

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Anne-Liis Tänav 178116IABM

Kaoelektri prognoosimine Elektrilevi OÜ näitel

Magistritöö

Juhendajad: Juri Belikov, PhD
Margarita Matson,
PhD

Tallinn 2022

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Anne-Liis Tänav

07.05.2022

Annotatsioon

Magistritöö raames vaatame kaelektrit kui võrku siseneva ja tarbijate poolt tarbitud elektrienergia vahet. Lõputöö eesmärkideks on:

- Koostada võimalikult hea mudel, mis prognoosib kao kogust tunnipõhiselt päev ette.
- Võrrelda, kas lõputöö raames koostatud kaelektri prognoosimise mudel on täpsem kui olemasolev kaelektri prognoosimise meetoodika.
- Hinnata milline on pakutud mudeli kasutamise finantsmõju.
- Uurida, kas kaelektri kogust mõjutavad tuul ja/või päike.

Analüüsi tulemusena pakutakse lõputöö eesmärkide täitmiseks kaks lahendust. Esimene lahendus kasutab SLR (Sliding Linear Regression) meetoodikat tunnipõhistel andmetel ja iga ööpäeva tunni kohta leitakse eraldi SLR mudel uuritavaks päevaks. Teine lahendus kasutab SLR meetoodikat päevapõhistel andmetel ja prognoositud päeva kogus jaotatakse hiljem eelneva aasta sama nädalapäeva kaelektri kõvera alusel. Vaatlusaluste andmete põhjal saab väita, et nii tunnipõhine SLR mudel kui ka päevapõhine SLR-il põhinev mudel saavutavad madalamad MAPE-d võrreldes hetkel kasutatava kaelektri prognoosimise meetoodikaga.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 36 leheküljel, 5 peatükki, 13 joonist, 13 tabelit.

Abstract

Forecasting electricity losses: Elektrilevi OÜ case study

In this master's thesis, we look at electricity losses as the difference between the electricity entering the network and the electricity consumed by consumers. The purchase of electricity losses entails significant costs for Elektrilevi OÜ, which in turn are passed on to the network service tariffs. The more accurate the forecast of electricity losses, the lower the additional costs generally associated with its purchase.

The existing methodology for forecasting electricity losses was developed years ago. It has not changed significantly over time and is based on expert judgment on an almost daily basis. Previous years have shown that the predictions of the current model are quite successful, but due to the COVID-19 crisis, there have been setbacks in the accuracy of the model.

The goals of this thesis are:

- To develop the best possible model that predicts the amount of loss electricity on an hourly basis for one day ahead.
- Compare the existing loss electricity forecast methodology to the new loss electricity model forecast.
- Assess the financial implications of using the proposed new model.
- Examine if wind or sunshine have an effect on the amount of electricity losses.

As a result of the analysis conducted on previous literature and of the data provided for this thesis, two solutions are proposed to meet the goals of this thesis. The first solution uses the SLR (Sliding Linear Regression) methodology on an hourly basis and a separate SLR model is found for each of the 24 hours of the day. The second solution uses the SLR methodology on a day basis and the estimated amount of a days loss of electricity is later distributed on the basis of the electricity loss curve for the same day of the week in the previous year. Based on the data under consideration, it can be stated that both the

hourly SLR model and the day-based SLR model achieve lower MAPE results compared to the currently used loss of electricity forecast.

The financial impact analysis for the alternative loss of electricity forecast models showed that the added expense associated with the hourly SLR model is more than 9000€ higher than that of the day-based SLR model for the period 16.02.2020–31.10.2021 (excluding special days).

Regarding the effects of wind and sunshine on loss of electricity quantities the analysis showed that only wind was statistically and practically significant and only in the daily SLR model.

In order to expand this work, a solution for forecasting hourly loss of electricity for special days should be explored. In addition, it should be noted that both proposed SLR models rely heavily on the assumption that the data for the model is available in a timely manner and is of high quality. In order to see how the model works in real life, it should be used on a daily basis during a test period in addition to the existing loss of electricity forecasting process.

The thesis is in Estonian and contains 36 pages of text, 5 chapters, 13 figures, 13 tables.

Lühendite ja mõistete sõnastik

| | |
|------------|--|
| AI | <i>Artificial Intelligents</i> ehk tehisintellekt |
| ANOVA | <i>Analysis of Variance</i> ehk dispersioonanalüüs |
| AGNES | <i>Agglomerative Hierarchical Clustering</i> ehk aglomeratiivne hierarhiline klasterdamine |
| CET | <i>Central European Time</i> ehk Kesk-Euroopa aeg |
| EET | <i>Eastern European Time</i> ehk Ida-Euroopa aeg |
| ELV | Elektrilevi OÜ |
| Kaoelekter | Võrku siseneva ja tarbitud elektrienergia vahe |
| LR | Lineaarne regressioon |
| MAE | <i>Mean Absolute Error</i> ehk keskmine absoluutne viga |
| MAPE | <i>Mean Absolute Percentage Error</i> ehk keskmine absoluutne protsentuaalne viga |
| ML | <i>Machine Learning</i> ehk masinõpe |
| MP1–MP9 | Päevapõhistel andmetel koostatud libiseva lineaarse regressiooni mudeli versioonid |
| MP8n45 | Päevapõhistel andmetel koostatud libiseva lineaarse regressiooni mudeli versioon 8, milles on kasutatud treeningandmete akent suurusega 45 |
| MPT1–MPT5 | Päevapõhise libiseva lineaarse regressiooni mudeli MP8n45 väljundi jaotusreeglitiku variandid tunnipõhiseks väljundiks |
| MSE | <i>Mean Square Error</i> ehk keskmine ruutviga |
| MT1–MT8 | Tunnipõhistel andmetel koostatud libiseva lineaarse regressiooni mudeli versioonid |
| MT8n45 | Tunnipõhistel andmetel koostatud libiseva lineaarse regressiooni mudeli versioon 8, milles on kasutatud treeningandmete akent suurusega 45 |
| R2 | <i>Squared Correlation Coefficient</i> ehk determinatsioonikordaja |
| RMSE | <i>Root Mean Square Error</i> ehk keskmise ruutvea juur |
| SLR | <i>Sliding Linear Regression</i> ehk libisev lineaarne regressioon |
| SWE | <i>Sliding Window Regression</i> ehk libiseva akna regressioon |
| UTC | <i>Coordinated Universal Time</i> ehk koordineeritud universaalae |

Sisukord

| | |
|---|----|
| 1 Sissejuhatus | 10 |
| 2 Kaoelektri prognoosimine | 12 |
| 2.1 Kasutatavad andmed, mudeli headuse mõõdikute ja tarkvara valik..... | 13 |
| 2.2 Kaoelektri kirjeldav statistika..... | 16 |
| 2.3 Hetkel kasutatava meetodika kirjeldus ja mudeli täpsus..... | 19 |
| 2.4 Olemasolevate meetodite ülevaade | 21 |
| 3 Uus kaoelektri prognoosimise mudel | 25 |
| 3.1 Kasutatud tunnused ja seos kaoelektriga | 26 |
| 3.2 Uute kaoelektri prognoosimudelite koostamise protsess..... | 30 |
| 3.3 Tunnipõhine SLR mudel | 32 |
| 3.4 Päevapõhine SLR mudel | 35 |
| 4 Tulemuste analüüs | 39 |
| 4.1 Olemasoleva ja pakutud kaoelektri prognoosimudelite võrdlus..... | 39 |
| 4.2 Pakutud mudelite käitumine seoses kellakeeramiselega | 41 |
| 4.3 Pakutud mudelite rahaline mõju | 42 |
| 5 Kokkuvõte | 44 |
| Kasutatud kirjandus | 46 |
| Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks | 49 |
| Lisa 2 – Töös kasutatavate tunnipõhiste arvtunnuste jaotused ja korrelatsioonimaatriks koos teisendatud muutujatega..... | 50 |
| Lisa 3 – Pilvisuse kategooria dendogramm 2020 ja 2021 aasta kohta | 51 |
| Lisa 4 – Nädalapäeva ANOVA tulemused koos Tukey testiga | 52 |
| Lisa 5 – Pilvisuse kategooria ANOVA tulemused koos Tukey testiga..... | 53 |
| Lisa 6 – Töös kasutatavate päevapõhiste arvtunnuste jaotused ja korrelatsioonimaatriks koos teisendatud muutujatega..... | 54 |

Jooniste loetelu

| | |
|---|----|
| Joonis 1. Mudeli loomise üldine protsess..... | 12 |
| Joonis 2. Kaoelektri kuised kogused (MWh, sinised tulbad) 01.2020–10.2021 koos Tallinna kuu keskmise temperatuuriga (C°, oranž joon)..... | 17 |
| Joonis 3. Kaoelektri päevane kogus (MWh, sinine joon) ja Tallinna päevane keskmine temperatuur (C°, oranž joon) perioodis 01.2020–10.2021..... | 17 |
| Joonis 4. Keskmine tunnipõhine kaoelektri kogus (MWh). Eripäevad on andmetest eemaldatud..... | 18 |
| Joonis 5. Keskmine kaoelektri kogus tunnis (MWh) erinevate nädalapäevade lõikes koos tava- ja eripäeva eristusega..... | 18 |
| Joonis 6. Keskmine tunnipõhine kaoelektri kogus (MWh) nädalapäeviti. Eripäevad ei ole kaasatud..... | 19 |
| Joonis 7. Olemasoleva kaoelektri prognoosi MAPE kuude lõikes 01.2020–10.2021.... | 21 |
| Joonis 8. Nädalapäeva klasteranalüüsi tulemus 2020 aasta andmetel. Vasakpoolsel dendogrammil on andmetest eraldatud riigipühad ja riiklikult lühendatud tööpäevad, parempoolsel dendogrammil on andmed koos nimetatud eripäevadega..... | 29 |
| Joonis 9. Tegelik ja MT8 _{n45} mudeli prognoositud kaoelektri kõverad perioodis 07.2020..... | 35 |
| Joonis 10. Tegelik ja MPT5 mudeli prognoositud kaoelektri kõverad perioodis 07.2020..... | 38 |
| Joonis 11. Kaoelektri päevapõhised tegelikud, MT8 _{n45} ja MPT5 põhjal prognoositud kogused 13.08.2020–13.09.2020..... | 41 |
| Joonis 12. 2020 aasta kevadise kellakeeramise seotud kaoelektri kõverad. Joonisel on toodud 22.03.2020, 29.03.2020 ja 5.04.2020 tegelik, MT8 _{n45} prognoositud ja MPT5 prognoositud kaoelektri kõverad..... | 41 |
| Joonis 13. 2020 aasta sügisese kellakeeramise seotud kaoelektri kõverad. Joonisel on toodud 18.10.2020, 25.10.2020 ja 1.11.2020 tegelik, MT8 _{n45} prognoositud ja MPT5 prognoositud kaoelektri kõverad..... | 42 |

Tabelite loetelu

| | |
|---|----|
| Tabel 1. Kokkuvõte magistritöös kasutatavatest tunnipõhistest töötlemata tunnustest koos nendega seotud lühenditest, ühikutest, päritolust ja nende kaasatusest olemasolevas metoodikas ($M_{\text{Olemasolev}}$) ja käesoleva töö raames uue mudeli leidmisel (M_{Uus})..... | 13 |
| Tabel 2. Töös kasutatavate tunnipõhiste arvtunnuste korrelatsioonimaatriks. | 27 |
| Tabel 3. Töös kasutatavate viitega ja teisendusega tunnipõhiste arvtunnuste korrelatsioonimaatriks. | 28 |
| Tabel 4. Päevapõhiste seletavate arvtunnuste korrelatsioon uuritava tunnusega. | 30 |
| Tabel 5. Treeningandmete akna leidmise skeem prognoositava päeva suhtes..... | 31 |
| Tabel 6. Uutes kaoelektri prognoosimudelites kasutatavad tunnused ja treeningandmete akende suurused..... | 31 |
| Tabel 7. Tunnipõhise SLR mudeli versioonid koos kasutatud tunnuste ja mudelite keskmise MAPE-ga erinevate treeningandmete akende juures..... | 33 |
| Tabel 8. Tunnipõhiste mudelite MT5–MT8 keskmistatud korrigeeritud R^2 , Mudeli olulisus (p), MAPE, MAE, MSE ja RMSE. | 34 |
| Tabel 9. Päevapõhise SLR mudeli versioonid koos kasutatud tunnuste ja mudelite tervikvalimi MAPE-ga erinevate treeningandmete akende juures. | 36 |
| Tabel 10. Päevapõhiste mudelite MP5–MP9 keskmistatud korrigeeritud R^2 , Mudeli olulisus (p), MAPE, MAE, MSE ja RMSE. | 37 |
| Tabel 11. Päevapõhise SLR mudeli $MP8_{n45}$ väljundi jaotusreeglitiku variandid koos nende tulemusel tekkinud tunnipõhise prognoosi MAPE-ga. | 38 |
| Tabel 12. ELV ehk hetkel kasutatava prognoosi ja mudelite $MT8_{n45}$ ning MPT5 prognoosi keskmine MAPE (hall taust: uue mudeli MAPE ületab ELV MAPE-t) kuude lõikes koos tegeliku ja pakutud mudelite poolt prognoositud kaoelektri kuise kogusega (MWh) (sinine taust: kogus prognoositud kogus on väiksem kui tegelik, punane taust: prognoositud kogus ületas tegelikku). *andmed alates 16.02.2020. | 40 |
| Tabel 13. Tunnipõhise SLR mudeli $MT8_{N45}$ ja päevapõhisel SLR põhineva mudeli MPT5 prognoosi veaga seotud rahaline mõju kuude lõikes.* andmed alates 16.02.2020. | 43 |

1 Sissejuhatus

Magistritöö raames vaatame kaelektrit kui võrku siseneva ja tarbijate poolt tarbitud elektrienergia vahet. Tegemist on energiavooga, mille saab jagada tehniliseks kaoks ja kommertskaoks. Pidev investeerimine võrku, sealhulgas sobivate materjalide kasutamine ja võrgu ilmastikukindlamaks muutmine aitab hoida tehnilise kao taseme optimaalselt madalal tasemel. Kommertskadu on olulisel määral alandanud ja ohjanud kaugloetavatele arvestitele üleminek ja tarbimisjärelvalve töö.

Vastavalt Elektrilevi OÜ (ELV) viimaste majandusaastate aruannetele on kaelektri kogused võrgus olnud pidevas väikeses languses, mis on seotud suuresti kaugloetavate arvestite kasutusega. Kaugloetavad arvestid võimaldavad elektrikoguste täpset ja operatiivset mõõtmist ning võrgus olevate mõõtepunktide pidevat monitoorimist. Kao kogused 2018–2020 olid järgnevad: 319,2 GWh (4,3% sisenevast elektrist), 298 GWh (4,1% sisenevast elektrist) ja 284,1 GWh (4,0% sisenevast elektrist) [1].

Võttes arvesse taastuenergia direktiivi [2] ja juba ajas suurenevat taastuenergia osatähtsust tarbimises [3] on põhjust arvata, et ka ilmastikunähtused nagu tuul ja päike võivad mõjutada kaelektri koguseid.

Kaelektri ostmisega kaasnevad ELV-le olulised kulud, mis suunatakse omakorda edasi võrguteenuse tariifi. Alates 2013.aastast ostab ELV võrgukadude katmiseks vajaliku elektri avatud turult [4].

Olemasolev kao prognoosi meetodika kujunes välja aastaid tagasi ja on üsnagi ressursikulukas protsess, mis ei ole ajas märkimisväärselt muutunud ja tugineb igapäevaselt ka olulisel määral eksperthinnangule. Varasemad aastad on näidanud, et olemasoleva mudeli ennustused on küllaltki edukad, kuid seoses koroonakriisiga ilmnedid mudeli täpsuses tagasilöögid.

Lõputöö eesmärkideks on:

- Koostada võimalikult hea mudel, mis prognoosib kao kogust tunnipõhiselt päev ette.

- Võrrelda, kas lõputöö raames koostatud kao prognoosi mudel on täpsem kui olemasolev kao prognoosi meetoodika.
- Hinnata milline on pakutud mudeli kasutamise finantsmõju.
- Seoses taastuvenergia toodangu suureneva osakaaluga uurida, kas kao kogust mõjutavad tuul ja/või päike.

Autori panus käesolevasse töösse hõlmab kogu tööga seotud andmetöötlust- ja analüüsi (sh andmete korjet erinevatest algallikatest, töötlust, teisendusi, kirjeldavat statistikat, korrelatsioonanalüüsi, klasteranalüüsi), hetkel kasutatava kaelektri prognoosi meetoodika kirjeldust ja täpsuse analüüsi, varasema kirjanduse analüüsi ja sünteesi, uue prognoosimudeli koostamiseks rakendatava meetoodika valikut ja rakendamist (sh sisendite ja mudeli parameetrite valikut, R-i koodi kirjutamist), mudelite arendamist ja valideerimist, tulemuste (sh finantsmõju) analüüsi ja kokkuvõtete ning järelduste tegemist.

Lõputöö koosneb kolmest põhiosast, millest esimeses osas defineeritakse uuritav probleem koos kasutatavate andmete ja tarkvara valikuga. Lisaks sellele antakse detailne ülevaade hetkel kasutatavast meetoodikast kao prognoosimisel ja seoses sellega antakse ka ülevaade mudeli täpsusest. Esimese põhiosa võtab kokku kirjanduse ülevaade, mille põhjal toimub ka uue kao koguse prognoosimise meetoodika valik antud töö raames.

Töö teises osas antakse detailne ülevaade olemasolevatest andmetest ja antud töö käigus kaasatud uutest tunnustest ning nende seosest kaelektriga ja uuritavate hüpoteesidega. Töö teises osas toimub ka uue mudeli rakendamine ja optimeerimine kaelektri koguse prognoosimiseks.

Töö viimases põhiosas toimub tulemuste analüüs, mis hõlmab olemasoleva ja pakutud mudeli täpsuse võrdlust, pakutud mudeli rakendamisega kaasnevat finantsmõju hinnangut ja järelduste tegemist pakutud mudeli rakendatavuse üle tulevikus.

