



TALLINNA TEHNIKAÜLIKOOL

INSENERITEADUSKOND

Elektroenergeetika ja mehhatroonika instituut

MAAGAASI TARBIMISE MATEMAATILISTE PROGNOOSIMEETODITE ANALÜÜS

ANALYSIS OF MATHEMATICAL FORECASTING METHODS
OF NATURAL GAS CONSUMPTION

MAGISTRITÖÖ

Üliõpilane: Alexander Varushchenkov

Üliõpilaskood: 177045AAVM

Juhendaja: Hardi Koduvere, nooremteadur

Konsultandid: Margus Kaasik, Teet Eelmere,
Muhammad Naveed Iqbal,
Noman Shabbir

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.

“.....” 201.....

Autor:

/ allkiri /

Töö vastab bakalaureusetöö/magistritööle esitatud nõuetele

“.....” 201.....

Juhendaja:

/ allkiri /

Kaitsmisele lubatud

“.....”201... .

Kaitsmiskomisjoni esimees

/ nimi ja allkiri /

LÕPUTÖÖ LÜHIKOKKUVÕTE

Autor: Alexander Varushchenkov

Lõputöö liik: Magistritöö

Töö pealkiri: Maagaasi tarbimise matemaatiliste prognoosimeetodite analüüs

Kuupäev: 24.05.2019

82 lk

Ülikool: Tallinna Tehnikaülikool

Teaduskond: Inseneriteaduskond

Instituut: Elektroenergeetika ja mehhatroonika instituut

Töö juhendaja: nooremteadur Hardi Koduvere

Töö konsultandid: Margus Kaasik, Teet Eelmer, Muhammad Naveed Iqbal, Noman Shabbir

Sisu kirjeldus:

Vaadeldav lõputöö on seotud maagaasi tarbimise prognoosimisega. Lõputöö on välja pakutud Eesti Gaasi poolt ja ülesandeks oli analüüsida erinevad maagaasi tarbimise prognoosi meetodeid ja koostada maagaasi tarbimise päev-ette prognoosimudel, mis oleks täpsem kui hetkel kasutuses olev mudel Eesti Gaasis. Töö käigus tuli analüüsida tarbimist mõjutavaid ilmastiku, ökonoomilisi ja kalendrilisi faktoreid, valida sobivad masinõppe ja tehisintellekti algoritmid ning koostada nende abil prognoosimudelid. Samuti tuli kontrollida nende mudelite kasutatavust tegelikkuses.

Töö käigus uuriti ka prognoosi täpsust parandavaid meetodeid, nagu mitme mudeli kombineerimine ja klientide jagamine klastritesse vastavalt nende tarbimisprofiilidele, kasutades aegridade analüüsimeetodit nimega dünaamiline ajadeformatsioon.

Mudelite verifitseerimiseks kasutati erinevaid prognoosi täpsuse hindamise meetodeid, võrreldes saadud prognoosandmeid reaalse tarbimisandmetega.

Töö lõpptulemuseks on suurema täpsusega maagaasi tarbimise päev-ette prognoosimudelid ning kasulikud meetodikaid prognooside täpsuse suurendamiseks, mis võivad tulla kasuks mitte ainult energeetikutele ja inimestele kes tegelevad maagaasi tarbimise prognoosimisega, vaid ka inimestele, kes uurivad aegridade prognoosimist ja klasterdamist üldisemalt.

Märksõnad: maagaas, tarbimine, prognoos, mudel, aegrida, aegread, masinõpe, närvivõrk, tehisintellekt, klasterdamine, parameetrid, täpsus, lineaar, regressioon, otsuspuu

ABSTRACT

<i>Author:</i> Alexander Varushchenkov	<i>Type of the work:</i> Master Thesis
<i>Title:</i> Analysis of mathematical forecasting methods of natural gas consumption	
<i>Date:</i> 24.05.2019	82 pages
<i>University:</i> Tallinn University of Technology	
<i>School:</i> School of Engineering	
<i>Department:</i> Department of Electrical Power Engineering and Mechatronics	
<i>Supervisor of the thesis:</i> junior researcher Hardi Koduvere	
<i>Consultants:</i> Margus Kaasik, Teet Eelme, Muhammad Naveed Iqbal, Noman Shabbir	
<i>Abstract:</i> <p>This thesis is related to forecasting natural gas consumption. The subject of the thesis was suggested by Eesti Gaas, and the task was to analyse different mathematical methods of natural gas consumption and produce natural gas consumption day-ahead forecasting model, which would be more accurate than the model currently in use at the Eesti Gaas. In the course of the research, the author had to analyse the weather, economy and calendar factors affecting the consumption, to choose suitable machine-learning and artificial intelligence algorithms and use them to create forecasting models. It was also necessary to verify the actual usability of these models in real life.</p> <p>The author also examined the prognosis accuracy-improving methods such as a combination of models and dividing customers into clusters according to their consumption profiles, using time series analysis method called dynamic time warping.</p> <p>Various methods of estimating forecast accuracy were used to verify the models, comparing the data obtained in the prognoses with the actual consumption data.</p> <p>The research resulted in greater accuracy day-ahead natural gas consumption forecasting models and useful methodology for increasing the accuracy of forecasts, which may be useful not only to power engineers and people engaged in forecasting natural gas consumption, but also those to people who study time series forecasting and clustering in general.</p>	
<i>Keywords:</i> natural gas, consumption, forecast, model, time, series, machine, learning, neural, network, artificial, intelligence, clustering, parameters, accuracy, linear, regression, decision, tree	

LÕPUTÖÖ ÜLESANNE

Lõputöö teema:	Maagaasi tarbimise matemaatiliste prognoosimeetodite analüüs
Üliõpilane:	Alexander Varushchenkov, 177045AAVM
Eriala:	Elektroenergeetika
Lõputöö liik:	magistritöö
Lõputöö juhendaja:	nooremteadur Hardi Koduvere
Lõputöö ülesande kehtivusaeg:	01.07.2019
Lõputöö esitamise tähtaeg:	24.05.2019 15:00

Üliõpilane (allkiri)

Juhendaja (allkiri)

Õppekava juht (allkiri)

1. Teema põhjendus

Teema on seotud gaasitarbimise lühiajalise prognoosimisega. Gaasi pakkumine ja tarbimine peavad kogu aeg bilansis olema, kui tarbimine ületab pakkumise, siis peab gaasimüüja osta gaasi kõrgema hinnaga, ja kui tarbimine on madalam kui pakkumine, peab gaasimüüja ülejääki müüja kahjumliku hinnaga. Nii ülejaak kui puudujaak on gaasimüüjale kahjulik. Seoses sellega, et gaasitarbimine sõltub paljudes muutujatest (nt. ilmastik, aastaeg, tarbija tüüp jne) muutub tarbimise prognoosimine üsna keerukamaks. Selleks, et võimalikult täpselt ennustada, peaks mis prognoosimismeetod on kõige tõhusam ja mis faktorid on kõige kriitilisemad. Prognoosiviga päev ette prognoosil on täna keskmiselt 5-6%, sihiks on jõuda keskmise veaga 2% alla.

2. Töö eesmärk

Töö käigus otsitakse parim lahendus gaasitarbijate klasterdamiseks vastavalt nende tarbimise iseloomule ja seejärel parim matemaatiline meetod klasterite tarbimise prognoosimiseks. Võrreldud lahenduste hindamise kriteeriumiks on prognoosiviga.

3. Lahendamisele kuuluvate küsimuste loetelu:

Mis faktorid mõjutavad gaasitarbimise kõige rohkem?

Millised matemaatilisi meetodeid kasutatakse gaasitarbimise ennustamiseks?

Kuidas klasterdada kliente tarbimise iseloomude alusel?

Milline gaasitarbimise prognoosimiseks kasutatav matemaatiline meetod annab parima tulemuse?

4. Lähteandmed

Eesti gaasi poolt antud tarbimise statistilised andmed. Eelnevate meetodite prognoosimisvead võrreldes reaalseste numbritega. Ilmateenistuse ilmastikuandmed.

5. Uurimismeetodid

Gaasitarbimise puudutava kirjanduse analüüs. Tarbimise mõõtmine, erinevate meetodite katsutamine ja võrdlemine. Tehisintellekti kasutamiseks sobiva keskkonna leidmine (Matlab, Microsoft Azure Machine Learning, v.m.) ja seda kasutamise otstarbekuse analüüs.

6. Graafiline osa

Graafikud, mis aitavad võrrelda erinevate meetodite täpsuse.

7. Töö struktuur

Ligikaudne:

1. Sissejuhatus
2. Mis mõjutab tarbimise
3. Tarbimise prognoosimise meetodid
4. Reaalsed andmed ja arvutused
5. Tulemuste analüüs
6. Kokkuvõte

8. Kasutatud kirjanduse allikad

- Forecasting natural gas consumption [1]
- Natural gas consumption forecast with MARS and CMARS models for residential users [2]
- Countywide natural gas consumption forecast, a machine learning approach [3]
- Predicting Natural Gas Consumption - A Literature Review [4]

Ja teised teadusartiklid, mis on seotud statistikaga, tehisintellektiga, masinõpega ja maagasi tarbimise prognoosimisega.

9. Lõputöö konsultandid

-

10. Töö etapid ja ajakava

01.03.2019 - 31.03.2019 – Kirjanduse läbitöötamine ja lähteandmete kogumine, teoreetilise osa kirjutamine

01.04.2019 – 30.04.2019 – Arvutused, järeldused, kuu lõpuks esimene versioon valmis

01.05.2019 – 24.05.2019 – Parandused, lõplik versioon valmis

SISUKORD

EESSÕNA.....	8
SISSEJUHATUS	9
1. MAAGAAS KUI ENERGIAALLIKAS	10
1.1 Maagaasist üldist	11
1.2 Maagaas Eestis.....	13
1.3 Bilansivastutus	17
1.4 Gaasitarbimise prognoosimise tähtsus.....	18
2. GAASITARBIMISE PROGNOOSIMINE	20
2.1 Maagaasi tarbimisega seotud varasemad uuringud.....	20
2.2 Aegridade prognoosimiseks traditsioonilised meetodid	22
2.3 Lõputöös kasutatud masinõppel põhinevad meetodid	26
2.4 Eesti Gaasi mudel.....	33
2.5 Aegridade klasterdamine	34
2.6 Peatüki kokkuvõte.....	39
3. MUDELITE VÄLJATÖÖTAMINE.....	40
3.1 Prognoosi täpsuse võrdlemise meetodid	40
3.2 Algandmed	43
3.3 Mudelite testimine.....	48
3.4 Mudelite kombineerimine	56
3.5 Klasterdamise mõju prognoositäpsusele.....	58
3.6 Katsetamine päev-ette prognoosi andmetega	64
3.7 Peatüki kokkuvõte.....	70
KOKKUVÕTE	73
SUMMARY	75
KASUTATUD KIRJANDUSE LOETELU	77

EESSÕNA

Käesolev lõputöö on seotud maagaasi tarbimise prognoosimisega. Magistritöö ülesanne pakuti välja Eesti Gaasi poolt, töö teema ja struktuur koostati üliõpilase ja juhendaja poolt. Käesoleva töö põhilised algandmed on saadud Eesti Gaasi ja Riigi Ilmateenistuse poolt. Muud avalikult kättesaadavad andmed on saadud internetist. Töö käigus on kasutatud MS Excelit, Matlabi, IBM SPSS-i ja programmeerimiskeelt R.

Käesolev lõputöö on mõeldud eeskätt energeetikutele, kes tegelevad maagaasi tarbimise prognoosimisega, aga ka inimestele, kes tegelevad muude energiaressursside tarbimise prognoosimisega.

Tänan juhendajat Hardi Koduvere, kes mind toetas ja juhendas. Tänan Eesti Gaasi juhatuse liiget Margus Kaasiku pakutud töö ja toetuse eest, teema uurimisel sain teada palju huvitavat. Tänan Teet Eelmere tehniliste aspektide selgitamise eest. Tänan Muhammad Naveed Iqbal ja Noman Shabbir masinõppe küsimustes konsulteerimise eest.

Alexander Varushchenkov

Telefon: +37256729108

Email: alexander.varushchenkov@gmail.com

SISSEJUHATUS

Maagaasi tarbimise prognoosimine on viimastel aastakümnetel infotehnoloogia ja energeetika arengu tõttu oluliselt muutunud. Tehnoloogia kiire areng võimaldab täpseid mõõtmisi, saada varem kättesaamatuid andmeid reaajas ning teha keerukaid arvutusi. Tänu nüüdisaegsetele meetoditele on võimalik koostada keerukaid prognoosimudeleid, mis suudavad võtta arvesse paljusid erinevaid tegureid. Maagaasi tarbimise prognoosimisel on peamiseks proovikiviks tarbimise kõrge volatiilsus ning sõltuvus erisugustest välisparameetritest nagu välisõhu temperatuur, õhuniiskus, päikesepaiste, nädalapäevad, riigipühad jmt.

Käesoleva töö põhieesmärk oli luua ja võrrelda erinevaid matemaatilisi prognoosimudeleid, mis võimaldaksid ennustada maagaasi tarbimist järgmisel päeval. Töö oluline osa oli ka leida tõhus meetod klientide jagamiseks klastritesse ja kontrollida, kuidas see mõjutab prognoosi täpsust. Samuti analüüsiti nende mudelite toimimist tegelikes tingimustes. Töö käigus kasutati erinevaid programme, sh Microsoft Excelit (Data Analysis Add-in, Solver Add-in), IBM SPSS-i, Matlabi ja programmeerimiskeelt R.

Lõputöö koosneb kolmest peatükist. Esimeses peatükis kirjeldatakse maagaasiga seotud põhifakte, nagu selle tootmine ja jaotamine, keemilised parameetrid, maagaasi tarbimise suundumused maailmas, tarbimise ajalugu Eestis ning prognooside olulisus ettevõtete, klientide ja energiasüsteemi seisukohast.

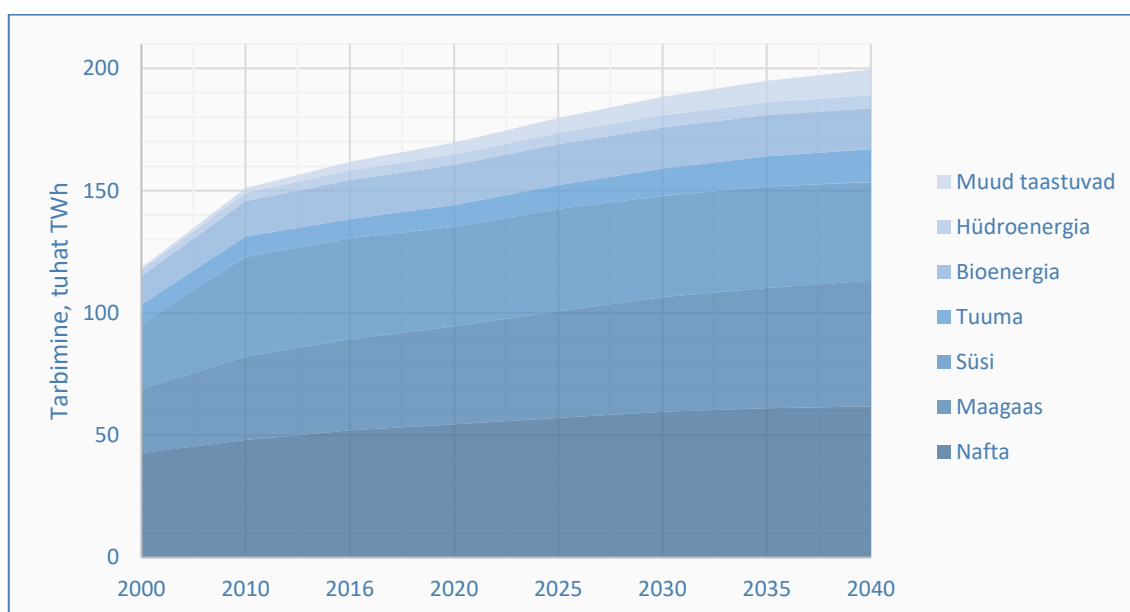
Teises peatükis üldistatakse varasemaid uuringuid: milliseid parameetreid ja mudeleid neis kasutati, milliseid prognoositäpsuse hindamismeetodeid rakendati. Edasi käsitletakse traditsioonilisi meetodeid aegridade prognoosimiseks ning näidatakse ka nende võimetust mittelineaarsete aegridade prognoosimisel. Teises peatükis kirjeldatakse mudelite koostamisel kasutatud meetodite põhimõtteid ja aegridade analüüsimeetodit, mida kasutatakse klientide klasterdamiseks.

Kolmas peatükk on otseselt seotud lõputöös püstitatud ülesandega. Siin analüüsitakse parameetreid, tuvastatakse mudelite loomiseks olulised näitajad nagu korrelatsiooni- ja osakorrelatsioonikoefitsiendid ja Studenti t-testi tegurid. Samuti on siin esitatud erinevate prognoosimudelite hindamise ja võrdlemise meetodid. Antud peatükis analüüsitakse ja võrreldakse tulemusi, võrreldes saadud andmeid omavahel ja algse mudeliga tehtud andmetega. Järgnevalt ülalpool nimetatud aegridade analüüsi meetodit rakendatakse klientide jaotamiseks klastritesse, mida edaspidi kasutatakse arvutustes. Peatüki lõpus kontrollitakse mudelite rakendatavust tegelikkuses eri mudelite analüüsi kaudu, kasutades ilmaprognoosi andmeid (järgmise päeva temperatuuri prognoos).

1. MAAGAAS KUI ENERGIAALLIKAS

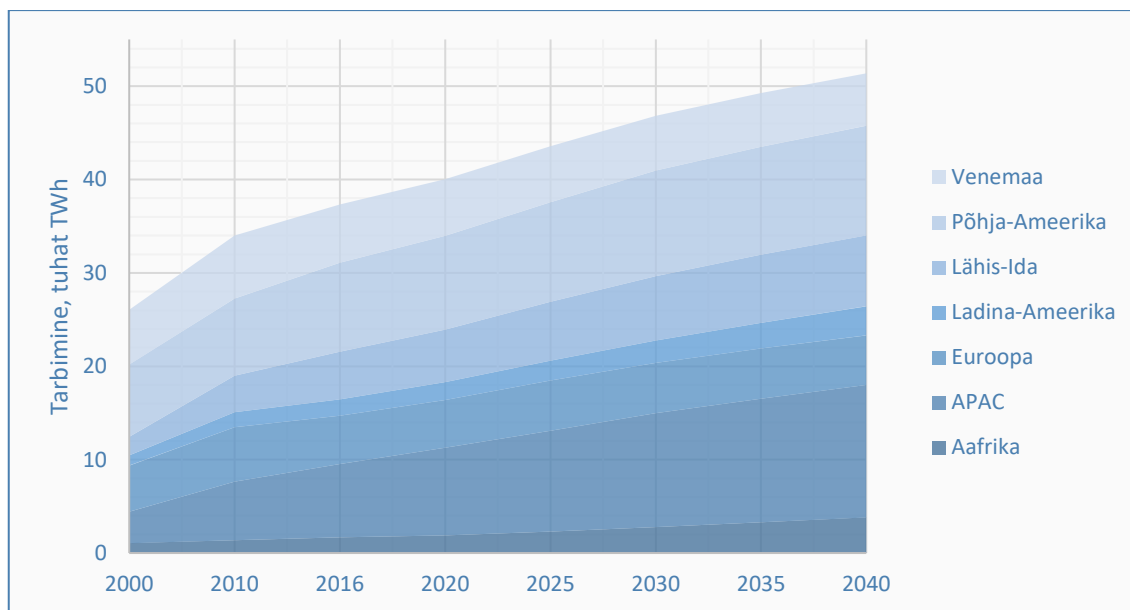
Energia tarbimine kogu maailmas kasvab pidevalt, eelkõige arenevate tööstusmaade arvelt. Arenenud riikides kasvab energiatarbimine aeglaselt, kuid nad tarbivad suurema osa toodetud energiast. Vaatamata rohelise energia kiirele arengule jäävad fossiilkütused (sh maagaas, nafta ja kivisüsi) siiski veel peamiseks energiaallikateks. [1] [2]

USA Energia Informatsiooni Administratsiooni (EIA, *U.S. Energy Information Administration*) viimane “International Energy Outlook 2017” (IEO 2017) on seisukohal, et energiatarbimine maailmas suureneb ajavahemikus 2015–2040 28% võrra. Suurem osa nimetatud kasvust tuleneb eeldatavasti riikidest, mis ei kuulu Majanduskoostöö ja Arengu Organisatsiooni (*Organisation for Economic Co-operation and Development, OECD*) ja eriti riikidest, kus nõudlust mõjutab tugev majanduskasv, iseäranis Aasias. OECD-sse mittekuuluv Aasia (sh Hiina ja India) moodustab aastatel 2015–2040 kogu maailma energiatarbimise kasvust 60%. [1]



Joonis 1.1. Maailma energiatarbimine allikate kaupa [2]

Kuigi IEO 2017 loodab, et mittefossiilkütuste (taastuvad ja tuumkütus) osakaal kasvab kiiremini kui fossiilkütuste kasutamine, moodustavad viimased endiselt 2040. aastaks üle kolme neljandiku maailma energiatarbimisest. Maagaas, mille süsiniku jalajälg on madalam kui söel ja naftal, on väljavaadete kohaselt kõige kiiremini kasvav fossiilkütus, seejuures kasvab maagaasi kogutarbimine 1,4% aastas. Suhteliselt kõrge maagaasi tarbimise kasv on tingitud rikkalikest maagaasiressurssidest ja kasvavast tootmisest, sealhulgas varustamisest kinnisgaasi, põlevkivigaasi ja söekihi metaaniga. [1]



Joonis 1.2. Maailma maagaasi tarbimine [2]

Jooniselt 1.2 on näha, et maagaasi tarbimine kasvab peamiselt arenevates riikides. Aafrikas, APAC-is (*Asia Pacific*), Lääne-Ameerikas ja Lähis-Idas prognoositud maagaasi tarbimise kasv perioodil 2016-2040 on 129%, 80%, 77% ja 49%. Põhja-Ameerika, Euroopa ja Venemaa kasv on vastavalt 24%, 3% ja -10%.

1.1 Maagaasist üldist

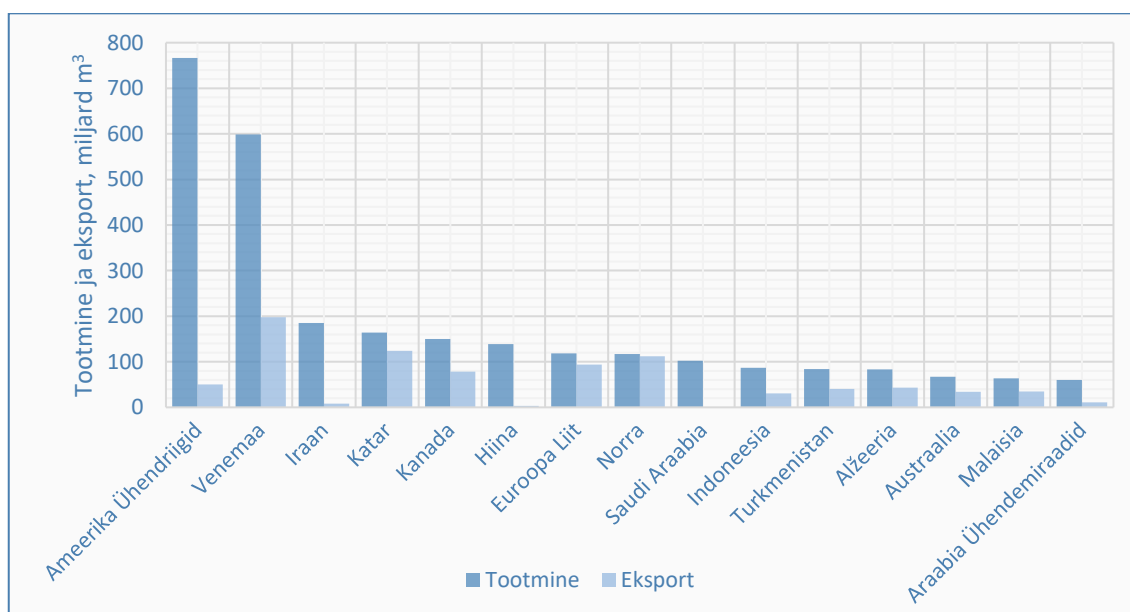
Maagaas on fossiilne energiaallikas, mis on tekkinud sügaval maapinna all. Maagaas sisaldab palju erinevaid ühendeid, peamine komponent on metaan, millel on üks süsiniku aatom ja neli vesiniku aatomit (CH₄). Maagaas sisaldab vähesel määral ka vedelikke (*NGL*; mis on ühtlasi süsivesinikgaasi vedelikud) ja mittesüsivesinikgaase, näiteks süsinikdioksiidi ja veeauru. [3]

Miljoneid (kuni 100 miljonit) aastaid tagasi ja väga pika aja jooksul ladestusid taimede ja loomade jäänused (näiteks diaatomid ehk ainuraksed ränivetikad) paksude kihtidena maakoore kurdudesse ja ookeanipõhja, kus nad vahel segunesid liiva, muda ja kaltsiumkarbonaadiga. Aegade jooksul mattusid need kihid liiva, muda ja kivide alla. Rõhu ja soojuse toimel muutus osa sellest süsiniku- ja vesinikurikkast materjalist kivisöeks, osa naftaks ja osa maagaasiks. Mõnes kohas liikus maagaas suurtesse pragudesse ja kivimikihtide katte vahekohtadesse. Seda tüüpi moodustistes leiduvat maagaasi nimetatakse mõnikord tavapäraseks maagaasiks. Mitmes kohas on maagaas kogunenud pisikestesse pooridesse (tühikutesse) savikiltkivi, liivakivi ja muud liiki settekivimites. Seda maagaasi nimetatakse kildagaasiks või tihedaks gaasiks ning ka ebatavaliseks maagaasiks. Maagaasi tekib ka toornafta ladestumisel ning seda maagaasi nimetatakse naftagaasiks. Maagaasivarusid

leidub maapõues ning avamere põhjas ja sügaval ookeanipõhja all. Söekihtide vahelistes tühimikes leiduvat maagaasi nimetatakse söekihi metaaniks. [3]

Maagaasi või toornafta kaevudest ammutatud maagaasi nimetatakse märjaks maagaasiks, sest koos metaaniga sisaldab see tavaliselt NGL-d – etaani, propaani, butaane ja pentaane ning veeauru. Puuraukudest saadav maagaas võib sisaldada ka mittesüsivesiniku gaase nagu vääveldioksiid, heelium, lämmastik, vesiniksulfiid ja süsihappegaas, millest suurem osa tuleb enne tarbijale müümist maagaasist eemaldada. Looduslikest allikatest / puuraukudest saadud maagaas saadetakse töötlemisettevõttesse, kus gaasist eemaldatakse veeaur ja mittesüsivesinike ühendid ning NGL eraldatakse märggaasist ja need müüakse eraldi. Osa etaani jäetakse sageli töödeldud maagaasi sisse. Eraldatud NGL-e nimetatakse maagaasitehase vedelikeks (NGPL) ning töödeldud maagaasi nimetatakse kuivaks, tarbijaklassiks ehk kvaliteetseks torujuhtme maagaasiks. Mõnest puuraugust saadav maagaas on piisavalt kuiv ja vastab torujuhtme transpordistandarditele ilma töötlemiseta. Maagaasiile lisatakse kemikaale (lõhnaaineid), et oleks võimalik maagaasijuhtmetes lekkeid avastada. Kuiv maagaas saadetakse torujuhtmete kaudu maa-alustele ladustamisaladele või gaasijaotusettevõtetele ja seejärel tarbijatele. [3]

Kohtades, kus pole maagaasitrasse, et toimetada edasi puuraukudest toodetud naftasisaldusega maagaas, võib seda looduslikku reservuaari maagaasi uuesti sisse pritsida, või seda ventileerida või põletada („lõkkele lüüa“). Turustamatu maagaasi uuesti sisse pritsimine võib aidata säilitada rõhku naftaaukudes, et tõhustada naftatootmist. Söekihi metaani võib söelademest kaevandada enne söekaevandamist või ka söekaevandamise ajal ning seda võib saata maagaasijuhtmetesse ilma eritöötluseta. [3] Peamised maagaasi tootjad ja eksportijad on esitatud allpool Joonisel 1.3.



Joonis 1.3. Maagaasi tootvad ja eksportivad riigid ja nende tootmismahud [4] [5]

Jooniselt 1.3 selgub, et peamised maagaasi tootjad on Ameerika Ühendriigid, Venemaa ja Iraan. Peamised eksportijad on Venemaa, Katar ja Norra.

1.2 Maagaas Eestis

Käesolevas alapeatükis kirjeldatakse gaasitarbimise ajalugu Eestis ja vaadeldakse maagaasi ülekandevõrgu hetkeolukorda. Räägitakse maagaasi impordist Eestisse ja maagaasi tarbimise suundumustest viimastel aastatel. Siin räägitakse ka bilansivastutusest, samuti maagaasi tarbimise prognoosimise tähtsusest energiasüsteemi, ettevõtte ja kliendi seisukohalt.

