.
- [21] I. Pardoe, *Applied Regression Modeling*, Hoboken: John Wiley & Sons, 2013.
- [22] Riigi Ilmateenistus, „Vaatlusandmed,” [Võrgumaterjal]. Available: <http://www.ilmateenistus.ee/ilm/ilmavaatlused/vaatlusandmed/>. [Kasutatud 15 02 2019].
- [23] Elering AS, „Summaarne sisemaine gaasivoog ülekandevõrgust - Elering Live,” [Võrgumaterjal]. Available: <https://dashboard.elering.ee/et/gas-system/index>. [Kasutatud 28 02 2019].

- [24] M. Akpınar ja N. Yumusak, „ESTIMATING HOUSEHOLD NATURAL GAS CONSUMPTION WITH MULTIPLE REGRESSION: EFFECT OF CYCLE,“ Dept. of Computer Engineering Sakarya University , Sakarya, Turkey, 2013.
- [25] K. Van Steen, „Using R for Linear Regression,“ Montefiore Institute, Department of Electrical Engineering and Computer Science, Liege, Belgium.
- [26] R Documentation, „lm function | R Documentation,“ [Võrgumaterjal]. Available: <https://www.rdocumentation.org/packages/stats/versions/3.5.3/topics/lm>. [Kasutatud 28 03 2019].
- [27] R Documentation, „predict function | R Documentation,“ [Võrgumaterjal]. Available: <https://www.rdocumentation.org/packages/raster/versions/2.8-19/topics/predict>. [Kasutatud 28 03 2019].

LISAD

Lisa 1. Gaasitarbimise prognoosimise mudeli kood

```
# mudel gaasitarbimise prognoosimiseks

mudel          <- "mudeli_nimi"
mudelnr        <- 45
esimene_indeks <- 1  #mitmendast indeksist algab modelleerimine
treeningrida   <- 45 #prognoosimisel kasutatav treeningrea pikkus
forecast_lag   <- 2  #ajalooliste andmete ning prognoositava päeva vahe

# andmete sisestamine R-i

regressioon <-
read.csv2("C:/Users/sander.randver/Desktop/Loputoo_magister/R_sisend.csv"
)

#tööpäevadele ja puhkepäevadele faktortunnuse määramine

regressioon$TP <- as.factor(regressioon$TP)

# regressioonanalysi koostamine määratud treeningreale ja parameetritele

regressioon_treening <- regressioon %>%
slice(esimene_indeks:treeningrida)

#lisatreeningrea lisamine

if (esimene_indeks > 365){
esimene_indeks365 <- esimene_indeks-365
treeningrida365 <- esimene_indeks365+30
regressioon_treening365 <- regressioon %>%
slice(esimene_indeks365:treeningrida365)
regressioon_treening <- rbindlist(list(
regressioon_treening,regressioon_treening365))
}

regressioonanalyyys <- lm(`Gaasitarbimine` ~
`TP`+
`Temp`+
`Tuule.Kiirus` +
`Tuule.Suund` +
`Sademed` +
`Nahtavus`
, data = regressioon_treening)

# määrame andmestikust treeningrea vahemiku

slice_index = treeningrida + forecast_lag

# regressioonkoefitsentide lisamine prognoosi väljundile

coefid <- data.frame(coef(regressioonanalyyys))
coefid <- t(coefid)
coefid <- data.frame(coefid)

# prognoosi alusandmete sisestamine

uusdata <- regressioon
```

```

uusdata <- uusdata %>% slice(slice_index:slice_index)

# regressioonmudeli koostamine prognoosimiseks

uus_tabel <- predict(regressioonanalyy, data.frame(
  `TP` = uusdata$TP,
  `Temp` = uusdata$Temp,
  `Tuule.Kiirus` = uusdata$Tuule.Kiirus,
  `Tuule.Suund` = uusdata$Tuule.Suund,
  `Sademed` = uusdata$Sademed,
  `Nahtavus` = uusdata$Nahtavus
),
  level = 0.9, interval = "confidence")

# koostame prognoosi tabeli koos usaldusvahemike ja
regressioonkoefitsentidega

uus_tabel <- data.frame(uus_tabel)
uus_tabel <- uus_tabel %>% slice(1:1)
uus_tabel <- merge(uus_tabel, coefid)

# väljundi saamine .csv formaadis

write.csv2(uus_tabel,
paste0("C:/Users/sander.randver/Desktop/Loputoo_magister/Forecastid/forec
ast_training_", mudelnr, "_", mudel, ".csv"), row.names = F)

# prognoosi simuleerimine kogu andmestikule FOR-tsükli abil

for(i in 1:length(regressioon$time_EET)){

# andmete sisestamine R-i

regressioon <-
read.csv2("C:/Users/sander.randver/Desktop/Loputoo_magister/R_sisend.csv"
)

# regressioonanalysi koostamine määratud treeningreale ja parameetritele

regressioon_treening <- regressioon %>%
slice(esimene_indeks:treeningrida)

#lisatreeningrea lisamine

if (esimene_indeks > 365){
esimene_indeks365 <- esimene_indeks-365
treeningrida365 <- esimene_indeks365+30
regressioon_treening365 <- regressioon %>%
slice(esimene_indeks365:treeningrida365)
regressioon_treening <- rbindlist(list(
regressioon_treening, regressioon_treening365))
}

regressioonanalyy <- lm(`Gaasitarbimine` ~
  `TP` +
  `Temp` +
  `Tuule.Kiirus` +
  `Tuule.Suund` +
  `Sademed` +
  `Nahtavus`
, data = regressioon_treening)

```

```

# määratakse igal tsüklil uued indeksid kasutatavatele treeninggridadele
ning ilmastikuandmetele

esimene_indeks = esimene_indeks + 1
treeningrida   = treeningrida + 1
slice_index    = treeningrida + forecast_lag

# regressioonkoefitsentide lisamine prognoosi väljundile

coefid <- data.frame(coef(regressioonanalyy))
coefid <- t(coefid)
coefid <- data.frame(coefid)

# prognoosi alusandmete sisestamine

uusdata <- regressioon
uusdata <- uusdata %>% slice(slice_index:slice_index)

# regressioonmudeli koostamine prognoosimiseks

uus_tabel <- predict(regressioonanalyy, data.frame(
  `TP`           = uusdata$TP,
  `Temp`        = uusdata$Temp,
  `Tuule.Kiirus` = uusdata$Tuule.Kiirus,
  `Tuule.Suund`  = uusdata$Tuule.Suund,
  `Sademed`     = uusdata$Sademed,
  `Nahtavus`    = uusdata$Nahtavus
),
  level = 0.9, interval = "confidence")

# koostame prognoosi tabeli koos usaldusvahemike ja
regressioonkoefitsentidega

uus_tabel <- data.frame(uus_tabel)
uus_tabel <- uus_tabel %>% slice(1:1)
uus_tabel <- merge(uus_tabel, coefid)

# vana tabeli sisestamine R-i, et liita see eelnevatele prognoosidele

vana_tabel <-
read.csv2(paste0("C:/Users/sander.randver/Desktop/Loputoo_magister/Foreca
stid/forecast_training_", mudelnr, "_", mudel, ".csv"))

# vana ja uue tabeli kokku liitmine

uus_tabel <- rbindlist(list(vana_tabel, uus_tabel))

# väljund

write.csv2(uus_tabel,
paste0("C:/Users/sander.randver/Desktop/Loputoo_magister/Forecastid/forec
ast_training_", mudelnr, "_", mudel, ".csv"), row.names = F)
}

```