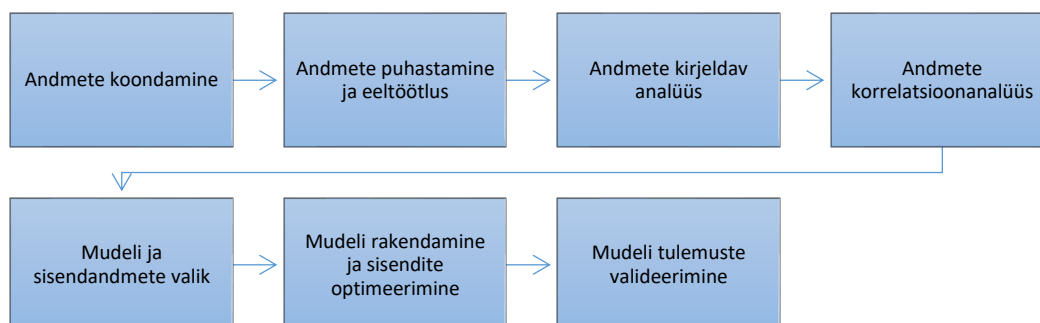
Autor soovib tänada Elektrilevi OÜ-d andmete kasutuse õiguse eest.

2 Kaoelektri prognoosimine

Alates 2013.aastast ostab ELV võrgukadude katmiseks vajaliku elektri avatud turult [4]. Kao ost koosneb määratud ja avatud tarnest. Määratud tarne on ELV poolne ost otse Nord Pooli turult ja avatud tarne on seotud bilansitehingutega, mida teostab ELV eest hanke korras Eesti Energia [4]. Mida täpsem on kaoelektri prognoos, seda väiksemad on üldjuhul selle ostuga seoses tehtavad lisakulud.

Nord Pool on Euroopa juhtiv päev ette elektriturg, mille osalejatele (kokku 16 Euroopa riiki) antakse võimalus igal päeval kella 12.00-ni (CET) esitada järgmise ööpäeva tunnipõhiseid elektri ostu- ja müügipakkumisi, mille põhjal tuvastatakse pärast päeva sulgemist elektri pakkumise ja nõudluse tasakaalupunktipõhine elektrihind, kusjuures hinnad arvutatakse tunnipõhiselt ja iga hinnapiirkonna jaoks eraldi ning avaldatakse hiljemalt sama päeva kella 13-ks (CET) [5].

Üldise mudeli loomise protsessi sammud on toodud Joonisel 1 ning antud sammudest lähtutakse ka käesolevas töös kaoelektri prognoosimudeli loomisel. Määratud tarne ost Nord Pool turult määrab ära kaoelektri prognoosi formaadi, täpsemalt selle, et prognoos tuleb teha tunnipõhiselt ja vähemalt 24 tundi ette. Kaugloetavate arvestite kasutamine ELV võrgus annab eelise võimalikult täpseks kao ennustuseks, kuna eelnev annab võimaluse kasutada kõrge detailsuse ja kvaliteediga andmeid.



Joonis 1. Mudeli loomise üldine protsess.

Järgnevas alampeatükides antakse ülevaade kasutatavatest andmetest, kaoelektrist kui prognoositavast tunnusest koos selle kirjeldava statistikaga, mudeli headuse mõõdikutest

ja tarkvara valikust, olemasolevast metoodikast ja selle senisest edust ning lõpetatakse ülevaatega maailmas kasutatavatest alternatiividest sarnase probleemi lahendamisel. Eelnevaga seoses läbitakse mudeli loomise üldise protsessi (Joonis 1) esimesed 3 sammu.

2.1 Kasutatavad andmed, mudeli headuse mõõdikute ja tarkvara valik

Käesoleva lõputöö eesmärkide täitmiseks on koondatud tunnipõhised andmed kolmest andmekogust: ELV koondatud bilansiselgitused, Tartu füüsikahoone e-ilmajaama arhiivist [6] ja rp5.ru ilmaandmete arhiivist [7]. Täiendav info riigipühade ja riiklikult lühendatud tööpäevade kohta on võetud Riigiteatajast [8].

Siinkohal tuleb ära märkida asjaolu, et kao andmed on ajas muutuvad kuna bilansiselgituse koostamine ja seega ka kaoelektri koguse arvutus toimub aruandluskuu ja sellele eelneva 12 kuu kohta. Arvestades eelnevat ja andmete saamise kuupäeva on töös kasutatavatest andmetest lõplikud perioodi 01.2020–11.2020 tunniandmed.

Kõigis andmekogudes on ajaga seotud andmed kohalikus ajavööndis, mis tähendab aga seda, et analüüsimisel tuleb arvesse võtta suve- ja talveajale üleminekutega. Eelnevaga seoses on analüüsi lihtsustamise ja kooskõla huvides üle mindud UTC+2 ajavööndile.

Tabel 1 võtab kokku töös kasutatavad andmed nii olemasoleva kao prognoosi metoodika kui ka pakutud uue metoodika vaates. Tabeli 1 koostamise eeldusena on läbitud mudeli loomise üldise protsessi esimesed kaks sammu (Joonis 1).

Tabel 1. Kokkuvõte magistritöös kasutatavatest tunnipõhistest töötlemata tunnustest koos nendega seotud lühenditest, ühikutest, päritolust ja nende kaasatusest olemasolevas metoodikas ($M_{\text{Olemasolev}}$) ja käesoleva töö raames uue mudeli leidmisel (M_{Uus}).

| Tunnuse kirjeldus (nimetus) | Arvestuspõhimõte /mõõtühik | Andmete päritolu | Kasutatud $M_{\text{Olemasolev}}$ | Kasutatud M_{Uus} |
|---|-------------------------------|---------------------|-----------------------------------|----------------------------|
| Võrku sisenev elekter kokku (Sisenev) | Summaarne, MWh | ELV bilansiselgitus | x | |
| Ostetav elekter =kadu + omatarve (Kadu) | Summaarne, MWh | ELV bilansiselgitus | | x |
| Kaoelektri osakaal | Osakaal, % | ELV bilansiselgitus | x | |
| Ostu prognoosi viga | Tegeliku ja prognoosi vahe, % | ELV bilansiselgitus | | |

| Tunnuse kirjeldus (nimetus) | Arvestuspõhimõte /mõõtühik | Andmete päritolu | Kasutatud $M_{Olemasolev}$ | Kasutatud M_{Uus} |
|------------------------------------|-----------------------------------|--------------------------|--|---------------------------------------|
| Temperatuur Tartus (Mtemp) | Keskmine, C° | Allikas [6] | x | x |
| Tuule kiirus Tartus (Mtuul) | Keskmine, m/s | Allikas [6] | | x |
| Sademe hulk Tartus (Mvihm) | Summaarne, mm | Allikas [6] | | x |
| Kiirgusvoo hulk Tartus (Mkiirgus) | Summaarne, W/m ² | Allikas [6] | | x |
| Temperatuur Tallinnas (Rtemp) | Keskmine, C° | Allikas [7] | | x |
| Tuule kiirus Tallinnas (Rtuul) | Keskmine, m/s | Allikas [7] | | x |
| Sademe hulk Tallinnas (Rvihm) | Summaarne, mm | Allikas [7] | | x |
| Pilvisuse aste Tallinnas (Rpilv) | Kategooriline, % | Allikas [7] | | x |
| Kuupäev | Kuupäev, DMY | ELV bilansiselgitus | x | x |
| Nädalapäev | Kategooriline | Exceli funktsioon | x | x |
| Tund | Lõputund, EET / UTC+2 | ELV bilansiselgitus | x | x |
| Päeva pikkus | Tundide arv | R funktsioon daylength() | | x |
| Eripäev | Kategooriline, Y/N | Allikas [8] | | x |
| Prognoositud temperatuur | Päeva keskmine temperatuur, C° | ELV andmed | | x |

Olemasolevas mudelis kasutatavad tunnused ja nende kasutamise viis on kujunenud läbi aja ja põhinevad valdavalt katse-eksitus meetodil ning ekspertide hinnangutel. Olemasoleva kao prognoosi mudeli puhul mängivad olulist rolli aeg (nädalapäev, tund, eripäev) ja temperatuur. Lühidalt kokkuvõttes prognoositakse päeva võrku sisenevat energiat ja kao protsenti, ning see päevakogus jagatakse laiali 24 tunni vahel.

Uue mudeli leidmisel võetakse arvesse nii olemasolevas mudelis edukalt kasutatud tunnuseid kui ka kirjanduses leiduvaid asjakohaseid ettepanekuid elektrienergia kogust

mõjutavatest teguritest. Siinkohal saab välja tuua Eesti lühiajalist elektritarbimist mõjutavaid tegureid uuriva artikli [9] tulemused, kus olulisteks mõjuriteks osutusid temperatuur, nädalapäev, päeva pikkus, tuule kiirus, pilvisus ja lähimineviku tarbimine.

Täiendavate ilmastikutegurite nagu tuule kiirus ja pilvisus kaasamine uude mudelisse on seotud ka taastuvenergiaga. Eestis on päikese ja tuuleenergiaga seotud ja avalikkusele tarbimiseks võrku antud kogused aastatel 2018–2020 vastavalt statistikaameti andmebaasile tõusnud tuuleenergia puhul 635,0 GWh-lt 842,0 GWh-le ja päikeseenergia puhul 30,8 GWh-lt 122,5 GWh-le [10].

Kirjanduses leidub hulgaliselt erinevaid meetodeid ja näidikuid mudelite headuse hindamiseks ja mudelite omavaheliseks võrdlemiseks. Valdavalt põhinevad nimetatud näidikud vea hinnangutel ehk prognoositava ja tegeliku väärtuse erinevusel. Valik selliseid näidikuid allikate [11], [12] ja [13] põhjal koos arvutusvalemitega, kus n on vaatluste arv, k on sõltumatute muutujate arv, \hat{y}_i on prognoositav väärtus ja y_i on tegelik väärtus, on:

- determinatsioonikordaja (R^2) ja korrigeeritud determinatsioonikordaja (R_{adj}^2), mis kirjeldavad kui suur osa sõltuva muutuja dispersioonist on mudeli poolt kaetud. Viimast eelistatakse, kui mudelis osaleb mitu sõltumatut muutujat. R^2 ja R_{adj}^2 arvutamiseks kasutatakse allolevaid valemeid :

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

$$R_{adj}^2 = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1} \right] \quad (2)$$

- keskmine absoluutne viga (MAE), mis on hindab absoluutset keskmist erinevust mudeli ja tegelike tulemuste vahel kuid ei ole tundlik suurte vigade olemasolule tulemis. MAE arvutamiseks kasutatakse allolevat valemit :

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3)$$

- keskmine ruutviga (MSE) ja keskmise ruutvea juur (RMSE), mis sobivad hästi erindite leidmiseks, kuna võimendavad suuri vigu. MSE ja RMSE arvutamiseks kasutatakse allolevaid valemeid :

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5)$$

- keskmine absoluutne protsentuaalne viga (MAPE), mis sobib eelnevatest paremini erinevate mudelite võrdlemisel, kuna erinevus leitakse ühiku tasemel. MAPE arvutamiseks kasutatakse allolevat valemit :

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (6)$$

Uue mudeli versioone võrreldakse kõigi ülalnimetatud mõõdikute alusel.

Olemasoleva ja uue prognoosimudeli headuse võrdlemiseks kasutatakse keskmise absoluutse protsentuaalse vea näitajat ehk *mean absolute percentage error* (MAPE), mis võtab lisaks prognoositava ja tegeliku väärtuse vahele arvesse ka tegeliku väärtuse suurust andmaks lihtsamini mõistetava tulemuse erinevate mudelite võrdlemisel. Lisaks MAPE arvutusmetoodikast tulenevale, kasutatakse käesolevas töös just MAPE-t võrdluses olemasoleva mudeliga, kuna need väärtused olemasoleva mudeli kohta sisalduvad andmestikus, mille alusel käesolevat analüüsi teostatakse.

Rakendatava tarkvara valik on tehtud lähtudes Elektrilevi OÜ-s kasutatavate IT-halduse ja IT-turbe piirangutest, arvestades võimalusega, et uut mudelit oleks võimalik ettevõttes ka tulevikus sobivuse korral rakendada hakata. Sisendandmete koondamiseks ja puhastamiseks ning analüüsiks sobivale kujule viimiseks kasutatakse Microsoft Exceli tabelitöötuse tarkvara ja uue mudeli leidmiseks on kasutatud R-i koos kasutajaliidesega R Studio.

2.2 Kaoelektri kirjeldav statistika

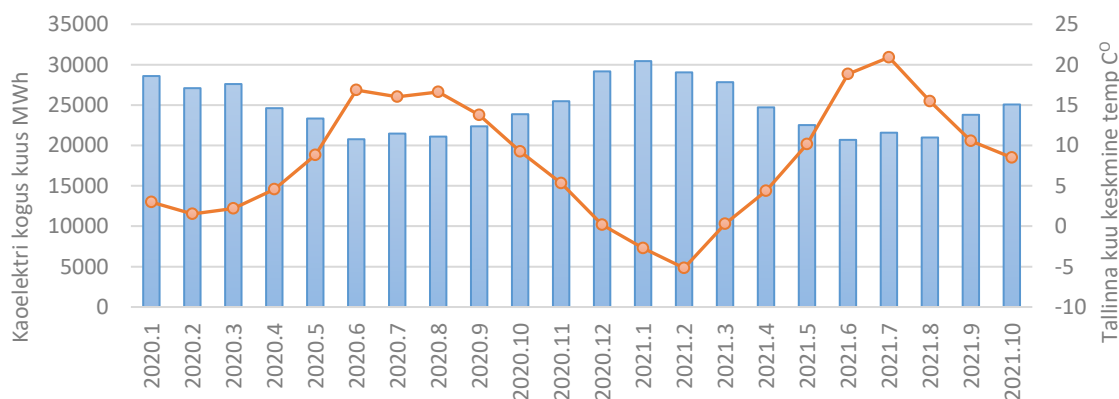
Käesolevas peatükis antakse ülevaade kaoelektri kirjeldavast statistikast ja läbitakse mudeli loomise üldise protsessi samanimeline samm (Joonis 1). Uuritava tunnuse kirjeldav statistika annab võimaluse tutvuda võimalike trendide ja mõjuritega.

Kaoelektri uurimiseks on autori kasutuses kaoelektri tunnipõhise detailsusega andmed perioodist 01.2020–10.2021 (perioodis sisaldub 27 riigipüha ja riiklikult lühendatud tööpäeva). Kaoelektri kohta on vaatluse all kokku 16080 rida. Vaatlusaluse perioodi keskmine kaoelektri tunnikogus on 33,75 MWh ja standardhälve on 5,72 MWh, seega ei

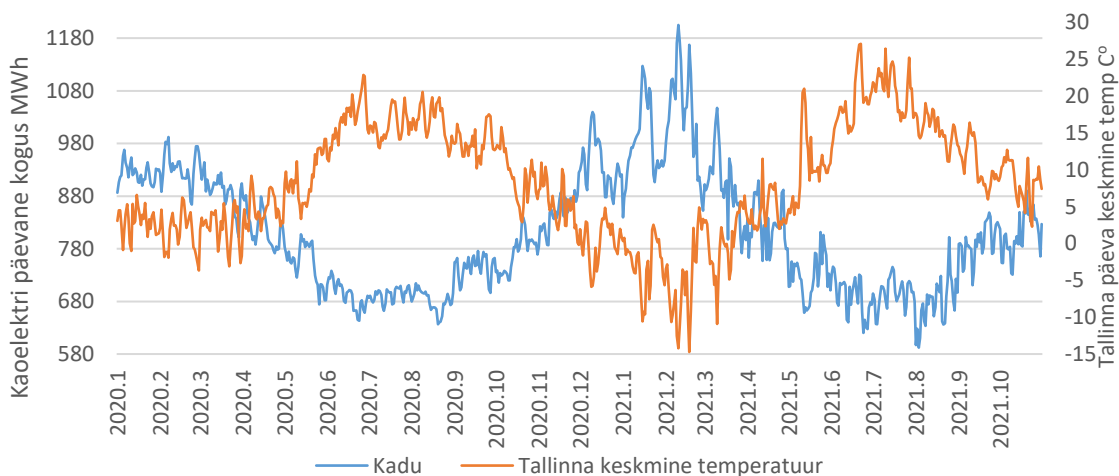
ole kõikumine ümber keskväärtuse ülemäära suur. Andmestiku minimaalne koelektri tunnikogus on 20,15 MWh (15.06.2020 kell 04:00 UTC+2) ja maksimaalne tunnikogus on 56,00 MWh (10.02.2021 kell 19:00 UTC+2).

Vaadates Joonised 2 ja 3 näeme, et koelektri andmetes esineb arvestatav perioodiline kõikumine, mis on üsnagi ootuspärase mustriga näidates külmematel kuudel suuremaid koelektri koguseid kui soojematel kuudel. Jooniselt 3 näeme, et ka kuusiseselt esineb suuri kõikuvusi koelektri päevastes kogustes, seda eriti perioodil 12.2020–03.2021.

Kuna koelektri kogus kujuneb kahe suuruse vahena, siis võivad suured kõikumised ja ebatavalised kogused olla kas ühe või mõlema suurusega seotud eripärast, olgu see siis viga andmetes või suurusi mõjutavate tegurite toimest.

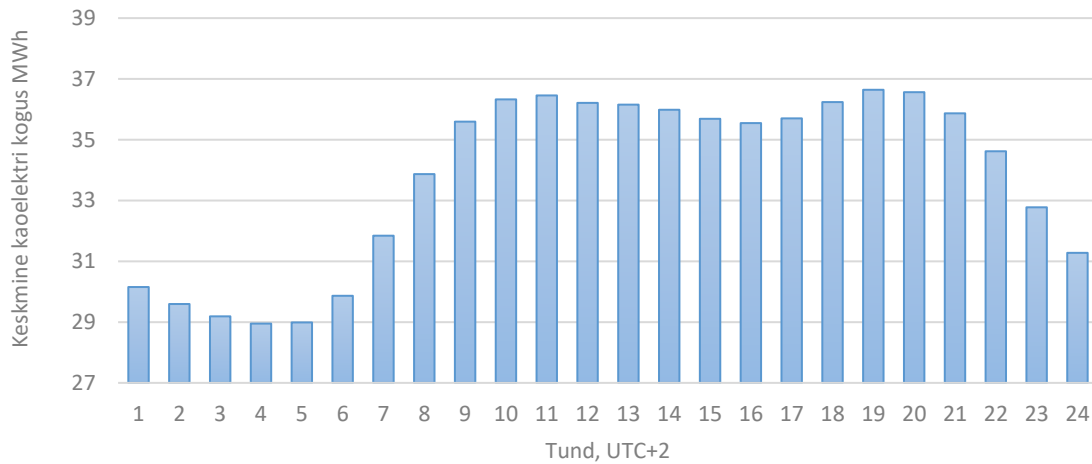


Joonis 2. Koelektri kuus kogused (MWh, sinised tulbad) 01.2020–10.2021 koos Tallinna kuu keskmise temperatuuriga (C°, oranž joon).