1.2.1 Ajalooline taust

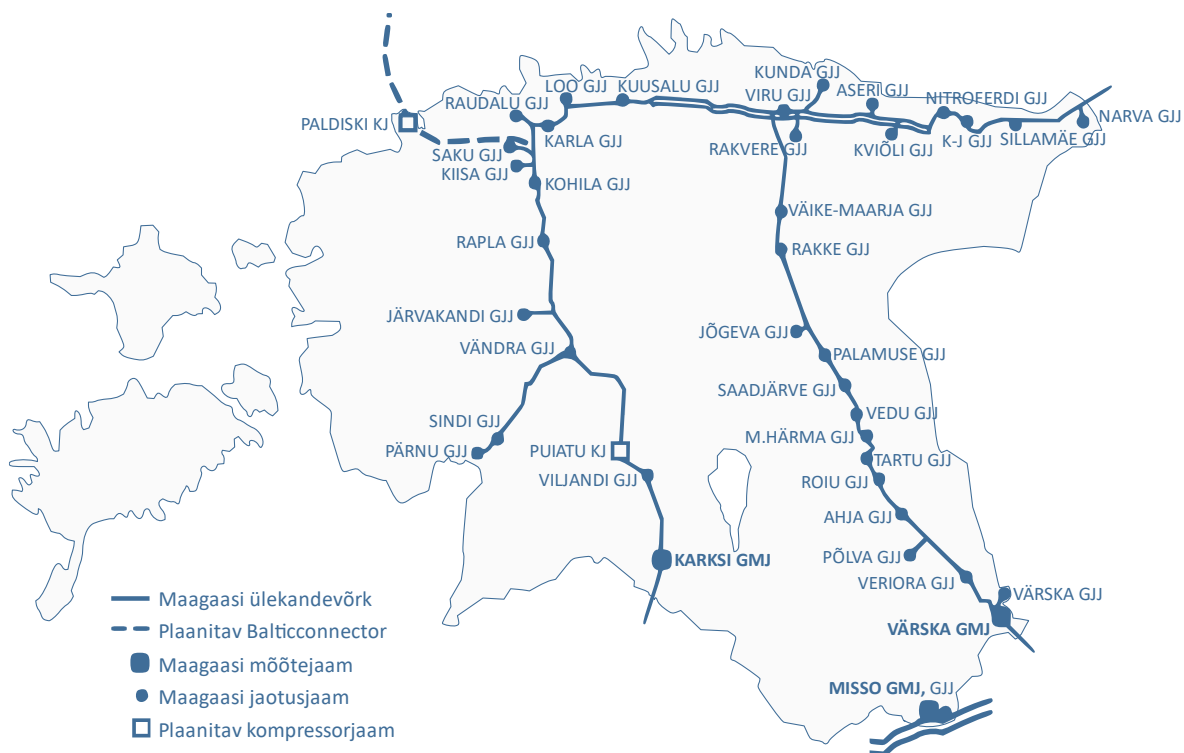
Gaasitarbimise ajalugu Eestis algab juba 19. sajandist. Eesti esimene gaasivabrik (Tallinna Gaasijaam, ka Tallinna Linna Gaasivabrik) valmis 1865. aastal ja tehiskaasi toodeti seal Inglismaalt imporditud kivisöest. Gaasiga varustati peamiselt kesklinna tänavalaternaid, aga ka avalikke hooneid, hiljem Balti Manufaktuuri, Lutheri vineerivabrikut ning Kadrioru lossi ja parki. 1880. aastal alustas tööd ka Tartu Gaasivabrik. 1948. aastal valmis Kohtla-Järve Põlevkivitöötlemise Kombinaat, mis 1949. aastal hakkas Kohtla-Järve – Leningradi gaasitorustiku kaudu transportima gaasi Leningradi linna. 1953. aastal avati pidulikult esimene Kohtla-Järvelt Tallinna suubuv kõrgsurve gaasitrass. Vaid mõne aasta jooksul suurenes gaasi tarbimine Tallinnas mitukümmend korda – 1955. aastal oli see juba 65 miljonit kuupmeetrit. Esimeste seas said Kohtla-Järve põlevkivigaasi Riigi Viinatehase (praegune Liviko) Mere puisteel asuv tsehh ja Raua tänava saun. [6]

1969. aastal hakati Kohtla-Järve – Leningradi torustiku kaudu tarnima Eestisse Venemaalt maagaasi, mis Jõhvis segati põlevkivigaasiga. 1976. aastal alustas tööd Irboska–Tartu ja 1978 Tartu–Rakvere magistraalgaasitorustik. 1979. aastal tarbiti Eestis juba 708 miljonit kuupmeetrit maagaasi ja 273 miljonit kuupmeetrit põlevkivigaasi, ja nõudlus gaasi järele üha kasvas. 1988. aastal hakati ehitama Vireši–Tallinna gaasitrassi ja Vireši kompressorjaama, tänu millele võis transportida gaasi ka Lätis Inčukalnsis asuvast maa-alusest gaasihoidlast. Ja seejärel järgnes lausa riburada uusi torustikke Eesti asulate varustamiseks gaasiga: Loo, Viljandi, Kohila, Kehra, Kunda, Narva-Jõesuu, Saue, Muuga, Narva, Vändra ja Raudalu. [6]

1.2.2 Eesti ülekandevõrk tänapäeval

Täna Eesti maagaasi ülekandevõrku kuulub 885 km gaasitorustikku, 3 gaasimõõtejaama (GMJ), milles mõõdetakse võrku siseneva gaasi koguseid ning uuritakse gaasi kvaliteeti, ja 36 gaasijaotusjaama (GJJ), milles toimub võrgust väljuva gaasi rõhu redutseerimine, mõõtmine, lõhnastamine ning kokkulepitud tarbimisrežiimi tagamine. Gaasi liikumiseks ühest punktist teise peab kahe punkti vahel olema rõhkude erinevus, mis tekitatakse kompressorjaamades, kus gaas

surutakse kokku ja sel moel tõstetakse gaasi rõhku. Eestis täna oma kompressorjaama veel pole, gaasi ülekandeks vajaliku rõhu tagavad Venemaa ülekandesüsteemi kompressorjaamad või siis tehakse seda Inčukalnsis (Lätis) asuvas maa-aluses gaasihoidlas. [7]



Joonis 1.4. Eesti maagaasi ülekandevõrk [7]

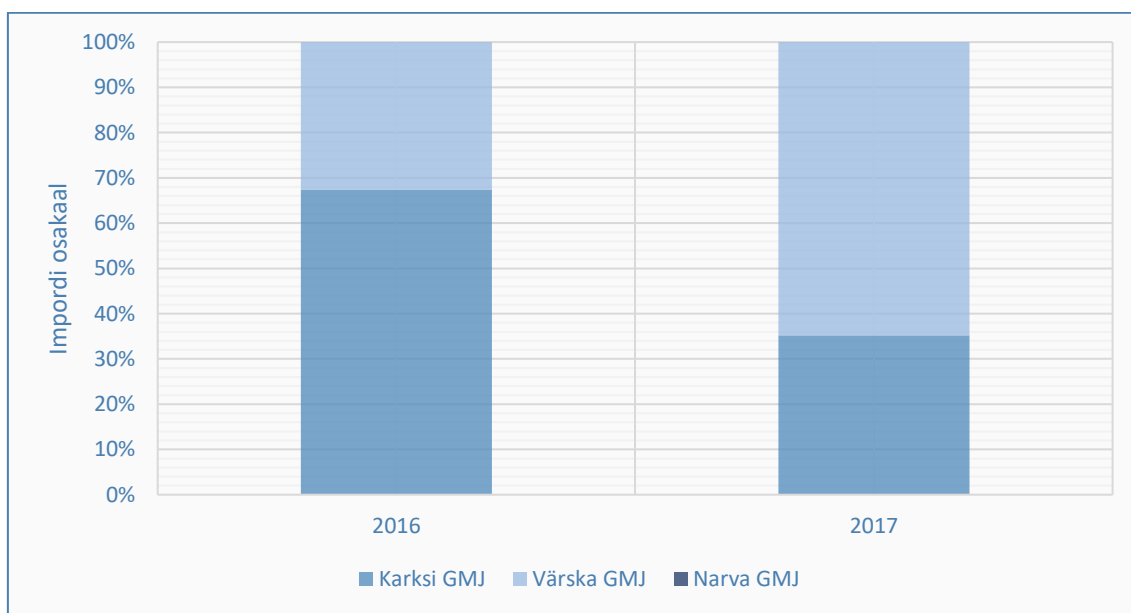
Eesti gaasi ülekandevõrk koosneb paljudest torustikest, mis erinevad üksteisest maksimaalse lubatud töö rõhu (*MOP*), läbimõõdu ja ka vanuse poolest. [8]

Tabel 1.1. Eesti maagaasi ülekandevõrgu torustik [8]

Torustik	Pikkus	DN (Nominaal-diaameeter)	Maksimaalne töö rõhk (MOP)	Vanus
Vireši - Tallinn	202,4	700	49,6	26
Vändra - Pärnu	50,2	250	54	12
Tallinn - Kohtla-Järve I	97,5	200	≤ 30	65
Tallinn - Kohtla-Järve II	149,1	500	≤ 30	50
Kohtla-Järve - Narva	45,1	350/400	≤ 30	58
Irbooska - Värskas GMJ	10,1	500	53,7	43
Värskas GMJ – Tartu	75,6	500	45,9	43
Tartu - Rakvere	133,2	500	45,2	39
Irbooska - Inčukalns	21,3	700	49,2	34
Pihkva - Riia	21,3	700	51,4	46
Harutorustikud	79,2			
Kokku	885,0			

1.2.3 Maagaasi import

Ei Eestis ega ka Lätis, Leedus ja Soomes maagaasi ei toodeta. Neis maades tarbitav gaas ostetakse sisse Venemaalt ja LNG (*Liquefied Natural Gas*, veeldatud looduslik gaas) maailmaturult. Venemaal toodetud gaas tarnitakse Balti riikidesse ja Soome kõrgsurvevõrgustike kaudu. LNG transporditakse laevadega Leedu Klaipeda LNG terminali, mis on ühendatud piirkonna maagaasivõrgustikuga. Alltoodud joonisel on näha gaasi maaletoomine sisendpunktide lõikes aastatel 2016 ja 2017. [9]

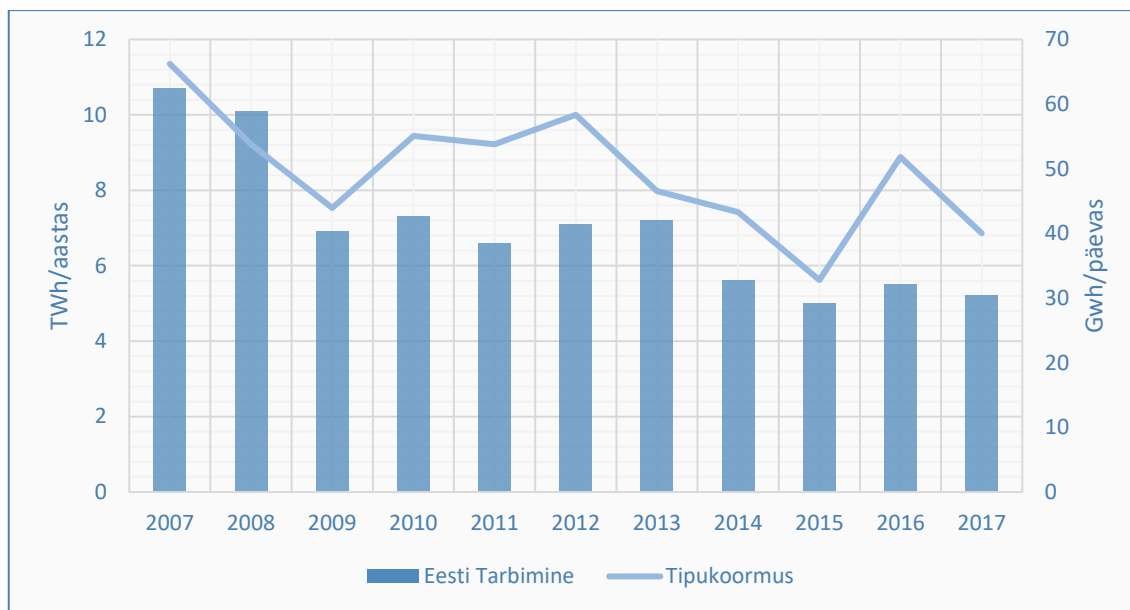


Joonis 1.5. Gaasi import sisendpunktide lõikes [9]

Jooniselt 1.5 on näha, et kogu Eestis tarbitav gaas läheb läbi Karksi ja Värskas gaasimõõtejaamade. Läbi Narva gaasiülekannet ei toimu. Võrreldes 2016. aastaga on 2017. aastal suurenenud Värskas gaasimõõtejaama kaudu saadava gaasi osakaal.

1.2.4 Maagaasi tarbimine Eestis

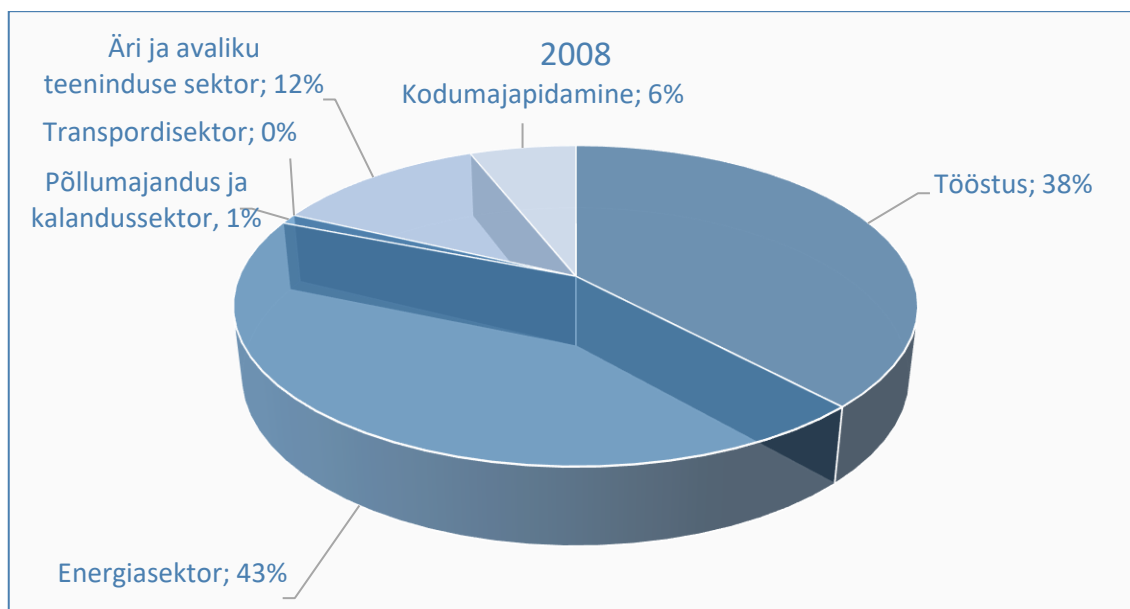
Viimastel aastatel on maagaasi tarbimine olnud Eestis üldiselt languses. 2016. aastal tarbimine külmema talve tõttu küll suurenes 2015. aastaga võrreldes, ent 2017. aastal toimus eelmise aastaga võrreldes jällegi 5,1-protsendiline langus. [10]



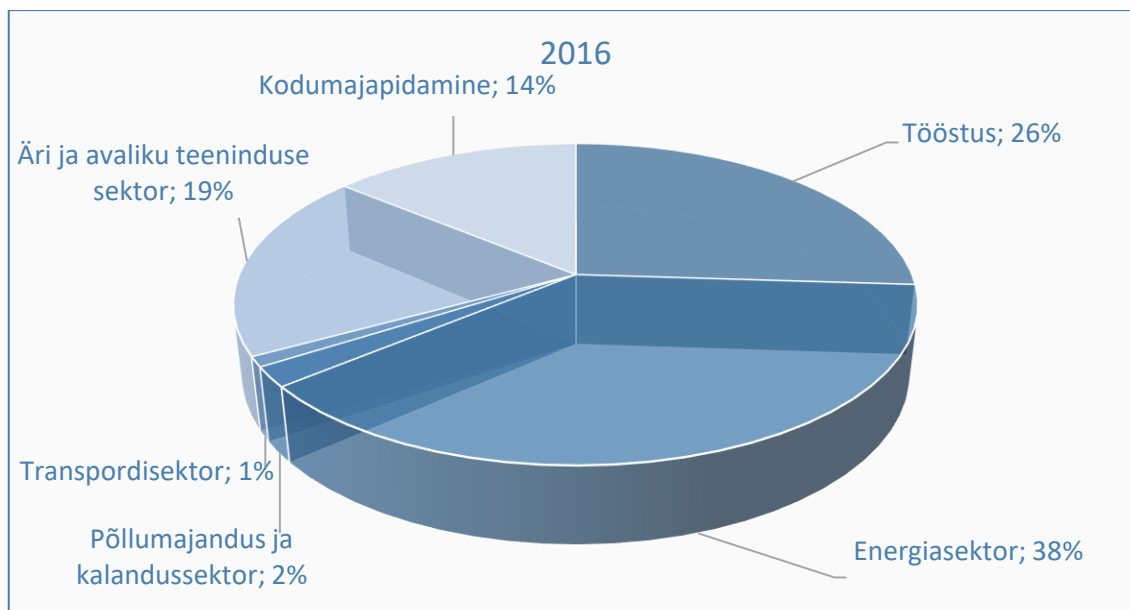
Joonis 1.6. Eesti maagaasi aasta tarbimine ja tipukoormus aastatel 2007-2017 [10]

2007–2017. a on maagaasi tarbimine vähenenud meil umbes poole võrra. [10]

Maagaasi tarbimise vähenemise põhjusteks on nii muutused tarbimise struktuuris (mitmed elektri- ja soojatootjad ning tööstustarbijad ei kasuta enam gaasi kütusena, ka vähendab nõudlust gaasi järele energiakasutuse tõhustamine), gaasiosuga seotud halb maine (gaas kui nn poliitiline kütus) kui ka puudujäägid taristuses (kaugus tarbijast) (Joonised 1.7 ja 1.8). [10]



Joonis 1.7. Gaasi tarbimise jaotus sektorite kaupa aastal 2008 [10]



Joonis 1.8. Gaasi tarbimise jaotus sektorite kaupa aastal 2016 [10]

Joonistelt 1.7 ja 1.8 on näha, et tööstussektoris on maagaasi tarbimine oluliselt vähenenud.

1.3 Bilansivastutus

Bilansijuhtimise eesmärgiks on tagada Eesti gaasisüsteemi bilanss reaajas. Maagaasiseaduse järgi on igal turuosalisel bilansivastutus. Seejuures tuleb silmas pidada, et kui turuosaline sõlmib gaasi müüjaga avatud tarne lepingu, on bilansivastutuse kandjaks selle turuosalise eest vastutav avatud tarnija. Ka kodutarbijate puhul on bilansivastutuse kandjaks isik, kes neile gaasi müüb. [11] [12]

Bilansivastutus kujutab endast tasakaalu tagamist turuosalisele bilansiperioodiks kokku lepitud gaasikoguse ja bilansiperioodil turuosalisele realselt tarnitud gaasikoguse vahel. Turuosalisele bilansiperioodiks kokkulepitud gaasikoguse tarnimist nimetatakse määratud tarneks ja turuosalisele bilansiperioodil gaasi müümist vastavalt mõõteandmetele, puudujääva gaasi müümist eabilansina või ülejääva gaasi ostmist nimetatakse avatud tarneks. Bilansiperioodi kestus on 24 tundi ehk üks ööpäev (alguseks on vööndiaja järgi hommik kell 07.00 ja lõpuks järgmise päeva hommik kell 07.00). Suveajale üleminekul on bilansiperioodi päev 23-tunnine ja talveajale üleminekul on bilansiperioodi päev 25-tunnine. [12]

Süsteemihaldur arvutab igaks bilansiperioodiks hinnad nii ostetud kui müüdü bilansigaasile. Süsteemihaldur avaldab bilansiperioodi bilansigaasi hinnad bilansiperioodile järgneval tööpäeval kella 17.00-ks. Süsteemihaldur kujundab bilansigaasi hinna selliselt, et see võimaldaks tal:

- katta tasakaalustamisgaasi soetamise ja hoidmise kulud;
- katta bilansigaasi ostmisega seotud põhjendatud kulud;

- katta bilansi selgitamiseks sooritatud kulud;
- katta bilansigaasi ostuks ja müügiks kasutatava reguleeritava vara kapitali kulud, mille arvestus ei kajastu võrguteenuse tariifides;
- tagada põhjendatud tulukus. [13]

Bilansigaasi müügihind (süsteemihalduri müük) bilansiperioodi kohta põhineb vastavaks bilansiperioodiks tasakaalustamisgaasi soodsaima ostupakkumise hinnal (gaasi ost ülekandevõrku), millele süsteemihaldur lisab marginaali põhjendatud kulude katteks. [13]

Bilansigaasi ostuhind (süsteemihalduri ost) bilansiperioodi kohta põhineb vastavaks bilansiperioodiks tasakaalustamisgaasi kalleima müügipakkumise hinnal (gaasi müük ülekandevõrgust välja), millele süsteemihaldur lisab marginaali põhjendatud kulude katteks. Juhul, kui vastavaks bilansiperioodiks tasakaalustamisgaasi müügipakkumisi ei ole, on vastava bilansiperioodi bilansigaasi ostuhinna võrdlushinnaks gaasibörsi vastava bilansiperioodi hind. [13]

Allolevas tabelis on võrreldud süsteemihalduri ostu- ja müügihinda ning lisaks ka börsihinda kahepäevase perioodil (01.12.2018 – 14.12.2018).

Tabel 1.2. Süsteemihalduri ostu ja müügihind ja Baltic Trades börsihind [14] [15]

Kuupäev	SH Ostuhind, €/MWh	SH Müügihind, €/MWh	BT Börsihind, €/MWh
01.12.2018	23,5	26,2	25,8
02.12.2018	23,5	26,2	25,8
03.12.2018	23,5	26,2	25,7
04.12.2018	23,8	26,2	25,7
05.12.2018	23,8	26,2	25,7
06.12.2018	23,8	26,2	25,7
07.12.2018	23,8	26,2	25,7
08.12.2018	23,8	26,2	25,8
09.12.2018	23,8	26,2	25,7
10.12.2018	23,8	26,2	25,6
11.12.2018	23,8	26,2	25,6
12.12.2018	23,8	26,2	25,8
13.12.2018	23,8	26,2	25,8
14.12.2018	23,8	26,2	25,8

1.4 Gaasitarbimise prognoosimise tähtsus

Maagaasinõudluse prognoosimine on energiasektori tõhusa logistika jaoks väga oluline. Maagaas on EL-i eluasemesektori peamine energiaressurss, mis moodustab 37% kogu tarbitavast energiast. Selle tarbimise prognoosimine on seetõttu energiamajanduse ja -transpordi tähtis osa. Oluline näide, kus sellist prognoosimist vaja läheb, on maagaasi transportimine torujuhtmeid pidi. Sellist

transporti tuleb korraldada vastavalt prognoositavale nõudlusele, mida saab ennustada paar päeva ette. Kui tööstusliku gaasinõudluse prognoose on võimalik saada igalt kliendilt eraldi, siis kodukasutajate puhul see nii ei ole. Seda osa gaasitarbimisest tuleb prognoosida mudelite abil. [16]

Gaasinõudluse prognoosimine on tihedalt seotud küttekoormuse prognoosimisega. Seda toetab asjaolu, et 76% EL-i elamumajanduse maagaasist kasutatakse ruumide kütmiseks, samal ajal kui 19% ja 5% kasutatakse vastavalt vee soojendamiseks ja toiduvalmistamiseks. [16]

On väga oluline, et maagaasi tarbimist oleks võimalik prognoosida vastuvõetava täpsusega, et parandada tegevustõhusust, säästa energiat, vähendada kulusid erinevatel tasanditel, hallata tarnelepinguid, omamaist tootmist ja infrastruktuuride planeerimist. Kui üks neist eesmärkidest jääks saavutamata, põhjustaks see kindlasti ebastabiilsust riigi energiasüsteemis. [17] [18]

Seega on ülitähtis täpsete maagaasihindade prognoosimise mudelite väljatöötamine, sest need prognoosid on olulised maagaasi kogu pakkumist ja nõudlust määravate regulatiivsete otsuste tegemisel ehk turuosaliste jaoks. [17] [18]

Gaasi tarnivate ettevõtete ja klientide seisukohalt on täpne gaasitarbimise prognoos ülioluline. Ettevõtetel võimaldab täpne prognoos vähendada bilansikulusid. Tarbijatele aga tähendab hea prognoositavus madalamat gaasihinda, kuna tavaliselt sõltub tarbijatele pakutava maagaasi hind otseselt sellest, kui hästi prognoositav on tarbija. Kui tarbija on raskesti prognoositav, tähendab, et temaga on seotud rohkem bilansikulusid, mis omakorda tähendab kõrgemat hinda.

2. GAASITARBIMISE PROGNOOSIMINE

Käesolevas peatükis kirjeldatakse maagaasi tarbimise prognoosimisega seotud varasemate uurimuste tulemusi. Kirjeldatakse, milliseid meetodeid ja algandmeid kasutati, samuti räägitakse raskustest erinevate meetodite võrdlemisel. Antud peatükis vaadeldakse ka traditsioonilisi meetodeid aegridade ennustamiseks ja nende nõrka jõudlust mittelineaarsete mudelite kirjeldamisel. Lisaks kirjeldatakse masinõppe meetodeid ja neurovõrku, mida kasutati prognoosimudelite loomiseks selles töös. Lisaks on kirjeldatud aegridade analüüsimeetodit, mida hiljem kasutatakse klientide jagamisel klastritesse.

2.1 Maagaasi tarbimisega seotud varasemad uuringud

Aastate jooksul on energianõudlus, eriti maagaasinõudlus, pälvinud üha suuremat huvi, kuna see on muutunud üha olulisemaks. Paljudes uurimustes on prognoositud maagaasi tootmist, tarbimist või nõudlust, hindu ja sissetulekuelastsust, turu volatiilsust ja hinnatõusu mitmetes erinevates sfäärides: maailma tasandil, piirkondlikul tasandil, riiklikul tasandil, linna tasandil, tööstus- ja eluasemesektoris ning üksiktarbijate tasandil. Nendes uurimustes kasutati erinevaid andmeid, mida võib liigitada kolme põhikategooriasse: meteoroloogilised andmed, mineviku andmed ja majandusandmed. [17] [18]

Erinevates töödes on kasutatud erinevaid andmeid. Ka olid andmete allikad ja hulgad erinevad. Seda seetõttu, et vajaminevad andmed määrab rakendatav mudel. [17] [18]

Maagaasi tarbimise prognoosimisega on tegeletud erinevatel aegadel. Mõned autorid keskendusid kõrge eraldusvõimega prognoosimisele nagu tunni-, päeva- ja nädalapõhine prognoosimine, samal ajal kui teised autorid valisid keskmises ja pikas perspektiivis prognoosimise ühekuulistest kuni kvartali- ja aastaprognosideni. Sõltuvalt mudelitüübist kasutati erinevaid lisaandmeid. [17] [18] Varasemates uuringutes kasutati erinevatel ajavahemikel erinevaid andmeid, mis on esitatud allpool.

Aastaprognooosi mudeli jaoks:

Fossiilkütusevarude prognoos, nafta/maagaasi suhe teadaolevates puurkaevudes, keerukad institutsionaalsed tegurid (nt olulised muudatused seadustes ja eeskirjades jne), SKT (sisemajanduse kogutoodang), kütte kraadpäevad (*Heating Degree Days, HDD*), kodumajapidamiste arv linnades, villitud gaasi tarbimine, leibkonna tubade arv ja aastane sissetulek, fossiilkütuse tarbimine, maagaasi hind, kütuse- ja elektrienergia hind, söe hind, naftatoodete hind. [17] [18]

Kuuproгноosi mudeli jaoks:

Kütte kraadpäevad (*Heating Degree Days, HDD*), jahutuse kraadpäevad (*Cooling Degree Days, CDD*), maagaasi hind, sisetulekuindeks, masuudi hind, temperatuur, klientide arv, nafta hind, suhteline õhuniiskus, tuulekiirus, nädalapäev, puhke- või tööpäeva muutuja, igakuine keskmine maagaasi tarbimine kliendi kohta, keskmine palk, leibkonna sisetulek, soojendatava(te) ruumi(de) suurus, rentnike arv, kütteseadme tõhusus, kütteseadme vanus. [17] [18]

Päevaproгноosi mudeli jaoks:

Kütte kraadpäev (*Cooling Degree Day, CDD*), temperatuur, tuulekiirus, nädalapäev, kuupäev, kuu, eelmise päeva tarbimine, üle-eelmise päeva tarbimine, eelmise päeva temperatuur, üle-eelmise päeva temperatuur, eelmise päeva tuulekiirus, üle-eelmise päeva tuulekiirus, tööpäevad – nädalavahetused, pühad, eksperdi hinnangud küttehooaja algus-/lõppkuupäeva kohta, päevane miinimumtemperatuur, päevane maksimumtemperatuur, klientide koguarv, eelmise aasta sama päeva tarbimine. [17] [18]

Tunniproгноosi mudeli jaoks:

Temperatuur, tuulekiirus, aastaaeg, nädalapäev, pühad. [17] [18]

Testitud on erinevaid mudeleid ja meetodeid, nagu:

- Hubberti mudel ja selle variatsioonid;
- statistilised mudelid;
- tehislikud närvivõrgud;
- ökonomeetrilised mudelid;
- matemaatilised mudelid;
- kombineeritud/hübriidmudelid. [17] [18]

Maagaasi tarbimise prognoosimise meetodid on erinevad, seega on erinevad ka hindamiskriteeriumid. Teadlased on eriarvamusel selles, mida pidada heaks tulemusnäitajaks ja mida mitte, kuid kõige olulisem mõõdupuu on prognoosimise täpsus. Sobivat täpsuse näitajat, mis oleks kõigi teadlaste poolt ühtviisi tunnustatud, pole antud küsimuses siiski veel leitud. Mõõtetäpsust defineeritakse tulevase tegeliku ja prognoositud väärtuse vahena. Teadlased on välja toonud mitmeid tulemusnäitajaid, millest igal on oma eelised ja puudused. [17] [18]

Vaatamata sellele, et prognoosimisvahendid tehniliselt aina täiustuvad, ei ole teadlased veel kindlad, millised mudelid on teistest paremad. Lisaks puuduvad selged kriteeriumid asjakohaste muutujate ja iseloomustavate tunnuste valimiseks prognoosimudelite loomisel. Ka on mudelite

prognoosimise täpsuse kohta saadaolev teave mitmetes aspektides ebarahuldav saadud tulemuste selge mõistmise tagamiseks. Enamikul juhtudel on väga raske ja vahel isegi võimatu võrrelda objektiivselt kaht või enam uurimistulemust. Seda seetõttu, et iga uurimus lähtub erinevatest hindamiskriteeriumidest. [17] [18]

Erinevad autorid on käsitlenud erinevaid parameetreid ja lähenemisviise. Ühe autori poolt tähelepanuta jäetud parameetrit peetakse teises uuringus esmatähtsaks. Igal juhul ei ole eesmärgid alati ühesugused ja lisaks sellele ei ole andmed alati nii kättesaadavad nagu võiksid, mis toob kaasa lahknevad lõplikud seisukohad. Siiski töötatakse usinalt välja uusi mudeleid ja võetakse arvesse veelgi enam parameetreid, et tulevikusuundumusi oleks võimalik täpsemalt prognoosida. [17] [18]

2.2 Aegridade prognoosimiseks traditsioonilised meetodid

Traditsioonilised meetodid aegridade prognoosimiseks on libisev keskmine (*Moving Average, MA*), autoregressioon (*Autoregression, AR*), nende kombinatsioon (*Autoegressive–Moving–Average, ARMA*) ja nende variatsioonid. Lisaks ka eksponentsiaalne silumine (*Exponential smoothing, ES*). Selguse huvides on kõik allpool esitatud mudelite näited tehtud ühe mittelineaarse juhuslikult genereeritud aegrea alusel. [19]

2.2.1 Autoregressiivne mudel

Autoregressiivne mudel õpib üksteisele järgnevatest ajamomentidest, kasutades eelmistes hetkedes toimunu andmeid regressioonimudeli sisenditena selleks, et prognoosida järgmise ajamomendi väärtust. Autoregressiooni modelleerimine keskendub eelmiste ajamomentide (ajamuutujate) vaatluste vahelise korrelatsiooni mõõtmisele järgmise ajahetke väärtuse (väljundi) prognoosimiseks. Kui mõlemad muutujad muutuvad samas suunas, näiteks kasvavad või kahanevad koos, on tegemist positiivse korrelatsiooniga. Kui muutujad liiguvad väärtuste muutudes vastassuunas, näiteks üks kasvab ja teine kahaneb, siis on korrelatsioon negatiivne. Mõlemal juhul on võimalik väljundi ja eelmise muutuja vahelist korrelatsiooni statistika abil hõlpsasti mõõta. Mida kõrgem on see korrelatsioon, olgu positiivne või negatiivne, seda tõenäolisem on, et minevik ennustab tulevikku. Kuna see korrelatsioon esineb muutuja ja tema enda vahel eelmistes ajamomentides, nimetatakse seda autokorrelatsiooniks. [20]

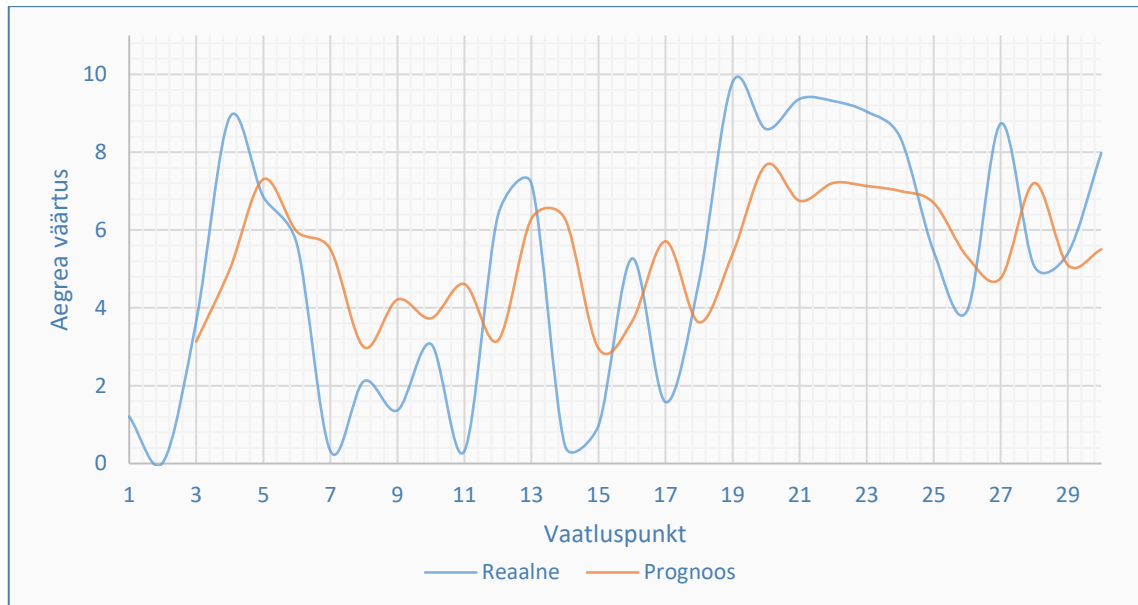
Tähistus AR(p) tähistab autoregressiivset mudelit järguga p. AR(p), see mudel kirjutatakse kujul:

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \quad (2.1)$$

kus $\varphi_1, \dots, \varphi_p$ – mudeli parameetrid,

c – konstant,

ε – valge müra.