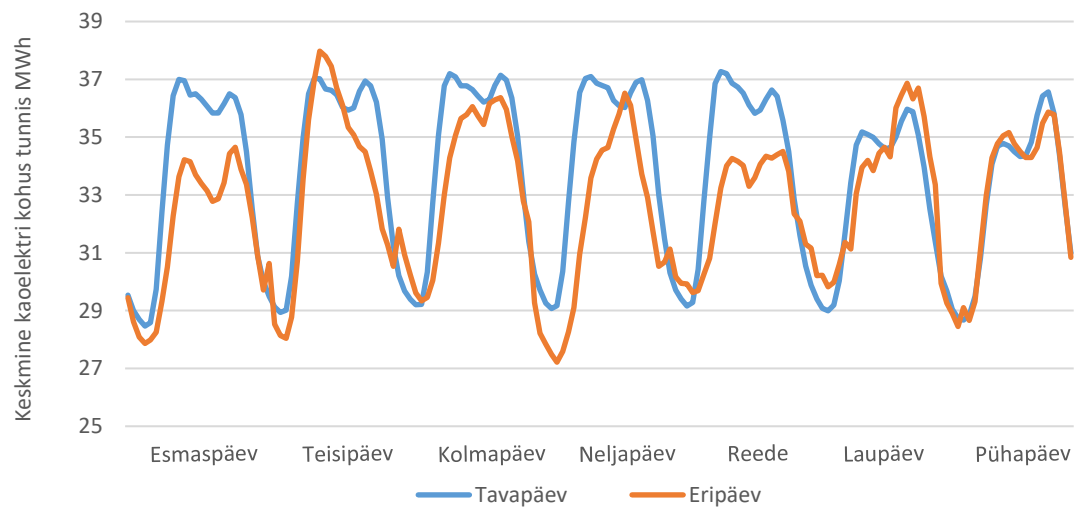
Joonis 3. Koelektri päevane kogus (MWh, sinine joon) ja Tallinna päevane keskmine temperatuur (C°, oranž joon) perioodis 01.2020–10.2021.

Jooniselt 4 näeme, et ka koelektri kogused kõiguvad suures ulatuses ka ühe ööpäeva siseselt, olles madalaimad vahemikus ajavahemikel 01:00–06:00, mil inimesed on oma tegevustes üldjuhul kõige passiivsemad ja saavutab suurimad väärtused lõunastel (10:00–11:00) ja õhtustel tundidel (19:00–20:00), mis langevad kokku aegadeka, mil inimesed on kõige aktiivsemad (tööle minek, koju tulek).



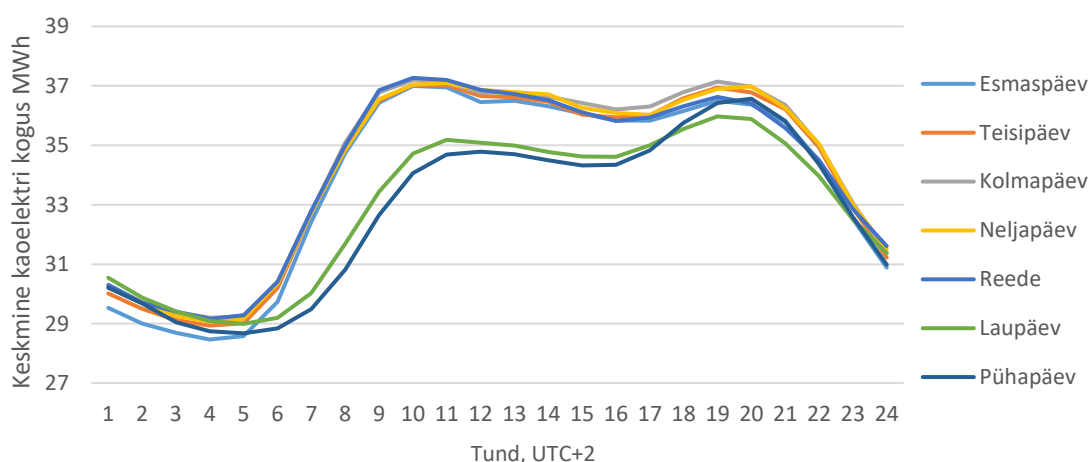
Joonis 4. Keskmine tunnipõhine koelektri kogus (MWh). Eripäevad on andmetest eemaldatud.

Jooniselt 5 näeme, et riigipühadel ja nende eelnevatel päevadel ehk eripäevadel on selgelt erinev koelektri päevasisene jaotus võrreldes tavapäevaga, kui tegemist on tööpäevaga. Tööpäevale langeva eripäeva graafik ei sarnane ka nädalavahetuse tavapäeva graafikule. Nädalavahetuse päevade puhul märkimisväärset erinevust tava- ja eripäeva graafikute vahel visuaalselt ei esine.



Joonis 5. Keskmine koelektri kogus tunnis (MWh) erinevate nädalapäevade lõikes koos tava- ja eripäeva eristusega.

Joonisel 6 näeme, et koelektri tarbimisel on selge erinevus tööpäeva ja nädalavahetuse päevasisestes graafikutes. Joonisel 6 jääb silma nädalavahetuse graafikute alane ala, mis on märkimisväärselt madalam kui tööpäeva graafikute alane ala ning tuleb ka märkida, et tööpäeva ja nädalavahetuse graafikute kujud ei ole paralleelsed ehk tegemist ei ole vaid lihtsalt madalama koelektri kogusega vaid erinevusega üldises dünaamikas. Tööpäeva grupisisene varieeruvus on võrdlemisi väike ehk tööpäevade keskmised graafikud ei erine omavahel visuaalselt märkimisväärselt. Puhkepäevade puhul on graafikult näha, et laupäevane graafik on tiputundide osas laugem kui seda on pühapäevane graafik.



Joonis 6. Keskmine tunnipõhine koelektri kogus (MWh) nädalapäeviti. Eripäevad ei ole kaasatud.

2.3 Hetkel kasutatava metoodika kirjeldus ja mudeli täpsus

Hetkel on igapäevane koelektri prognoosimine võrdlemisi keerukas ja detailirohke protsess. Eelnevalt nimetatud koosneb kahest suuremast alamprotsessist, milleks on koelektri kuu mudeli koostamine ja igapäevase tunnipõhise koelektri prognoosi tegemine.

Koelektri kuu mudeli koostamisel kasutatakse ajaloolisi andmeid, mille põhjal leitakse:

- Kalendripäeva võrgu siseneva energia tunnipõhine jaotus. Eristatakse viite erinevat gruppi: esmaspäev, nädala keskpäev (TKN), reede, laupäev ja pühapäev. Päevasised jaotused põhinevad prognoositava perioodi eelmise aasta kolme vaatlusaluse nädala keskmisel.

- Võrku siseneva energia päevakõvera lähendus kaoelektri kõverale, kasutades hooajalist koefitsienti.
- Kasutades Microsoft Excelisse sisseehitatud mudeldamise võimalusi leitakse temperatuurist sõltuv päevapõhine võrku siseneva elektrienergia mudel tööpäevadeks, laupäevaks ja pühapäevaks (TP, L ja P) ning temperatuurist sõltuv päevapõhise kaoelektri osakaalu mudel (TP, L ja P). Matemaatilise mudeli sobivust kasutuseks hinnatakse R^2 ja graafiku kuju alusel.

Kuu mudeli koostamise protsessi väljundiks on fail, kus temperatuuri alusel prognoositakse igapäevaselt võrku sisenevat energiat ja kao osakaalu sellest. Viimane jagatakse 24 tunni vahel vastavalt ajaloolistel andmetel põhinevatele koefitsientidele.

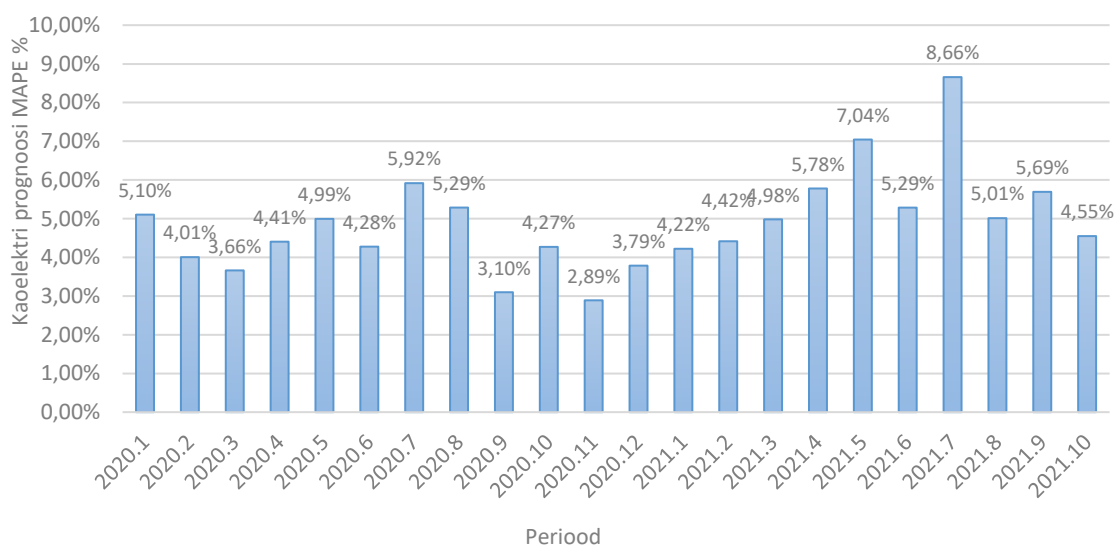
Igapäevase prognoosi tegemisel võetakse aluseks kaoelektri kuu mudel ja täiendatakse seda järgnevalt:

- Igal tööpäeval lisatakse faili eelmis(t)e päeva(de) võrku siseneva energia mille alusel korrigeeritakse tulevaste nädalate võrku siseneva energia päevasiseseid jaotusi.
- Prognoosi aluseks oleva temperatuuri määramiseks kasutatakse erinevate ilmajaamade temperatuuri prognooside andmed, kusjuures korrigeeritakse andmeid vähemalt järgneva 5 päeva kohta. Kasutatakse EMHI, Accuweather, Gismeteo ja yr.no prognoose ja andmeid uuendatakse igal tööpäeval.
- Lõpliku prognoosi tegemisel mängib olulist rolli ka eksperthinnang. Prognoosi tegev ekspert määrab nii kasutatava temperatuuri põhimõtte (kas kasutada ühe või mitme ilmteenistuse andmeid) kui ka päeva võrku siseneva koguse ja päeva kao osakaalu korrigeerimise koefitsiendi. Korrigeerimise koefitsientide juures on määravaks võimalikud temperatuuri mõjutused prognoosimisele eelnevatel päevadel, mudeli ja tegeliku tarbimise erinevus eelmistel analoogpäevadel ning samuti arvestatakse muude oluliste muutustega nagu eripäevad pühade eel ja vahel.

Igapäevase prognoosi tegemise protsessi väljundiks on järgneva 3–5 päeva tunnipõhised kaoelektri kogused, mille alusel ostetakse elektrit kaoelektri katteks võrgus.

Riigipühased ja eritähapäevi käsitletakse eranditena mõlema alamprotsessi juures ja tunnipõhised väärtused põhinevad suuresti eelmiste analoogpühade andmetel, milles on määrav ka eksperthinnang.

Joonisel 7 on toodud ülaltoodud meetodika alusel tehtava prognoosi MAPE kuude lõikes perioodis 10.2020–10.2021. Nimetatud perioodi kuu keskmine MAPE on 4,89%. Siinkohal tuleb ära märkida asjaolu, et kao andmed on ajas muutuvad kuna bilansiselgituse koostamine ja seega ka kaelektri koguse arvutus toimub aruandluskuu ja sellele eelneva 12 kuu kohta. Arvestades eelnevat ja andmete saamise kuupäeva on alloleval joonisel (Joonis 7) toodud andmetest lõplikud perioodil 01.2020–11.2020.



Joonis 7. Olemasoleva kaelektri prognoosi MAPE kuude lõikes 01.2020–10.2021.

Jooniselt 7 näeme olulist MAPE tõusu alates 2021 aasta algusest. Nimetatud MAPE tõus võib olla tingitud mitmest erinevast põhjusest, milleks võivad olla näiteks veel parandamata vead tarbimisandmetes (vaatlusaluse perioodi andmed ei ole veel lõplikud) või COVID-19 alguse andmete mõju kao olemasolevale mudelile, kuna olemasolev mudel kasutab ka pikaajalisi andmeid prognoosimudeli koostamisel.

2.4 Olemasolevate meetodite ülevaade

Käesoleva töö raames vaatame kaelektrit kui võrku siseneva ja tarbijate poolt tarbitud elektrienergia vahet. Olemasolev kirjandus kaelektri prognoosimise kohta on äärmiselt piiratud, ning fookus on rohkem kao allikate kui koguse prognoosimisel. Kaelektri prognoosimisega otseselt seotud varasematest artiklitest on märkimisväärsim Serbia

võrguettevõttega seotud artikkel [14], milles kasutatakse kaelektri tunnipõhiseks prognoosimiseks osaliselt mitmest regressiooni ja mis toetub oluliselt eksperthinnangule.

Järgnev prognoosimise meetodika kokkuvõte ja valik on tehtud elektrienergia tarbimise ja koormuse prognoosimist uurivate artiklite põhjal, kuna need andmed sarnanevad kõige enam kaelektrile. Elektrienergia tarbimise ja koormuse vaatlus ja prognoosimine jaotatakse valdavalt kolmeks ajahorisondiks, millest käesoleva töö raames keskendutakse lühiajalisele (kuni üks nädal) ja keskmisele (kuni üks aasta) [15].

Artiklid [12] ja [16] annavad põhjaliku ülevaate elektrienergia tarbimise prognoosimist kajastavatest varasematest artiklitest, hõlmates kokku üle 120 varasema artikli. Suures vaates võib eelnevalt viidatud artiklite põhjal jagada elektrienergia tarbimise lühiajalise prognoosime mudelid kaheks: klassikalised statistilised mudelid ja tehisintellektil (AI), sealhulgas masinõppel, põhinevad mudelid. Klassikaliste statistiliste mudelite hulka kuuluvad näiteks lineaarsed ja mittelineaarsed regressiooni mudelid, autoregressiivsed (integreeritud) libiseva keskmise mudelid, aegridade mudelid ning nende kombinatsioonid teiste mudelitega. Tehisintellektil põhinevate mudelite hulka kuuluvad sellised mudelid nagu tugivektor regressioon, mis põhineb tugivektor masina meetodile ja kunstliku närvivõrgu raamistik, mis kombineerib erinevaid masinõppe algoritme.

Artiklis [12] toodud analüüsi tulemuse alusel saavutab pikaajalise prognoosi mudel madalaima keskmise absoluutse protsentuaalse vea (MAPE) mittelineaarse regressiooniga, kuid lühiajalise elektrienergia tarbimise modelleerimisel klassikaliste ja tehisintellektil põhinevate mudelite kasutus suhteliselt võrdne.

Klassikalistest statistilistest mudelistest on elektri tarbimise lühiajalise prognoosimise jaoks kasutatud kõige enam (lineaarset) regressiooni. Lineaarne regressioon (LR) on mudel, mis eeldab põhjuslikku suhet uuritava ja seletava tunnuse vahel ja kasutatakse, kui andmetes ilmneb ajalooline trend tunnuste vahel [17]. Regressiooni kasutavad uurimused on näidanud silmapaistvaid tulemusi artiklites [17], kus prognoosimise eesmärgiks on leibkonna tarbimine ja [18], kus lineaarset regressiooni on kasutatud Itaalias elektri prognoosimiseks. Teiseks näiteks selles kategoorias on autoregressiivsed integreeritud libiseva keskmise mudelid, mida on kasutatud näiteks artiklis [19] selleks, et prognoosida elektri tarbimist võrdluses autoregressiivse närvivõrgu mudeliga.

Tehisintellekti alamdistsipliin masinõpe (ML) on tuleviku prognoose tegevate algoritmide kogumid, mis on suutelised suurtest andmemahtudest üle aja automaatselt ja autonoomselt mustreid tuvastama [17]. Elektrienergia tarbimise prognoosimisel kasutatavatest ML mudelitest on kõige enam levinud erinevad närvivõrkude mudelid, mida on edukalt kasutatud näiteks koormuse lühiajaliseks prognoosiks kahe USA elektrivõrgu näitel [20] ning Ateenas oleva Energia Regulatsioonibüroo hoone 24 tunnise elektritarbimise ennustamiseks [21], otsustusmetsa mudel, mida on näiteks edukalt kasutatud lühiajalise koormuse prognoosimisel Tuneesias, kus prognoositud järgmise 24 tunni keskmine viga ületas harva 2,3% [22], tugivektor regressioon, mida on kasutatud Türgi kuise elektritarbimise ennustamiseks, milles kasutati nii tugivektor regressiooni kui ka tehisnärvivõrkude mudeleid, millest just tugivektor regressioon kujunes täpsemaks [23] ning ka hoonete päevase elektritarbimise prognoosimiseks Jaapanis [24] ja Gaussi protsessi regressioon, mida on kasutatud näiteks hoonete elektritarbimise prognoosimisel [25].

Eduka meetodikana lühiajalise elektrienergia koguse prognoosimisel saab välja tuua ka närvivõrkude hulka kuuluvat Long short term memory (LSTM) meetodit. LSTM meetodit on rakendatud näiteks artiklis [26] Itaalia lühiajalise elektritarbimise prognoosimiseks, artiklis [27] Jaapanis oleva tootmishoone tarbimise lühiajaliseks prognoosimiseks ja artiklis [28] Indias tavahoone elektritarbimise prognoosimiseks.

Ühe huvitava ja eduka hübriidlahendusena võib välja tuua artiklis [29] väljapakutud libiseva akna regressiooni (SWE, *Sliding window regression*), mis võimaldab kasutada vabalt valitud regressiooni algoritmi koos leitud optimaalse treeningandmete aknaga. Nimetatud allikas prognoositi New Yorgi viie erineva piirkonna lühiajalist elektrienergia tarbimist mitme algoritmi abil, millest parimaks osutus libiseva akna regressioon, kus regressiooni algoritmiks oli tugivektor regressioon, kusjuures parim MAPE oli igas piirkonnas pea kaks korda väiksem võrreldes teiste rakendatud mudelitega (nt. LR, otsustusmets).

Olemasolev kirjandus annab hea ülevaate ka elektri ja koormuse prognoosimise mudelite võrdlusest koos nende tugevuste ja nõrkustega. Siinkohal saab välja tuua artiklid [12], [17], [19] ja [30]. Artiklis [19] võrreldakse autoregressiivse integreeritud libiseva keskmise ja mittelineaarse autoregressiivse närvivõrgu mudeleid, ning näidatakse, et kuigi mõlema mudeli tulemused on head, osutub autoregressiivse integreeritud libiseva

keskmise mudel siiski mittelineaarse autoregressiivse närvivõrgu mudelist paremaks nii prognoosi vea kui ka mudeli rakendatavuse lihtsuse poolest. Artiklis [30] võrreldakse omavahel kunstlike närvivõrke ja MOGUL metodoloogial põhinevat hägusa loogika süsteemi hägusaks reeglite õppimiseks tunnipõhise tarbimis prognoosimiseks, mille tulemus näitas hägusal loogikal põhineva mudeli ülimuslikkust kunstliku närvivõrgu süsteemi ees. Artikkel [17] toob välja, et kasutatud andmestiku puhul, milleks oli väikemajapidamine, andis parima ennustustulemuse LR ja tugivektor regressiooni mudelid, kuid see ei tähenda nende mudelite ülimuslikkust teiste mudelite ees. Artikkel [12] annab väga põhjaliku ülevaate aastani 2019 teostatud elektrienergia tarbimise prognoosimisega seotud töödest ja meetoditest, ning näitab, et lühiajalise ennustushorisoni puhul annavad klassikalised statistilised meetodid ja AI-l põhinevad meetodid ennustustäpsuse puhul võrreldavaid tulemusi.

Kokkuvõtvalt saab väita, et elektri koguste prognoosimisel on edukalt kasutatud väga erinevaid meetodeid kuid tänaseni ei ole kujunenud ühte kindlat meetodit, mis oleks teistega võrreldes universaalselt paremini rakendatav või mis annaks oluliselt parema ennustustäpsusega tulemusi.

Võttes arvesse lõputöö eesmärke, kasutatavaid andmeid ja varasemast kirjandusest selgunud asjaolu, et n.ö lihtsamad ja keerukamad mudelid annavad ennustustäpsuse osas võrreldavaid tulemusi ning võimalikku soovi töö tulemusi ka hilisemalt praktikas rakendada on autor otsustanud kasutada koelektri prognoosimiseks lineaarset regressiooni libiseval treeningandmete aknal.

Lineaarne regressioon libiseval andmete aknal on meetodina valitud kuna kasutada olevad andmed koos nende võimalike teisendustega ja taustainfo toetavad antud meetodika rakendamist. Lisaks saab siinkohal välja tuua, et sama meetodikat on juba varasemalt edukalt Eesti andmete peal kasutatud elektrikoguste prognoosiks allikates [9] ja [31], kus sarnaselt käesolevale tööle uuriti elektrikoguste ja ilmastikunäitajate vahelisi seoseid käesolevas töös kasutatava detailsusega. Ka olemasolevas koelektri prognoosimise protsessis rakendatakse osaliselt nimetatud meetodit, kuid seda kaudselt koelektri koguse leidmiseks ehk koelektri kogus ei ole uuritav tunnus, vaid selleks on võrku sisenev elektrienergia kogus ja koelektri päevane osakaal sellest.

3 Uus kaoelektri prognoosimise mudel

Sissejuhatuses toodud lõputöö eesmärkide, milleks on võimalikult hea kaoelektri tunnipõhise prognoosi mudeli koostamine ja ilmastikutegurite (tuul, päike) mõju kaoelektri kogusele uurimise lahendavad mõlemad uue kaoelektri prognoosimudeli koostamine.

Eelnevalt nimetatud eesmärkide täitmiseks testitakse uue mudeli loomiseks kahte erinevat lahendust:

- Lahendust, mille aluseks võetakse allikad [9], [31] ja [29], millest esimesed kaks keskenduvad Eesti andmete põhjal elektritarbimise lühiajalisele prognoosimisele ja mõjurite valikule ja allikas [29] tutvustab täpsemalt libiseva akna regressiooni meetodikat. Sarnaselt viidatud allikatele koostatakse käesoleva töö raames kaoelektri tunnipõhiseks prognoosimiseks mudel iga tunni kohta eraldi, seega kokku koostatakse 24 eraldiseisvat mudelit kasutades lineaarset regressiooni libiseval treeningandmete aknal.
- Lahendust, mille aluseks võetakse olemasolev kaoelektri prognoosimise protsess. Kaoelektrit prognoositakse päevapõhiselt kasutades lineaarse regressiooni meetodit libiseval treeninandmete aknal ja jaotatakse hiljem 24 tunni vahel.

Järgnevates peatükkides antakse esmalt ülevaade kõikidest töös kasutatud tunnuste omavahelistest seostest kasutades vastavalt tunnuse tüübile kas korrelatsioonanalüüsi või siis klasterdamist koos ANOVA dispersioonanalüüsiga. Lisaks eelnevale uuritakse tunnuste jaotust ja tuuakse sisse teisendusi või viitaega vajavad tunnused. Peatüki teistes osas koostatakse eelnevalt leitud tunnuste alusel esialgsed mudelid kaoelektri prognoosimiseks vastavalt ülaltoodud meetodikale ning optimeeritakse esmaseid mudeleid kuni saavutatakse parim võimalik ennustustäpsus ja mudeli sobivus.

Käesoleva peatüki alampeatükkides läbitakse järjest mudeli loomise üldise protsessi sammud alates korrelatsioonanalüüsist kuni mudeli tulemuste valideerimiseni (vt Joonis 1).

3.1 Kasutatud tunnused ja seos kaoelektriga

Uue mudeli leidmisel on kaoelektri kogust mõjutavate tunnuste kaasamisel arvestatud nii hetkel kaoelektri prognoosiks kasutatavate tunnuste kui ka teaduskirjanduses oluliseks osutunud teguritega. Järgnevalt antakse ülevaade töös kasutatavatest arvtunnustest ja nende korrelatsioonist nii kaoelektriga kui ka teiste tunnustega ning läbitakse mudeli koostamise protsessi korrelatsioonanalüüsi samm. Tekstis kasutatud tunnuste nimetustele vastavad kirjeldused on toodud Tabelis 1, täpsed korrelatsioonikordajad on toodud Tabelis 2.

Vahemaa Tallinna ja Tartu vahel on piisavalt väike, et nende kahe asukohaga seotud ilmastikunähtused oleks omavahel tugevalt seotud. Eelnevat tõestab tugev korrelatsioon Tallinna ja Tartu temperatuuride ($r_{Mtemp-Rtemp}=0,97$) ja keskmisest tugevam korrelatsioon ($r_{Mtuul-Rtuul}=0,61$) nimetatud linnade tuulekiiruste vahel. Selleks, et vältida nendest tunnustest tulenevat multikollineaarsusega seotud probleemi prognoosimudeli koostamisel on nende tunnuste jaoks prognoosimisel kasutatud nende kaalutud keskmisi. Kaalude aluseks on võetud vaatlusaluse perioodi ELV võrgu elektritarbimise geograafiline jaotus Põhja- ja Lõuna-Eestis, mille tulemusena kujunes Tallinna temperatuuri ja tuulekiiruse kaaluks $W_{Tallinn}=0,69$ ja Tartu temperatuuri ja tuulekiiruse kaaluks vastavalt $W_{Tartu}=(1-W_{Tallinn})=0,31$.

Kaoelekter on tugevas positiivses korrelatiivses seoses võrku siseneva kogusega ($r_{Kadu-Sisenev}=0,92$), keskmisest tugevamas negatiivses seoses temperatuuriga ($r_{Kadu-Temp}=-0,67$), keskmisest tugevamas negatiivses seoses päevapikkusega ($r_{Kadu-Paevapikkus}=-0,65$) ja nõrgas positiivses korrelatiivses seoses tuule kiirusega ($r_{Kadu-Tuul}=0,25$). Korrelatsioonikordaja kaoelektri ja kiirguse ning vihma vahel jääb absoluutväärtuselt alla 0,1, mis viitab väga nõrgale seosele tunnuste vahel.

Temperatuuri korrelatsioon kaoelektriga on tugevam kui temperatuuri korrelatsioon võrku siseneva kogusega ning sama märgiga ($r_{Temp-Sisenev}=-0,51$). Andmetest ilmneb ka keskmise tugevusega positiivne korrelatiivne seos temperatuuri ja kiirguse ning päeva pikkuse ja kiirguse vahel ($r_{Temp-Kiirgus}=0,48$, $r_{Paevapikkus-Kiirgus}=0,44$) ning keskmisest tugevam positiivne korrelatiivne seos temperatuuri ja päevapikkuse vahel ($r_{Paevapikkus-Temp}=0,71$). Võrku siseneva elektrienergia ja päeva pikkuse vahel esineb keskmise tugevusega negatiivne korrelatiivne seos ($r_{Sisenev-Paevapikkus}=-0,55$).

Sademetede andmed korreleeruvad teiste töös kasutatavate tunnustega kõige vähem, kusjuures korrelatsioonikordajad teiste tunnustega jäävad kõik absoluutväärtuselt alla 0,1, kusjuures isegi Tallinna ja Tartu sademetede andmete vaheline korrelatsioon on nõrk ($r_{Rvihm-Mvihm}=0,22$).

Tabel 2. Töös kasutatavate tunnipõhiste arvtunnuste korrelatsioonimaatriks.

| | | | | | | | | | | | | |
|-------------|----------|-------------|-------|-------|-------|-------|-------|---------|-------|-------|-------|-------|
| Mkiirgus | 1,00 | 0,44 | 0,46 | 0,48 | 0,48 | -0,05 | -0,04 | 0,05 | -0,09 | 0,06 | 0,16 | 0,14 |
| Paevapikkus | 0,44 | 1,00 | 0,73 | 0,69 | 0,71 | 0,03 | 0,00 | -0,55 | -0,65 | -0,18 | -0,19 | -0,21 |
| Mtemp | 0,46 | 0,73 | 1,00 | 0,97 | 0,99 | 0,03 | 0,05 | -0,52 | -0,67 | -0,10 | -0,09 | -0,10 |
| Rtemp | 0,48 | 0,69 | 0,97 | 1,00 | 1,00 | 0,04 | 0,03 | -0,49 | -0,65 | -0,08 | -0,03 | -0,05 |
| Temp | 0,48 | 0,71 | 0,99 | 1,00 | 1,00 | 0,04 | 0,03 | -0,51 | -0,67 | -0,09 | -0,05 | -0,07 |
| Mvihm | -0,05 | 0,03 | 0,03 | 0,04 | 0,04 | 1,00 | 0,22 | 0,00 | -0,01 | 0,04 | 0,04 | 0,04 |
| Rvihm | -0,04 | 0,00 | 0,05 | 0,03 | 0,03 | 0,22 | 1,00 | 0,02 | 0,01 | 0,12 | 0,06 | 0,08 |
| Sisenev | 0,05 | -0,55 | -0,52 | -0,49 | -0,51 | 0,00 | 0,02 | 1,00 | 0,92 | 0,19 | 0,26 | 0,26 |
| Kadu | -0,09 | -0,65 | -0,67 | -0,65 | -0,67 | -0,01 | 0,01 | 0,92 | 1,00 | 0,20 | 0,24 | 0,25 |
| Mtuul | 0,06 | -0,18 | -0,10 | -0,08 | -0,09 | 0,04 | 0,12 | 0,19 | 0,20 | 1,00 | 0,61 | 0,80 |
| Rtuul | 0,16 | -0,19 | -0,09 | -0,03 | -0,05 | 0,04 | 0,06 | 0,26 | 0,24 | 0,61 | 1,00 | 0,96 |
| Tuul | 0,14 | -0,21 | -0,10 | -0,05 | -0,07 | 0,04 | 0,08 | 0,26 | 0,25 | 0,80 | 1,00 | 1,00 |
| | Mkiirgus | Paevapikkus | Mtemp | Rtemp | Temp | Mvihm | Rvihm | Sisenev | Kadu | Mtuul | Rtuul | Tuul |

Tulenevalt andmete ajalisest kättesaadavusest ja uuritava tunnuse iseloomust kaasatakse käesolevasse analüüsi ka viitega tunnused. Siinkohal tuleb esile tõsta asjaolu, et prognoositava päeva (d) kohta on prognoosi tegemise päeval ($d-1$) teada kõige värskem seis võrku siseneva elektrienergia tegeliku tunnipõhise koguse kohta prognoositava päeva suhtes üle-eelmise päeva ($d-2$) kohta, sisuliselt sama kehtib ka tegeliku temperatuuri kohta. Ajaliselt kõige värskem aruandeline teave tegeliku kaelektri kohta on prognoositava päeva suhtes kuue nädala tagune, mis on tingitud bilansiselgituse koostamise protsessist.

Arvestades eelnevat ja kao nädalasisest jaotust on seletavate tunnuste hulka lisatud kaks viitega tunnust võrku siseneva elektrienergia kohta (2 päeva ja 7 päeva), üks 2-päevase viitega tunnus keskmise tegeliku temperatuuri kohta ja üks kuuenädalase viitega tunnus

kaoelektri kohta. Tabelist 3 näeme, et kõigi nimetatud tunnuste korrelatsioon koelektriga on tugevusest üle keskmise. Nimetatud tunnused esinevad järgnevas tekstis vastavalt eelnevalt toodud loetelule järgnevalt: Sis_lag2, Sis_lag7, Temp_lag2, Kadu_lag6w.

Ajalooliste prognoositud temperatuuride vaates on kasutatud ELV andmekogusid, mis sisaldavad vaatlusaluse perioodi päevakeskmisi prognoositud temperatuure, mis on hiljem omistatud päeva kõigile tundidele. Kuigi antud lahendus ei kirjelda tunni tegelikku temperatuuri kaugeltki nii hästi kui tegelike keskmise temperatuuri kasutamine usub autor, et antud andmete kasutamine omab praktilist väärtust, kuna igapäevaselt on neid andmeid lihtsam kasutada. Eelnevat toetab ka tugev negatiivne korrelatsioon koelektri ja prognoositud temperatuuri vahel ($r_{\text{Kadu-Prog_Temp}}=-0,77$).

Töös kasutatavate arvtunnuste jaotuse uurimisel selgus, et osade tunnuste teisendamine suurendab korrelatsiooni uuritava ja seletava tunnuse vahel (vt Tabel 3, Lisa 2). Eelnevast tulenevalt kaasatakse töösse olemasolevate tunnuste alternatiivina ruutjuurt kaalutud keskmisest tuulekiirusest ja kuupjuurt nii Tallinna kui ka Tartu sademete hulgast.

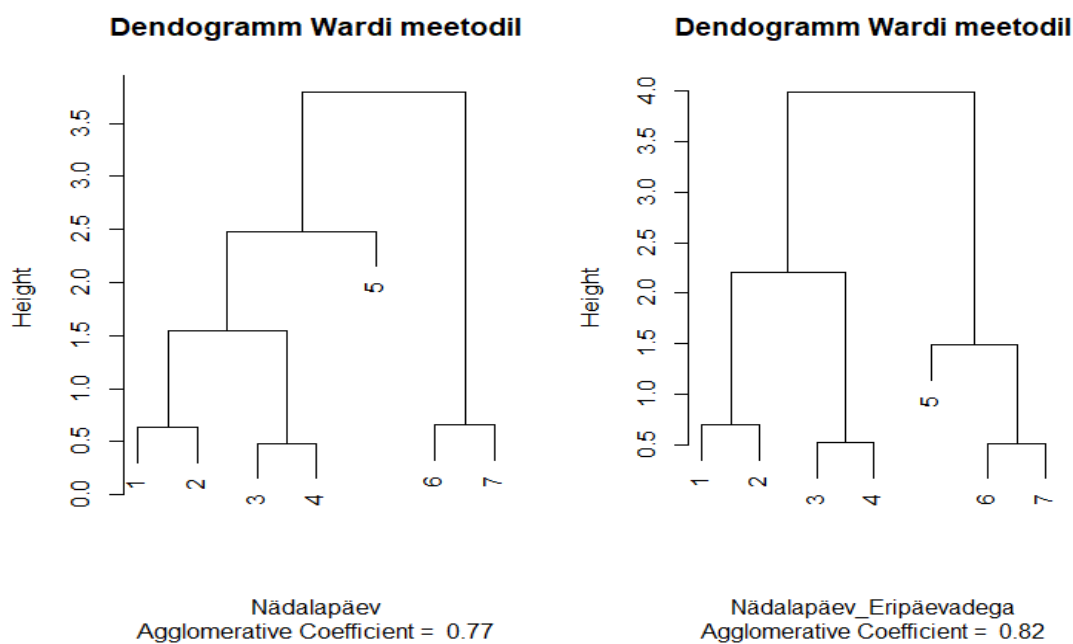
Tabel 3. Töös kasutatavate viitega ja teisendusega tunnipõhiste arvtunnuste korrelatsioonimaatriks.

| | | | | | | | | | |
|------------|-----------|-----------|---------|---------|------------|-------|----------|----------|----------|
| Temp_lag2 | 1,00 | 0,91 | -0,07 | -0,05 | -0,47 | -0,64 | -0,51 | -0,46 | -0,13 |
| Prog_temp | 0,91 | 1,00 | -0,06 | -0,04 | -0,57 | -0,77 | -0,61 | -0,59 | -0,17 |
| KJRvihm | -0,07 | -0,06 | 1,00 | 0,34 | 0,04 | 0,10 | 0,08 | 0,06 | 0,17 |
| KJMvihm | -0,05 | -0,04 | 0,34 | 1,00 | 0,03 | 0,06 | 0,05 | 0,04 | 0,12 |
| Kadu_lag6w | -0,47 | -0,57 | 0,04 | 0,03 | 1,00 | 0,71 | 0,66 | 0,75 | 0,25 |
| Kadu | -0,64 | -0,77 | 0,10 | 0,06 | 0,71 | 1,00 | 0,85 | 0,86 | 0,26 |
| Sis_lag2 | -0,51 | -0,61 | 0,08 | 0,05 | 0,66 | 0,85 | 1,00 | 0,82 | 0,28 |
| Sis_lag7 | -0,46 | -0,59 | 0,06 | 0,04 | 0,75 | 0,86 | 0,82 | 1,00 | 0,26 |
| SQRTtuul | -0,13 | -0,17 | 0,17 | 0,12 | 0,25 | 0,26 | 0,28 | 0,26 | 1,00 |
| | Temp_lag2 | Prog_temp | KJRvihm | KJMvihm | Kadu_lag6w | Kadu | Sis_lag2 | Sis_lag7 | SQRTtuul |

Järgnevalt antakse ülevaade uuritava tunnuse ja töös kaasatud kategooriliste tunnuste (nädalapäev, kuu, pilvisuse kategooria) omavahelistest seostest. Seoste kirjeldamiseks rakendati AGNES klasterdamise algoritmi Wardi meetodikaga vastavalt allikale [32] ja

klasteranalüüsi täiendavaks selgitamiseks rakendati ANOVA dispersioonanalüüsi koos Tukey testiga vastavalt allikale [33].