Joonis 2.1. Autoregressiivse mudeli graafiline näide

2.2.2 Libiseva keskmise mudel

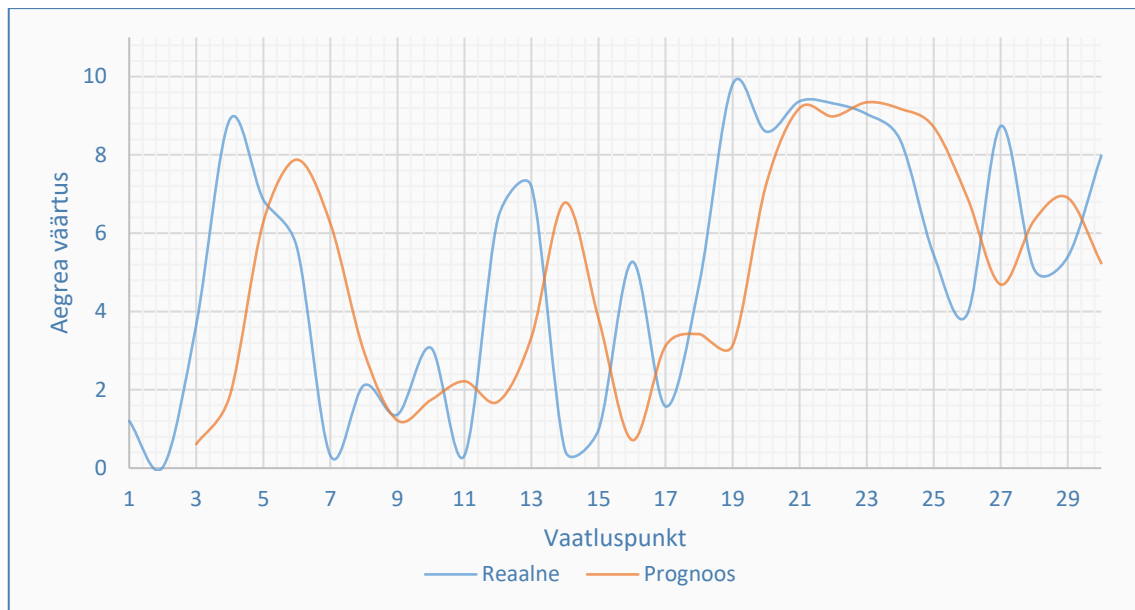
Aegridade analüüsis kasutatakse libiseva keskmise mudelit, mida tuntakse ka libiseva keskmise protsessi nime all, ühemõõtmeliste aegridade modelleerimiseks. Libiseva keskmise mudeli kohaselt sõltub väljundi muutuja lineaarselt juhusliku (ebatäpselt ennustatava) protsessi praegusest ja mitmetest mineviku väärtustest. [21]

Tähis MA(q) tähistab järku q libiseva keskmise mudelit:

$$X_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2.2)$$

kus $\theta_1, \dots, \theta_q$ – mudeli parameetrid,

$\varepsilon_{t-i}, \varepsilon_t$ – valge müra.



Joonis 2.2. Libiseva keskmise mudeli graafiline näide

2.2.3 Autoregressiivne libiseva keskmise mudel

Aegridade statistilises analüüsis saab autoregressiivse libiseva keskmisega mudelitega lühiajaliselt kirjeldada (nõrgalt) statsionaarseid juhuslikke protsesse kahe polünoomi – autoregressiooni (AR) ja libiseva keskmise (MA) – suhtes. Peter Whittle kirjeldas ARMA mudelit 1951. aastal oma dissertatsioonis „Hüpoteeside kontrollimine aegridade analüüsis“ ja laiemalt tutvustasid seda George E. P. Box ja Gwilym Jenkins oma 1970. aastal ilmunud teoses. [21] [22] [23]

Tähis $ARMA(p,q)$ viitab p järku autoregressiivse komponendi ja q järku libiseva keskmise komponendiga mudelile. See mudel sisaldab $AR(p)$ ja $MA(q)$ mudeleid:

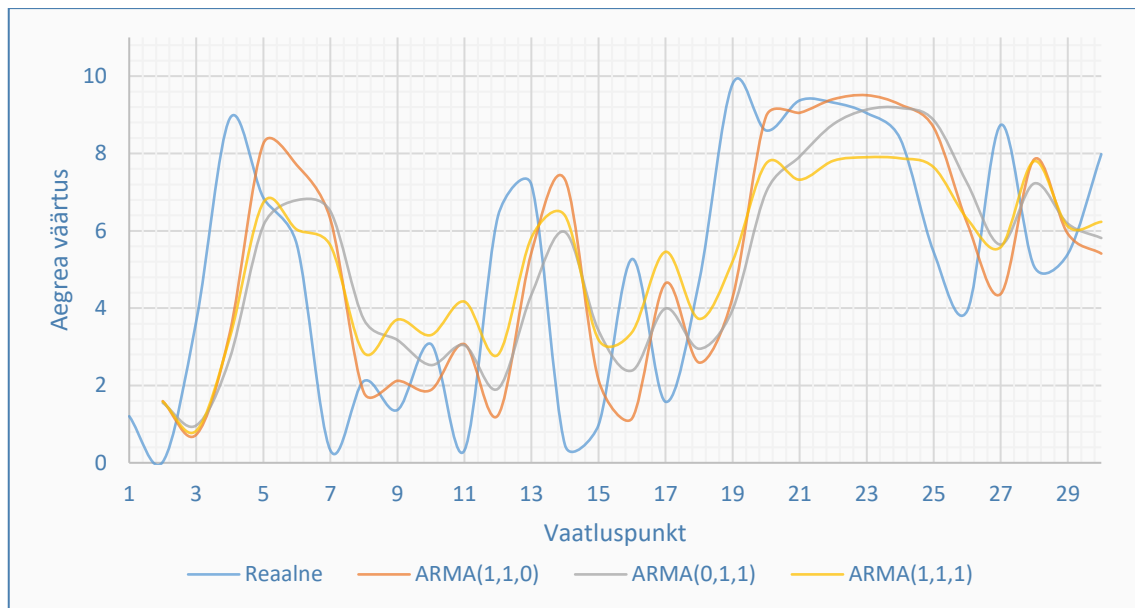
$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (2.3)$$

kus $\varphi_1, \dots, \varphi_p$ – autoregressiooni parameetrid,

$\theta_1, \dots, \theta_q$ – libiseva keskmise parameetrid,

c – konstant,

$\varepsilon_{t-i}, \varepsilon_t$ – valge müra.



Joonis 2.3. Kolme erineva ARMA mudeli graafiline näide

2.2.4 Eksponentsiaalne silumine

Eksponentsiaalne silumine on tavatehnika aegridade andmete silumiseks eksponentsiaalse aknafunktsiooni abil. Kui lihtsa libiseva keskmise puhul käsitletakse varasemaid vaatlusi võrdsetena, siis ajas eksponentsiaalselt vähenevate suuruste määramiseks kasutatakse eksponentsiaalseid funktsioone.

See on kergesti õpitav ja hõlpsasti rakendatav protseduur varasemate eelduste põhjal mõnede otsuste tegemiseks näiteks hooajalisuse kohta. Eksponentsiaalset silumist kasutatakse aegridade analüüsis palju. [21] [24]

Üldine võrrand prognoosi arvutamiseks eksponentsiaalse silumise abil:

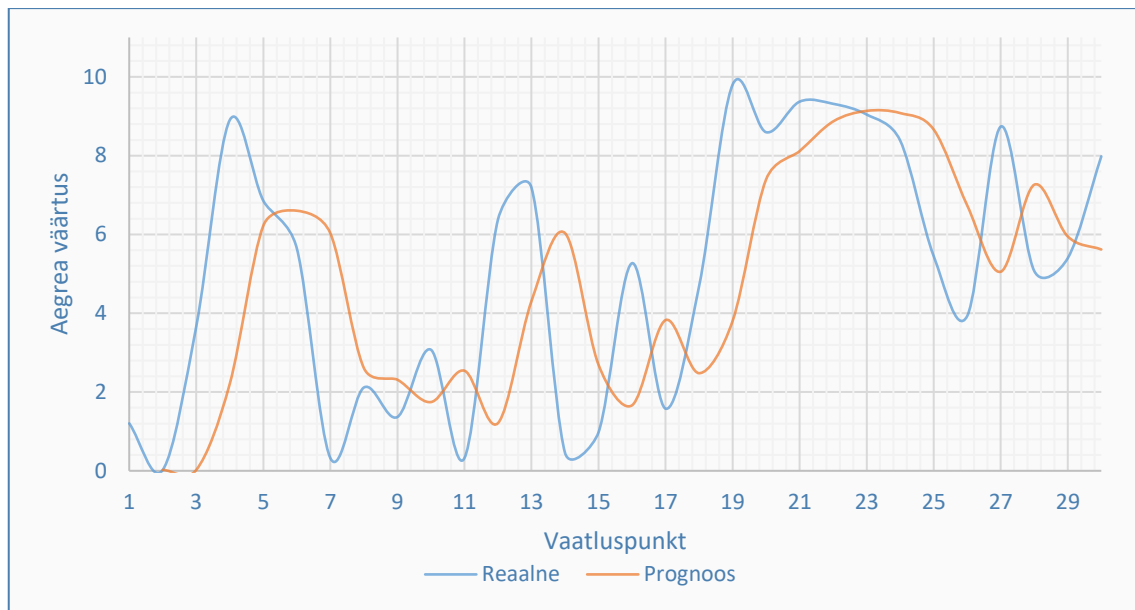
$$P_{t+1} = P_t + \alpha(R_t - P_t) \tag{2.4}$$

kus P_{t+1} – prognoositud väärtus ajal $t + 1$,

P_t – prognoositud väärtus ajal t ,

R_t – reaalne väärtus ajal t ,

α - silumise koeffitsient.



Joonis 2.4. Eksponentsiaalse silumise mudeli graafiline näide

2.2.5 Järeldus

Nagu ülaltoodud näidetest näha, ei suuda klassikalised mudelid mittelineaarset aegrida täpselt ennustada. Tihti prognoosivead ületavad 50%. Need mudelid on lineaarsed ja sageli ei arvesta mõjutavaid parameetreid, vaid tuginevad ainult aegride varasematele väärtustele, seepärast on nende võimekus puudulik isegi lihtsate mittelineaarsete mudelite jaoks. See muudab nende kohaldamise äärmiselt ebapraktiliseks. Kuna gaasitarbimine sõltub paljudest välistest parameetritest, peaksid tarbimist prognoosivad mudelid võimaldama nende parameetrite kasutamist. [19]

2.3 Lõputöös kasutatud masinõppel põhinevad meetodid

Käesolevas peatükis kirjeldatakse üldisi põhimõtteid meetoditest, mida antud töös kasutati, et luua mudeleid maagaasi tarbimise prognoosimiseks.

2.3.1 Lineaarne regressioonimudel

Lineaarne regressioonimudel (*Linear Regression model, LR*) kirjeldab sõltuva tunnuse y -i ja ühe või rohkema sõltumatu tunnuse X -i suhet. Sõltuvat tunnust kutsutakse ka uuritavaks tunnuseks või funktsiooniks. Sõltumatuid tunnuseid nimetatakse ka seletavateks tunnusteks või prediktoriteks. [25]

Mitme muutujaga lineaarsed regressioonimudelid kirjeldatakse võrrandi 2.5 abil:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad (2.5)$$

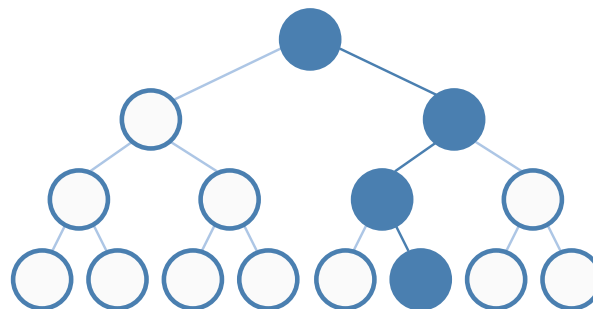
$$i = 1, \dots, n \quad (2.6)$$

kus y_i – uuritav tunnus,
 b_i – mudeli tegurid,
 X_{ij} – prediktori väärtus,
 ε_i – valge müra.

Kui mudel sisaldab ainult ühte prediktori ($p = 1$), siis mudelit nimetatakse lihtsaks lineaarse regressioonimudeliks. [25]

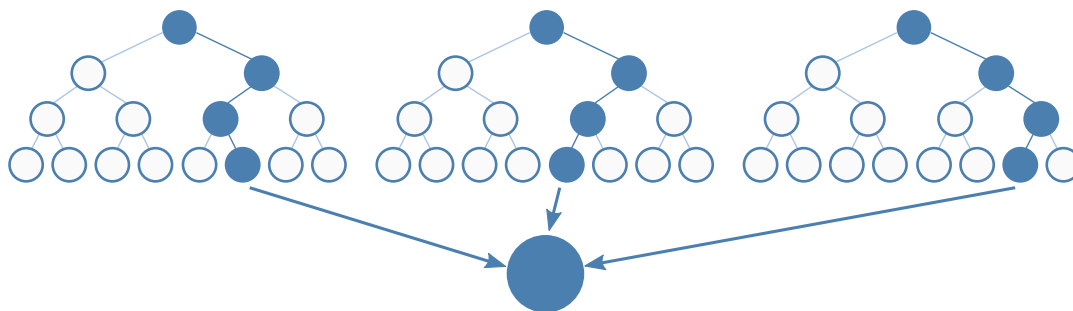
2.3.2 Otsuspuu ja otsuspuu ansambli mudelid

Otsustuspuu (*Decision Tree, DT*) on juhendatud õppimise hierarhiline mudel, mille algusala märgib vähemate sammude hulk rekursiivsete hargnemiste järjestikuses reas. Otsustuspuu koosneb sisemistest otsustussõlmedest ja terminaalsõlmedest ehk lehtedest. [26]



Joonis 2.5. Otsuspuu mudeli põhimõte

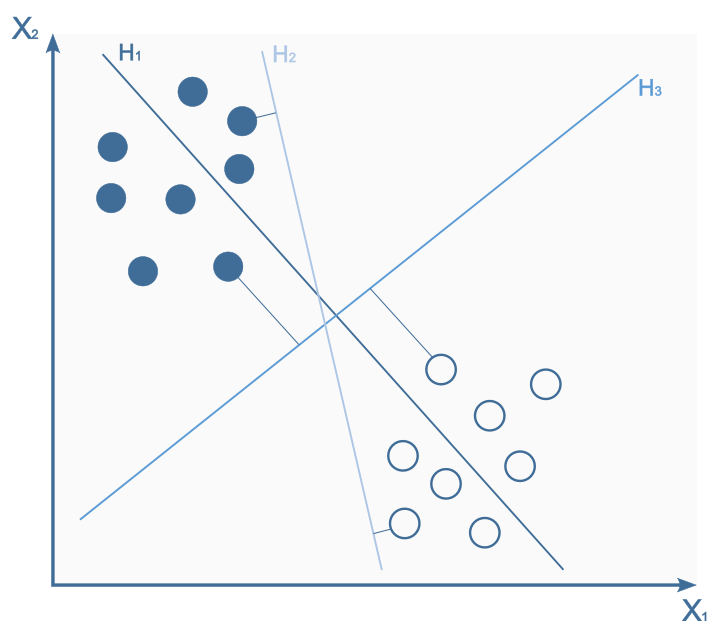
Otsuspuu ansambelmeetodid (*Decision Tree Ensembles*) kasutavad mitmeid erinevaid õppimise algoritme, saavutamaks paremat prognoosimisvõimet kui see oleks võimalik üksiku õppimisalgoritmi abil. Ansamblid kombineerivad paljusid hüpoteese, et esitada (loodetavasti) parem hüpotees. [27] [28] [29]



Joonis 2.6. Kolme otsuspuuga ansambli mudeli näide

2.3.3 Tugivektoriregressiooni mudel

Tugivektori masina (*Support Vector Machine, SVM*) idee seisneb selles, et ta konstrueerib hüpertasandi või hüpertasanditeede kogumi, mida saab kasutada klassifitseerimiseks, regressiooniks või muudeks ülesanneteks. SVM põhimõte seisneb hüpertasandi leidmisel, millel on suurim vahekaugus lähima treening-andmepunktini. Klasteritesse jagamine tugivektori abil on näidatud joonisel 2.7. [30]



Joonis 2.7. Kahe klassi hüpertasanditega jagamine [30]

Jooniselt on näha kolm hüpertasandit, siit me näeme, et esimene hüpertasand H_1 ei suuda klasse eraldada. H_2 teeb seda, kuid väikese vahekaugusega. H_3 eraldab klasse maksimaalse vahekaugusega.

Vladimir N. Vapnik, Harris Drucker, Christopher J. C. Burges, Linda Kaufman ja Alexander J. Smola pakkusid 1996. aastal välja regressioonis kasutatava tugivektor-masina versiooni. Seda meetodit nimetatakse tugivektorregressiooniks (*Support-Vector Regression, SVR*) või TVR. [31] [32]

Algupärase TVR-i treenimine seisneb järgnevas:

$$\text{võrrandi jaoks: } y = wx + b \quad (2.7)$$

$$\text{minimeerida: } \frac{1}{2} \|w\|^2 \quad (2.8)$$

$$\text{arvestades: } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \quad (2.9)$$

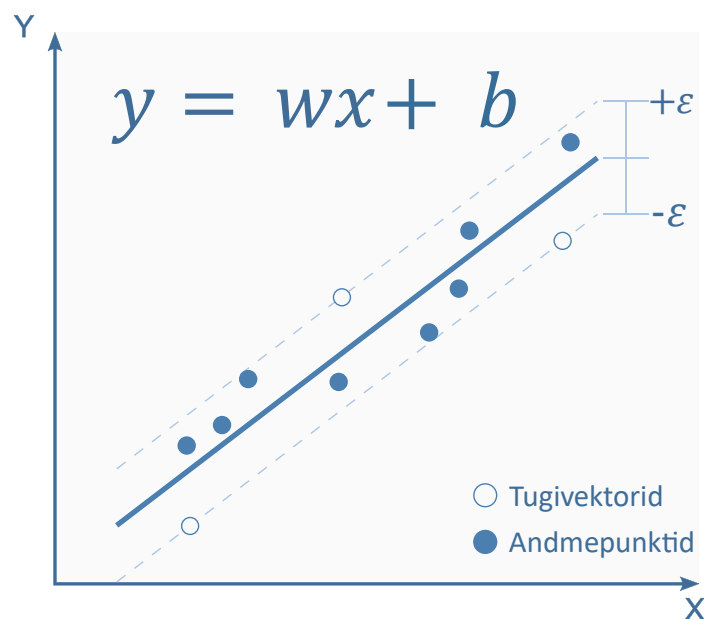
kus x_i – treeningandmed,

y_i – sihtväärtus,

$\langle w, x_i \rangle + b$ – antud andmete prognoos,

ε – vaba parameeter, mis toimib lävena.

Kõik prognoosid peavad jääma tõeste prognooside ε vahemikku. [31] [32]



Joonis 2.8. TVR mudeli põhimõte [31] [32]

2.3.4 Gaussi protsessi regressiooni mudel

Geostatistikast pärit *kriging*-meetod ehk Gaussi protsessi regressioon (*Gaussian process regression*) on statistikas kasutatav interpolatsioonimeetod, kus interpoleeritud väärtusi modelleeritakse eelnevate kovariatsioonidega määratud Gaussi protsesside abil. Gaussi protsess on täielikult määratletud selle keskvärtuse ning kovariatsioonifunktsiooniga ehk tuumaga. Gaussi protsessi valim annab iga kord erineva funktsiooni. [33] Krigimise üldpõhimõtet kirjeldatakse allpool koos näitega.

Oletame, et meil on 30 punkti, millest ainult 18 väärtuseid on teada, nende väärtuste Y parameetrid:

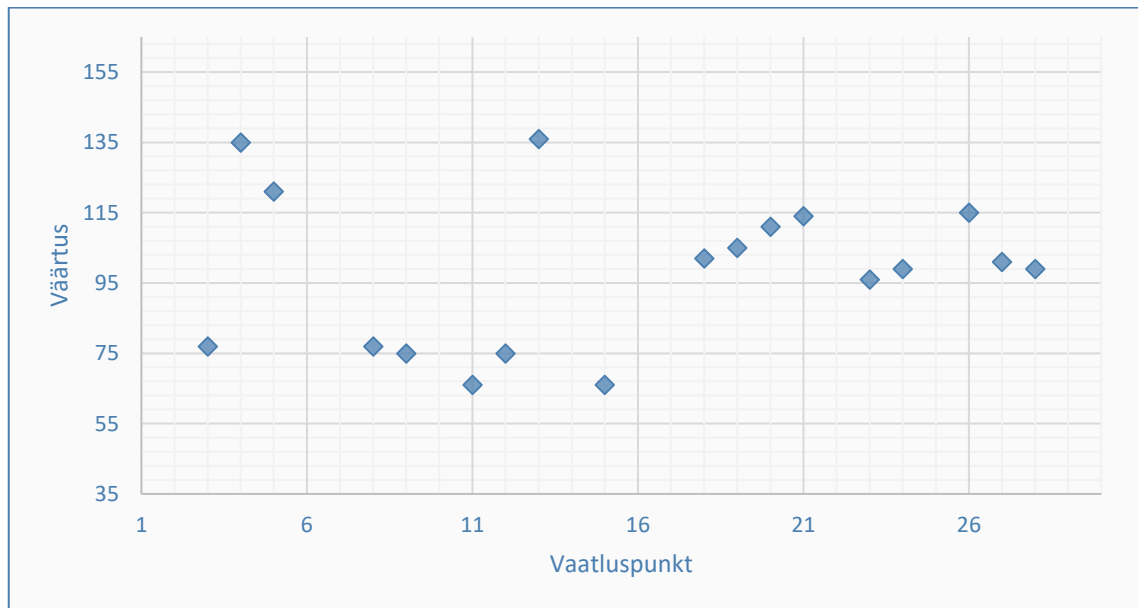
$$\bar{Y} = 100 \quad (2.10)$$

$$\delta = 20 \quad (2.11)$$

kus \bar{Y} - Y väärtuste keskmine,

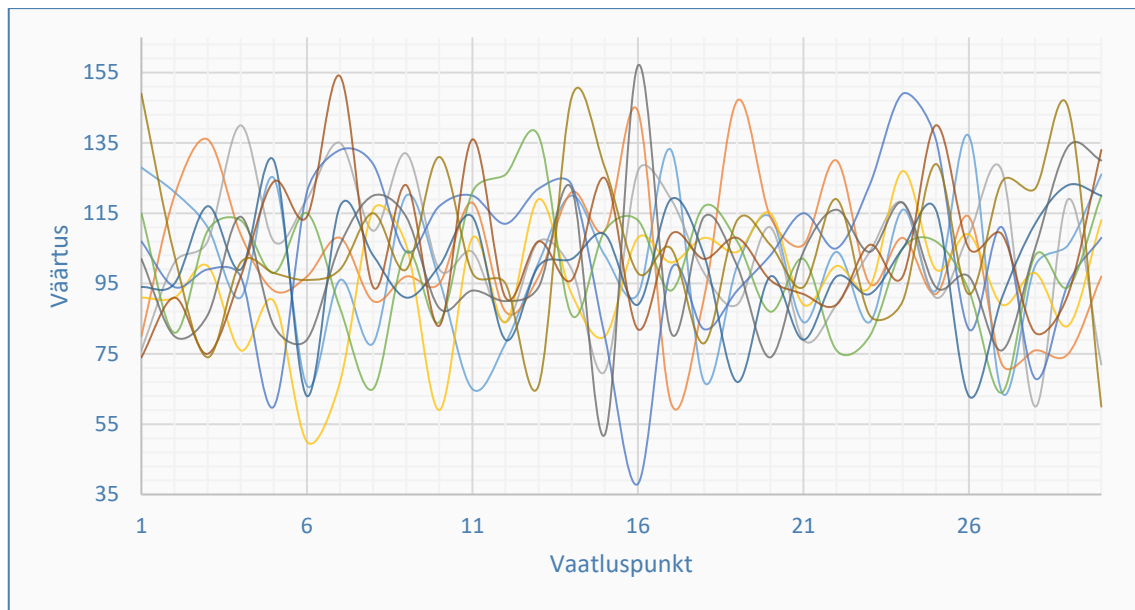
δ – standarthalve.

Kõik teadaolevad punktid on esitatud graafikus (Joonis 2.9).



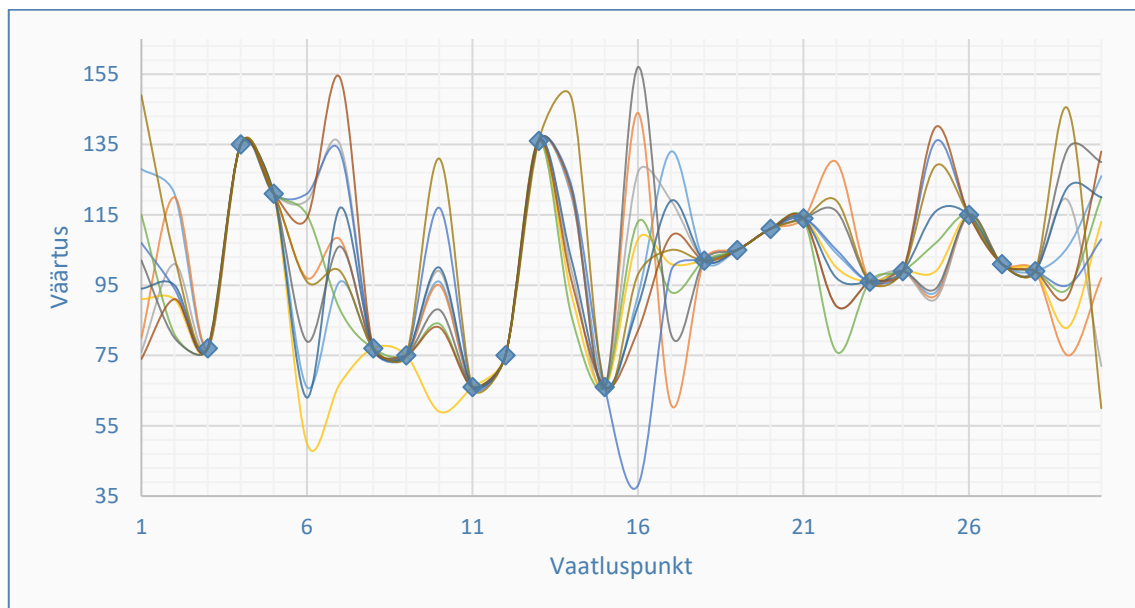
Joonis 2.9. Teadaolevad punktid

Vahepunktide leidmiseks genereeritakse juhuslikult normaaljaotustega aegread, mille parameetrid on samad mis teadaolevates punktides. Allpool on esitatud graafik 10 gaussi meetodiga genereeritud aegridadega (Joonis 2.10).