Joonisel 8 on toodud nädalapäevade klasteranalüüsi dendogrammid, mis on leitud kasutades AGNES (*agglomerative hierarchical clustering*) algoritmi Wardi meetodikaga 2020 aasta andmetel. Vasakpoolne dendogramm annab eeldatud klasterduse, kus ühte gruppi kogunevad tööpäevad ja teise nädalavahetuse päevad, millele viitasid ka kirjeldava analüüsi tulemused. Parempoolne dendogramm illustreerib eripäevade mõju muutes reedest päeva oluliselt sarnasemaks nädalavahetuse päevale. 2021 aasta andmetel põhinev klasteranalüüs andis samad tulemused, mis 2020 ilma eripäevadeta. Täiendav ANOVA analüüs (Lisa 4) näitab, et statistiliselt oluliselt erineva keskmise kaelektri tunnikogusega on vaid nädalavahetuse päevad tööpäevadest.



Joonis 8. Nädalapäeva klasteranalüüsi tulemus 2020 aasta andmetel. Vasakpoolsel dendogrammil on andmetest eraldatud riigipühad ja riiklikult lühendatud tööpäevad, parempoolsel dendogrammil on andmed koos nimetatud eripäevadega.

Kuude jaoks teostatud ANOVA analüüs näitab, et statistiliselt oluliselt mitte erineva keskmise kaelektri tunnikogusega on omavahel jaanuar ja detsember ning juuni ja juuli kuu.

Pilvisuse kategooria (Tallinna andmed) kujutab endast taeva hinnangulist kaetust (%) pilvedega ja on esitatud andmestikus 10% vahemikena. Antud tunnuse ja uuritava tunnuse omavahelise seose analüüs (vt Lisa 3 ja Lisa 5) näitab, et kõikidest teistest kategooriatest

eristub statistiliselt oluliselt vaid 100% pilvkattega kategooria. Statistiliselt olulised keskmised kaelektri tunnikogused on ka selge ilma ja üle 50% pilvkattega ilma puhul ning 90% ja enama pilvkatte ja all 50% pilvkattega ilma puhul.

Päevapõhiste tunnuste väärtuse leidmisel on kasutatud tunnipõhiste arvtunnuste agregeerimisel vastavalt tunnuste sisust kas tunnipõhiste väärtuste summeerimist (nt. võrku sisenev elektrienergia) või keskmistamist (nt. temperatuur). Kategoorilise tunnuse Rpilv puhul on tekitatud uus tunnus, mis sisaldab endas päeva tundide domineerivat pilvisuse kategooriat.

Allolevas tabelis (vt. Tabel 4, Lisa 6) on toodud päevapõhiste arvtunnuste korrelatsioonid uuritava tunnuse ehk kaelektriga. Tabelist näeme, et päevapõhiste arvtunnuste korrelatsioonid kaelektriga sarnanevad suurel määral tunnipõhise korrelatsioonimaatriksiga, kuid valitud tunnuste teisendused ei suurendanud korrelatsiooni kaelektriga.

Tabel 4. Päevapõhiste seletavate arvtunnuste korrelatsioon uuritava tunnusega.

| | | | | | | | | | | | |
|------|-------|-----------|-----------|-------------|----------|-------|-------|------|------------|----------|----------|
| Kadu | -0,91 | -0,89 | -0,92 | -0,79 | -0,67 | -0,05 | -0,10 | 0,25 | 0,64 | 0,85 | 0,85 |
| | temp | temp_lag2 | Prog_temp | paevapikkus | Mkiirgus | Rvihm | Mvihm | tuul | Kadu_lag6w | Sis_lag7 | Sis_lag2 |

3.2 Uute kaelektri prognoosimudelite koostamise protsess

Uute mudelite koostamiseks on töös rakendatud SLR-i meetodikat. SLR (Sliding Linear Regression) kasutab prognoosi leidmiseks liikuvat treeningandmete akent ja meetodikana lineaarset regressiooni.

Lineaarne regressioon (LR) on mudel, mis eeldab põhjuslikku suhet uuritava ja seletava(te) tunnuse vahel ja kasutatakse, kui andmetes ilmneb ajalooline trend tunnuste vahel [17]. Käesoleva töö raames rakendatakse mitmest lineaarset regressiooni, mis võimaldab uuritava tunnuse y prognoosimist mitme seletava tunnuse x abil ja avaldub lihtsustatud kujul alloleva valemi abil, kus β on mudeli parameetrid ja ε on mudeli juhuslik viga [34]:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon. \quad (7)$$

Käesoleva töö raames leitakse treeningandmed vastavalt Tabelis 5 toodule, kus prognoositava päeva d mudeli aluseks on treeningandmed perioodist $d-n$ kuni $d-2$.

Tabel 5. Treeningandmete akna leidmise skeem prognoositava päeva suhtes.

| Treeningandmed | | | | | | Pole andmeid | Prognoositav päev | Tulevik |
|----------------|-------|-------|---------|-----|-----|--------------|-------------------|-----------|
| d-n | d-n-1 | d-n-2 | d-n-... | d-3 | d-2 | d-1 | d | d+1 d+... |

Olles paika pannud treeningandmete leidmise skeemi, rakendatakse Tabelis 5 kirjas olevaid akende suurusi (n) vastavalt tunni- ja päevapõhisele mudelile. Liiga pikk treeningandmete aken toob mudelisse kaasa liiga tugeva hooajalise mõju, liiga lühike aken ei anna piisavalt vaatlusi adekvaatse mudeli loomiseks. Akende suurused põhinevad nii katse-eksituse meetodil kui ka varasemal kirjandusel (nt. [9] ja [31]). Tabelis 6 on toodud ka kokkuvõtte kõigist mudelites rakendatavatest sõltumatutest tunnustest.

Tabel 6. Uutes koelektri prognoosimudelites kasutatavad tunnused ja treeningandmete akende suurused.

| Mudel | Tunnipõhine SLR | | | Päevapõhine SLR | | | |
|-------------|----------------------|----|----|-----------------|----|----|----|
| | Tunnus/päevade arv n | 30 | 45 | 60 | 30 | 45 | 60 |
| Sis_lag2 | | x | x | x | x | x | x |
| Sis_lag7 | | x | x | x | x | x | x |
| Kadu_lag6w | | x | x | x | x | x | x |
| Mkiirgus | | x | x | x | x | x | x |
| Rpilv | | x | x | x | x | x | x |
| Rpilv_grupp | | x | x | x | x | x | x |
| Päev | | x | x | x | x | x | x |
| paevapikkus | | x | x | x | x | x | x |
| temp | | x | x | x | x | x | x |
| temp_lag2 | | x | x | x | x | x | x |
| Prog_temp | | x | x | x | x | x | x |
| tuul | | x | x | x | x | x | x |
| SQRTtuul | | x | x | x | | | |
| Mvihm | | x | x | x | x | x | x |
| Rvihm | | x | x | x | x | x | x |
| KJMvihm | | x | x | x | | | |
| KJRvihm | | x | x | x | | | |

Kuna eripäevad erinevad olulisel määral tavapäevadest on analüüsi tulemuste parandamise huvides eripäevade andmed valimist eemaldatud.

Esimesel iteratsioonil kaasatakse mudelisse kõik kasutada olevad tunnused (v.a teisendatud tunnused) ja igal järgmisel eemaldatakse kas üks tunnus, mis osutus tekitatud mudelites statistiliselt kõige vähem oluliseks või mitu tunnust korraga, kui nende tunnuste kasutamine osutus statistiliselt oluliseks vähem kui 10% loodud mudelitest. Iteratsioone korratakse kuni mudel saavutab minimaalse keskmistatud MAPE, kusjuures kõikide mudelite kaalud on võrdsed. Periood, mille alusel mudeleid tekitatakse on 01.01.2020–31.10.2021, millest pärast eripäevade eemaldamist jääb valimisse 643 päeva. Mudeli headuse näitajate arvutamisest on eemaldatud päevad, mille treeningandmete aken sisaldab vähem objekte kui treeningandmete akna suuruseks määratud n objekti.

Pilvisuse kategooria puhul on lisaks tekitatud täiendava grupeerimise tunnus, mis põhineb eelmise peatüki ANOVA analüüsi tulemustel. Tunnuste täiendav grupeerimine osutus vajalikuks pilvisuse kategooria puhul, kuna vähendas olukordade teket, kus vaadeldav treeningandmete aken ei garanteeri uuritava päeva pilvisuse kategooria olemasolu ja seega ei osutu sellised mudelid uuritava päeva jaoks rakendatavaks.

Tunnipõhise SLR mudeli puhul koostatakse SLR mudel iga tunni jaoks eraldi, seega kokku koostatakse 24 mudeli. Päevapõhise SLR mudeli puhul prognoositakse päeva summaarset kadu ning hiljem jaotatakse antud suurus 24 tunni vahel. Mudelite tähistamiseks kasutatud lühendites tähistab: M-mudel, T-tunnipõhine, P-päevapõhine, PT-tundideks jaotatud päevapõhine, n -treeningandmete aken.

Alampeatüki kokkuvõttena läbitakse mudeli koostamise üldise protsessi samm, kus toimub mudeli ja sisendandmete valik.

3.3 Tunnipõhine SLR mudel

Tunnipõhise SLR mudeli aluseks on võetud allikad [9], [31] ja [29], millest esimesed kaks keskenduvad Eesti andmete põhjal elektritarbimise lühiajalisele prognoosimisele ja mõjurite valikule ja allikas [29] tutvustab täpsemalt libiseva akna regressiooni (SWR), millel põhineb kasutatud SLR meetodika. Sarnaselt viidatud allikatele koostatakse koelektri tunnipõhiseks prognoosimiseks mudel iga tunni kohta eraldi, seega kokku

koostatakse 24 eraldiseisvat mudelit kasutades lineaarset regressiooni libiseval treeningandmete aknal vastavalt eelmises peatükis kirjeldatud protsessile.

Kokkuvõtte tunnipõhistest SLR mudelitest on toodud Tabelis 7. Näeme, et mudelite keskmine MAPE kahaneb märkimisväärselt, kui mudelist eemaldatakse kõik ilmastikuga seotud tunnused peale temperatuuri (vt mudelid MT1–MT4 vs MT5–MT8). Tabelist on näha ka seda, et madalaimad keskmised MAPE-d saavutavad mudelid, mille treeningandmete aken on n=45.

Tabel 7. Tunnipõhise SLR mudeli versioonid koos kasutatud tunnuste ja mudelite keskmise MAPE-ga erinevate treeningandmete akende juures.

| Tunnus/Mudel | MT1 | MT2 | MT3 | MT4 | MT5 | MT6 | MT7 | MT8 |
|--------------|--------|--------|-------|-------|-------|-------|-------|-------|
| Sis_lag2 | x | x | x | x | x | x | x | x |
| Sis_lag7 | x | x | x | x | x | | | x |
| Kadu_lag6w | x | | | | | | | |
| Mkiirgus | x | x | | | | | | |
| Rpilv | x | x | | | | | | |
| Rpilv_grupp | | | x | x | | | | |
| Päev | x | x | x | x | x | x | x | x |
| paevapikkus | x | x | x | x | x | x | x | x |
| temp | x | x | x | x | x | x | | |
| temp_lag2 | x | x | x | | | | | |
| Prog_temp | | | | | | | x | x |
| tuul | x | x | | | | | | |
| SQRTtuul | | | x | | | | | |
| Mvihm | x | | | | | | | |
| Rvihm | x | | | | | | | |
| KJMvihm | | x | | | | | | |
| KJRvihm | | x | | | | | | |
| MAPE n=30 | 38,22% | 29,17% | 6,39% | 6,10% | 3,67% | 3,62% | 3,94% | 3,97% |
| MAPE n=45 | 26,27% | 19,20% | 4,65% | 4,58% | 3,57% | 3,58% | 3,72% | 3,72% |
| MAPE n=60 | 14,75% | 14,00% | 4,36% | 4,29% | 3,66% | 3,67% | 3,80% | 3,79% |

Arvestades, et tuule ja päikesega seotud tegurid olid suures osas mudelitest statistiliselt ebaolulised, siis võib teha järelduse, et tunnipõhiselt puudub seos nimetatud tegurite ja kaoelektri koguste vahel.

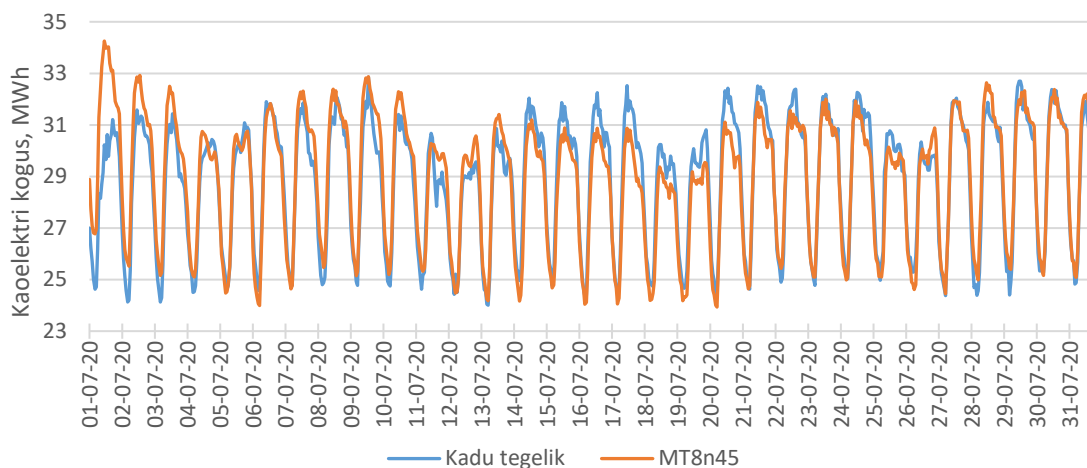
Võrreldes mudeleid MT5–MT8 näeme, et kuigi tegelikku keskmist temperatuuri kasutavad mudelid annavad parema MAPE, siis erinevus tegeliku ja prognoositava temperatuuriga mudelite ennustustäpsuse vahe ei ole märkimisväärselt suur, jäädes alla 0,3% kui võrrelda parimat tegelikku temperatuuri ja sama prognoositud temperatuuri kasutavat mudelit. Lisaks on prognoositava temperatuuri kasutamine päriselus oluliselt suurema tõenäosusega realiseeritav kui tunnipõhiste tegelike andmete lähedaste andmete rakendamine.

Tabelis 8 on toodud mudelite MT5–MT8 seotud headuse näitajad, mille alusel saab loodud mudeleid omavahel võrrelda. Tabelis 8 toodud andmetest selgub, et treeningandmete akende suuruse (n) tasemel on kõik mudeli headuse näitajad sarnasel tasemel. Kõige paremate näidikutega mudel versioon koostatutest on MT5, mis kasutab kaoelektri prognoosimiseks 2- ja 7-päevase viitega võrku siseneva elektrienergia kogust, nädalapäeva, päevapikkust ja kaalutud keskmist tegelikku temperatuuri.

Tabel 8. Tunnipõhiste mudelite MT5–MT8 keskmistatud korrigeeritud R^2 , Mudeli olulisus (p), MAPE, MAE, MSE ja RMSE.

| Mudel | n | Korrigeeritud R^2 | Mudeli p | MAPE | MAE | MSE | RMSE |
|--------------|----------|---------------------------------------|-----------------|-------------|------------|------------|-------------|
| MT5 | 30 | 0,675 | 0,020 | 3,67% | 1,199 | 2,924 | 1,710 |
| MT5 | 45 | 0,693 | 0,003 | 3,57% | 1,162 | 2,637 | 1,624 |
| MT5 | 60 | 0,709 | 0,001 | 3,66% | 1,188 | 2,664 | 1,632 |
| MT6 | 30 | 0,652 | 0,024 | 3,62% | 1,181 | 2,841 | 1,685 |
| MT6 | 45 | 0,669 | 0,009 | 3,58% | 1,168 | 2,685 | 1,639 |
| MT6 | 60 | 0,687 | 0,006 | 3,67% | 1,194 | 2,700 | 1,643 |
| MT7 | 30 | 0,624 | 0,027 | 3,94% | 1,299 | 3,247 | 1,802 |
| MT7 | 45 | 0,647 | 0,007 | 3,72% | 1,222 | 2,874 | 1,695 |
| MT7 | 60 | 0,671 | 0,004 | 3,80% | 1,243 | 2,871 | 1,694 |
| MT8 | 30 | 0,646 | 0,022 | 3,97% | 1,314 | 3,286 | 1,813 |
| MT8 | 45 | 0,671 | 0,003 | 3,72% | 1,220 | 2,836 | 1,684 |
| MT8 | 60 | 0,691 | 0,001 | 3,79% | 1,240 | 2,846 | 1,687 |

Eelneva analüüsi kokkuvõttes on hinnatud käesoleva töö raames kõige paremaks ja rakendatavamaks mudeliks aga MT8, kus $n=45$ (edaspidi MT8_{n45}). Mudeli MT8_{n45} põhjal prognoositud väärtuste näide koos kaelektri tegelike kõveratega juhuslikult valitud perioodi 07.2020 kohta on toodud Joonisel 9.



Joonis 9. Tegelik ja MT8_{n45} mudeli prognoositud kaelektri kõverad perioodis 07.2020.

Edasises analüüsis on kasutatud mudeli MT8_{n45} põhjal prognoositud väärtusi.

3.4 Päevapõhine SLR mudel

Päevapõhine SLR mudel on lahendus, mille aluseks võetakse olemasolev kaelektri prognoosimise protsess, kus prognoositavad kogused on päevapõhised ja nimetatud päevapõhine kogus jaotatakse valitud reeglite alusel 24 tunni vahel. Lõputöö raames kasutatakse päevapõhise kaelektri koguse leidmiseks SLR mudelit.

Kokkuvõtte päevapõhistest SLR mudelitest on toodud Tabelis 9, millest näeme, et parimad tulemused MAPE vaates annavad mudelid MP5–MP9, mis kaasavad võrku sisenevat elektrienergiat, nädalapäeva, päevapikkust, temperatuuri ja tuulekiirust. Tuulekiirus on statistiliselt oluline muutuja ligi kolmandikus päevapõhise SLR mudeli versiooni MP5 väljunditest, mis viitab päevapõhisele seosele keskmise tuulekiiruse ja kao summaarse koguse vahel.

Tabelist 9 näeme ka seda, mudelite puhul, mis sisaldasid pilvisuse kategooria tunnust ei anna mudel väljundit, kui $n=30$, kuna treeningandmete aken ei sisalda kõiki pilvisuse kategooriaid. Tabelist 9 tuleb esile ka asjaolu, et madalaimad MAPE-d saavutavad mudelid, mille treeningandmete aken on $n=45$.

Sarnaselt tunnipõhise SLR mudeliga näeme võrreldes mudeleid MP5–MP9, et kuigi tegelikku keskmist temperatuuri kasutavad mudelid annavad parema MAPE, siis erinevus tegeliku ja prognoositava temperatuuriga mudelite ennustustäpsuse vahel ei ole märkimisväärselt suur, jäädes alla 0,2% kui võrrelda parimat tegelikku temperatuuri ja sama prognoositud temperatuuriga mudelit.

Tabel 9. Päevapõhise SLR mudeli versioonid koos kasutatud tunnuste ja mudelite tervikvalimi MAPE-ga erinevate treeningandmete akende juures.

| Tunnus/ Mudel | MP1 | MP2 | MP3 | MP4 | MP5 | MP6 | MP7 | MP8 | MP9 |
|--------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Sis_lag2 | x | x | x | x | x | x | x | x | x |
| Sis_lag7 | x | x | x | x | x | x | | x | |
| Kadu_lag6w | x | x | | | | | | | |
| Mkiirus | x | | | | | | | | |
| Rpilv | x | | | | | | | | |
| Rpilv_grupp | | x | x | x | | | | | |
| Päev | x | x | x | x | x | x | x | x | x |
| paevapikkus | x | x | x | x | x | x | x | x | x |
| temp | x | x | x | x | x | x | x | | |
| temp_lag2 | x | x | x | | | | | | |
| Prog_temp | | | | | | | | x | x |
| tuul | x | x | x | x | x | | | | |
| Mvihm | x | | | | | | | | |
| Rvihm | x | | | | | | | | |
| MAPE n=30 | | | | | 2,89% | 2,84% | 2,88% | 3,06% | 3,09% |
| MAPE n=45 | 5,29% | 4,57% | 4,48% | 4,45% | 2,74% | 2,77% | 2,89% | 2,85% | 2,94% |
| MAPE n=60 | 4,50% | 4,10% | 4,09% | 4,07% | 2,98% | 2,95% | 3,01% | 3,03% | 3,07% |

Tabelis 10 on toodud mudelite MP5–MP9 keskmistatud headuse mõõdikud: korrigeeritud R^2 , mudeli statistiline olulisus, MAPE, MAE, MSE ja RMSE. Kõigi mõõdikute (v.a korrigeeritud R^2) puhul loetakse paremaks tulemust, mis on kõige väiksem. Allolevast tabelist (Tabel 10) näeme, et parimate näidikutega on mudel MP5, mille treeningandmete aken on n=45. Tabelist 10 on näha ka seda, et mudelite headuse näitajad ei ole suurusjärgult omavahel märkimisväärselt erinevad.

Tabel 10. Päevapõhiste mudelite MP5–MP9 keskmistatud korrigeeritud R^2 , Mudeli olulisus (p), MAPE, MAE, MSE ja RMSE.

| Mudel | n | Korrigeeritud R^2 | Mudeli p | MAPE | MAE | MSE | RMSE |
|-------|----|---------------------|----------|-------|--------|------|--------|
| MP5 | 30 | 0,788 | 0,002 | 2,89% | 22,389 | 1002 | 31,647 |
| MP5 | 45 | 0,787 | 0,000 | 2,74% | 21,252 | 847 | 29,100 |
| MP5 | 60 | 0,786 | 0,000 | 2,98% | 23,034 | 919 | 30,314 |
| MP6 | 30 | 0,775 | 0,003 | 2,84% | 21,989 | 993 | 31,513 |
| MP6 | 45 | 0,774 | 0,001 | 2,77% | 21,518 | 869 | 29,476 |
| MP6 | 60 | 0,775 | 0,000 | 2,95% | 22,872 | 903 | 30,056 |
| MP7 | 30 | 0,747 | 0,004 | 2,88% | 22,335 | 959 | 30,968 |
| MP7 | 45 | 0,741 | 0,002 | 2,89% | 22,437 | 926 | 30,437 |
| MP7 | 60 | 0,748 | 0,000 | 3,01% | 23,312 | 954 | 30,883 |
| MP8 | 30 | 0,762 | 0,003 | 3,06% | 24,175 | 1133 | 33,655 |
| MP8 | 45 | 0,766 | 0,001 | 2,85% | 22,323 | 901 | 30,024 |
| MP8 | 60 | 0,771 | 0,000 | 3,03% | 23,683 | 945 | 30,747 |
| MP9 | 30 | 0,734 | 0,004 | 3,09% | 24,353 | 1086 | 32,959 |
| MP9 | 45 | 0,736 | 0,001 | 2,94% | 23,035 | 959 | 30,968 |
| MP9 | 60 | 0,746 | 0,000 | 3,07% | 23,927 | 994 | 31,532 |

Kuigi kõige madalam keskmine MAPE on mudelil MP5, mis kasutab sõltumatute muutujatena teiste hulgas nii tunnipõhist tegelikku kaalutud keskmist temperatuuri kui päeva keskmist tuule kiirust, siis on eelnevalt nimetatud tunnuste igapäevane täpne prognoosimine ebatõenäoline ja rakendatavuse huvides on järgnevates peatükkides toodud analüüsis kasutatud mudelit MP8 treeningandmete aknaga $n=45$ (edaspidi $MP8_{n45}$).

Tabelis 11 on toodud päevapõhise SLR mudeli väljundi jaotusreeglitiku variandid koos nende tulemusel tekkinud tunnipõhise prognoosi MAPE-ga. Mudelite MPT4 ja MPT5 jaotuse skeemi täiendamine nädalataguse võrku siseneva elektrienergia jaotusega (kaal vahemikus 0,05–0,10) vähendab mudelite saavutatud MAPE-t kuni 0,03%.

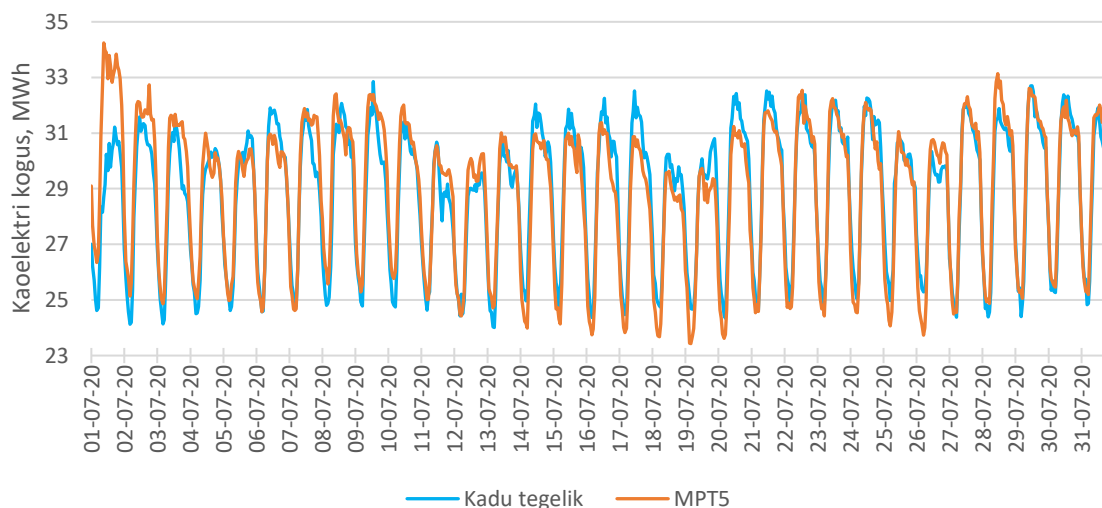
Madalaima MAPE saavutab skeem, mille puhul kasutatakse tunnipõhise koelektri jaotuse leidmiseks tunnipõhise SLR mudeli $MT8_{n45}$ jaotuskõverat uuritava päeval. MAPE suhtes paremuselt teise tulemuse annab aga skeem, kus koelektri jaotus põhineb koelektri tegelikel kogustel aasta tagasi samal nädalapäeval. Kuna MPT4 mudeli

kasutamiseks võib kaasneda mudeli vigade kuhjumine olukorras, kus MT_{8n45} ja MP_{8n45} mudeli sisendites esineb suuri kõrvalekaldeid tegelikest lõplikest väärtustest ja erinevus MPT4 ja MPT5 MAPE-de vahel on vaid 0,2%, siis on edasises analüüsis eelistatud mudelit MPT5.

Tabel 11. Päevapõhise SLR mudeli MP_{8n45} väljundi jaotusreeglistiku variandid koos nende tulemusel tekkinud tunnipõhise prognoosi MAPE-ga.

| Mudel | Jaotuse skeem | MAPE |
|-------|---|-------|
| MPT1 | Eelmise nädala sama nädalapäeva võrku siseneva elektrienergia jaotus | 7,15% |
| MPT2 | Kaoelektri kõver 6 nädalat tagasi uuritavast päevast samal nädalapäeval | 4,65% |
| MPT3 | Kaalutud keskmine 6 nädala tagune kao ($w=0,85$) ja nädalatagune võrku siseneva elektrienergia ($w=0,15$) kõver | 4,47% |
| MPT4 | MT_{8n45} mudeli jaotuskõver uuritaval päeval | 3,75% |
| MPT5 | Kaoelektri kõver samal nädalapäeval aasta tagasi | 3,94% |

Joonisel 10 on toodud näide mudeli MPT5 prognoositud kaoelektri kogustest võrdluses kaoelektri tegeliku kõveraga juhuslikult valitud perioodist 07.2020.



Joonis 10. Tegelik ja MPT5 mudeli prognoositud kaoelektri kõverad perioodis 07.2020.

Edasises analüüsis on kasutatud mudeli MPT5 põhjal prognoositud kaoelektri väärtusi.

Nii tunnipõhise kui ka päevapõhise SLR mudeli peatükkide koostamisel läbitakse mudeli loomise üldise protsessi sammud, milles toimub mudeli rakendamine ja sisendite optimeerimine ning mudeli tulemuste valideerimine.

4 Tulemuste analüüs

Töö sissejuhatuses toodud eesmärgid, millele selles peatükis vastused antakse on järgnevad:

- Võrrelda, kas lõputöö raames koostatud kao prognoosi mudel on täpsem kui olemasolev kao prognoosi meetodika.
- Hinnata milline on pakutud mudeli kasutamise finantsmõju.

Seoses mudeli algandmete ja mudeldamise protsessiga on mudeli väljundiks andmed perioodist 16.02.2020–31.10.2021, millest on veel omakorda eemaldatud eripäevad. Käesoleva töö järeldused põhinevad ja laienevad nimetatud kirjelduse alla kuuluvatele andmetele.

Käesoleva peatüki alampeatükkides viiakse läbi olemasoleva ja pakutud kaelektri prognoosimudelite täpsuse võrdlus, antakse ülevaade sellest, kuidas pakutud mudelid tulevad toime kellakeeramisega seotud probleemidega ja hinnatakse uute kaelektri prognoosimudelite kasutamisega kaasnevat finantsmõju.

4.1 Olemasoleva ja pakutud kaelektri prognoosimudelite võrdlus

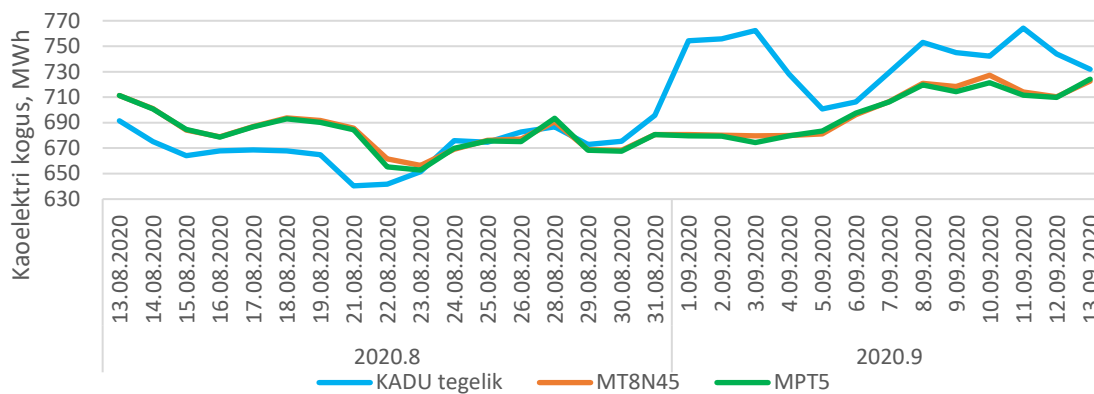
Olemasoleva ja pakutud mudelite võrdluse aluseks on kaelektri prognoosimudelite MAPE-d. Mudelite võrdlemise näidikuks on valitud MAPE, kuna see on ainus, mis ei ole kallutatud mudeli disaini poolt, viies võrreldavad suurused ühele alusele.

Tabelis 12 on toodud ELV prognoosi ja mudelite $MT8_{n45}$ ning MPT5 prognoosi keskmine MAPE kuude lõikes ja vaatlusalusel perioodil koos tegeliku kaelektri koguse ja pakutud mudelite poolt prognoositud kaelektri kuise kogusega (MWh). Tabelist näeme, et pakutud mudelid saavutavad vaatlusalusel perioodil kokku ja kuude lõikes enamikel juhtudest madalamad MAPE-d kui hetkel kasutatav mudel. Arvestades, et pakutud uued mudelid on leitud sama meetodika alusel, ei ole üllatav, et kuised kaelektri kogused jäävad samasse suurusjärku ja MAPE-d liiguvad peaaegu paralleelselt.

Tabel 12. ELV ehk hetkel kasutatava prognoosi ja mudelite MT8_{n45} ning MPT5 prognoosi keskmine MAPE (hall taust: uue mudeli MAPE ületab ELV MAPE-t) kuude lõikes koos tegeliku ja pakutud mudelite poolt prognoositud koelektri kuise kogusega (MWh) (sinine taust: kogus prognoositud kogus on väiksem kui tegelik, punane taust: prognoositud kogus ületas tegelikku). *andmed alates 16.02.2020.

| Periood | ELV MAPE | MT8 _{n45} MAPE | MPT5 MAPE | Kadu tegelik MWh | MT8 _{n45} MWh | MPT5 MWh |
|---------|-------------|----------------------------|--------------|---------------------|---------------------------|-------------|
| 2.2020* | 3,82% | 1,91% | 2,55% | 11208 | 11175 | 11135 |
| 3.2020 | 3,82% | 2,59% | 3,02% | 27599 | 27956 | 27980 |
| 4.2020 | 4,18% | 2,70% | 3,62% | 22988 | 22773 | 22747 |
| 5.2020 | 4,66% | 3,17% | 3,59% | 21867 | 21616 | 21695 |
| 6.2020 | 3,99% | 4,00% | 4,33% | 18816 | 18389 | 18394 |
| 7.2020 | 5,51% | 2,36% | 2,49% | 21468 | 21532 | 21532 |
| 8.2020 | 5,37% | 2,45% | 2,62% | 20458 | 20728 | 20737 |
| 9.2020 | 3,09% | 4,25% | 4,64% | 22366 | 21898 | 21905 |
| 10.2020 | 4,50% | 3,33% | 2,80% | 23866 | 24056 | 24085 |
| 11.2020 | 2,99% | 2,31% | 2,41% | 25474 | 25238 | 25214 |
| 12.2020 | 4,05% | 2,90% | 2,90% | 24699 | 24525 | 24545 |
| 1.2021 | 4,33% | 2,57% | 2,61% | 29599 | 29790 | 29845 |
| 2.2021 | 4,63% | 3,43% | 3,53% | 27117 | 26975 | 26951 |
| 3.2021 | 5,41% | 3,83% | 4,07% | 27831 | 27876 | 27903 |
| 4.2021 | 5,82% | 5,57% | 5,84% | 23150 | 22918 | 22905 |
| 5.2021 | 7,84% | 7,25% | 6,28% | 21081 | 20815 | 20903 |
| 6.2021 | 4,87% | 5,51% | 5,66% | 18681 | 18537 | 18546 |
| 7.2021 | 7,76% | 3,21% | 4,15% | 21577 | 21700 | 21692 |
| 8.2021 | 4,92% | 5,99% | 6,41% | 20340 | 20266 | 20290 |
| 9.2021 | 5,30% | 4,02% | 3,91% | 23791 | 23654 | 23676 |
| 10.2021 | 4,26% | 4,01% | 4,70% | 25070 | 25094 | 25099 |
| Kokku | 4,85% | 3,72% | 3,94% | 479044 | 477511 | 477779 |

SLR mudelite miinuseks on see, et lühikeste treeningandmete akende puhul ei tule nad hästi toime lühiajaliste suurte kõikumistega uuritava tunnuse ehk koelektri kõverates. Joonis 11 illustreerib SLR mudeli kehva kohanemist lühiajaliste suurte kõikumistega andmetes. Näeme, et kumbki SLR mudelid ei suuda prognoosida kooliaasta algusega (01.09.2020–04.09.2020) seotud järsku tõusu koelektri kogustes.