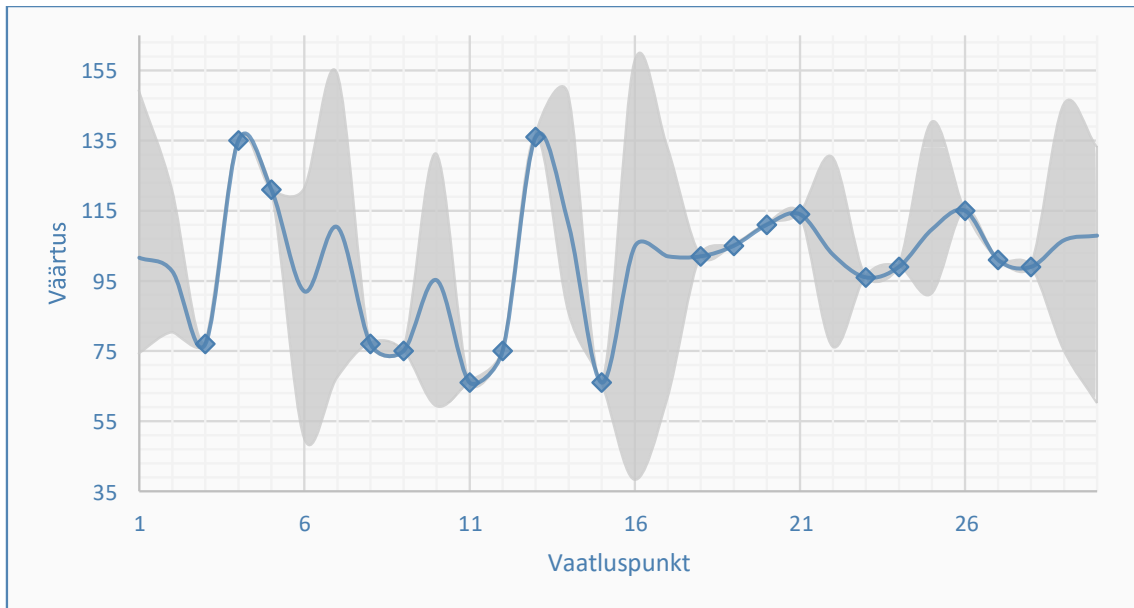
Joonis 2.10. Normaalkaotusega juhuslikult genereeritud aegread

Kombineerime Gaussi protsessiga saadud väärtused ja teatud väärtused ja saame graafiku, mis on esitatud joonisel 2.11.



Joonis 2.11. Teatud punktid ja juhuslikult genereeritud aegread kombineeritud

Gaussi meetodi abil saadud tulemuste maksimaalsed ja minimaalsed väärtused on vastavalt usaldusvahemiku maksimaalne ja minimaalne piir ning nende keskmine väärtus on prognoos.

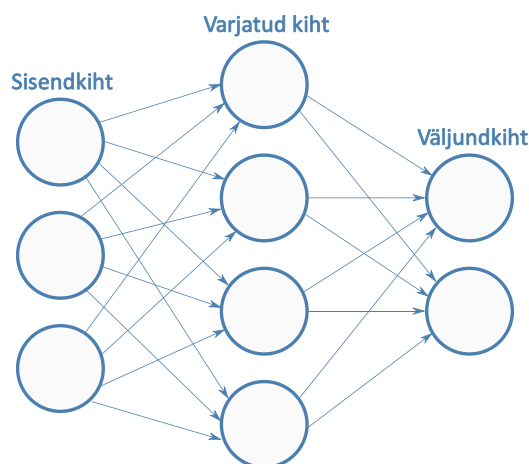


Joonis 2.12. Gaussi protsessi abil tehtud prognoos

Joonisel 2.12 hall ala on usaldusvahemik ja sinine joon on Gaussi protsessi abil tehtud prognoos.

2.3.5 Tehislik närvivõrk

Tehislik närvivõrk (*Neural Network, NN*) on süsteem, mis põhineb bioloogilisel närvivõrgul (nagu aju). Sellised süsteemid „õpivad“ ülesandeid täitma näidete alusel, olemata üldjuhul programmeeritud mingite konkreetsete ülesannete täitmisega seotud reeglitega. Tavaliselt vaadeldakse närvivõrku kui musta kasti, millega saab peaaegu iga süsteemi jaoks leida ligikaudse lahenduse. [34] [35]



Joonis 2.13. Tehisliku närvivõrgu põhimõte

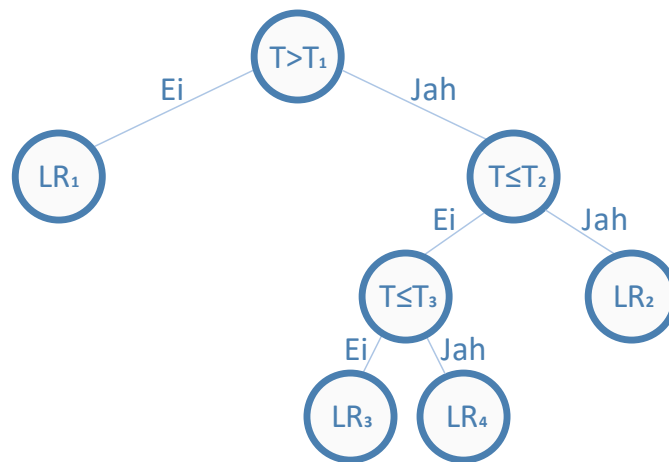
Tehislik närvivõrk on omavahel seotud sõlmede kogum, mille aluseks on ajus olevate neuronite lihtsustatud mudel. Iga ringikujuline sõlm tähistab skeemil tehislikku neuronit ja nooleke näitab ühe tehisliku neuroni väljundi ühendust teise sisendiga. [36] [37]

Mudelit, mida kasutati, nimetatakse NARX RNN-mudeliks (*Nonlinear Autoregressive Recurrent Neural Networks with Exogenous Input*), see on mittelineaarne autoregressiooni mudel, millel on välised sisendid. See tähendab, et antud mudel seostab aegrea hetkeväärtust nii

- sama rea minevikuväärtuste;
- mõjutavate (välise) ridade — s.o väliselt determineeritud ridade, mis mõjutavad kõnealust rida — hetke- ja minevikuväärtustega. [36] [37]

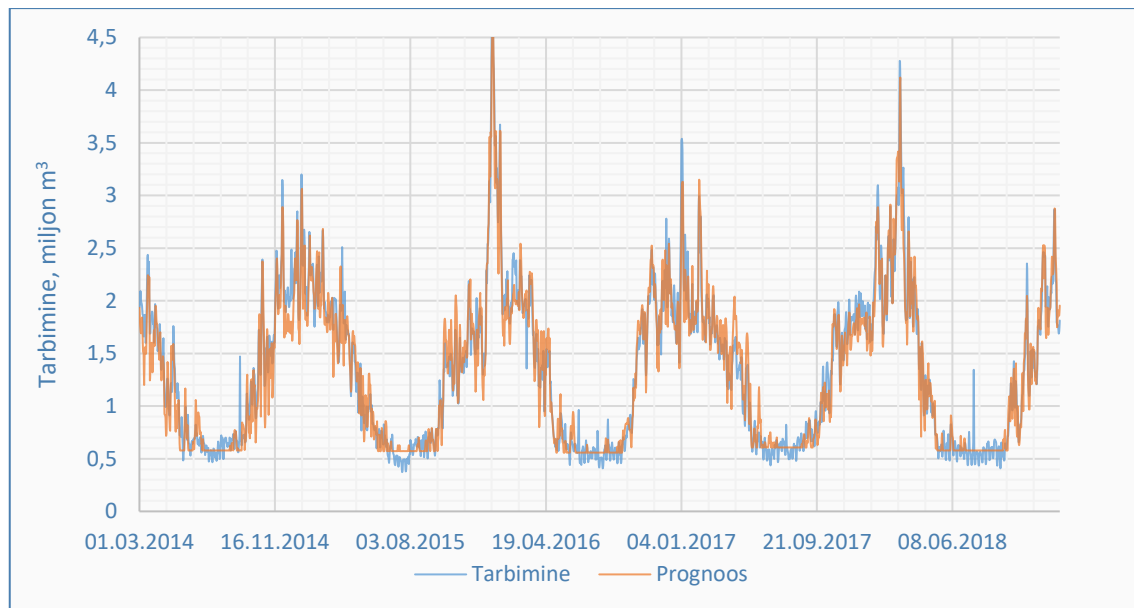
2.4 Eesti Gaasi mudel

Eesti Gaasi mudel on ühe parameetriga (temperatuur) otsuspuu mudel, kus on lineaarse regressiooni mudelid terminaallehtedes. Seda mudeli tihti nimetakse lineaarse regressiooni mudelite puuks (*Linear Regression Model Tree, LRMT, LMT*). Tegemist on kütteseadmete töö juhtimise automaatikas kasutatava printsiibiga, mis eeldab, et eksisteerib mingi välisõhu temperatuur, millest suuremate väärtuste juures seos välisõhu temperatuuri ja kütteks kuuluva energia vahel puudub ning energia kogus on ajas konstantne. Sellest temperatuurist väiksemat väärtuste juures eksisteerib aga lineaarne seos välisõhu temperatuuri ja kütteks kuluva energia vahel. Antud mudel teeb päev-ette prognoosi, tuginedes temperatuuri prognoosile.



Joonis 2.14. Eesti Gaasi mudeli põhimõte

Joonisel 2.14 on T – välisõhu temperatuuri prognoos, T_1, T_2, T_3 – teatud temperatuuri piirväärtused, LR_1, LR_2, LR_3, LR_4 – lineaarsed prognoosimudelid.



Joonis 2.15. Eesti Gaasi mudeliga prognoos ja tegelik tarbimine 01.03.2014 – 31.12.2018

Jooniselt 2.15 on näha, et soojematel kuudel prognoositud tarbimine on konstantne, see on põhjendatav ülalpool selgitatud asjaoluga, et nendel perioodidel on välisteõhu temperatuur kõrgem kui teatud väärtus, ja mudel kasutab ainult baaskoormust. Mudeli matemaatiline ja statistiline analüüs on toodud kolmandas peatükis.

2.5 Aegridade klasterdamine

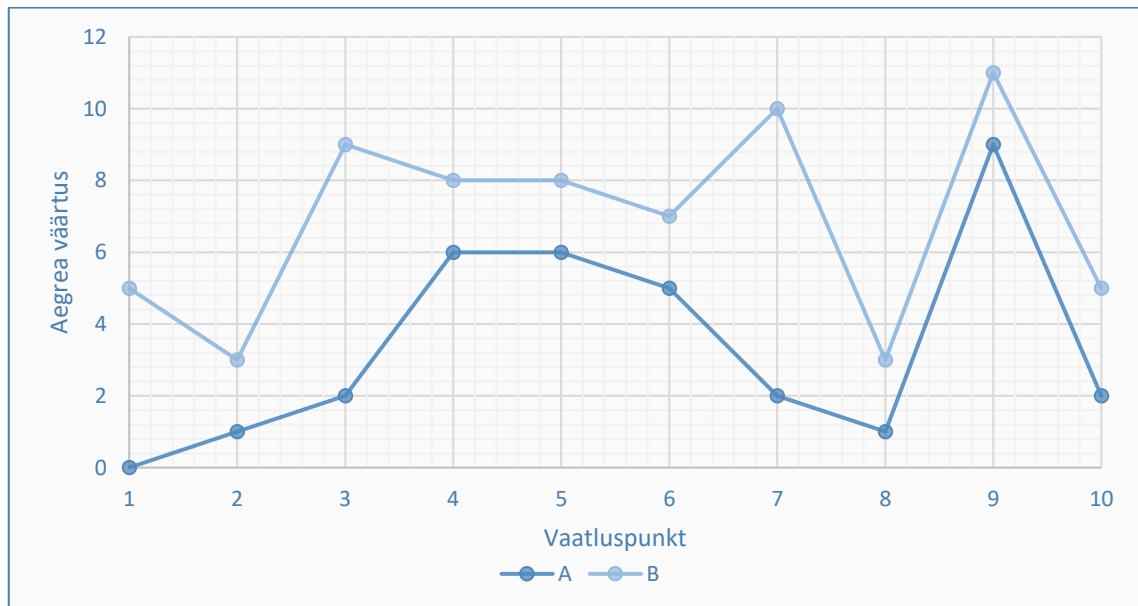
Kliendid jaotatakse tarbimise alusel enamasti eraklientideks, äriklientideks ja tööstusklientideks. See meetod on väga lihtne, sest gaasitarbija teab alati, millisesse klienditüüpi ostja kuulub. See meetod aga pole siiski alati kasulik, sest sageli võib eri kliendirühmade käitumine gaasi tarbimisel olla sarnane. Näiteks võib äriklientide gaasitarbimine sarnaneda eraklientide omaga. Klientide rühmitamine mitte ainult ei lihtsusta iga rühma tarbimise prognoosimist, vaid aitab ka täpsemalt hinnata, millised kliendid on paremini prognoositavad ja millised mitte. Sellest oleneb hind, mida gaasitarbija kliendile pakub. Mida raskem on kliendi gaasitarbimise prognoosimine, seda suuremad on sellega seotud bilansikulud ja seega on ka on gaasi hind kliendile kõrgem. Aegridade analüüsimiseks kasutatud meetodi nimeks on dünaamiline ajadeformatsioon (*Dynamic Time Warping, DTW*) ja see meetod on ette nähtud ajajärjestuste muutmiseks, tuginedes suurima sarnasuse otsimisele. Antud meetod põhineb distantssmaatriksi koostamisel esimese ja teise aegjärjestuse vahel. DTW meetodi eripäraks on, et see võimaldab võrrelda aegridu, mis on ajaskaalal nihutatud. [38] Meetodi põhimõtet on selgitatud allpool.

Kujutame ette, et meil on kaks aegrida, A ja B , mille väärtused on toodud tabelis 2.1.

Tabel 2.1. A ja B aegread

A	B
0	5
1	3
2	9
6	8
6	8
5	7
2	10
1	3
9	11
2	5

Ülaltoodud tabelis esitatud andmete põhjal koostati graafik (Joonis 2.16).



Joonis 2.16. A ja B aegride graafikud

Seda meetodit kasutades peame leidma minimaalsed vahekaugused kahe graafiku punktide vahel ja iga punkti saab kasutada mitu korda, kuid pole ühtki punkti, mis oleks ühendamata. Minimaalsete vahekauguste arvutamiseks ehitatakse nn distantssmaatriks. Täitmata distantssmaatriks on esitatud tabelis 2.2.

Tabel 2.2. Täitmata distantssmaatriks

10	2										
9	9										
8	1										
7	2										
6	5										
5	6										
4	6										
3	2										
2	1	A_2B_1	A_2B_2								
1	0	A_1B_1	A_1B_2								
i	A/B	5	3	9	8	8	7	10	3	11	5
	j	1	2	3	4	5	6	7	8	9	10

Distantssmaatriksi täitmiseks kasutatav üldvalem on järgmine:

$$A_iB_j = |A_i - B_j| + \min([A_{i-1}B_{j-1}], [A_{i-1}B_j], [A_iB_{j-1}]) \quad (2.12)$$

kus A_iB_j – leitav suurus matriksi täitmiseks,

A_i – aegrea A väärtus punktis i ,

B_j – aegrea B väärtus punktis j .

Nelja esimese numbri A_1B_1 , A_1B_2 , A_2B_1 ja A_2B_2 arvutamise näited:

$$A_1B_1 = |A_1 - B_1| + \min([0], [0], [0]) = |A_1 - B_1| = |0 - 5| = 5 \quad (2.13)$$

$$A_1B_2 = |A_1 - B_2| + \min([A_1B_1], [0], [0]) = |A_1 - B_2| + A_1B_1 = |0 - 3| + 5 = 8 \quad (2.14)$$

$$A_2B_1 = |A_2 - B_1| + \min([A_1B_1], [0], [0]) = |A_2 - B_1| + A_1B_1 = |1 - 5| + 5 = 9 \quad (2.15)$$

$$A_2B_2 = |A_2 - B_2| + \min([A_1B_1], [A_1B_2], [A_2B_1]) = |1 - 3| + 5 = 7 \quad (2.16)$$

Sel viisil täidetakse distantssmaatriksi täielikult, nagu on näidatud järgnevas tabelis 2.3.

Tabel 2.3. Täidetud distantsmaatriks

10	2	28	24	24	23	24	24	29	23	32	29
9	9	25	23	17	18	19	21	22	28	26	30
8	1	21	17	23	29	29	27	30	24	32	35
7	2	17	15	22	22	22	21	25	22	31	31
6	5	14	16	18	16	16	17	21	22	28	28
5	6	14	14	14	13	15	16	20	23	28	29
4	6	13	11	11	13	15	16	20	23	28	29
3	2	12	8	14	20	26	31	39	40	49	52
2	1	9	7	15	22	29	35	44	46	56	60
1	0	5	8	17	25	33	40	50	53	64	69
<i>i</i>	<i>A/B</i>	5	3	9	8	8	7	10	3	11	5
	<i>j</i>	1	2	3	4	5	6	7	8	9	10

Järgmisena ühendame andmed kahanevas järjekorras, valides lähima numbriga madalaima väärtusega, nagu tabelis 2.4 näidatud.

Tabel 2.4. Väärtuste ühendamine

10	2	28	24	24	23	24	24	29	23	32	29
9	9	25	23	17	18	19	21	22	28	26	30
8	1	21	17	23	29	29	27	30	24	32	35
7	2	17	15	22	22	22	21	25	22	31	31
6	5	14	16	18	16	16	17	21	22	28	28
5	6	14	14	14	13	15	16	20	23	28	29
4	6	13	11	11	13	15	16	20	23	28	29
3	2	12	8	14	20	26	31	39	40	49	52
2	1	9	7	15	22	29	35	44	46	56	60
1	0	5	8	17	25	33	40	50	53	64	69
<i>i</i>	<i>A/B</i>	5	3	9	8	8	7	10	3	11	5
	<i>j</i>	1	2	3	4	5	6	7	8	9	10

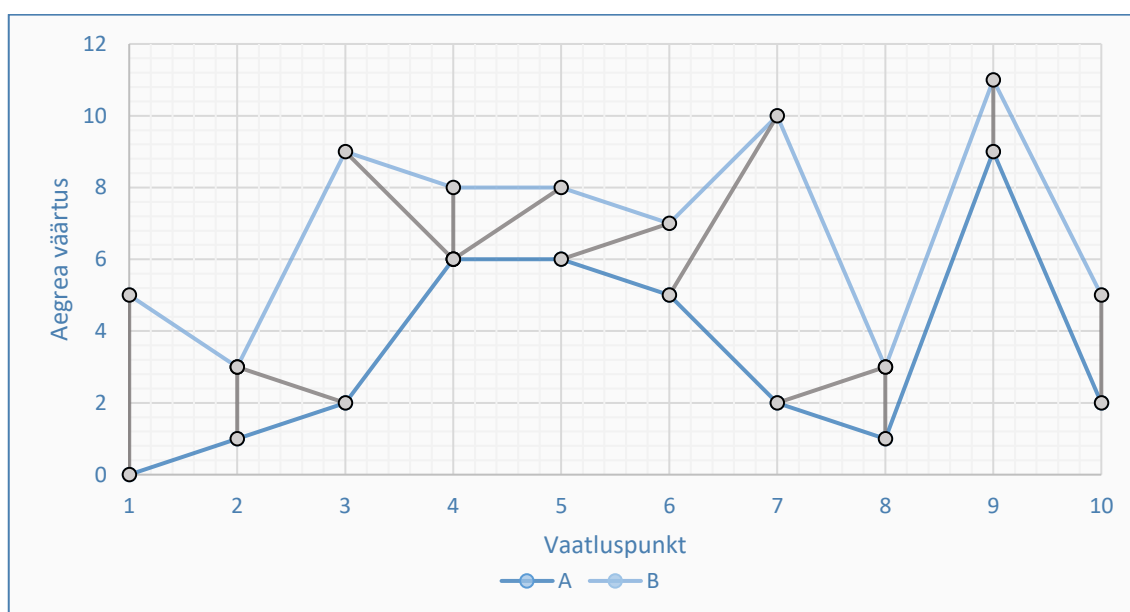
NB! Sageli on punktide ühendamiseks mitu võimalust, sest vahekaugused on sarnased. Mõlemad variandid on õiged ja neid saab kasutada.

Ühendades punktid ülaltoodud meetodi abil, saame järgmise tabeli 2.5.

Tabel 2.5. Ühendatud väärtused

10	2	28	24	24	23	24	24	29	23	32	29
9	9	25	23	17	18	19	21	22	28	26	30
8	1	21	17	23	29	29	27	30	24	32	35
7	2	17	15	22	22	22	21	25	22	31	31
6	5	14	16	18	16	16	17	21	22	28	28
5	6	14	14	14	13	15	16	20	23	28	29
4	6	13	11	11	13	15	16	20	23	28	29
3	2	12	8	14	20	26	31	39	40	49	52
2	1	9	7	15	22	29	35	44	46	56	60
1	0	5	8	17	25	33	40	50	53	64	69
i	A/B	5	3	9	8	8	7	10	3	11	5
	j	1	2	3	4	5	6	7	8	9	10

Nüüd koostame tabelist 2.5 saadud tulemuste põhjal graafiku ja saame ühenduspunktide visualiseeringu (Joonis 2.17).



Joonis 2.17. Seotud A ja B aegridade graafikud

Seega saab seda meetodit kasutades võrrelda aegridu ja leida sarnaseid. Sarnased aegread rühmitatakse hiljem klasteriteks edasiseks kasutamiseks prognoosimisel.

2.6 Peatüki kokkuvõte

Gaasitarbimise prognoosimine on keeruline protsess, sest tarbimine sõltub erinevatest ilmastikuteguritest ja on nii lühiajaliselt kui ka pikaajaliselt sesoonne. Gaasitarbimine eristub teistest energiaallikate kasutamisest ka selle poolest, et on suurema volatiilsusega. Kuna enamik traditsioonilisi meetodeid ei võimalda arvestada väliseid mõjureid ja hooajalisust, tuleb otsida võimsamaid meetodeid, mis seda võimaldaksid. Varasemad uuringud on näidanud, et igale prognoosimudelile tuleb läheneda individuaalselt ning arvestada erinevaid faktoreid, nagu nt: milliseks perioodiks prognoos tehakse; millised lähteandmed on kättesaadavad; millist matemaatilist meetodit kasutatakse. Ja mudelite võrdlemine omavahel on samuti keerukas, kuna erinevates uuringutes on kasutatud erinevaid mudeli täpsuse võrdlusmeetodeid.

3. MUDELITE VÄLJATÖÖTAMINE

Selle töö kontekstis tuli analüüsida erisuguseid meetodeid maagaasi tarbimise prognoosimiseks Eestis ja luua mudel, mis oleks tulemuslikum kui olemasolev mudel. Põhirõhk oli masinõppel ja närvivõrkudel, sest need meetodid võimaldavad luua mitmete välisparameetritega mitmetasandilisi mudeleid ning kiiresti tuvastada parameetrite vahelised seosed ja luua mudeleid, mis suudavad täpselt kirjeldada käsitletavat süsteemi. Samuti oli oluline ülesanne leida tõhus meetod klientide jaotamiseks klastritesse ja selgitada välja, kas see parandab prognoosinäitajaid.

Sobiva mudeli loomiseks analüüsiti erisuguseid välisparameetreid, mis mõjutavad maagaasi tarbimist, ning nimetatud parameetrite eri kombinatsioone. Sobiva meetodi otsingul uuriti 5 masinõppe algoritmi ja 1 tüüpi närvivõrku. Algselt loodi mudelid reaalse ilmaandmete põhjal ja hiljem, et testida mudeli jõudlust tegelikkuses, kasutati ilmaprognoosi andmeid (järgmise päeva prognoos). Klientide jaotamiseks klastritesse kasutati mittestandardset aegridade analüüsimeetodit, mis eristub suure täpsuse poolest.

Uurimistöö käigus kasutati järgmisi programme:

- Microsoft Excel (Data Analysis Add-in, Solver Add-in);
- IBM SPSS;
- Matlab;
- R.

3.1 Prognoosi täpsuse võrdlemise meetodid

Varasemaid uurimusi analüüsid selgus, et prognoosi täpsuse analüüsimeetodi osas puudub senini üksmeel. Seetõttu otsustati kasutada nii klassikalisi meetodeid, nagu ruutkeskmine viga ja sümmeetriline keskmine absoluutviga protsentides, kui ka vähem kasutatud, kuid tõestatud meetodeid, nagu Thieli indeks, ja üht suhteliselt uut meetodit – keskmine skaleeritud absoluutviga. Lõppkokkuvõttes hinnatakse mudeli täpsust mitme näitaja analüüsimise abil. Allpool on esitatud kõik töös kasutatud prognoositäpsuse hinnatavad parameetrid ning nende arvutusmeetodid, arvestades reaalselt väärtust R_i , prognoositud väärtust P_i ja vaatluspunktide koguarvu n .

Keskmine viga (*Mean Error, ME*)

Keskmine viga on kõigi vigade keskmine hulgas.

$$ME = \frac{1}{n} \sum_{i=1}^n (R_i - P_i) \quad (3.1)$$

Keskmine ruutviga (*Mean Squared Error, MSE*)

Keskmine ruutviga näitab, kui lähedal on regressioonijoon punktihulgale. See toimub nii, et võetakse kaugused punktidest regressioonijooneni ja ruudustatakse need. Ruudustamine on vajalik miinusmärkide eemaldamiseks. Samuti lisab see kaalu suuremate vigade puhul. Seda nimetatakse keskmiseks ruutveaks, sest leitakse keskmine vigade hulgas. [39]

$$MSE = \frac{1}{n} \sum_{i=1}^n (R_i - P_i)^2 \quad (3.2)$$

Ruutkeskmine viga (*Root-Mean-Square Error, RMSE*)

Ruutkeskmine viga on jääkide standardhälve (ennustusvead). Jäägid on mõõt, kui kaugel on regressioonijooone andmepunktid; RMSE on mõõt, kui palju need jäägid on levinud. Teisisõnu annab see teada, kui koondunud on andmed kõige sobivama joone ümber. Ruutkeskmist viga kasutatakse palju katsetulemuste kontrollimiseks klimatoloogias, prognoosimisel ja regressioonianalüüsis. [40]

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_i - P_i)^2} \quad (3.3)$$

Keskmine absoluutne viga (*Mean Absolute Error, MAE*)

Keskmine absoluutne viga on kõigi absoluutsete vigade keskmine. [41]

$$MAE = \frac{1}{n} \sum_{i=1}^n |R_i - P_i| \quad (3.4)$$

Keskmine protsendiviga (*Mean Percentage Error, MPE*)

Keskmine protsendiviga on prognoositud ja tegelike väärtuste vahelise protsendivigade arvatud keskmine.

$$MPE = \frac{100\%}{n} \sum_{i=1}^n \left(\frac{R_i - P_i}{R_i} \right) \quad (3.5)$$

Keskmine absoluutne protsendiviga (*Mean Absolute Percentage Error, MAPE*)

Keskmine absoluutne protsendiviga on statistiline mõõt, mis näitab prognoosisüsteemi täpsust. MAPE mõõdab seda täpsust protsendina ja seda saab arvutada iga perioodi jaoks prognoosina, millest on lahutatud tegelikud väärtused, jagatud tegelike väärtustega. [42]

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{R_i - P_i}{R_i} \right| \quad (3.6)$$

Sümmeetriline keskmine absoluutne protsendiviga (*Symmetric Mean Absolute Percentage Error, SMAPE, sMAPE*)

Sümmeetriline keskmine absoluutne protsendiviga on alternatiiv keskmisele absoluutsele protsendiveale (MAPE), kui artiklite puhul on null- või nullilähedane nõudlus. SMAPE ise piirdub veamääraga 200%, vähendades nende väikesemahuliste artiklite mõju. Väikesemahulised artiklid on probleemsed, sest neil võivad muidu olla veamäärad, mis moonutavad üldist veamäära, ülimalt kõrged. [43]

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|R_i - P_i|}{\frac{(|R_i| + |P_i|)}{2}} \quad (3.7)$$

Theili indeks U_1 (prognoosi täpsus)

U_1 lähtekoht on Theil uuringud 1958. aastal, kus ta nimetab U -d prognoosi täpsuse mõõduks. Ri esindab tegelikke vaatlusi ja P_i vastavaid prognoose. Ta jättis selle lahtiseks, kas R ja P tuleks kasutada kui absoluutväärtusi või kui vaadeldud ja prognoositud muutusi. Mõlema võimalusega on uurimustes tegeletud. [44]

$$U_1 = \frac{\left[\frac{1}{n} \sum_{i=1}^n (R_i - P_i)^2 \right]^{\frac{1}{2}}}{\left[\frac{1}{n} \sum_{i=1}^n R_i^2 \right]^{\frac{1}{2}} + \left[\frac{1}{n} \sum_{i=1}^n P_i^2 \right]^{\frac{1}{2}}} \quad (3.8)$$

Theili indeks U_2 (prognoosi kvaliteet)

1966. aastal Theil pakkus välja U_2 prognoosikvaliteedi mõõtmiseks, "kus R_i ja P_i tähistavad prognoositud ja vaadeldud muutuste paari." [44]

$$U_2 = \frac{\left[\sum_{i=1}^n (P_i - R_i)^2 \right]^{\frac{1}{2}}}{\left[\sum_{i=1}^n R_i^2 \right]^{\frac{1}{2}}} \quad (3.9)$$

Keskmine absoluutne skaleeritud viga (*Mean Absolute Scaled Error, MASE*)

Statistikas on keskmine absoluutne skaleeritud viga prognooside täpsuse mõõt. See on prognoosi väärtuste keskmine absoluutne viga, mis on jagatud valimisisesel ühesammulise naiivse prognoosi keskmise absoluutse veaga. Selle ettepaneku tegid 2005. a statistik Rob J. Hyndman ja otsuseteaduse professor Anne B. Koehler, kes kirjeldasid seda kui "üldiselt rakendatavat prognoosi täpsuse mõõtmist ilma teiste mõõtmiste käigus ilmnenud probleemideta". [45]

$$MASE = \frac{MAE}{MAE_{naivne}} \quad (3.10)$$

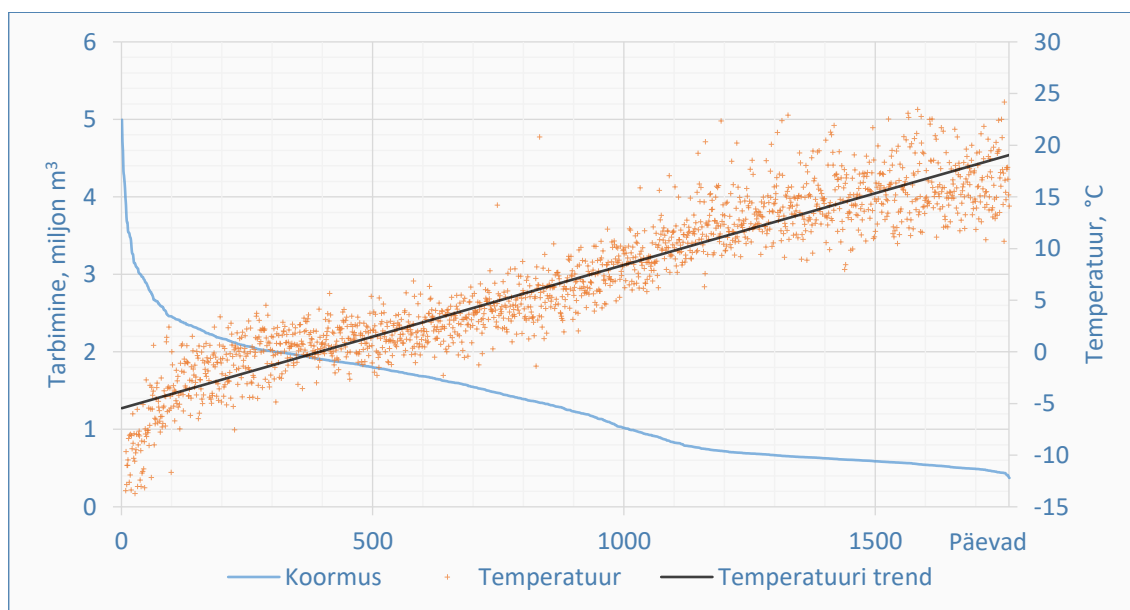
3.2 Algandmed

Peamine parameeter, mille abil mudel koostakse, on kogu Eesti eelmine tarbimine alates 01.03.2014 kuni 31.12.2018, sellekohased andmed on saadud Eesti Gaasist. Lisaks oli antud 654 kliendist koosnev valim, iga kliendi puhul on igapäevane tarbimine alates 01.01.2016 kuni 30.09.2017.