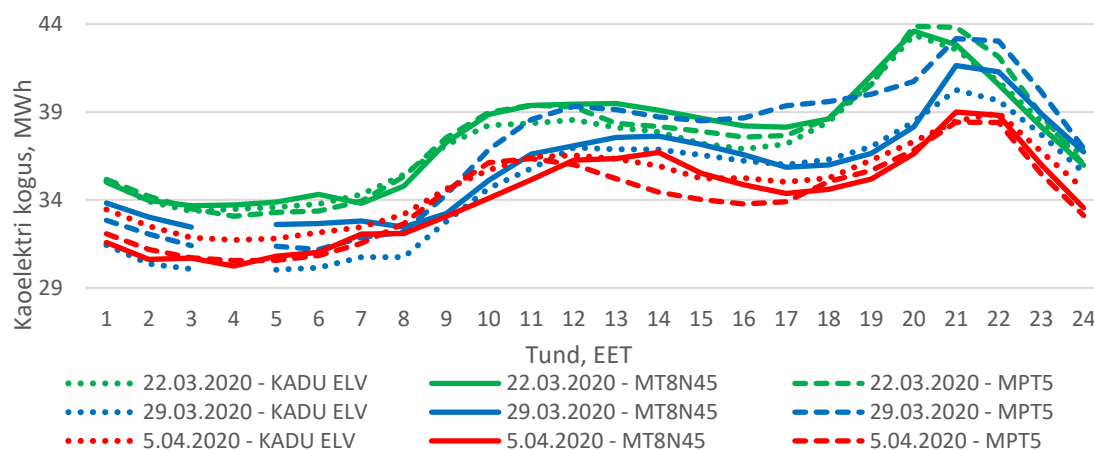


Joonis 11. Kaoelektri päevapõhised tegelikud, MT8_{n45} ja MPT5 põhjal prognoositud kogused 13.08.2020–13.09.2020.

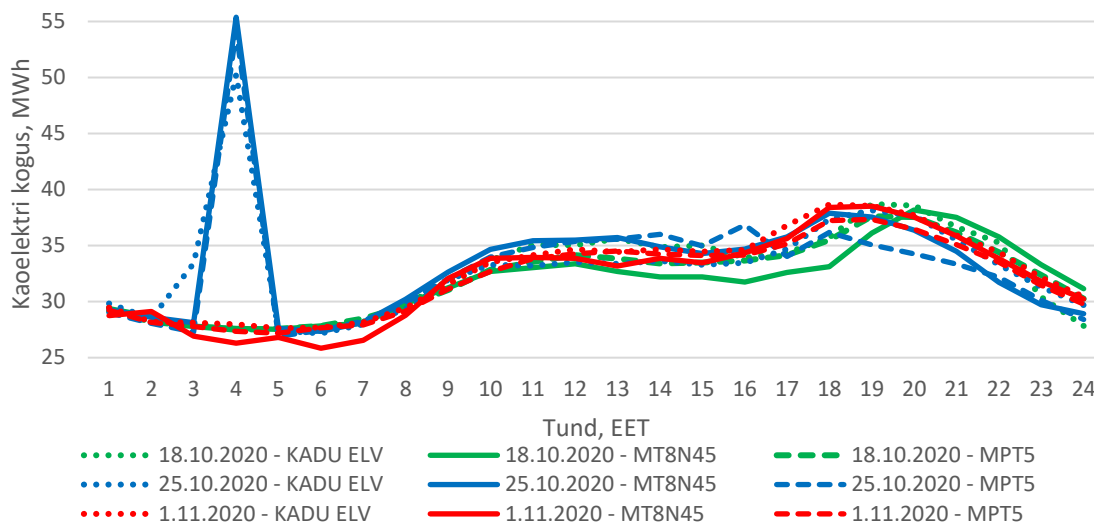
4.2 Pakutud mudelite käitumine seoses kellakeeramisega

Üheks oluliseks aspektiks tunnipõhiste andmete prognoosimisel on mudeli võimekus toime tulla kellakeeramise seotud andmete nihkega.

Joonistel 12 ja 13 on toodud 2020 aasta kellakeeramisele eelneva nädala sama päeva, kellakeeramise päeva ja kellakeeramisele järgneva nädala sama päeva tegelik, MT8_{n45} prognoositud ja MPT5 prognoositud kaoelektri kõverad. Nimetatud joonised demonstreerivad hästi seda, kuidas nii tunnipõhine SLR kui ka päevapõhisel SLR-il põhinev lahendus suudavad kohanduda kellakeeramisele kui jälgida tegeliku ja prognoositud kaoelektri kõveraid. On näha, et mõlemad prognoositud kaoelektri kõverad nihkuvad vastavalt tegelikule kaoelektri kõverale ja suuri kõrvalekaldeid tegelikkusest ei esine.



Joonis 12. 2020 aasta kevadise kellakeeramise seotud kaoelektri kõverad. Joonisel on toodud 22.03.2020, 29.03.2020 ja 5.04.2020 tegelik, MT8_{n45} prognoositud ja MPT5 prognoositud kaoelektri kõverad.



Joonis 13. 2020 aasta sügise kellakeeramisega seotud kaoelektri kõverad. Joonisel on toodud 18.10.2020, 25.10.2020 ja 1.11.2020 tegelik, MT8_{n45} prognoositud ja MPT5 prognoositud kaoelektri kõverad.

Joonisel 12 toodud näide põhineb 2020 aasta kevadisel kellakeeramisel, 2021 aasta kevadine kellakeeramise graafik näitab sama head prognoosimudelite kohanemist.

4.3 Pakutud mudelite rahaline mõju

Kaoelektri ost koosneb määratud ja avatud tarnest. Määratud tarne on ELV poolne ost otse Nord Pooli turult ja avatud tarne on seotud bilansitehingutega, mida teostab ELV eest hanke korras Eesti Energia AS [4]. Selleks, et hinnata pakutud mudelite kasutamisega seotud finantsmõju, leitakse prognoositud kaoelektri koguseid kasutades mudeli veast tingitud tulu ja kulu.

Mudeli veast tingitud tulu arvutatakse tunnis t , kui y_t ja \hat{y}_t on vastavalt kaoelektri tegelik ja prognoositud kogus, h_{NPS} , $h_{BE\ müük}$ ja $h_{BE\ ost}$ on vastavalt NPS, bilansienergia müügi ja bilansienergia ostuhind, järgneva valemi abil :

$$\text{Mudeli veast tingitud tulu}_t = y_t * h_{NPS} - (\hat{y}_t * h_{NPS} + (y_t - \hat{y}_t) * f(h)),$$

$$\text{kus } f(h) = \begin{cases} h_{BE\ müük}, & y_t - \hat{y}_t > 0 \\ h_{BE\ ost}, & y_t - \hat{y}_t < 0 \end{cases}$$

Ülaloleva valemi abil leitakse sisuliselt bilansienergia tehingutega seotud täiendav tulu/kulu tunnis, mis saab tekkida vaid juhul, kui prognoositud kaoelektri kogus tunnis ei kattu tegeliku kaoelektri kogusega samas tunnis.

Tabelis 13 on toodud ülaloleva valemi alusel arvatatud tunnipõhise SLR mudeli MT8_{n45} ja päevapõhise SLR põhineva mudeli MPT5 prognoosi veaga seotud rahaline mõju kuude lõikes. Tabeli viimases veerus on välja toodud kumb kasutatud mudelist on tulusam, mida on hinnatud kas suurema tulu ehk ka madalama kulu alusel.

Tabel 13. Tunnipõhise SLR mudeli MT8_{N45} ja päevapõhise SLR põhineva mudeli MPT5 prognoosi veaga seotud rahaline mõju kuude lõikes.* andmed alates 16.02.2020.

| Periood | MT8 _{n45} mudeli veast tingitud tulu | MPT5 mudeli veast tingitud tulu | Tulusam mudel |
|---------|---|---------------------------------|--------------------|
| 2.2020* | 1159 € | 600 € | MT8 _{n45} |
| 3.2020 | -3967 € | -5639 € | MT8 _{n45} |
| 4.2020 | -10025 € | -12169 € | MT8 _{n45} |
| 5.2020 | -9331 € | -9266 € | MPT5 |
| 6.2020 | -5339 € | -2132 € | MPT5 |
| 7.2020 | -4059 € | -3669 € | MPT5 |
| 8.2020 | -1463 € | -156 € | MPT5 |
| 9.2020 | -7596 € | -8388 € | MT8 _{n45} |
| 10.2020 | -6219 € | -5893 € | MPT5 |
| 11.2020 | -7908 € | -8353 € | MT8 _{n45} |
| 12.2020 | -13312 € | -11014 € | MPT5 |
| 1.2021 | -8759 € | -9013 € | MT8 _{n45} |
| 2.2021 | -6682 € | -6480 € | MPT5 |
| 3.2021 | 2639 € | 780 € | MT8 _{n45} |
| 4.2021 | -8203 € | -10495 € | MT8 _{n45} |
| 5.2021 | 2702 € | 6881 € | MPT5 |
| 6.2021 | 402 € | 4784 € | MPT5 |
| 7.2021 | -4257 € | -1443 € | MPT5 |
| 8.2021 | 2207 € | 7137 € | MPT5 |
| 9.2021 | -5840 € | -2327 € | MPT5 |
| 10.2021 | -9707 € | -18163 € | MT8 _{n45} |
| Kokku | -103559 € | -94420 € | MPT5 |

Tabelist 13 nähtub, et nii 2020 kui ka 2021 suvistel kuudel annab just mudel MPT5 parema rahalise tulemuse kui MT8_{n45}. Selleks, et hinnata tegeliku trendi olemasolu ja teha põhjapidavaid järeldusi oleks vaja mudelit testida pikema perioodi vältel.

5 Kokkuvõte

Töö sissejuhatuses on toodud neli eesmärki, millele käesoleva töö raames keskenduti ja käesolevas peatükis vastused antakse.

Esimeseks töös püstitatud eesmärgiks oli koostada võimalikult hea mudel, mis prognoosib kao kogust tunnipõhiselt. Nimetatud eesmärgi täitmiseks rakendati töös kasutatava andmestiku peal SLR meetodikat, ning leiti kaks võrdlemisi head lahendust kaelektri koguste tunnipõhiseks prognoosimiseks. Esimene lahendus kasutab SLR meetodikat tunnipõhistel andmetel ja iga ööpäev tunni kohta leitakse eraldi SLR mudel uuritavaks päevaks. Teine lahendus kasutab SLR meetodikat päevapõhistel andmetel ja prognoositud päeva kogus jaotatakse hiljem eelneva aasta sama nädalapäeva kaelektri kõvera alusel.

Parimaks tunnipõhiseks SLR mudeliks osutus mudel $MT8_{n45}$ (vaatlusaluse perioodi keskmine $MAPE=3,72\%$), mis kasutab kaelektri prognoosimiseks võrku siseneva elektrienergia kogust 2- ja 7-päevase viitajaga, nädalapäeva, päeva pikkust ja prognoositud päeva keskmist temperatuuri 45-päevase treeningandmete aknaga.

Parimaks päevapõhiseks SLR mudeliks osutus $MP8_{n45}$, mis kasutab sarnaselt parima tunnipõhise SLR mudeliga kaelektri prognoosimiseks võrku siseneva elektrienergia kogust 2- ja 7-päevase viitajaga, nädalapäeva, päeva pikkust ja prognoositud päeva keskmist temperatuuri 45-päevase treeningandmete aknaga. Optimaalseks päevapõhise kaelektri koguse jaotusreeglistikuks kujunes kaelektri kõver samal nädalapäeval aasta tagasi, mille tulemus kajastub käesolevas töös kui mudel $MPT5$ (vaatlusaluse perioodi keskmine $MAPE=3,94\%$).

Teiseks eesmärgiks oli võrrelda lõputöö raames koostatud mudelite täpsust olemasoleva mudeliga. Vaatlusaluste andmete põhjal saab väita, et nii tunnipõhine SLR mudel $MT8_{n45}$ kui ka päevapõhise SLR-il põhinev mudel $MPT5$ saavutavad vastavalt 1,13% ja 0,92% võrra madalamad $MAPE$ -d võrreldes hetkel kasutatava kaelektri prognoosiga.

Kolmandaks eesmärgiks oli pakutud kaelektri mudelite finantsmõju analüüs. Pakutud mudelite kasutamise seotud finantsmõju hindamiseks on kasutatud mudeli veast tulenevat tulu määra, mille arvutuse tulemused on kuvatud kuupõhiselt eelmise peatüki Tabelis 13. Lõputöö raames kasutatud andmete alusel oleks perioodis 16.02.2020–31.10.2021 (v.a eripäevad) mudeli veast tingitud täiendav kulu olnud mudeli MT8_{n45} puhul üle 9000€ võrra suurem kui mudeli MPT5 puhul.

Viimaseks eesmärgiks oli uurida kaelektri seost tuule ja päiksega, millega seoses kaasati sõltumatute tunnuste hulka nii keskmine tuulekiirus, sademete hulk, kiirguse tase kui ka pilvisuse kategooria. Tunnipõhise SLR meetodika puhul näitasid leitud mudelid, et üksi nimetatud tuule ja päikesepaistega seotud tunnus ei mängi statistiliselt ega praktiliselt olulist rolli kaelektri prognoosimisel. Päevapõhine SLR mudel näitas aga huvitaval kombel, et päeva keskmine tuulekiirus omab statistiliselt olulist rolli kaelektri koguse prognoosimisel.

Käesoleva töö laiendamiseks tuleks tulevikus uurimise alla võtta eripäevade kaelektri prognoosimise lahendus. Lisaks vajab märkimist asjaolu, et mõlemad pakutud SLR mudelid tuginevad tugevalt eeldusel, et mudeli kasutuseks olevad andmed on õigeaegselt kättesaadavad ja kvaliteetsed. Selleks, et hinnata mudeli käitumist praktikas, tuleks seda testperioodi vältel rakendada igapäevaselt paralleelselt olemasoleva kaelektri prognoosimise protsessiga.

Kasutatud kirjandus

- [1] Majandusaasta aruanded. Elektrilevi kodulehekülj [Online]. Kättesaadav: <https://www.elektrilevi.ee/et/ettevotest/elektrilevi-tutvustus?modal=aruanded> Kasutatud: 10.11.2021
- [2] Taastuvenergia. Teabelehed Euroopa Liidu kohta [Online]. Kättesaadav: <https://www.europarl.europa.eu/factsheets/et/sheet/70/taastuvenergia> Kasutatud: 06.05.2022
- [3] Energeetika. Statistikaameti koduleht [Online]. Kättesaadav: <https://www.stat.ee/et/avasta-statistikat/valdkonnad/energia-ja-transport/energeetika> Kasutatud: 06.05.2022
- [4] Elektrilevi korraldas hanke elektribörsilt kaoenergia ostuks. Elektrilevi kodulehekülj [Online]. Kättesaadav: https://www.elektrilevi.ee/uudised/avaleht/-/newsv2/news_28122012_hanked Kasutatud: 10.11.2021
- [5] Elektriturg. Energiatalgud koduleht. Kättesaadav: <https://energiatalgud.ee/Elektriturg> Kasutatud: 15.01.2022
- [6] Tartu füüsikahoone e-ilmajaama arhiiv. Tartu füüsikahoone e-ilmajaama kodulehekülj [Online]. Kättesaadav: <https://meteo.physic.ut.ee/> Kasutatud: 28.04.2022
- [7] Tallinna (lennujaama) ilmaandmete arhiiv. rp5.ru kodulehekülj [Online]. Kättesaadav: [https://rp5.ru/Weather_archive_in_Tallinn_\(airport\)](https://rp5.ru/Weather_archive_in_Tallinn_(airport)) Kasutatud: 28.04.2022
- [8] Pühade ja tähtpäevade seadus. Riigiteataja koduleht. Kättesaadav: <https://www.riigiteataja.ee/akt/109032011007?leiaKehtiv> Kasutatud: 20.02.2022
- [9] M. Sinimaa, M. Spichakova, J. Belikov, and E. Petlenkov, "Feature Engineering of Weather Data for Short-Term Energy Consumption Forecast," *2021 IEEE Madrid PowerTech*, pp. 1-6, Jun. 2021, doi: <https://doi.org/10.1109/powertech46648.2021.9494920>
- [10] KE033:Elektrijaamade toodang ja energia tootmiseks tarbitud kütus. Statistikaameti: Statistika andmebaas [Online]. Kättesaadav: https://andmed.stat.ee/et/stat/majandus__energeetika__energia-tarbimine-ja-tootmine__aastastatistika/KE033/table/tableViewLayout2 Kasutatud: 06.05.2022
- [11] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, p. e623, 2021, doi: <https://doi.org/10.7717/peerj-cs.623>
- [12] N. Wei, C. Li, X. Peng, F. Zeng, and X. Lu, "Conventional models and artificial intelligence-based models for energy consumption forecasting: A review," *Journal of Petroleum Science and Engineering*, vol. 181, p. 106187, 2019, doi: <https://doi.org/10.1016/j.petrol.2019.106187>
- [13] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B.-W. On, W. Aslam, and G. S. Choi, "COVID-19 Future Forecasting Using Supervised Machine Learning Models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020, doi: <https://doi.org/10.1109/access.2020.2997311>

- [14] J. Janjanin, A. Kurcubic, and Z. Jovanovic, "Forecast of Transmission System Losses by EMS," 2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET), 2019, doi:<https://doi.org/10.23919/smagrimet.2019.8720372>
- [15] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352–1372, 2015, doi: <https://doi.org/10.1016/j.rser.2015.04.065>
- [16] M. A. Mat Daut, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, and F. Hussin, "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review," *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 1108–1118, 2017, Doi: <https://doi.org/10.1016/j.rser.2016.12.015>
- [17] A. Gonzalez-Briones, G. Hernandez, J. M. Corchado, S. Omatu, and M. S. Mohamad, "Machine Learning Models for Electricity Consumption Forecasting: A Review," *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, 2019, doi: <https://doi.org/10.1109/cais.2019.8769508>
- [18] V. Bianco, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy*, vol. 34, no. 9, pp. 1413–1421, 2009, Doi: <https://doi.org/10.1016/j.energy.2009.06.034>
- [19] C. Nichiforov, I. Stamatescu, I. Fagarasan, and G. Stamatescu, "Energy consumption forecasting using ARIMA and neural network models," *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*, pp. 1-4, Oct. 2017, Doi: <https://doi.org/10.1109/iseee.2017.8170657>
- [20] A. Khotanzad, A. Abaye, and D. Maratukulam, "An adaptive and modular recurrent neural network based power system load forecaster," *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 2, pp. 1032-1036, 1995, doi: <https://doi.org/10.1109/icnn.1995.487563>
- [21] A. L. Katsatos and K. P. Moustris, "Application of Artificial Neuron Networks as energy consumption forecasting tool in the building of Regulatory Authority of Energy, Athens, Greece," *Energy Procedia*, vol. 157, pp. 851–861, 2019, doi: <https://doi.org/10.1016/j.egypro.2018.11.251>
- [22] A. Lahouar and J. Ben Hadj Slama, "Day-ahead load forecast using random forest and expert input selection," *Energy Conversion and Management*, vol. 103, pp. 1040–1051, Oct. 2015, doi: <https://doi.org/10.1016/j.enconman.2015.07.041>
- [23] G. Oğcu, O. F. Demirel, and S. Zaim, "Forecasting Electricity Consumption with Neural Networks and Support Vector Regression," *Procedia - Social and Behavioral Sciences*, vol. 58, pp. 1576–1585, 2012, doi: <https://doi.org/10.1016/j.sbspro.2012.09.1144>
- [24] F. Magoules, M. Piliougine, and D. Elizondo, "Support Vector Regression for Electricity Consumption Prediction in a Building in Japan," *2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES)*, pp. 189-196, 2016, doi: <https://doi.org/10.1109/cse-euc-dcabes.2016.184>
- [25] A. Zeng, H. Ho, and Y. Yu, "Prediction of building electricity usage using Gaussian Process Regression," *Journal of Building Engineering*, vol. 28, p. 101054, 2020, doi: <https://doi.org/10.1016/j.jobbe.2019.101054>
- [26] C. Tian, J. Ma, C. Zhang, and P. Zhan, "A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural

- Network,” *Energies*, vol. 11, no. 12, p. 3493, 2018, doi: <https://doi.org/10.3390/en11123493>
- [27] W. Chandramitasari, B. Kurniawan, and S. Fujimura, “Building Deep Neural Network Model for Short Term Electricity Consumption Forecasting,” *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, pp. 43-48, Aug. 2018, doi: <https://doi.org/10.1109/sain.2018.8673340>
- [28] N. Somu, G. R. M R, and K. Ramamritham, “A hybrid model for building energy consumption forecasting using long short term memory networks,” *Applied Energy*, vol. 261, p. 114131, 2020, doi: <https://doi.org/10.1016/j.apenergy.2019.114131>
- [29] I. A. Khan, A. Akber, and Y. Xu, “Sliding Window Regression based Short-Term Load Forecasting of a Multi-Area Power System,” *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pp. 1-5, 2019, doi: <https://doi.org/10.1109/ccece.2019.8861915>
- [30] A. Jozi, T. Pinto, I. Praca, Z. Vale, and J. Soares, “Day ahead electricity consumption forecasting with MOGUL learning model,” *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-6, Jul. 2018, doi: <https://doi.org/10.1109/ijcnn.2018.8489134>
- [31] M. Spichakova, J. Belikov, K. Nõu, and E. Petlenkov, “Feature Engineering for Short-Term Forecast of Energy Consumption,” *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 2019, doi: <https://doi.org/10.1109/isgteurope.2019.8905698>
- [32] F. Murtagh and P. Legendre, “Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion?,” *Journal of Classification*, vol. 31, no. 3, pp. 274–295, 2014, doi: <https://doi.org/10.1007/s00357-014-9161-z>
- [33] S. Lee and D. K. Lee, “What is the proper way to apply the multiple comparison test?,” *Korean Journal of Anesthesiology*, vol. 73, no. 6, pp. 572–572, 2020, doi: <https://doi.org/10.4097/kja.d.18.00242.e1>
- [34] D. Maulud and A. M. Abdulazeez, “A Review on Linear Regression Comprehensive in Machine Learning,” *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, 2020, doi: <https://doi.org/10.38094/jastt1457>

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

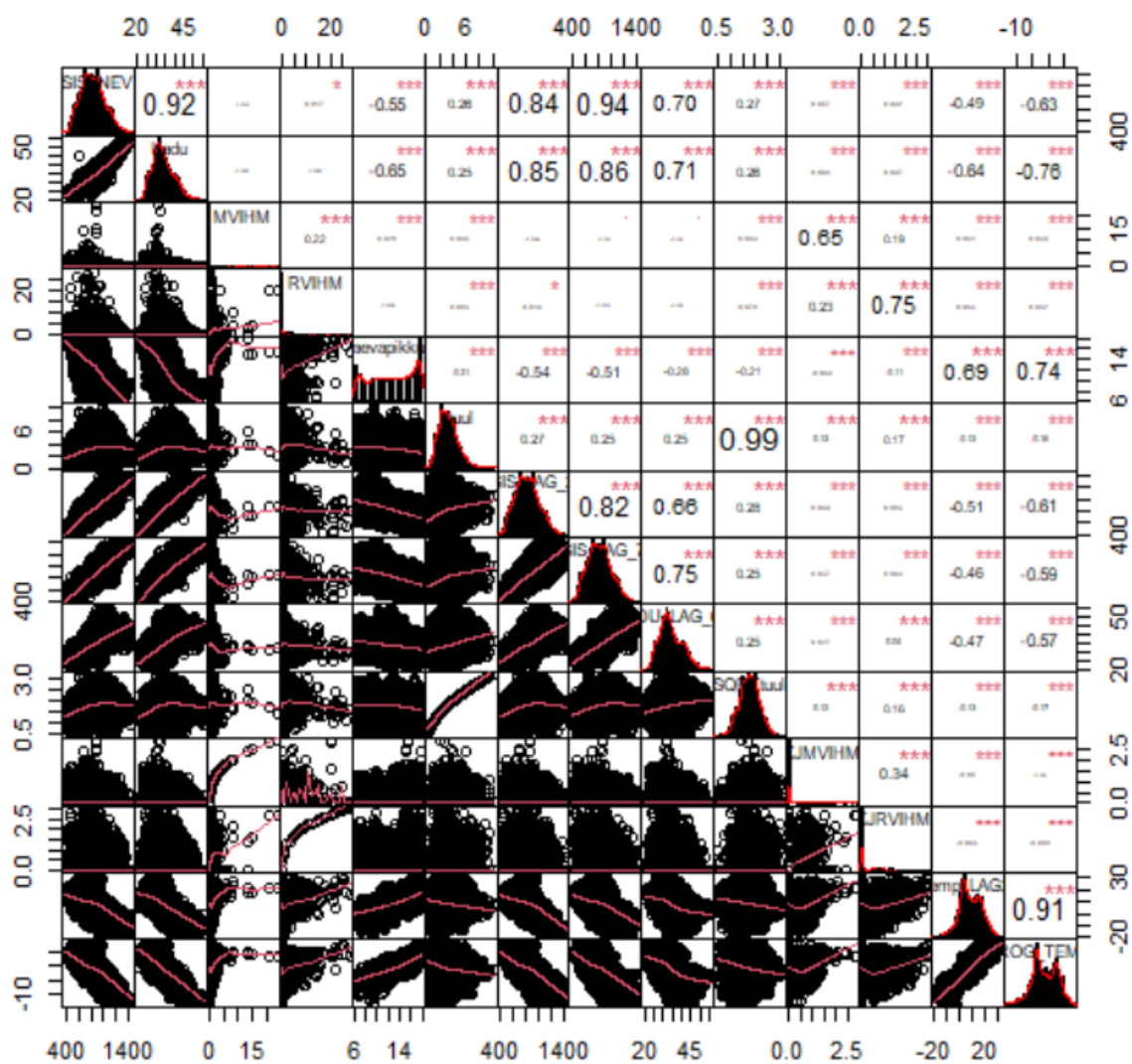
Mina, Anne-Liis Tänav

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Kaoelektri prognoosimine Elektrilevi OÜ näitel“, mille juhendajad on Juri Belikov ja Margarita Matson
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

05.05.2022

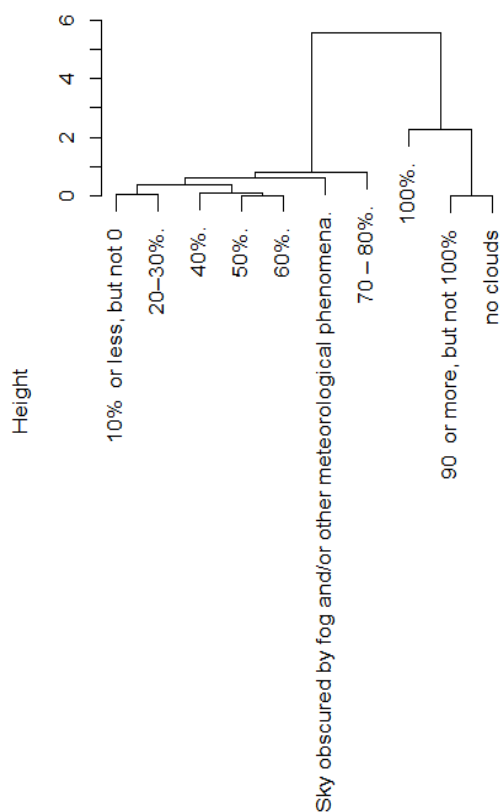
¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2 – Töös kasutatavate tunnipõhiste arvtunnuste jaotused ja korrelatsioonimaatriks koos teisendatud muutujatega



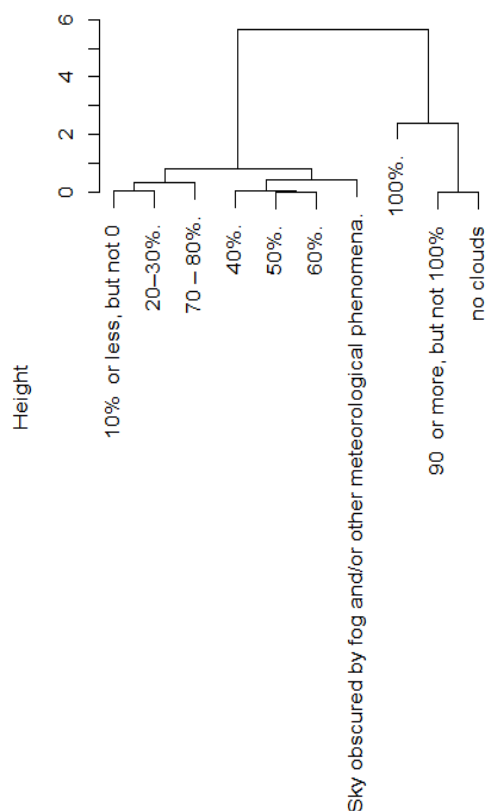
Lisa 3 – Pilvisuse kategooria dendogrammide 2020 ja 2021 aasta kohta

Dendogramm Wardi meetodil 2020



Pilvisus
agnes (*, "ward")

Dendogramm Wardi meetodil 2021



Pilvisus
agnes (*, "ward")

Lisa 4 – Nädalapäeva ANOVA tulemused koos Tukey testiga

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|-------|--------|---------|---------|-----------|
| PAEV | 6 | 5046 | 840,9 | 25,9 | <2e-16*** |
| Residuals | 15425 | 500750 | 32,5 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tukey multiple comparisons of means

95% family-wise confidence

level

Fit: aov(formula = Kadu ~ PAEV, data = Anova_andmed)

| | diff | lwr | upr | p adj |
|-----|-----------|-----------|-----------|----------|
| 2-1 | 0,296930 | -0,207354 | 0,801213 | 0,591502 |
| 3-1 | 0,481407 | -0,024267 | 0,987082 | 0,074164 |
| 4-1 | 0,410612 | -0,093671 | 0,914896 | 0,198029 |
| 5-1 | 0,335088 | -0,172004 | 0,842181 | 0,448028 |
| 6-1 | -0,819391 | -1,320971 | -0,317812 | 0,000030 |
| 7-1 | -1,014894 | -1,520569 | -0,509220 | 0,000000 |
| 3-2 | 0,184478 | -0,322554 | 0,691510 | 0,936109 |
| 4-2 | 0,113683 | -0,391962 | 0,619327 | 0,994515 |
| 5-2 | 0,038159 | -0,470287 | 0,546605 | 0,999990 |
| 6-2 | -1,116321 | -1,619269 | -0,613373 | 0,000000 |
| 7-2 | -1,311824 | -1,818856 | -0,804792 | 0,000000 |
| 4-3 | -0,070795 | -0,577827 | 0,436236 | 0,999627 |
| 5-3 | -0,146319 | -0,656145 | 0,363507 | 0,979996 |
| 6-3 | -1,300799 | -1,805141 | -0,796456 | 0,000000 |
| 7-3 | -1,496302 | -2,004717 | -0,987886 | 0,000000 |
| 5-4 | -0,075524 | -0,583970 | 0,432922 | 0,999467 |
| 6-4 | -1,230003 | -1,732951 | -0,727055 | 0,000000 |
| 7-4 | -1,425506 | -1,932538 | -0,918475 | 0,000000 |
| 6-5 | -1,154480 | -1,660244 | -0,648716 | 0,000000 |
| 7-5 | -1,349983 | -1,859808 | -0,840157 | 0,000000 |
| 7-6 | -0,195503 | -0,699846 | 0,308839 | 0,914752 |

Lisa 5 – Pilvisuse kategooria ANOVA tulemused koos Tukey testiga

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|-------|--------|---------|---------|-----------|
| RPILV | 9 | 44940 | 4993 | 169.7 | <2e-16*** |
| Residuals | 15088 | 443879 | 29 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

334 observations deleted due to missingness

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = Kadu ~ RPILV, data = Anova_andmed)

| | diff | lwr | upr | p adj |
|---|-----------|-----------|-----------|----------|
| 100%.-10% or less, but not 0 | 3,968911 | 3,303026 | 4,634797 | 0,000000 |
| 20–30%.-10% or less, but not 0 | 0,261897 | -0,614485 | 1,138280 | 0,994932 |
| 40%.-10% or less, but not 0 | 0,344201 | -0,681684 | 1,370086 | 0,988193 |
| 50%.-10% or less, but not 0 | 0,745519 | -0,239840 | 1,730879 | 0,329308 |
| 60%.-10% or less, but not 0 | 0,702689 | -0,289345 | 1,694723 | 0,428248 |
| 70 – 80%.-10% or less, but not 0 | 1,211147 | 0,410277 | 2,012018 | 0,000075 |
| 90 or more, but not 100%-10% or less, but not 0 | 1,702333 | 1,007930 | 2,396735 | 0,000000 |
| no clouds-10% or less, but not 0 | -0,415379 | -1,106825 | 0,276067 | 0,668587 |
| Sky obscured by fog and/or other meteorological phenomena.-10% or less, but not 0 | -0,000131 | -2,578873 | 2,578611 | 1,000000 |
| 20–30%.-100%. | -3,707014 | -4,374402 | -3,039626 | 0,000000 |
| 40%.-100%. | -3,624710 | -4,478993 | -2,770427 | 0,000000 |
| 50%.-100%. | -3,223392 | -4,028559 | -2,418225 | 0,000000 |
| 60%.-100%. | -3,266222 | -4,079544 | -2,452901 | 0,000000 |
| 70 – 80%.-100%. | -2,757764 | -3,322347 | -2,193181 | 0,000000 |
| 90 or more, but not 100%-100%. | -2,266579 | -2,666022 | -1,867135 | 0,000000 |
| no clouds-100%. | -4,384291 | -4,778572 | -3,990009 | 0,000000 |
| Sky obscured by fog and/or other meteorological phenomena.-100%. | -3,969043 | -6,484449 | -1,453636 | 0,000027 |
| 40%.-20–30%. | 0,082304 | -0,944556 | 1,109164 | 1,000000 |
| 50%.-20–30%. | 0,483622 | -0,502753 | 1,469997 | 0,871126 |
| 60%.-20–30%. | 0,440792 | -0,552251 | 1,433835 | 0,926235 |
| 70 – 80%.-20–30%. | 0,949250 | 0,147130 | 1,751370 | 0,006972 |
| 90 or more, but not 100%-20–30%. | 1,440435 | 0,744592 | 2,136279 | 0,000000 |
| no clouds-20–30%. | -0,677276 | -1,370169 | 0,015616 | 0,061767 |
| Sky obscured by fog and/or other meteorological phenomena.-20–30%. | -0,262028 | -2,841159 | 2,317102 | 0,999999 |
| 50%.-40%. | 0,401318 | -0,719989 | 1,522625 | 0,981445 |
| 60%.-40%. | 0,358488 | -0,768689 | 1,485665 | 0,991970 |
| 70 – 80%.-40%. | 0,866946 | -0,096272 | 1,830164 | 0,120809 |
| 90 or more, but not 100%-40%. | 1,358131 | 0,481438 | 2,234825 | 0,000042 |
| no clouds-40%. | -0,759580 | -1,633933 | 0,114773 | 0,154548 |
| Sky obscured by fog and/or other meteorological phenomena.-40%. | -0,344332 | -2,978019 | 2,289355 | 0,999994 |
| 60%.-50%. | -0,042830 | -1,133253 | 1,047593 | 1,000000 |
| 70 – 80%.-50%. | 0,465628 | -0,454309 | 1,385565 | 0,847845 |
| 90 or more, but not 100%-50%. | 0,956813 | 0,127908 | 1,785719 | 0,009807 |
| no clouds-50%. | -1,160898 | -1,987329 | -0,334468 | 0,000377 |
| Sky obscured by fog and/or other meteorological phenomena.-50%. | -0,745650 | -3,363818 | 1,872517 | 0,996475 |
| 70 – 80%.-60%. | 0,508458 | -0,418624 | 1,435540 | 0,775290 |
| 90 or more, but not 100%-60%. | 0,999643 | 0,162814 | 1,836472 | 0,006102 |
| no clouds-60%. | -1,118068 | -1,952446 | -0,283691 | 0,000937 |
| Sky obscured by fog and/or other meteorological phenomena.-60%. | -0,702820 | -3,323507 | 1,917866 | 0,997780 |
| 90 or more, but not 100%-70 – 80%. | 0,491185 | -0,106765 | 1,089136 | 0,217861 |
| no clouds-70 – 80%. | -1,626526 | -2,221041 | -1,032012 | 0,000000 |
| Sky obscured by fog and/or other meteorological phenomena.-70 – 80%. | -1,211278 | -3,765738 | 1,343181 | 0,892480 |
| no clouds-90 or more, but not 100% | -2,117712 | -2,558448 | -1,676975 | 0,000000 |
| Sky obscured by fog and/or other meteorological phenomena.-90 or more, but not 100% | -1,702464 | -4,225570 | 0,820642 | 0,502468 |
| Sky obscured by fog and/or other meteorological phenomena.-no clouds | 0,415248 | -2,107046 | 2,937542 | 0,999959 |

Lisa 6 – Töös kasutatavate päevapõhiste arvturnuste jaotused ja korrelatsioonimaatriks koos teisendatud muutujatega