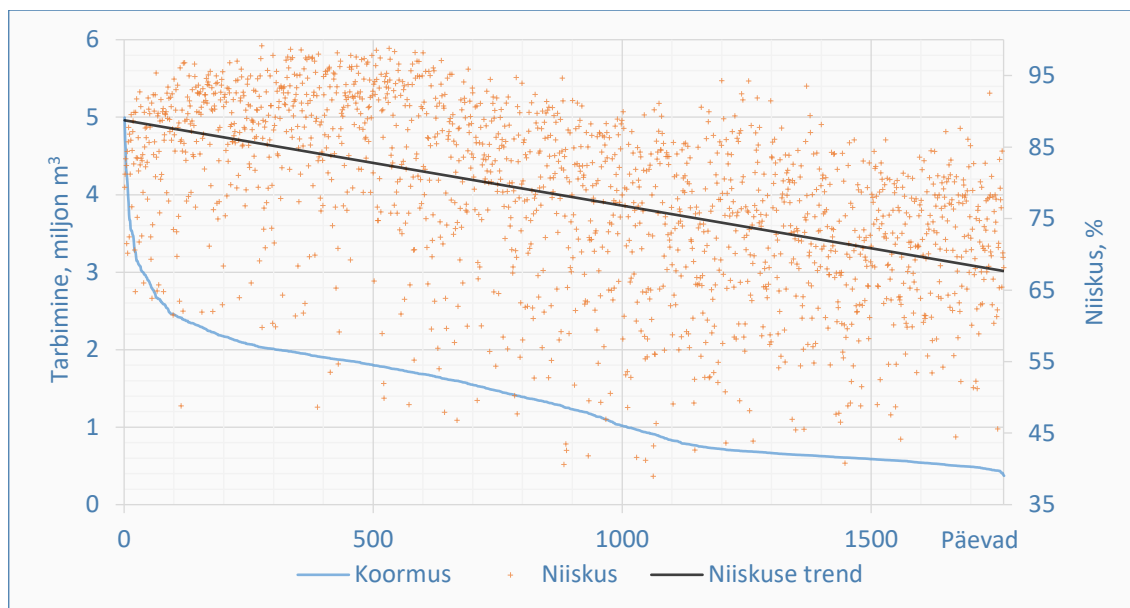
Lisaks otsustati analüüsida järgmisi lisaparameetreid/mõjureid:

- **Ilmastiku andmetest** on sademed, niiskus, temperatuur, tuul ja päikesekiirgus;
- **Majandus andmetest** on Brent'i nafta päevahind, Baltic Trades'i gaasi päevahind ja NordPool Spot'i (NPS) elektri päevahind;
- **Kalendri andmetest** on nädalapäev/nädalavahetus, kuupäev, kuu, riigipühad ning koolivaheajad.

Kõigepealt koostati koormuskeskusgraafikud koos erinevate parameetritega, et visuaalselt hinnata seoseid mõjuri ja tarbimise vahel. Kaks koormusekõverate näidet on toodud allpool.



Joonis 3.1. Tarbimise ja temperatuuri vaheline seos maagaasi tarbimise statistilistes andmetes



Joonis 3.2. Tarbimise ja niiskuse vaheline seos maagaasi tarbimise statistilistes andmetes

Visuaalselt analüüsid näeme ülaltoodud jooniselt, et tarbimise ja temperatuuri vahel on tugev otsene korrelatsioon, tarbimise ja niiskuse vahel aga pöördkorrelatsioon.

Järgmiseks sammuks oli matemaatiliselt analüüsida seoseid tarbimise ja väliste mõjurite vahel. Et kontrollida eespool nimetatud faktorite mõju tarbimisele, viidi läbi korrelatsioonianalüüs ning leiti korrelatsiooni. Pearsoni korrelatsiooni arvutusmeetod on esitatud allpool.

On kaks muutujat, mille korrelatsiooni tuleb kontrollida:

$$x^m = (x_1, \dots, x_m) \quad (3.11)$$

$$y^m = (y_1, \dots, y_m) \quad (3.12)$$

Nende muutujate vaheline korrelatsioon leitakse võrrandi 3.13 abil:

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}} = \frac{cov(x, y)}{\sqrt{s_x^2 s_y^2}} \quad (3.13)$$

kus $\bar{x}, \bar{y} - x^m$ ja y^m keskmised väärtused,

s_x^2, s_y^2 – dispersioonid.

Kõige olulisemad korrelatsiooniparameetrid, mida tuleb korrelatsiooni analüüsimisel arvesse võtta:

$$r_{xy} \in [-1, 1] \quad (3.14)$$

$$|r_{xy}| = 1 - \text{lineaarselt sõltuvad} \quad (3.15)$$

$$r_{xy} = 0 - \text{lineaarselt sõltumatud} \quad (3.16)$$

$$|r_{xy}| < 0,1 \dots \text{tähtsusetu korrelatsioon} \quad (3.17)$$

$$0,1 < |r_{xy}| < 0,3 \dots \text{väike korrelatsioon} \quad (3.18)$$

$$0,3 < |r_{xy}| < 0,5 \dots \text{keskmine korrelatsioon} \quad (3.19)$$

$$0,5 < |r_{xy}| < 1 \dots \text{suur korrelatsioon} \quad (3.20)$$

Leitud korrelatsioonid mõjurite ja tarbimise vahel on esitatud tabelis 3.1.

Tabel 3.1. Korrelatsioonitegurid

Korrelatsioonitegur	Väärtus
r_{sademed}	-0,08
r_{niiskus}	0,29
$r_{\text{temperatuur}}$	-0,94
r_{tuul}	0,17
$r_{\text{päike}}$	-0,65
r_{BRENT}	-0,20
r_{BT}	0,14
r_{NPS}	0,06
$r_{\text{riigipüha}}$	0,00
r_{kool}	-0,33

Võrreldes tabelis 3.1 esitatud andmed valemities (3.14) – (3.20) esitatud andmetega saab öelda, et peamised gaasi tarbimist mõjutavad tegurid on temperatuur, päikesekiirgus ja koolivaheajad, lisaks ka niiskus.

Tavaline korrelatsioon ei kirjelda sugugi alati reaalselt olukorda, kuna see korrelatsioon võib olla moonutatud kõrvalfaktoritega. Selleks et eemaldada väliste tegurite mõju, kasutatakse osakorrelatsioon. [46]

Funktsioon, mis näitab kahe muutuja vahelist korrelatsiooni, eemaldades kolmanda muutuju mõju, on järgmine:

$$r_{xy/z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (3.21)$$

kus $r_{xy/z}$ – x ja y vaheline korrelatsioon vältides z mõju,

r_{xy} – x ja y vaheline korrelatsioon,

r_{xz} – x ja z vaheline korrelatsioon,

r_{yz} – y ja z vaheline korrelatsioon.

Mitme faktori mõju eemaldamiseks koostakse maatriks R , kuhu on kantud kõikide parameetrite omavahelised seosed.

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ & 1 & r_{23} & \cdots & r_{2k} \\ & & 1 & \cdots & r_{3k} \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix} \quad (3.22)$$

$$R_{ij} = (-1)^{i+j} M_{ij} \quad (3.23)$$

kus $r_{1k} r_{2k} \dots r_{ik}$ – parameetrite vahelised korrelatsioonid,

M_{ij} – R maatriksi peamiinor.

Korrelatsioon i ja j vahel, vältides kõikide teiste parameetrite mõju leitakse järgmiselt:

$$r_{ij/faktor} = \frac{-R_{ij}}{\sqrt{R_{ii}R_{jj}}} \quad (3.24)$$

kus $r_{ij/faktor}$ – otsitav korrelatsiooni koefitsient i ja j vahel.

Kõikide parameetrite osakorrelatsiooni tegurid on esitatud tabelis 3.2.

Tabel 3.2. Osakorrelatsiooni tegurid

Korrelatsioonitegur	Väärtus
$r_{o.sademed}$	-0,06
$r_{o.niiskus}$	0,09
$r_{o.temperatuur}$	-0,88
$r_{o.tuul}$	0,09
$r_{o.päike}$	-0,18
$r_{o.BRENT}$	-0,18
$r_{o.BT}$	0,14
$r_{o.NPS}$	0,22
$r_{o.riigipüha}$	-0,11
$r_{o.kool}$	-0,22

Võrreldes tabeli 3.2 andmed tabeli 3.1 andmetega selgus, et ainus reaalne mõjur on temperatuur, vähemal määral ka koolivaheajad ja NPS elektri hind.

Et uurida korrelatsiooni statistilist olemasolu, kasutatakse Studenti t-testi. On loodud nullhüpotees:

$$r_{xy} = 0 \rightarrow H_0: \text{puudub seos } x \text{ ja } y \text{ vahel} \quad (3.25)$$

Nullhüpoteesi lükatakse tagasi juhul, kui:

$$|t_{arv}| > t_{kr} \quad (3.26)$$

kus t_{arv} – t arvutatud väärtus,

t_{kr} – t kriitiline väärtus.

t_{arv} väärtus leitakse järgmise valemiga:

$$t_{arv} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad (3.27)$$

kus r_{xy} – x ja y vaheline korrelatsiooni tegur,

n – vaatluspunktide koguarv.

Järgmiselt t_{kr} leitakse väärtustabelist, arvestades olulisuse tase ja vabaduse astmeid:

$$\alpha = 0,01 \quad (3.28)$$

$$n = 1767 \quad (3.29)$$

$$df = n - 2 = 1766 \quad (3.30)$$

kus α – veamäär,

n – vaatluspunktide koguarv,

df – vabaduse astmed.

Meie andmete puhul on väärtustabeli järgi t_{kr} väärtus järgmine:

$$t_{kr} = 2,579 \quad (3.31)$$

Tabel 3.3. Studenti testi t_{arv} tegurid erinevate parameetrite jaoks

t-testi tegur	Väärtus
$t_{arv.sademed}$	-2,69
$t_{arv.niiskus}$	3,63
$t_{arv.temperatuur}$	-75,93
$t_{arv.tuul}$	3,97
$t_{arv.päike}$	-7,69
$t_{arv.BRENT}$	-6,27
$t_{arv.BT}$	5,57
$t_{arv.NPS}$	9,61
$t_{arv.riigipüha}$	-4,48
$t_{arv.kool}$	-9,47

Kuna kõik $|t_{arv}| > t_{kr}$, saab nende andmete põhjal öelda, et kõik korrelatsioonid on statistiliselt olulised.

Mõjurite valikul oli arvestatud ka seda, et reaalses olukorras on tegemist nende prognoositud väärtustega, mitte reaalsete väärtustega (nt temperatuuri prognoos). Arvestades asjaolu, et elektri börsihinna korrelatsioonitegur ei ole suur, otsustati seda mitte kasutada.

Seega kasutati mudelite koostamisel järgmisi parameetreid:

- temperatuur;
- riigipühad;
- koolivaheajad;
- eelmise päeva tarbimine;
- nädalapäev;
- kuu.

Nendest parameetritest ainus, mida tuleb prognoosida, on temperatuur.

NB! Tuleb pöörata tähelepanu sellele, et saadud korrelatsioonitegurid on konkreetse andmekogumi jaoks! Iga andmekogumi puhul tuleks läbi viia eraldi parameetrianalüüs! Peab ka arvestama sellega, et saadud korrelatsioonitegurid ja Studenti t-tegurid vaid aitavad valida vajalikke parameetreid, lõplik valik tuleb teha katsetamise ajal.

3.3 Mudelite testimine

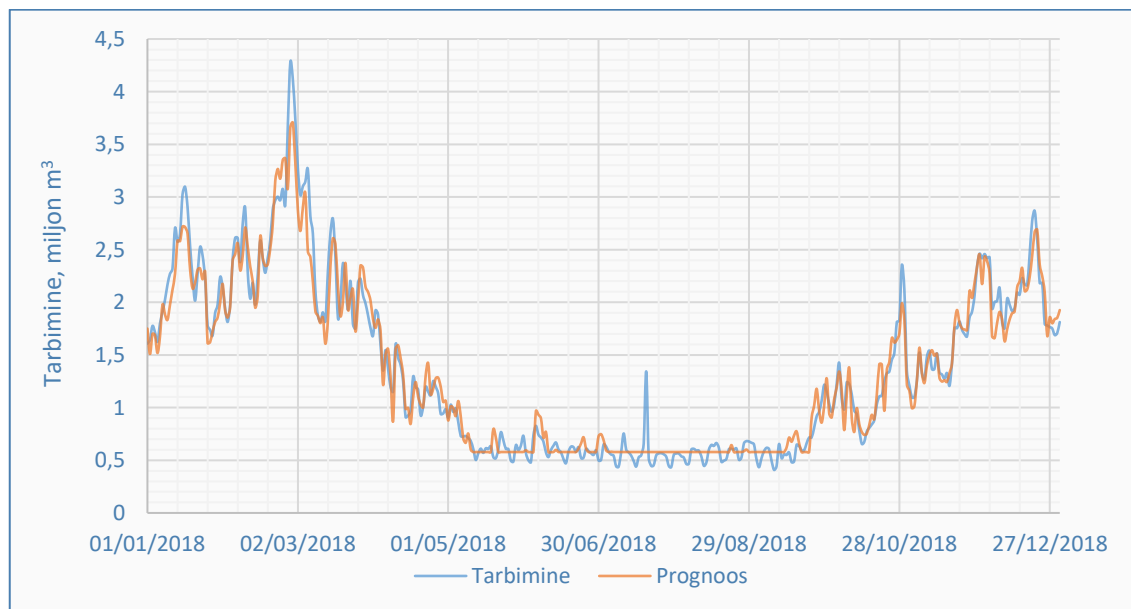
Töö käigus valiti katsetamiseks 5 masinõppe algoritmi ja tehisliku närvivõrgu algoritmi ning võrreldi neid Eesti Gaasi mudeliga. Masinõppe ja närvivõrgu eripäraks on see, et nende koostamisel ja katsetamisel kasutatakse kaht andmetüüpi:

- treenimisandmed
- katsetamisandmed

Treenimise andmetüüp on see, mille abil mudelid leiavad tendentsid ja seosed parameetrite vahel, nende andmete põhjal koostakse mudeli matemaatiline osa. Katsetamise andmetüüp on mõeldud mudeli täpsuse ja üldistamisoskuse hindamiseks. Teiste sõnadega: need on andmed, mille abil me hindame mudeliga saadud tulemusi. Algoritmide üldine kirjeldus ja nende abil saadud tulemused on toodud allpool. Kõiki mudeleid võrreldi omavahel samal perioodil, st 01.01.2018 – 31.12.2018. Kõiki masinõppe ja närvivõrgu mudeleid treeniti ühel ja samal perioodil, 01.03.2014 – 31.12.2017.

3.3.1 Eesti Gaasi mudel

Antud mudel on Eesti Gaasi poolt pakutud. Eesti Gaasi mudeli katsetamise tulemused on toodud joonisel 3.3. Prognositäpsuse parameetrid on toodud valemities 3.32 – 3.35. Antud mudelis on kasutatud tuleviku tegelikud temperatuuri andmed.



Joonis 3.3. Eesti Gaasi mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 166456,4 \quad (3.32)$$

$$SMAPE = 10\% \quad (3.33)$$

$$U_2 = 1 \quad (3.34)$$

$$MASE = 0,96 \quad (3.35)$$

Võrreldes naiivse meetodiga (meetod, mille puhul homseks prognoosiks kasutatakse tänaseid reaalseid andmeid) ei näita antud mudel paremust. Kõigile olemasolevatele andmetele tuginedes võrreldi Eesti Gaasi mudelit ja naiivset meetodit. Mudelite täpsust hindavad parameetrid on esitatud allpool tabelis 3.4

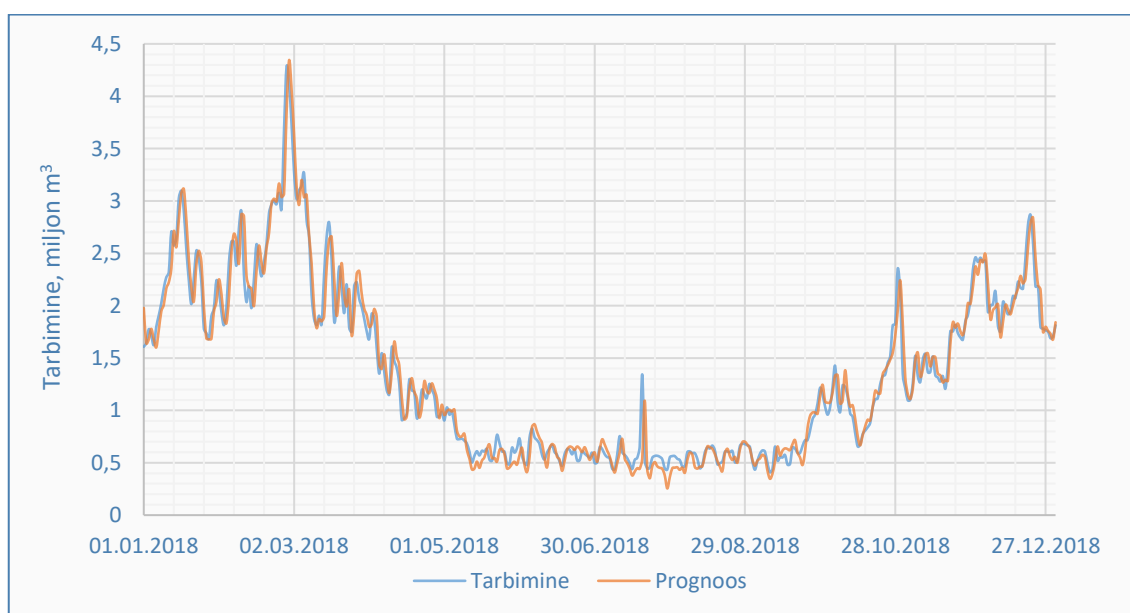
Tabel 3.4. Eesti Gaasi mudeli ja naiivse meetodi võrdlemine kõikide andmete (01.03.2014-31.12.2018) tuginedes.

Parameeter	EG	Naiivne
ME	$1,4 \times 10^4$	$-7,0E \times 10^1$
MSE	$3,0 \times 10^{10}$	$2,8 \times 10^{10}$
RMSE	$1,7 \times 10^5$	$1,7 \times 10^5$
MAE	$1,3 \times 10^5$	$1,1 \times 10^5$
MPE	-0,6%	-0,7%
MAPE	10,8%	9,0%
SMAPE	10,7%	8,9%
U ₁	0,06	0,05
U ₂	1,18	1,00
MASE	1,13	1,00

Tabelist 3.4 on näha, et Eesti Gaasi mudel ei toimi parem, kui naiivne meetod – RMSE on sama väärtusega, mõnede parameetrite järgi (MAE, SMAPE, U₂, MASE) toimib see mudel halvem kui naiivne meetod.

3.3.2 Lineaarne regressioonimudel

Lineaarse regressiooni mudeli jõudlus on sarnane Eesti Gaasi mudeli omale, mis on põhjendatav seega, et mõlemad on lineaarsed mudelid. Nende mudelite erinevus seisneb selles, et antud töö käigus koostatud mudel arvestab peale temperatuuri ka teisi mõjureid. Lineaarse regressioonimudeli abil tehtud prognoos on toodud joonisel 3.4. Prognoositäpsuse parameetrid on esitatud valemities 3.36 – 3.39.



Joonis 3.4. Lineaarse regressiooni mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 149823,0 \quad (3.36)$$

$$SMAPE = 9,3\% \quad (3.37)$$

$$U_2 = 0,95 \quad (3.38)$$

$$MASE = 0,85 \quad (3.39)$$

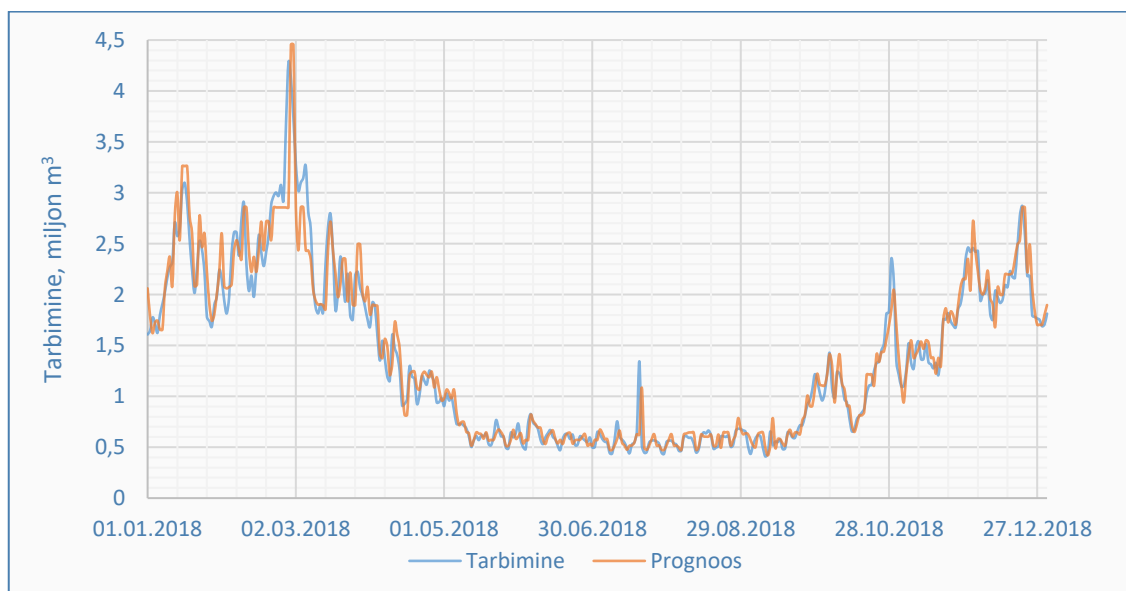
Võrreldes erinevaid prognoositäpsuse võrdlemise parameetreid, saab öelda, et RMSE läks 10% paremaks, SMAPE 6,4% paremaks, Thiel'I indeks U_2 muutus 8,9% paremaks ja MASE on 11% paremaks.

Nagu oodatud, selle mudeli jõudlus ei erine märgatavalt algsest mudelist. See põhineb ülalmainitud asjaolul, et mõlemad mudelid on lineaarsed.

3.3.3 Otsuspuu mudel

Kuigi see mudel on Otsuspuu mudel, nagu Eesti Gaasi poolt pakutud mudel on nende vahel märkimisväärsed erinevused. Kui Eesti Gaasi terminaallehtedes on lineaarsed mudelid, siis antud katses tehtud mudel on mittelineaarne. Lisaks, nagu mainitud lineaarse regressiooni mudeli kohta, antud mudelis kasutakse rohkem välisparameetreid kui Eesti Gaasi mudelis.

Otsuspuu mudeli abil saadud tulemus on esitatud allpool joonisel 3.5. Prognoosi täpsuse parameetrite võrreldakse valemities 3.40 – 3.43.



Joonis 3.5. Otsuspuu mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 193334,6 \quad (3.40)$$

$$SMAPE = 8,7\% \quad (3.41)$$

$$U_2 = 0,88 \quad (3.42)$$

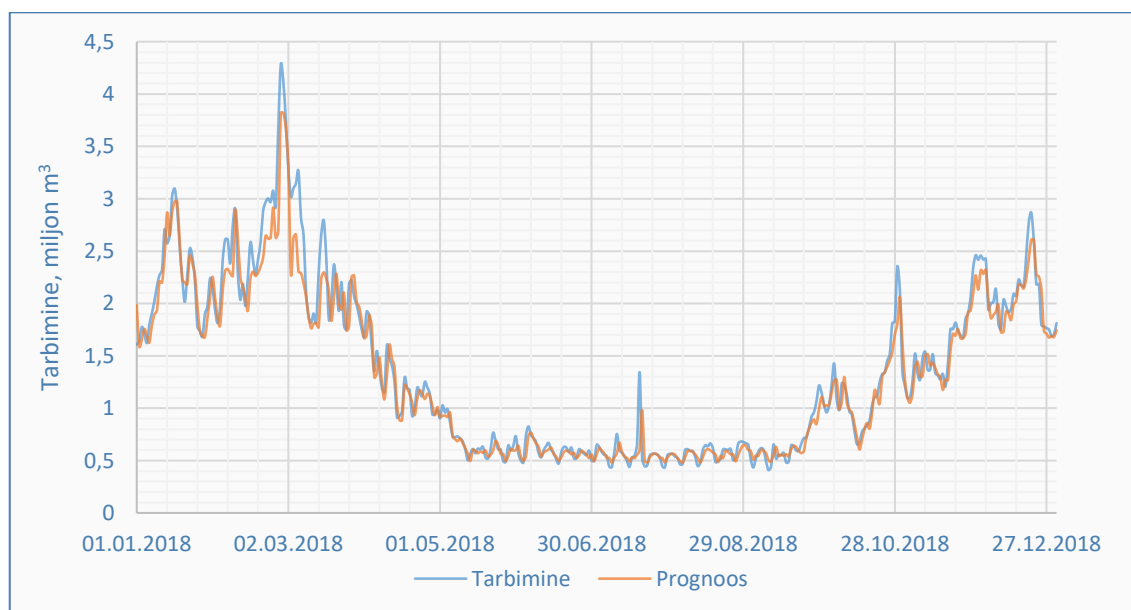
$$MASE = 0,97 \quad (3.43)$$

Sellise mudeli kasutamisel saadi segased tulemused – mõned täpsuse hindamise parameetrid küll paranesid, ent mõned läksid võrreldes algse mudeliga halvemaks. Näeme, et RMSE ja MASE muutusid 16,1% ja 1,6% halvemaks ning U_2 ja SMAPE muutusid 15,6% ja 12,4% paremaks. Ruutkeskmine viga annab rohkem kaalu suuremate vigadele, see tähendab, et antud mudeliga on keskmine viga väiksem kui algse mudeli puhul, kuid maksimaalne viga on suurem.

Kui me kasutaksime ainult ühte täpsuse võrdlemise parameetrit, näiteks SMAPE, siis tundub et antud mudel on parem, kui punktis 3.3.3 vaadeldud lineaarse regressioonimudel: 8,7% võrreldes 9,3%. Kuid teiste parameetrite võttes saab näha, et mõned (nt RMSE ja MASE) on palju halvemad. See tõestab mitme täpsuse võrdlemise meetodi kasutamise tähtsust.

3.3.4 Otsuspuu ansambli mudel

Otsuspuu ansambli mudeli katsetamise tulemused on toodud joonisel 3.6. Prognoositäpsuse analüüs on toodud valemities 3.44 – 3.47.



Joonis 3.6. Otsuspuu ansambli mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 173121,3 \quad (3.44)$$

$$SMAPE = 7,8\% \quad (3.45)$$

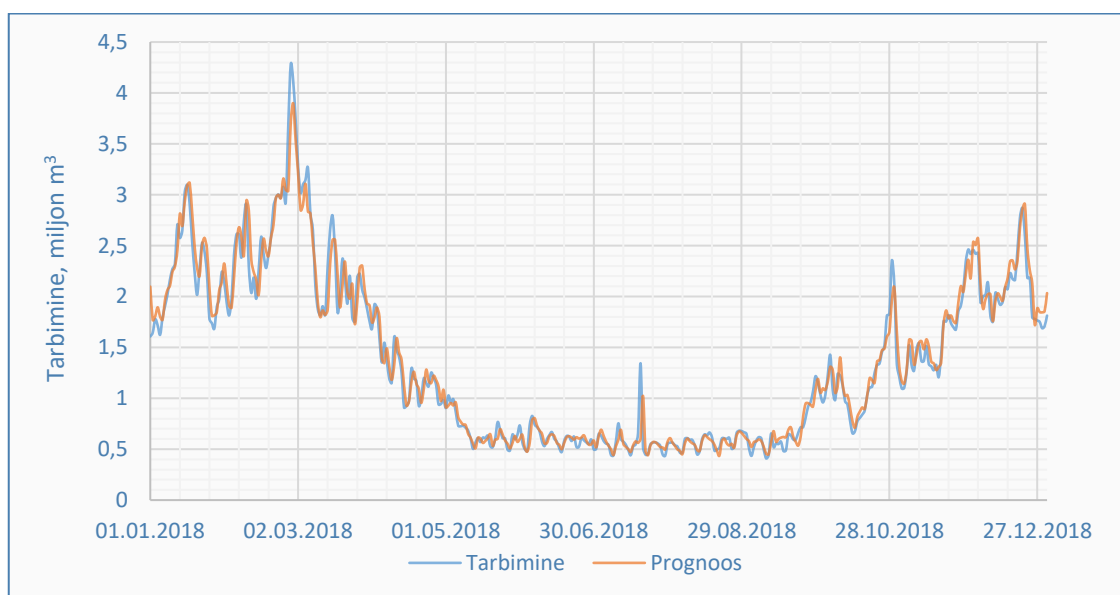
$$U_2 = 0,83 \quad (3.46)$$

$$MASE = 0,84 \quad (3.47)$$

Võrreldes Otsuspuu mudeliga selle mudeli jõudlus on oluliselt paranenud: RMSE on 10,5% parem, SMAPE on 10,3% parem, U_2 ja MASE vastavalt 5,7% ja 14,3% paremad. Kuid võrreldes Eesti Gaasi mudeliga SMAPE 21,5% paremaks, U_2 muutus 20,4% paremaks ja MASE on 11% paremaks, aga RMSE läks 4% halvemaks.

3.3.5 Tugivektoriregressiooni mudel

Tugivektoriregressiooni mudeli tulemused on joonisel 3.7. Prognoositäpsuse parameetrid on valemities 3.48 – 3.51.



Joonis 3.7. TVR mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 141474,5 \quad (3.48)$$

$$SMAPE = 7,3\% \quad (3.49)$$

$$U_2 = 0,78 \quad (3.50)$$

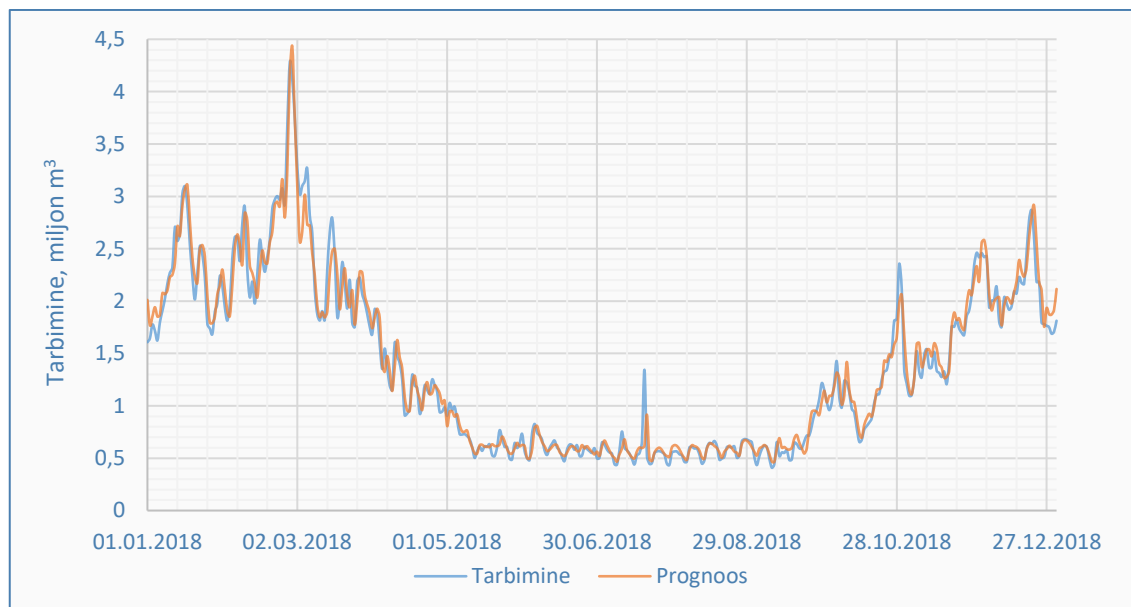
$$MASE = 0,75 \quad (3.51)$$

See mudel tõestas end väga tõhusana, kõik olulised täpsuse võrdlemise parameetrid muutusid paremaks võrreldes Eesti Gaasi mudeliga: RMSE on 15% parem, SMAPE on 26,5% parem, U_2 ja MASE vastavalt 20,4% ja 26,1% paremad.

Siinkohal on oluline tähelepanu pöörata sellele, et antud mudel koos allpool 3.3.7 punktis mainitud tehniliku närvivõrgu mudeliga näitavad kõige paremat MASE parameetri väärtust, mis tähendab seda, et nende mudelite jõudlus on kõige parem võrreldes naiivmeetodiga. Antud mudel ja gaussi regressiooni mudel mõlemad näitavad kõige väiksemat keskmise vea väärtust.

3.3.6 Gaussi protsessi regressiooni mudel

Joonisel 3.8 on esitatud Gaussi protsessi mudeli katsetamise tulemused. Valemites 3.52 – 3.55 saab näha mudeli prognoositäpsuse parameetrid.



Joonis 3.8. Gaussi protsessi regressiooni mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 139825,6 \quad (3.52)$$

$$SMAPE = 7,3\% \quad (3.53)$$

$$U_2 = 0,76 \quad (3.54)$$

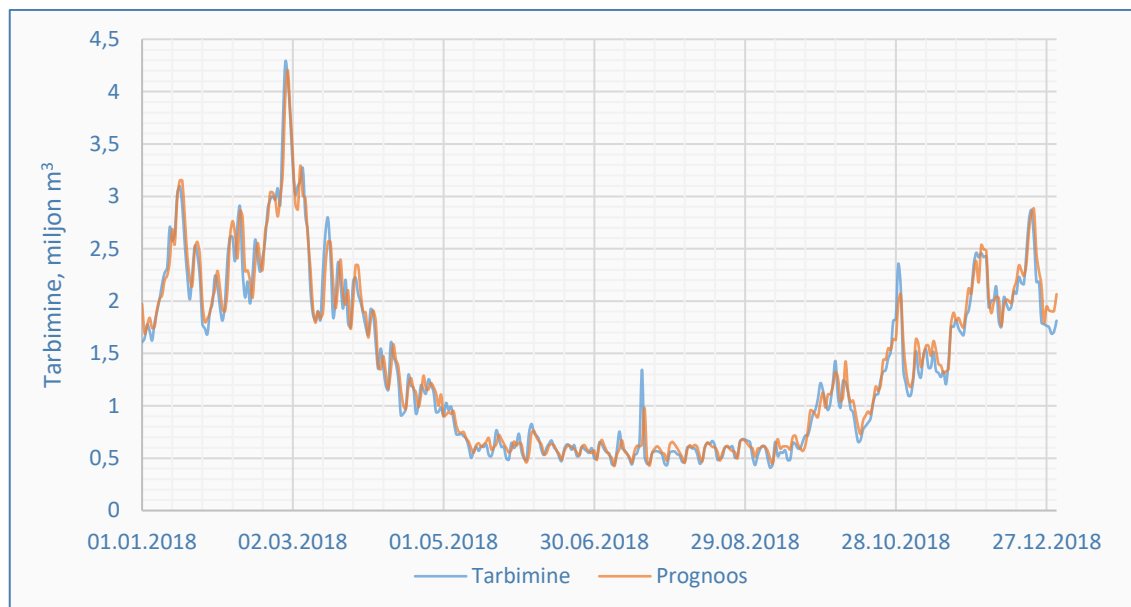
$$MASE = 0,76 \quad (3.55)$$

Antud mudeli jõudlus on sarnane TVR mudeliga, aga on siiski mõned erinevused. Antud mudelis on RMSE parameeter parem ja SMAPE parameeter on sama, see tähendab, et selle mudeli keskmine viga on sama kui TVM mudelil, kuid maksimaalne viga on väiksem.

Võrreldes Eesti Gaasi mudeliga on prognoositäpsuse parameetrid muutunud järgmiselt: RMSE on 16% parem, SMAPE on 26,5% parem, U_2 on 25,2% parem ja MASE on 20% parem.

3.3.7 Tehislik närvivõrk

Tehisliku närvivõrgu mudeli katsetamise tulemused on toodud joonisel 3.9. Prognositäpsuse parameetrid on toodud valemites 3.56– 3.59.



Joonis 3.9. Tehislik närvivõrku mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 133647,3 \quad (3.56)$$

$$SMAPE = 7,4\% \quad (3.57)$$

$$U_2 = 0,78 \quad (3.58)$$

$$MASE = 0,75 \quad (3.59)$$

Mudel tõestas end ühena parimatest: närvivõrgu mudel on väiksema ruutkeskmise veaga, mis tähendab, et maksimaalne viga on antud mudelis kõige väiksem. SMAPE parameeter on sarnane gaussi protsessi ja tugivektoriregressiooni mudelitele, kuid pisut suurem. Nagu oli tugivektoriregressiooni mudeli punktis 3.3.5 mainitud, tehisliku närvivõrgu ja tugivektoriregressiooni mudelid mõlemad näitavad kõige väiksema MASE parameetri.

Võrreldes Eesti Gaasi mudeliga on prognoosi täpsuse parameetrid muutunud järgmiselt: RMSE on 19,7% parem, SMAPE on 25,5% parem, U_2 on 25,2% parem ja MASE on 21,4% parem.

3.3.8 Mudelite võrdlus

Tabelist 3.5 on näha, et enamik masinõppe mudelid ja ka närvivõrgumudel annavad parema tulemuse kui algne mudel. Kõige parema tulemuse tagavad järgmised mudelid:

- TVR mudel;
- Gaussi protsessi mudel;
- närvivõrgu mudel.

Tabel 3.5. Mudelite prognooside täpsuseparameetrite võrdlemine

Parameeter	EG	LR	OP	TVR	G	OA	TN
ME	$-1,4 \times 10^2$	$-1,7 \times 10^3$	$-2,0 \times 10^4$	$-1,7 \times 10^4$	$-1,8 \times 10^4$	$5,4 \times 10^4$	$-2,7 \times 10^4$
MSE	$2,8 \times 10^{10}$	2,24101	$3,7 \times 10^{10}$	$2,0 \times 10^{10}$	$2,0 \times 10^{10}$	$3,0 \times 10^{10}$	$1,8 \times 10^{10}$
RMSE	$1,7 \times 10^5$	$1,5 \times 10^5$	$1,9 \times 10^5$	$1,4 \times 10^5$	$1,4 \times 10^5$	$1,7 \times 10^5$	$1,3 \times 10^5$
MAE	$1,2 \times 10^5$	$1,1 \times 10^5$	$1,2 \times 10^5$	$9,5 \times 10^4$	$9,6 \times 10^4$	$1,1 \times 10^5$	$9,5 \times 10^4$
MPE	-1,8%	0,3%	-2,8%	-2,2%	-2,6%	2,1%	-3,2%
MAPE	10%	9,2%	8,9%	7,4%	7,4%	7,6%	7,6%
SMAPE	10%	9,3%	8,7%	7,3%	7,3%	7,8%	7,4%
U_1	0,05	0,05	0,06	0,04	0,04	0,06	0,04
U_2	1	0,95	0,88	0,78	0,76	0,83	0,78
MASE	0,96	0,85	0,97	0,76	0,76	0,85	0,75

3.4 Mudelite kombineerimine

Uurimused on näidanud, et mitmete prognoosimudelite kombineerimine võib vähendada vigu, mis tulenevad vigastest eeldustest, eksiarvamustest või andmevigadest. Mitme prognoosimudeli kombineerimine vähendab prognoosimisvigu, kuna kombineeritud prognoosil on võrdne või sageli väiksem kõikumine kui iga üksikute mudelite prognoos. See on matemaatiliselt tõestatud. [47] [48] [49]

Väärtused leiti tavalise keskmise väärtuse valemiga:

$$P_i = \frac{EG_i + LR_i + OP_i + TVR_i + G_i + OA_i + TN_i}{7} \quad (3.60)$$

kus P_i – kõikide mudelite kombineerimisel saadud väärtus päeval i ,

EG_i – Eesti Gaasi mudeliga saadud prognoos päeval i ,

LR_i – Lineaarse regressioonimudeliga saadud prognoos päeval i ,

OP_i – Otsuspuu mudeliga saadud prognoos päeval i ,

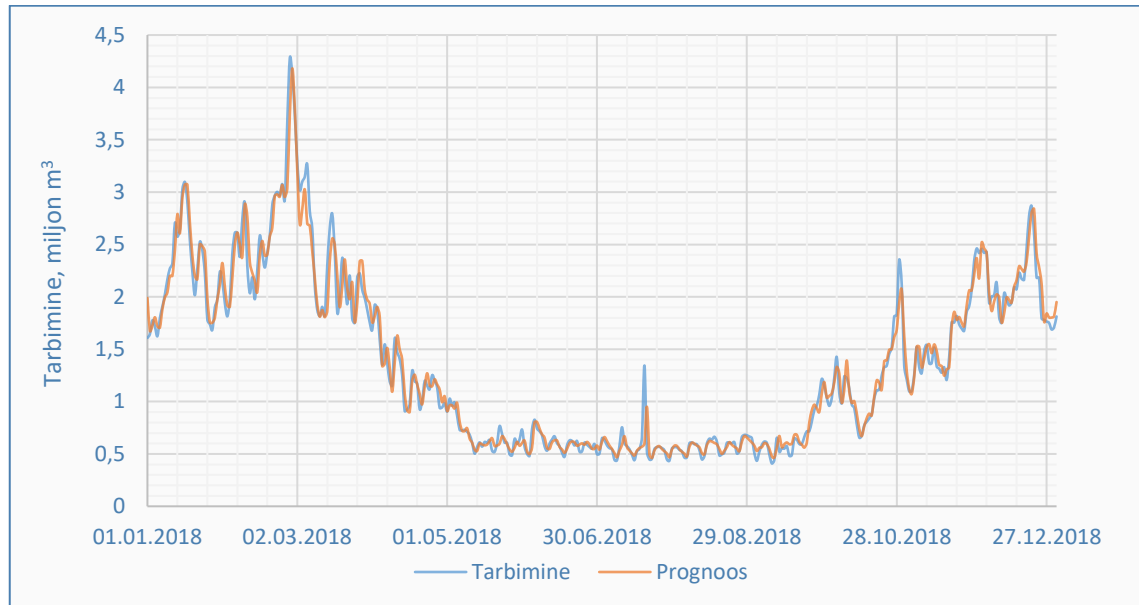
TVR_i – TVR mudeliga saadud prognoos päeval i ,

G_i – Gaussi protsessi regressiooni mudeliga saadud prognoos päeval i ,

OA_i – Otsuspuu ansambli mudeliga saadud prognoos päeval i ,

TN_i – Tehisliku närvivõrgu mudeliga saadud prognoos päeval i .

Kõiki seitsset mudelit kombineerides saadi tulemus, millest lähemalt allpool joonisel 3.10.



Joonis 3.10. Kombineeritud mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 133572,5 \quad (3.61)$$

$$SMAPE = 6,7\% \quad (3.62)$$

$$U_2 = 0,74 \quad (3.63)$$

$$MASE = 0,69 \quad (3.64)$$

Võrreldes eelmiste mudelitega muutus RMSE 0,1% paremaks, SMAPE läks 7,3% paremaks, U_2 läks 2,4% paremaks ja MASE paranes 7,3%. Võrreldes Eesti Gaasi mudeliga on RMSE 19,8% parem, SMAPE on parem 32,6%, U_2 ja MASE on paremad vastavalt 29% ja 27,7%. Nende näitajate põhjal on see meetod senini vaadeldud meetoditest kõige täpsem.

Seega oleme tõestanud, et mitme mudeli kombineerimine võimaldab prognooside täpsust parandada, nagu oli eelmistes uuringutes mainitud.

3.5 Klasterdamise mõju prognoositäpsusele

Enne kui andmeid saab klasterdamiseks kasutada, tuleb need normaliseerida, st viia ühele skaalale 0-st 1-ni. See võimaldab võrrelda erineva tarbimismahuga tarbijaid.

Normaliseerimise valem on järgmine:

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (3.65)$$

kus $X_i - X$ väärtus punktis i ,

$X_{max} - X$ maksimaalne väärtus,

$X_{min} - X$ minimaalne väärtus.

Selle valemi abil saame arvutada esimese tarbija esimene väärtuse:

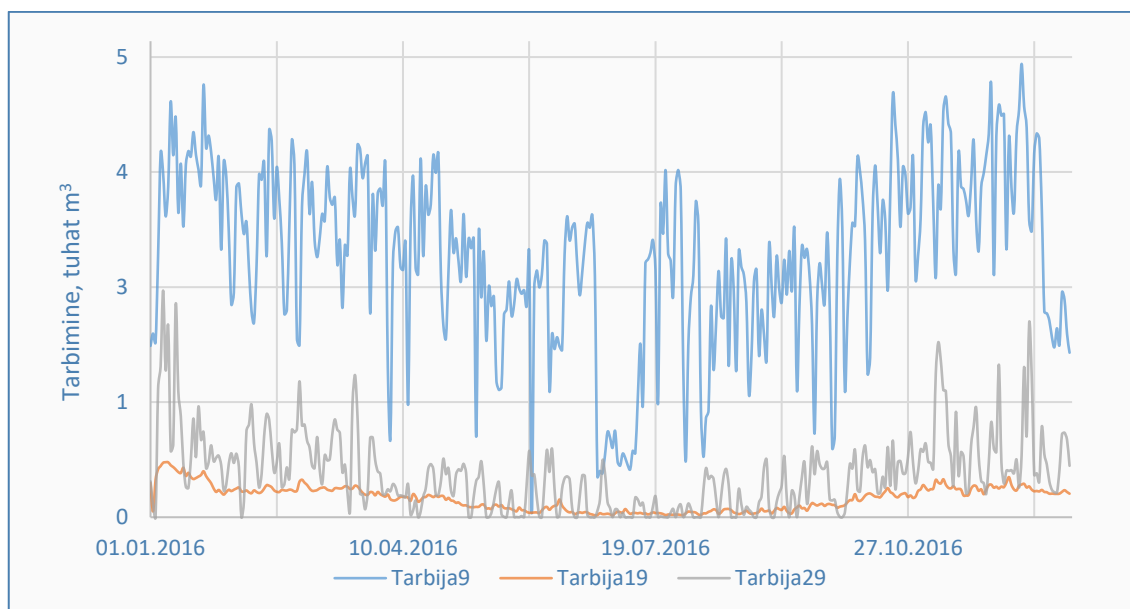
$$X_1 = 695 \text{ m}^3 \quad (3.66)$$

$$X_{max} = 870 \text{ m}^3 \quad (3.67)$$

$$X_{min} = 0 \text{ m}^3 \quad (3.68)$$

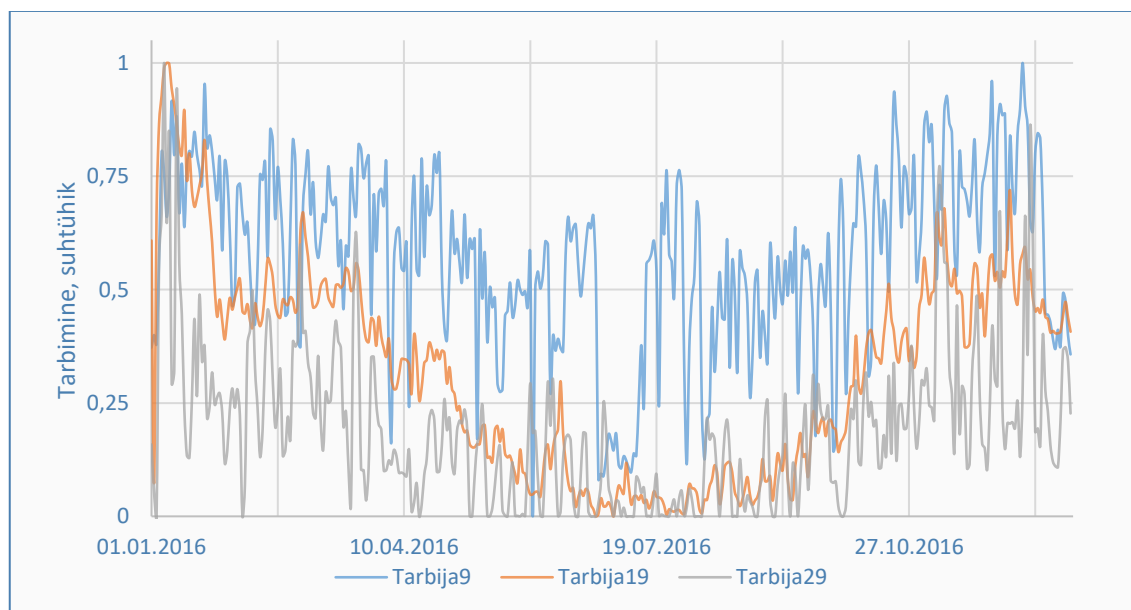
$$X_{norm} = \frac{695 - 0}{870 - 0} = 0,80 \quad (3.69)$$

Allpool on toodud mõnede klientide tarbimisandmed enne ja pärast nende normaliseerimist.



Joonis 3.11. Tarbimisandmed enne nende normaliseerimist

Jooniselt 3.11 on näha, et kõikide kolme kliendi tarbimismahud on erinevad ja nende tarbimise käitumise võrdlemine omavahel muutub keeruliseks. Ilma normaliseerimiseta ei saa neid tarbijaid efektiivselt grupeerida.

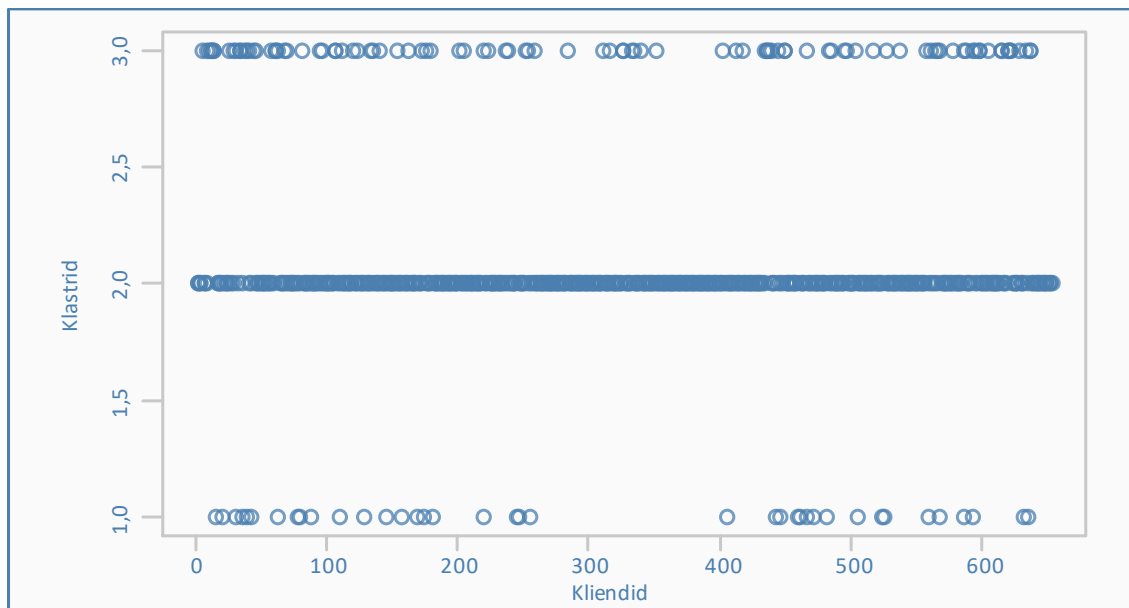


Joonis 3.12. Tarbimisandmed pärast nende normaliseerimist

Jooniselt 3.12 on näha, et andmete normaliseerimine võimaldab võrrelda klientide erinevaid tarbimismahte. Visuaalselt analüüsid saab öelda, et Tarbija19 ja Tarbija29 tarbimise käitumised on suhtelised sarnased ja neid võib viia ühe klastrisse.

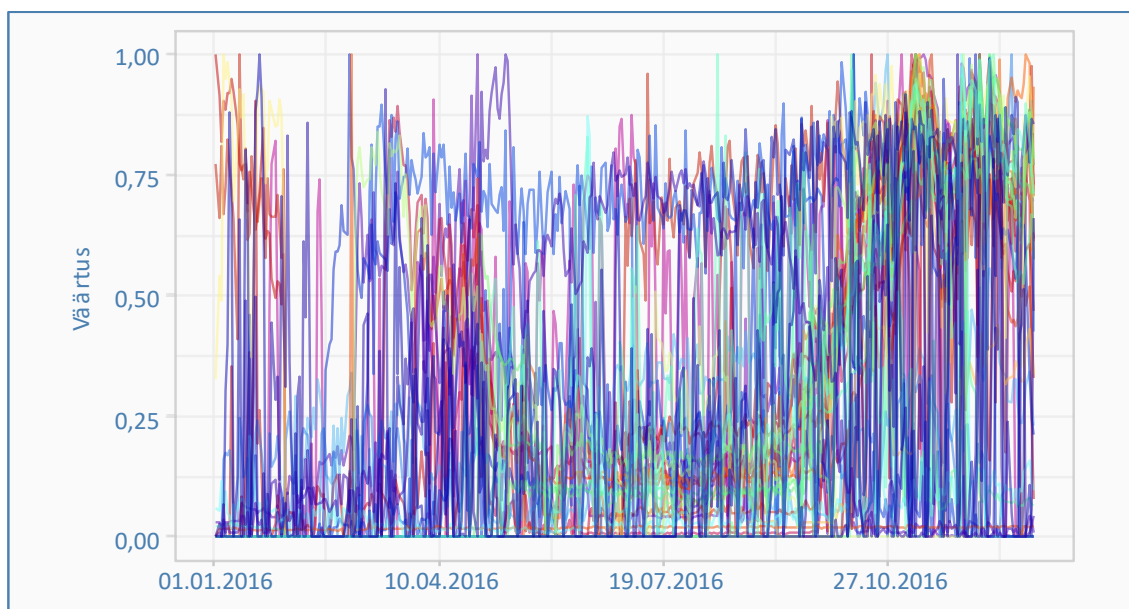
3.5.1 Gasitarbimise profiilide klasterdamine

R programmeerimiskeel võimaldab kasutada DTW meetodit suuremahuliste andmete puhul. Antud uurimistöös kontekstis otsustati jagada kliendid kolme klastrisse, mis kujutab endast sümboolset numbrit ja sõltub ainult prognoosija soovist. Klientide andmed on perioodist 01.01.2016–30.09.2017, seetõttu otsustati valida klasteri analüüsi ja mudeli treenimise jaoks periood 01.01.2016–31.01.2016 ja 01.01.2017–30.09.2017 on periood kontrollimiseks. Klientide jaotus klastrite vahel on esitatud joonisel 3.13.



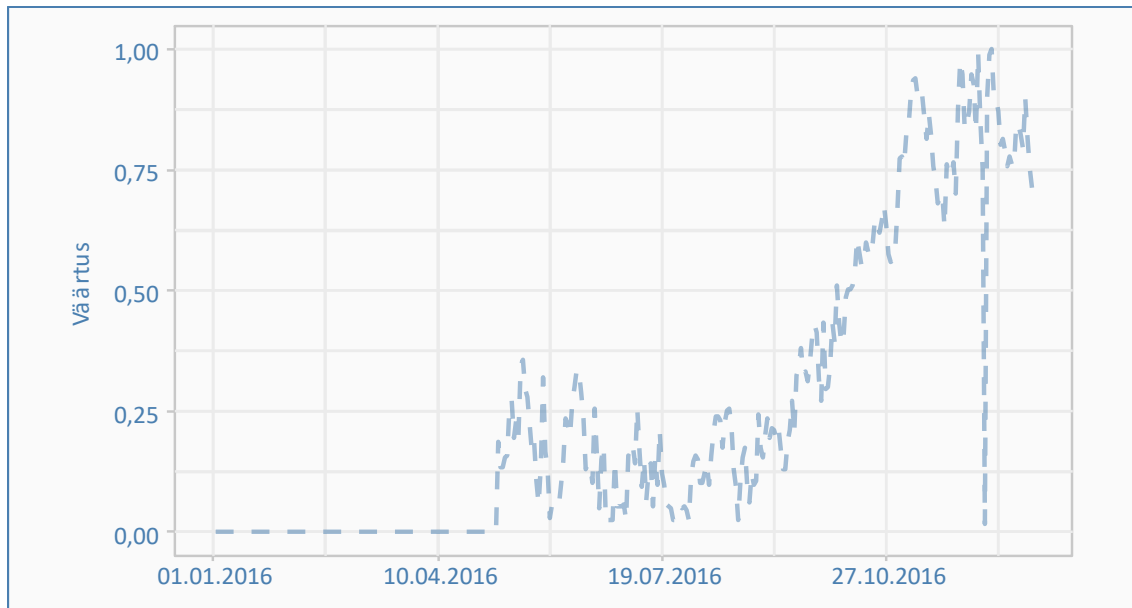
Joonis 3.13. Klientide jaotus klastrite vahel

Jooniselt 3.18 on näha, et enamus kliente on grupeeritud klastrisse 2, need on kõige kõrgema prognoositavusega kliendid, ja seda on näha allpool esitatud jooniselt 3.16. Esimesel klastril on väga suure volatiilsusega kliendid (Joonised 3.14 ja 3.15) ja kolmandas klastris on kliendid, kelle tarbimisprofiilides on märgata sesoonsust, aga nende tarbimisprofiilide volatiilsus on suurem kui klientidel teises klastris (Joonised 3.18 ja 3.19).



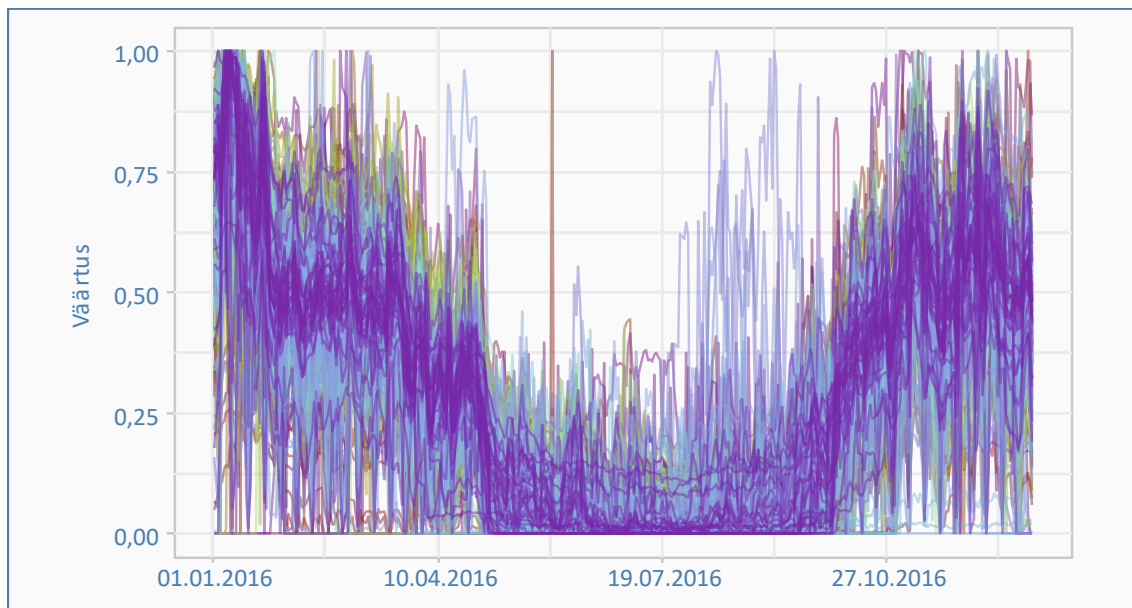
Joonis 3.14. Esimese klastri normaliseeritud aegread

Graafik näitab tarbimise väga suurt volatiilsust. On näha teravad hüpped minimaalsest väärtusest maksimumini.



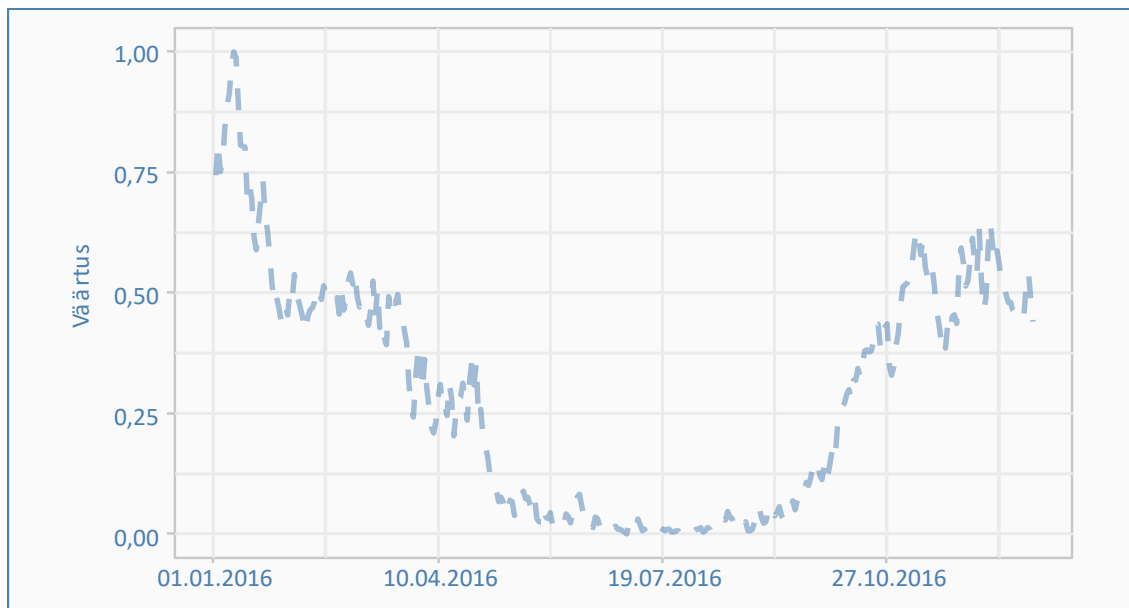
Joonis 3.15. Esimese klasteri normaliseeritud aegridade tsentroid

Tsentroid näitab, et antud klasteris on väga väike tarbimine aasta algusest kuni suvi alguseni, pärast näeme suurema volatiilsusega perioodi suvi lõpuni, ja siit märgatakse tarbimise kasvu.



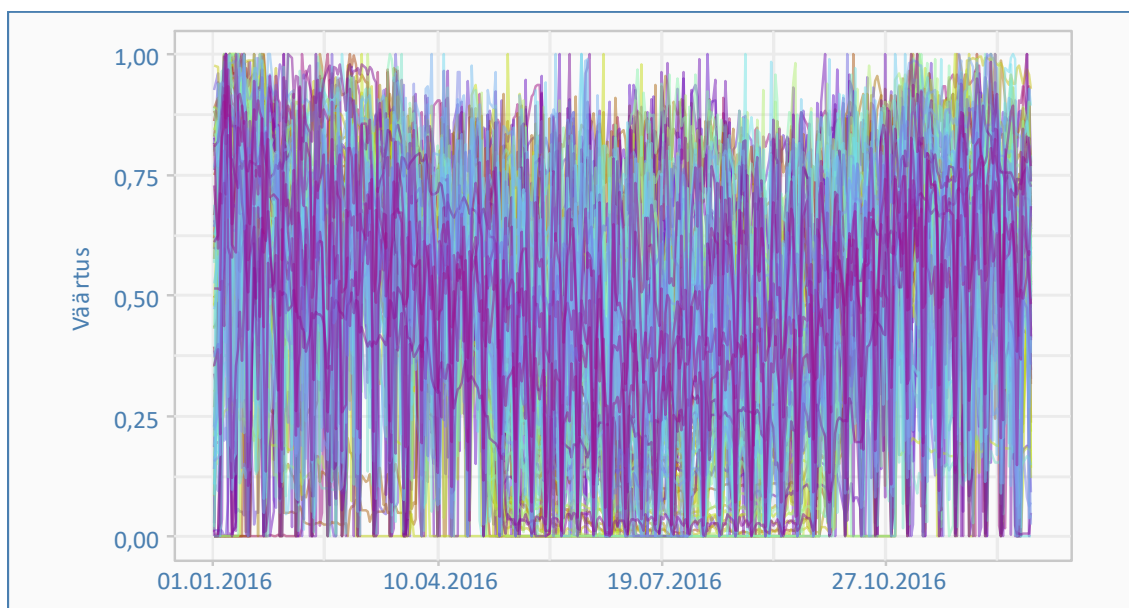
Joonis 3.16. Teise klasteri normaliseeritud aegread

Joonisel 3.16 esitatud teise klasteri andmed näitavad, et see on kõige väiksema tarbimise volatiilsusega klaster, ja seda on näha ka joonisel. Enamuste klientide tarbimise aegread on paigaldatud üksteise lähedal, kuigi on haruldasi kõrvalekaldeid ja teravaid hüppeid.



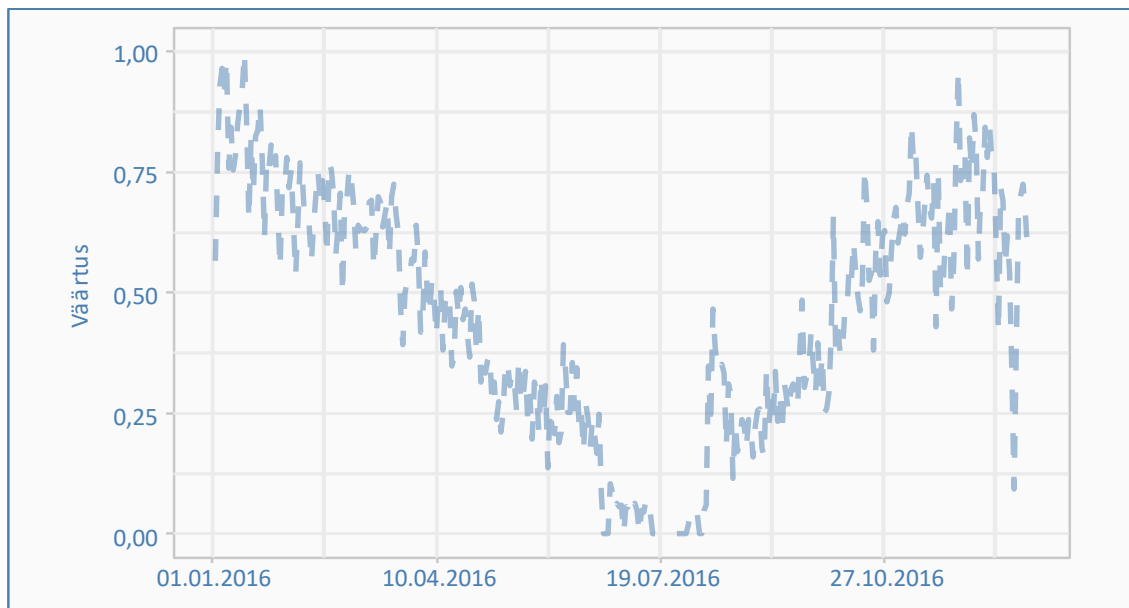
Joonis 3.17. Teise klatri normaliseeritud aegridade tsentroid

Joonisel 3.17 on näha, et tegelikkuses teine klaster on väikse tarbimise volatiilsusega ja hästi nähtavalt sesoonsusega.



Joonis 3.18. Kolmanda normaliseeritud aegread

Kolmas klaster näitab sesoonsuse, kuid omab ka suure tarbimise volatiilsuse.

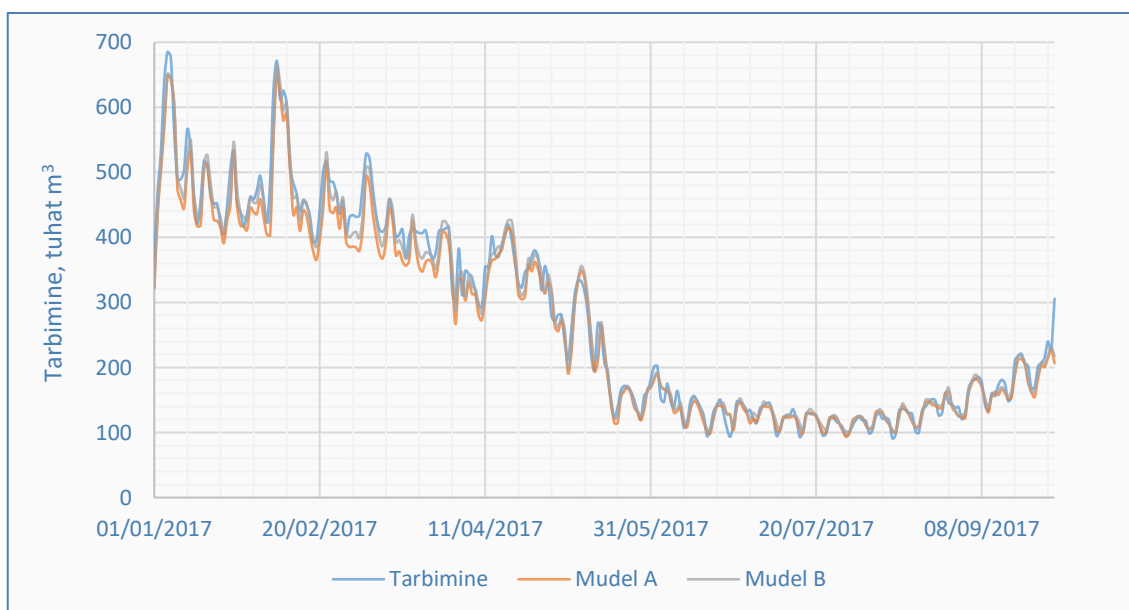


Joonis 3.19. Kolmanda klasteri normaliseeritud aegriidade tsentroid

Joonisel 3.19 esitatud kolmanda klasteri tsentroid on sarnane teise klasteri tsentroidiga (Joonis 3.17), kuid see tsentroid on märgatavalt vähem prognoositav. Saadud tulemused eksporditi *Exceli* formaati edaspidiseks kasutamiseks prognoosimisel.

3.5.2 Klasterdamise mõju prognoosi täpsusele

Viidi läbi kaks katset. Esimese katse käigus loodi mudel kõigi klientide tarbimissumma alusel ning prognoos tehti kohe kogutarbimise kohta (mudel A). Teine katse viidi läbi moodustatud klasterite abil ja iga klasteri jaoks loodi eraldi prognoosimudel (mudel B). Saadud tulemus on joonisel 3.20.



Joonis 3.20. Kahe mudeliga tehtud prognoosid ja reaalne tarbimine

Klastritel põhinev mudel on oluliselt tulemuslikum, nagu on näha tabelis 3.6.

Tabel 3.6. Mudelite prognooside täpsuseparameetrite võrdlus

Parameeter	Mudel A	Mudel B
ME	12491,6	3729,1
MSE	$5,9 \times 10^8$	$3,6 \times 10^8$
RMSE	24276,9	18869,1
MAE	17295,1	13403,8
MPE	3,2%	0,3%
MAPE	6,0%	5,2%
SMAPE	6,2%	5,2%
U_1	0,04	0,03
U_2	0,67	0,59
MASE	0,77	0,60

See tõestab teooriat, et klientide jagamine klastritesse vastavalt nende tarbimisharjumustele võib prognoosi täpsust oluliselt parandada. RMSE paranes 22,3%, SMAPE 15,4%, U_2 paranes 11,9% ja MASE 22,1%.

Tasub pöörata suurt tähelepanu asjaolule, et klastrite arv sõltub andmete liigist, mõnikord on vaja rohkem klastreid, mõnikord vähem. Selle töö kontekstis kasutati üldpõhimõtte demonstreerimiseks sümbolset numbrit, kuid see ei tähenda, et see on optimaalne klastrite arv antud andmetüübi jaoks. Optimaalse klastrite arvu leidmine on raske ülesanne ja sageli on see valik selle isiku käes, kes ennustuse teeb.

3.6 Katsetamine päev-ette prognoosi andmetega

Siiani viidi kõik uuringud läbi juba olemasolevate ilmastikuandmetega, kuid see ei kajasta tegelikkust täielikult, kuna gaasitarbimise prognoosimiseks tuleb teada ka tulevast temperatuuriväärtust (st ilmaprognoosi). Ilmaprognoos sisaldab viga, seega otsustati kontrollida, kui palju väheneb prognoosi täpsus, kui kasutada mitte reaalaja temperatuurinäitajaid, vaid järgmise päeva prognoosi.

Ilmaprognoosi andmed saadi ajavahemikul 01.07.2017–18.04.2019, mis oluliselt piirab andmeperioodi, mille jooksul mudelit saab treenida. Seega otsustati valida treenimisperioodiks ajavahemik 01.07.2017–30.06.2018 ja kontrollperioodiks 01.07.2018–31.12.2018.

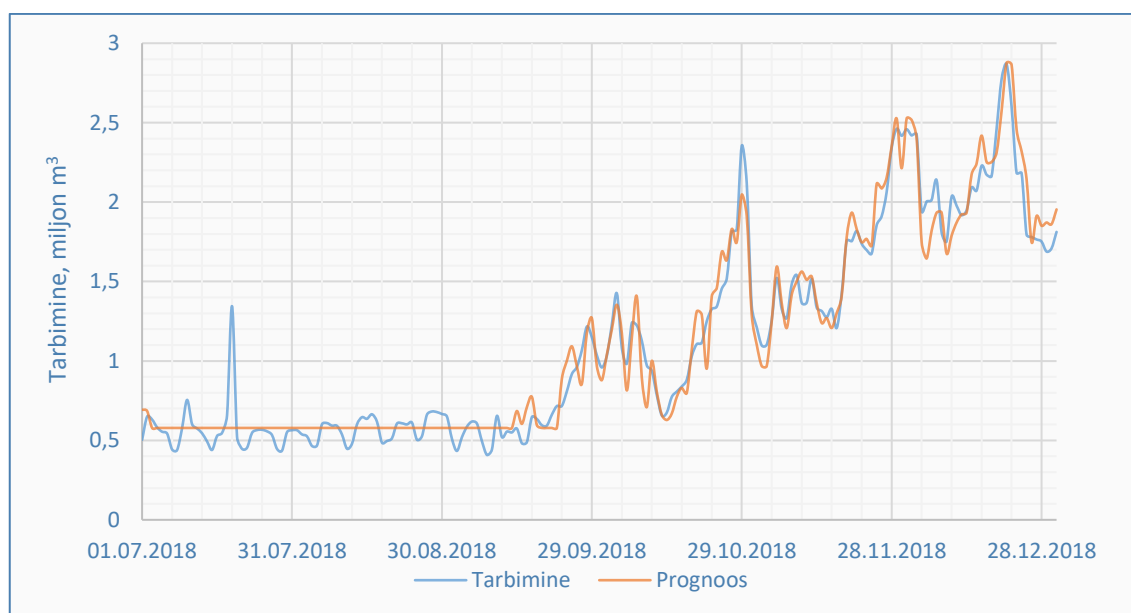
Viidi läbi 3 katset:

- Mudeli treenimine tegelike andmetega, prognoosimine keskmise ilmaprognoosi andmetega (mudel A)
- Treenimine ja prognoosimine ilmaprognoosi keskmist temperatuuri kasutades (mudel B)
- Treenimine ja prognoosimine ilmaprognoosi maksimumtemperatuuri kasutades (mudel C)

Andmeid võrreldi Eesti Gaasi mudeliga saadud andmetega ja mudeliga, mis tugines sama ajavahemiku reaalsele andmetele.

3.6.1 Eesti Gaasi mudel

Võrreldes mudeliga, mis oli kasutatud punktis 3.3.1 antud mudel kasutab päev-ette prognoosi temperatuuri andmed. Võrreldes reaalse andmete kasutatava Eesti Gaasi mudeliga, antud Eesti Gaasi mudeli täpsus ei muutunud palju: SMAPE, U_2 ja MASE parameetrid on sarnased ja RMSE läks parem. Allpool joonisel 3.12 on esitatud Eesti Gaasi mudeli tulemused perioodis 01.07.2018–31.12.2018.



Joonis 3.21. Eesti Gaasi mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 132383,6 \quad (3.70)$$

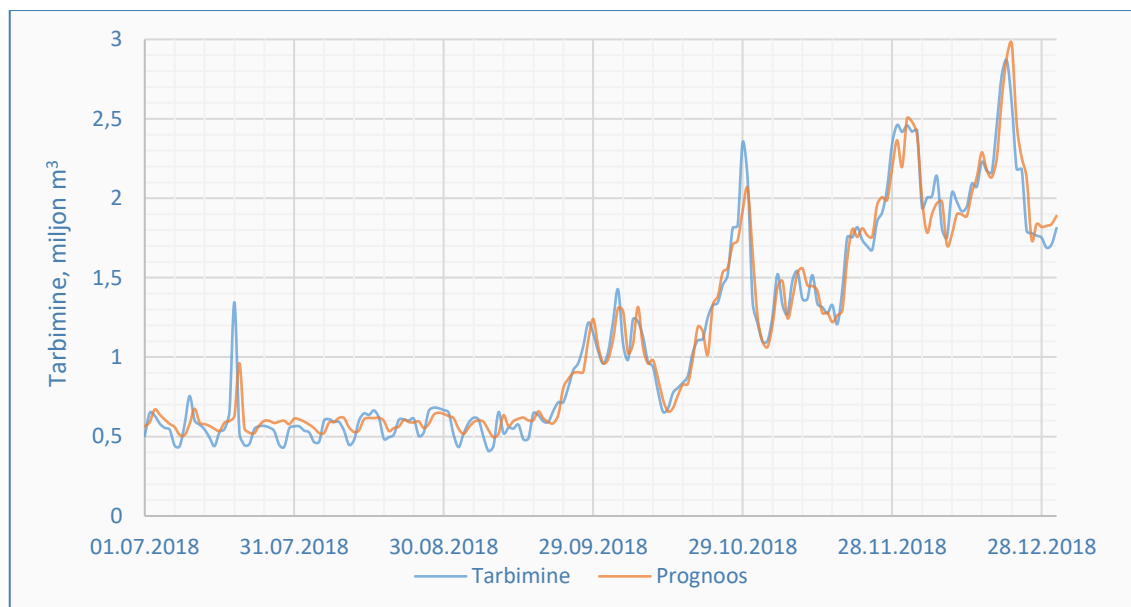
$$SMAPE = 10,2\% \quad (3.71)$$

$$U_2 = 1,01 \quad (3.72)$$

$$MASE = 0,96 \quad (3.73)$$

3.6.2 RA mudel

Antud mudelit treeniti tegeliku temperatuuri andmetega ja see koostab prognoosi kasutades tegeliku temperatuuri andmete põhjal. Antud juhul on see mudel mõeldud kontrollimiseks, kuidas muutub teiste mudelite täpsus võrreldes selle mudeliga. Täpsuse huvides on selle mudeli treenimisperiood sama mis allpool vaadeldavatel.



Joonis 3.22. „RA“ mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 120179,8 \quad (3.74)$$

$$SMAPE = 8,7\% \quad (3.75)$$

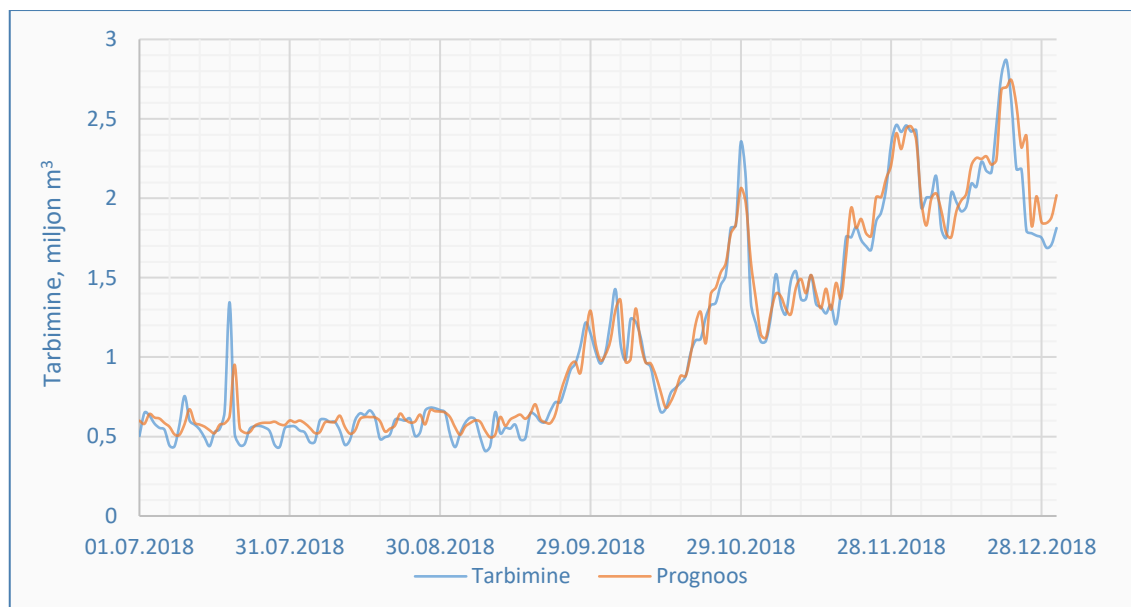
$$U_2 = 0,87 \quad (3.76)$$

$$MASE = 0,82 \quad (3.77)$$

See mudel on Gaussi regressiooni mudel, mida kirjeldati punktis 3.4.5. Ainus erinevus on see, et sellise mudeli treenimiseks oli ette nähtud lühem periood (01.07.2017–30.06.2018 võrreldes 01.03.2014–31.12.2017). Sel põhjusel halveneb prognoosimise täpsus selle mudeli abil.

3.6.3 Mudel A

Antud mudeli põhimõtte seisneb selles, et tihti on raske saada päev-ette prognoosi eelmisi andmeid. Tavaliselt on eelmised andmed reaallaja andmed ja tulevikuanndmeteks on ilmaprognoos. Seetõttu otsustati kontrollida, kuidas toimib mudel, mida on treenitud reaallaja andmetega, ja prognoos toimub päev-ette prognoosi andmete põhjal.



Joonis 3.23. „A“ mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 128675,9 \quad (3.78)$$

$$SMAPE = 8,9\% \quad (3.79)$$

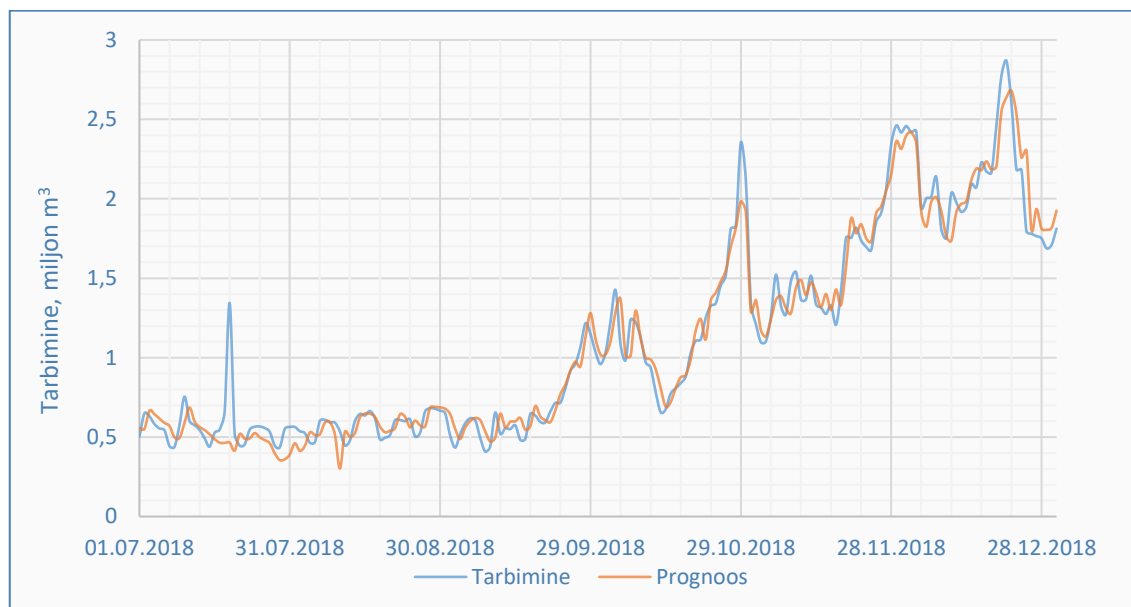
$$U_2 = 0,90 \quad (3.80)$$

$$MASE = 0,86 \quad (3.81)$$

Antud mudeli jõudlus on sarnane mudeliga, mida on treenitud ja koostab prognoosi reaallaja andmete põhjal. Nagu oli oodata, mudeli täpsus vähenes. Kuid täpsuse kahanemine osutus tähtsusetuks ja mudeli jõudlus on märgatavalt parem kui Eesti Gaasi mudel. Võrreldes RA mudeliga langesid RMSE, SMAPE, U_2 ja MASE vastavalt 7,1%, 2,7%, 3,5% ja 4,9% võrra, kuid võrreldes Eesti Gaasi mudeliga on antud mudeli jõudlus parem vastavalt 2,8%, 12,5%, 10,9% ja 10,4%.

3.6.4 Mudel B

Antud mudelit on treenitud ja see koostab prognoosi ilmaprognoosi keskmiste temperatuuriväärtuste põhjal.



Joonis 3.24. „B“ mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 127991,4 \quad (3.82)$$

$$SMAPE = 9,6\% \quad (3.83)$$

$$U_2 = 1,00 \quad (3.84)$$

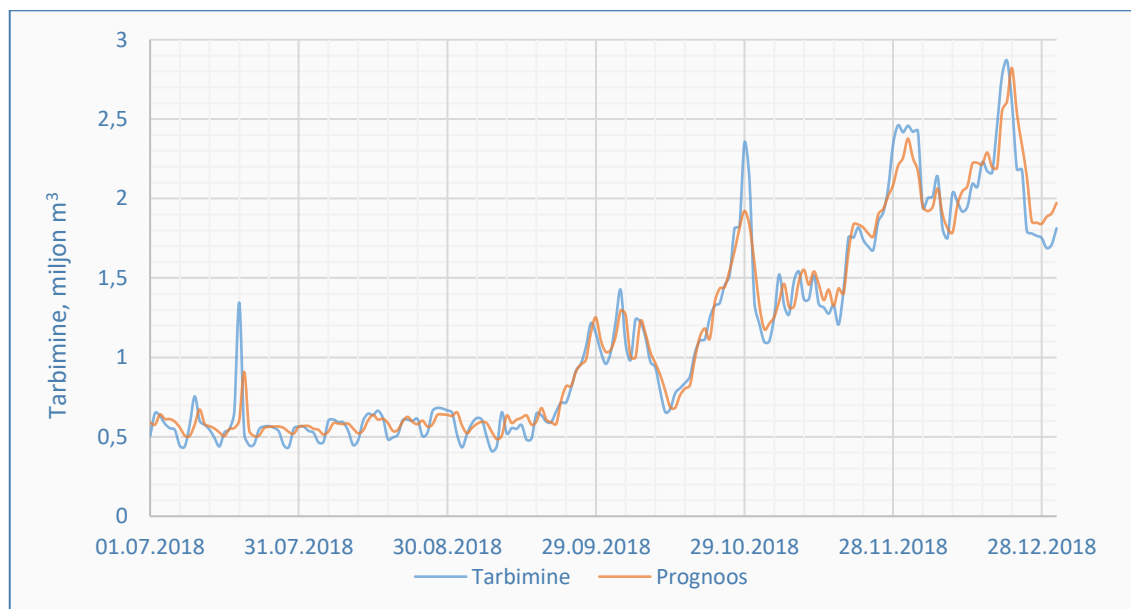
$$MASE = 0,85 \quad (3.85)$$

Selle mudeli täpsus vähenes oluliselt, võrreldes RA mudeliga. Eesti Gaasi mudeliga võrdlemisel saab öelda, et mudel B ei suuda olemasoleva mudeli tulemusi kuigi hästi parandada. Võrreldes RA mudeliga langesid RMSE, SMAPE, U_2 ja MASE vastavalt 6,5%, 10,3%, 22% ja 3,7% võrra ja võrreldes Eesti Gaasi mudeliga RMSE on 3,3% parem, SMAPE on 5,9% parem, U_2 on 1% parem ja MASE on 11,5% parem.

Mudeli täpsuse sellise tugeva halvenemise põhjus jäi selgusetuks, kuna mudel A kasutab prognoosimiseks samu andmeid.

3.6.5 Mudel C

Antud mudelit on treenitud ja see kasutab prognoosi koostamisel Ilmateenistuse päev-ette prognoosi temperatuuri maksimumväärtusi. Selle mudeliga saadud prognoos on toodud allpool joonisel 3.25.



Joonis 3.25. „C“ mudeliga tehtud prognoos ja reaalne tarbimine

$$RMSE = 127727,8 \quad (3.86)$$

$$SMAPE = 8,6\% \quad (3.87)$$

$$U_2 = 0,87 \quad (3.88)$$

$$MASE = 0,86 \quad (3.89)$$

Antud mudeli tulemus on väga sarnane RA mudeli tulemustega; RMSE ja MASE halvenesid vastavalt 6,3% ja 4,9% võrra, kuid U_2 ei muutunud ja SMAPE paranes 0,7%. Võrreldes Eesti Gaasi mudeliga muutusid RMSE, SMAPE, U_2 ja MASE paremaks vastavalt 3,5%, 15,5%, 13,9%, 10,4%.

3.6.6 Tulemuste analüüs ja järeldused

Kõikide mudelitega saadud tulemused on toodud allolevas tabelis.3.7

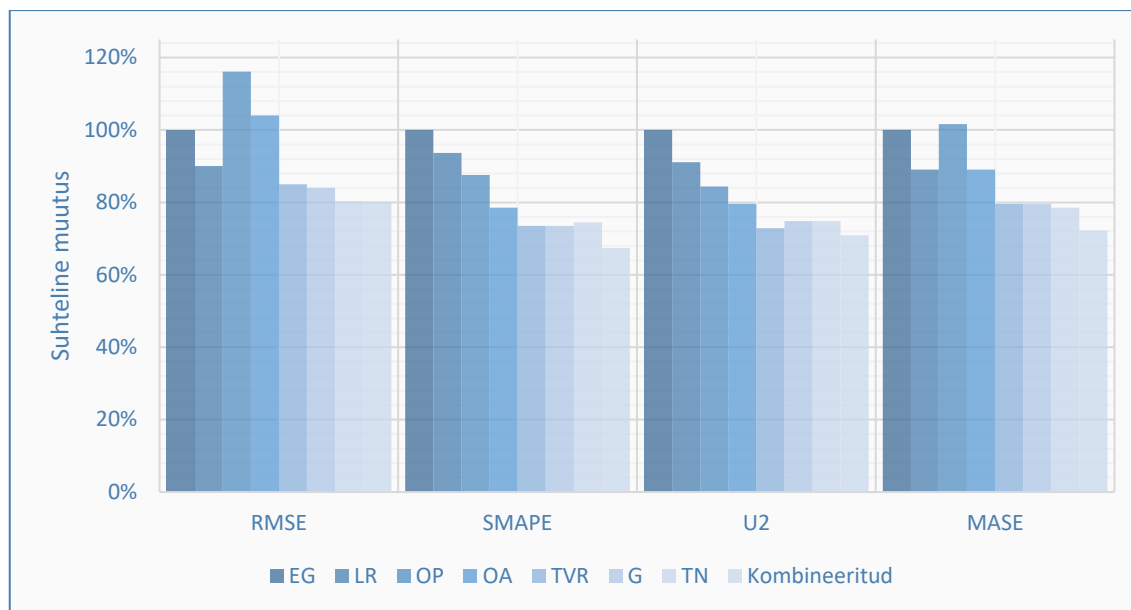
Tabel 3.7. Mudelite prognooside täpsuseparameetrite võrdlus

Parameeter	EG	RA	Mudel A	Mudel B	Mudel C
ME	$-1,3 \times 10^4$	$-7,2 \times 10^3$	$-2,7 \times 10^4$	$5,6 \times 10^3$	$-1,2 \times 10^4$
MSE	$1,8 \times 10^{10}$	$1,4 \times 10^{10}$	$1,7 \times 10^{10}$	$1,6 \times 10^{10}$	$1,6 \times 10^{10}$
RMSE	$1,3 \times 10^5$	$1,2 \times 10^5$	$1,3 \times 10^5$	$1,3 \times 10^5$	$1,3 \times 10^5$
MAE	$9,7 \times 10^4$	$8,3 \times 10^4$	$8,8 \times 10^4$	$8,7 \times 10^4$	$8,8 \times 10^4$
MPE	-3,1%	-2,8%	-4,3%	-0,1%	-2,9%
MAPE	10,4%	9,0%	9,3%	9,3%	8,8%
SMAPE	10,5%	8,7%	8,9%	9,6%	8,6%
U_1	0,05	0,05	0,05	0,05	0,05
U_2	1,01	0,87	0,90	1,00	0,87
MASE	0,96	0,82	0,86	0,85	0,86

Kõik kolm mudelit näitasid paremaid tulemusi kui Eesti Gaasi mudel. Ootuspäraselt nende täpsus kahanes, kuid see on tingitud peamiselt asjaolust, et nende kolme mudeli treenimisel kasutati märksa vähem algandmeid ja katsetamiseks palju lühemat ajavahemikku, mis kahtlemata mõjutas oluliselt mudelite täpsust. See leidis tõestust reaalaaja andmeid kasutava mudeli treenimisega ja katsetamisega samal ajavahemikul. Ebapiisavate andmete tõttu on raske kindlaks teha, milline mudel on kõige täpsem. Kuid võime siiski kindlalt öelda, et kõik eelpool vaadeldud mudelid toimivad täpsemini kui algne mudel. Ja veel: kui suureneb mudelite treenimiseks kasutatav andmehulk, muutuvad mudelid täpsemaks, nagu nähtub punktist 3.3, kus mudelite treenimiseks oli kasutatud 4 korda pikemat ajavahemikku.

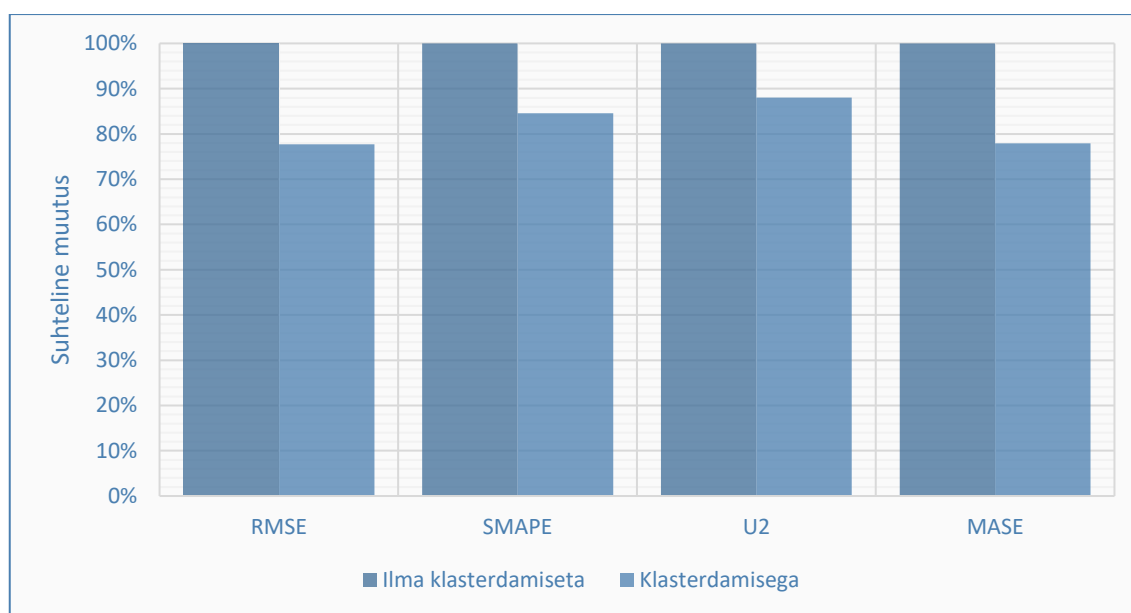
3.7 Peatüki kokkuvõte

Töö käigus uuriti erinevaid matemaatilisi meetodeid maagaasi tarbimise prognoosimiseks ja võrreldi neid algse (Eesti Gaas) mudeliga. Alapeatükis võetakse kokku kõikide katsetamise tulemused ning võrreldakse täpsusnäitajate suhtelist paranemist. Kõikide katsete tulemused visualiseeritakse allpool joonistel 3.26, 3.27 ja 3.28.



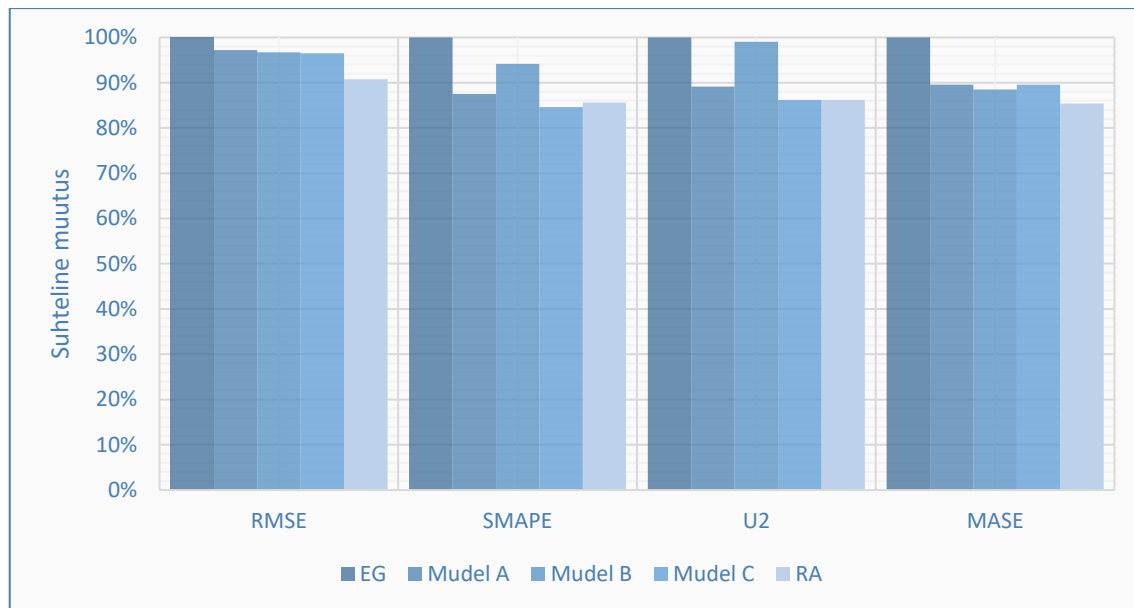
Joonis 3.26. Mudelite täpsusnäitajate võrdlus

Nagu diagrammilt 3.26 näha on enamiku masinõppemudelite jõudlus parem kui algne mudel. Parimad mudelid näitavad erinevate parameetrite paranemist 20% kuni 30%. Kõige täpsemad tulemused saavutati mudelite kombineerimisega.



Joonis 3.27. Mudeli täpsusparameetrite võrdlus ilma klasterdamiseta ja klasterdamisega

Nagu diagrammilt näha, võimaldab klientide jagunemine klastriteks parandada prognoosi täpsuse näitajaid kuni 20%.



Joonis 3.28. Mudeli täpsusparameetrite võrdlus päev-ette prognoosiandmetega treenitud mudelite puhul

Joonis 3.28 näitab, et hoolimata asjaolust, et mudelite üldine jõudlus on langenud, võrreldes täpsunäitajaid "RA" mudeliga võib kindlalt öelda, et see juhtus pigem mudelite treenimisperioodi lühendamise tõttu, mitte niivõrd ilmaprognooside andmete kasutamise tõttu. Kahjuks saadi selle katse jaoks liiga vähe andmeid, et treenida mudeleid pikema ajavahemiku jooksul.

KOKKUVÕTE

Lõputöö peamiseks ülesandeks oli maagaasitarbimise prognoosimudeli koostamine, millega saaks teha maagaasitarbimise päev-ette prognoose, võttes aluseks eelmine tarbimise ajalugu ja väliseid tarbimist mõjutavaid parameetreid. Töö aluseks on Eesti Gaasi poolt antud erinevate klientide maagaasi tarbimisandmed ning hetkel kasutusel olev prognoosimudel. Töö üks olulisemaid eesmärke oli leida prognoosimeetod mis lubaks täpsemaid tulemusi kui hetkel Eesti Gaasis kasutatav mudel.

Enne mudelite koostamist uuriti maagaasi tarbimise prognoosi tähtsust ettevõtte, kliendi ja energiasüsteemi seisukohalt, millised on traditsioonilised meetodid aegridade prognoosimiseks ning eelnevate sarnaste uuringute metoodikaid ja tulemusi. Mudelite koostamisel valiti katsetamiseks viis masinõpe algoritmi ja üks närvivõrgu algoritm. Enne mudelite koostamist uuriti ja analüüsiti järgnevaid prognoosi mõjutavaid aspekte:

- prognoositäpsuse võrdlemise meetodid;
- välised tarbimise mõjutavad parameetrid, korrelatsioonid parameetrite ja tarbimise vahel ja nende korrelatsioonide statistiline tähtsus;
- algoritmide põhimõtted.

Kasutatud mudelite prognooside täpsust võrreldi omavahel ja ka hetkel Eesti Gaasis oleva mudeli täpsusega. Tulemused viitasid, et masinõppe ja närvivõrkude mudelitega on võimalik saavutada tunduvalt suurem täpsus võrreldes hetkel kasutatava mudeliga. Masinõppe ja närvivõrkude mudelite kasutades märgati vigade suurust indikeerivate parameetrite paranemist keskmiselt 25% võrra, mõned parameetrid kuni 30%.

Eesmärgiga prognoositäpsust veelgi enam parandada, uuriti lähemalt kahte matemaatilist meetodit prognoosimudelite täiendamiseks: mitme prognoosimudeli tulemuse kombineerimine ja klientide klasterdamine rühmadesse nende tarbimisprofiilide järgi enne prognoosimudelite rakendamist.

Klientide klasterdamiseks oli kasutatud aegridade analüüsimeetodit, mis annab võimaluse jagada kliente nende tarbimisekäitumise järgi. Meetodi katsetamiseks kasutati R programmeerimiskeel. Mudelite testimine näitas, et klasterdamine omab positiivset mõju mudeli täpsusele, keskmiselt muutus mudeli täpsus 20% paremaks.

Kokkuvõttes osutusid mõlemad prognoositäpsuse parandamise meetodid tõhusaks.

Viimaseks etapiks oli katsetada kuidas toimib loodud mudel reaalses elus, arvestades ilmastikuprognoside ebatäpsust. Selleks treeniti mudelit ilmastiku mõõteandmete asemel ilmastiku prognoosiandmetega. Prognoosiandmetega mudeli testimine näitas, et mudeli täpsus muutus väiksemaks, kuid oli tõestatud, et täpsuse langus toimus seepärast, et masinaõpe mudelite treenimiseks oli kasutatud palju lühem treenimis periood piiratud andmete tõttu.

Tulemuseks on mudel, mille täpsus on suurem kui praegu Eesti Gaasis kasutatav mudel, ja mille eeliseks on see, et ta muutub veelgi täpsemaks tarbimise ajaloo andmete kasvamisega. Lisaks uuringu käigus saadud tulemused võivad olla kasulikud mitte ainult inseneridele, kes on huvitatud gaasitarbimise prognoosimisest, vaid ka inseneridele, keda huvitab teiste energiaallikate tarbimise prognoosimisest, kuna kasutatud meetodid on kohaldatavad teiste uuringute puhul.

Terve tarbimise prognoosimisel kõige parema mudeliga oli saavutatud sümmeetriline keskmine absoluutne protsendiviga 7,3%. Mudelite kombineerides viga läks 6,7% Klasterdamisega oli saavutatud viga 5,2%. Kokkuvõttes täideti lõputöö eesmärk leida matemaatiline meetod, mis oleks täpsem kui hetkel Eesti Gaasis kasutusel olev matemaatiline prognoosimudel.

SUMMARY

The main goal of the thesis was to create the forecasting model of natural gas consumption, which could be used for day-ahead prognoses based on previous consumption history and external consumption affecting parameters. The study is based on the natural gas consumption data of different customers provided by the Eesti Gaas and the forecasting model currently in use by the company. One of the most important research goals was to find a method of forecasting that would allow providing of more accurate results than those of the model currently in use at the Eesti Gaas.

Before creating models, the author studied the importance of natural gas consumption forecasting for the company, the customer and the energy system. Furthermore, the traditional methods of time series forecasting and previous similar studies, methodologies and results were analysed. Five machine-learning algorithms and one neural network algorithm were selected for testing.

Before the models were developed, the following aspects affecting the forecast were examined and analysed:

- methods of comparing forecasting accuracy;
- external parameters affecting consumption, correlations between parameters and consumption and statistical significance of such correlations;
- principles of algorithms.

The accuracy of the forecasts of the models used was compared with each other and with the accuracy of the model currently in use at the Eesti Gaas. The results indicated that models of machine-learning and of neural networks can achieve much higher precision compared to the model currently in use. When using the machine learning and neural network models, improvements on average by 25%, for some parameters – up to 30% were noted.

Aiming at further improvement of the forecasting accuracy, the author took a closer look at the two mathematical methods: the combination of several forecasting models and clustering of customers depending on their consumer profiles before the implementation of forecasting models.

A time series analysis method called dynamic time warping was used to cluster customers, which provides an opportunity to group customers according to their consumption behaviour. The R programming language was used to test the method. The model testing showed that clustering has a positive impact on model accuracy, with an average of 20% improvement in predictability.

To sum it up, both methods of improving forecasting accuracy proved to be effective.

The final step was testing the created model in a real-life situation, taking into account the inaccuracy of weather forecasts. To achieve the goal, the model was trained with weather forecast data instead of measured weather data. The test conducted for the model with weather forecast data showed that the model's accuracy decreased, but the test revealed that the drop in accuracy occurred in the presence of limited data as a significantly shorter training period was used to train the machine-learning models.

The research resulted in the creation of a model the accuracy of which is higher than that of the current model in use at the Eesti Gaas. Another advantage of the model created is that it becomes even more accurate with growth of historical consumption data. In addition, the outcomes of the research may be useful not only to engineers involved in gas consumption forecasting but also to people concerned with forecasting of consumption for other sources of energy since the methods used are applicable to other studies too.

The forecasting of the entire consumption with best model showed a symmetric mean absolute percentage error of 7.3%. When combining the models, the error went to 6.7%. With clustering, the error reached 5.2%. To sum up, the goal of the thesis to find a mathematical method that is more accurate than the mathematical prediction model currently in use in the Eesti Gaas was achieved.

KASUTATUD KIRJANDUSE LOETELU

- [1] L. Doman, „EIA projects 28% increase in world energy use by 2040,“ 01 09 2017. [Võrgumaterjal]. Available: <https://www.eia.gov/todayinenergy/detail.php?id=32912>. [Kasutatud 19 05 2019].
- [2] ExxonMobil, „ExxonMobil Outlook for Energy,“ 2018. [Võrgumaterjal]. Available: <https://corporate.exxonmobil.com/-/media/global/files/outlook-for-energy/2018-outlook-for-energy-data-pages.xlsx>. [Kasutatud 19 05 2019].
- [3] U.S. Energy Information Administration, „Natural Gas Explained,“ 11 12 2018. [Võrgumaterjal]. Available: https://www.eia.gov/energyexplained/print.php?page=natural_gas_home. [Kasutatud 19 05 2019].
- [4] Central Intelligence Agency, „COUNTRY COMPARISON :: NATURAL GAS - PRODUCTION,“ [Võrgumaterjal]. Available: <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2249rank.html>. [Kasutatud 19 05 2019].
- [5] Central Intelligence Agency, „COUNTRY COMPARISON :: NATURAL GAS - EXPORTS,“ [Võrgumaterjal]. Available: <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2251rank.html>. [Kasutatud 19 05 2019].
- [6] Elering, „Gaasisüsteemi ajalugu,“ [Võrgumaterjal]. Available: <https://elering.ee/gaasituru-kasiraamat/2-eesti-ja-regionaalne-maagaasisusteem/21-gaasisusteemi-ajalugu>. [Kasutatud 19 05 2019].
- [7] Elering, „Eesti maagaasi ülekandevõrk,“ [Võrgumaterjal]. Available: <https://elering.ee/gaasituru-kasiraamat/2-eesti-ja-regionaalne-maagaasisusteem/22-eesti-maagaasi-ulekandevork>. [Kasutatud 19 05 2019].
- [8] Elering, „Torustik,“ [Võrgumaterjal]. Available: <https://elering.ee/gaasituru-kasiraamat/2-eesti-ja-regionaalne-maagaasisusteem/22-eesti-maagaasi-ulekandevork/221>. [Kasutatud 19 05 2019].

- [9] Elering, „Maagaasi import,“ [Võrgumaterjal]. Available: <https://elering.ee/gaasituru-kasiraamat/2-eesti-ja-regionaalne-maagaasisusteem/25-maagaasi-import>. [Kasutatud 19 05 2019].
- [10] Elering, „Maagaasi tarbimine,“ [Võrgumaterjal]. Available: <https://elering.ee/gaasituru-kasiraamat/2-eesti-ja-regionaalne-maagaasisusteem/24-maagaasi-tarbimine>. [Kasutatud 19 05 2019].
- [11] Elering, „Bilansivastutuse protsessid,“ [Võrgumaterjal]. Available: <https://elering.ee/63-bilansivastutuse-protsessid>. [Kasutatud 19 05 2019].
- [12] Elering, „Gaasituru bilansivastutuse ahel,“ [Võrgumaterjal]. Available: <https://elering.ee/gaasituru-kasiraamat/6-bilansihaldus/61-gaasituru-bilansivastutuse-ahel>. [Kasutatud 19 05 2019].
- [13] Elering, „Bilansigaasi hinna määramise metoodika,“ [Võrgumaterjal]. Available: <https://elering.ee/gaasituru-kasiraamat/6-bilansihaldus/63-bilansivastutuse-protsessid/634-bilansigaasi-hinna>. [Kasutatud 19 05 2019].
- [14] Elering, „Bilansiteenus,“ [Võrgumaterjal]. Available: <https://elering.ee/bilansiteenus-0>. [Kasutatud 19 05 2019].
- [15] GET Baltic, „Baltic Gas Exchange,“ [Võrgumaterjal]. Available: <https://www.getbaltic.com/>. [Kasutatud 19 05 2019].
- [16] R. Hribar, P. Potočnik, J. Silc ja G. Papa, „A comparison of models for forecasting the residential natural gas demand of an urban area,“ 11 2018. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/328678341_A_comparison_of_models_for_forecasting_the_residential_natural_gas_demand_of_an_urban_area. [Kasutatud 19 05 2019].
- [17] J. G. Tamba, S. N. Essiane, E. F. Sapnken, F. D. Koffi, J. L. Nsouandélé, B. Soldo ja D. Njomo, „Forecasting Natural Gas: A Literature Survey,“ 28 03 2018. [Võrgumaterjal]. Available: <https://ideas.repec.org/a/eco/journ2/2018-03-28.html>. [Kasutatud 19 05 2019].
- [18] B. Soldo, „Forecasting natural gas consumption,“ 04 2012. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/257156711_Forecasting_natural_gas_consumption. [Kasutatud 19 05 2019].

- [19] N. K. Bose ja P. Liang, Neural Network Fundamentals with Graphs, Algorithms, and Applications, 1996.
- [20] DeepAI, „Autoregressive Model,“ [Võrgumaterjal]. Available: <https://deepai.org/machine-learning-glossary-and-terms/autoregressive-model>. [Kasutatud 19 05 2019].
- [21] J. Brownlee, „11 Classical Time Series Forecasting Methods in Python (Cheat Sheet),“ 6 08 2018. [Võrgumaterjal]. Available: <https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>. [Kasutatud 19 05 2019].
- [22] G. E. P. Box ja G. M. Jenkins, Time Series Analysis: Forecasting and Control, 5th Edition, 2015.
- [23] MathWorks, „Autoregressive Moving Average Model,“ [Võrgumaterjal]. Available: <https://se.mathworks.com/help/econ/arma-model.html>. [Kasutatud 19 05 2019].
- [24] J. Brownlee, „A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python,“ 20 08 2018. [Võrgumaterjal]. Available: <https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/>. [Kasutatud 19 05 2019].
- [25] MathWorks, „What Is a Linear Regression Model?,“ [Võrgumaterjal]. Available: <https://se.mathworks.com/help/stats/what-is-linear-regression.html>. [Kasutatud 19 05 2019].
- [26] E. Alpaydm, Introduction to Machine Learning, Third Edition, London: The MIT Press, 2014.
- [27] D. Opitz ja R. Maclin, „Popular Ensemble Methods: An Empirical Study,“ 12 1999. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/2244615_Popular_Ensemble_Methods_An_Empirical_Study. [Kasutatud 19 05 2019].
- [28] R. Polikar, „Ensemble based systems in decision making,“ 10 2006. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/3432431_Polikar_R_Ensemble_based_systems_in_decision_making_IEEE_Circuit_Syst_Mag_6_21-45. [Kasutatud 19 05 2019].

- [29] L. Rokach, „Ensemble-based classifiers,“ 02 2010. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/220637823_Ensemble-based_classifiers. [Kasutatud 19 05 2019].
- [30] scikit-learn, „Support Vector Machines,“ [Võrgumaterjal]. Available: <https://scikit-learn.org/stable/modules/svm.html>. [Kasutatud 19 05 2019].
- [31] H. Drucker, C. J. Burge, L. Kaufma, A. Smola ja V. Vapnik, „Support Vector Regression Machines,“ [Võrgumaterjal]. Available: <http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf>. [Kasutatud 19 05 2019].
- [32] C. Cortes ja V. Vapnik, „Support-Vector Networks,“ 1995. [Võrgumaterjal]. Available: http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf. [Kasutatud 19 05 2019].
- [33] F. Perez-Cruz, S. V. Vaerenbergh, J. J. Murillo-Fuentes, M. Lázaro-Gredilla ja I. Santamaria, „Gaussian Processes for Nonlinear Signal Processing: An Overview of Recent Advances,“ 07 2013. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/260637079_Gaussian_Processes_for_Nonlinear_Signal_Processing_An_Overview_of_Recent_Advances. [Kasutatud 19 05 2019].
- [34] M. R. Manne, R. K. Srivastava, D. Bisht, H. C. Sharma ja A. Kumar, „Development of Artificial Neural-Network-Based Models for the Simulation of Spring Discharge,“ 07 2011. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/220173163_Development_of_Artificial_Neural-Network-Based_Models_for_the_Simulation_of_Spring_Discharge. [Kasutatud 19 05 2019].
- [35] S. Asiri, „Meet Artificial Neural Networks,“ 22 12 2017. [Võrgumaterjal]. Available: <https://towardsdatascience.com/meet-artificial-neural-networks-ae5939b1dd3a>. [Kasutatud 19 05 2019].
- [36] „Stochastic Processes Glossary,“ [Võrgumaterjal]. Available: <http://www.stats.org.uk/stochastic-processes/glossary.html>. [Kasutatud 19 05 2019].
- [37] Z. Boussaada, O. Curea, R. Ahmed, H. Camblong ja M. B. Najiba, „A Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation,“ 03 2018. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/323717600_A_Nonlinear_Autoregressive_Ex

- ogenous_NARX_Neural_Network_Model_for_the_Prediction_of_the_Daily_Direct_Solar_Radiation. [Kasutatud 19 05 2019].
- [38] Д. И. Новохрестова, „Временная нормализация слогов алгоритмом динамической трансформации временной шкалы при оценке качества произнесения слогов в процессе речевой реабилитации,“ [Võrgumaterjal]. Available: <https://journal.tusur.ru/storage/97968/31-Новохрестова.pdf?1516337886>. [Kasutatud 19 05 2019].
- [39] Statistics How To, „Mean Squared Error: Definition and Example,“ [Võrgumaterjal]. Available: <https://www.statisticshowto.datasciencecentral.com/mean-squared-error/>. [Kasutatud 19 05 2019].
- [40] Statistics How To, „RMSE: Root Mean Square Error,“ [Võrgumaterjal]. Available: <https://www.statisticshowto.datasciencecentral.com/rmse/>. [Kasutatud 19 05 2019].
- [41] Statistics How To, „Absolute Error & Mean Absolute Error (MAE),“ [Võrgumaterjal]. Available: <https://www.statisticshowto.datasciencecentral.com/absolute-error/>. [Kasutatud 19 05 2019].
- [42] Statistics How To, „Mean absolute percentage error (MAPE),“ [Võrgumaterjal]. Available: <https://www.statisticshowto.datasciencecentral.com/mean-absolute-percentage-error-mape/>. [Kasutatud 19 05 2019].
- [43] Vanguard Software Corporation, „Forecast Fit,“ [Võrgumaterjal]. Available: <https://www.vanguardsw.com/business-forecasting-101/forecast-fit/>. [Kasutatud 19 05 2019].
- [44] F. Bliemel, „Theil's Forecast Accuracy Coefficient: A Clarification,“ [Võrgumaterjal]. Available: <https://pdfs.semanticscholar.org/cc66/033ed3047c1b3d5eb47d85ec52789b36324d.pdf>. [Kasutatud 19 05 2019].
- [45] A. B. Koehler ja R. J. Hyndman, „Another look at measures of forecast accuracy,“ 2006. [Võrgumaterjal]. Available: http://eva.fcea.edu.uy/pluginfile.php/109034/mod_resource/content/0/2006_Hyndman_Predicc.pdf. [Kasutatud 19 05 2019].

- [46] Распознавание, „Коэффициент корреляции Пирсона,“ [Võrgumaterjal]. Available: http://www.machinelearning.ru/wiki/index.php?title=Коэффициент_корреляции_Пирсона. [Kasutatud 19 05 2019].
- [47] S. R. Vitullo, R. H. Brown, G. F. Corliss ja B. M. Marx, „Mathematical Models for Natural Gas Consumption,“ 2009. [Võrgumaterjal]. Available: http://www.math.ualberta.ca/ami/CAMQ/pdf_files/vol_17/17_4/17_4i.pdf. [Kasutatud 19 05 2019].
- [48] R. L. Winkler ja S. Makridakis, „The Combination of Forecasts,“ 01 1983. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/261805839_The_Combination_of_Forecasts. [Kasutatud 19 05 2019].
- [49] J. P. Dickinson, „Some Statistical Results in the Combination of Forecasts,“ 06 1973. [Võrgumaterjal]. Available: https://www.researchgate.net/publication/245278376_Some_Statistical_Results_in_the_Combination_of_Forecasts. [Kasutatud 19 05 2019].