



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Vėjo jėginių generacijos laiko eilučių analizė ir prognozavimas

Baigiamasis magistro studijų projektas

Aivaras Slivikas

Projekto autorius

Doc. Dr. Evaldas Vaičiukynas

Vadovas

Doc. Dr. Vilda Gižienė

Vadovė

Kaunas, 2019



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Vėjo jėginių generacijos laiko eilučių analizė ir prognozavimas

Baigiamasis magistro studijų projektas

Didžiųjų verslo duomenų analitika (6213AX001)

Aivaras Slivikas

Projekto autorius

Doc. dr. Evaldas Vaičiukynas

Vadovas

Doc. dr. Vilda Gižienė

Vadovė

Dr. Mindaugas Kavaliauskas

Recenzentas

Prof. dr. Jurgita Bruneckienė

Recenzentė

Kaunas, 2019



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas
Aivaras Slivikas

Vėjo jėgainių generacijos laiko eilučių analizė ir prognozavimas

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Aivaro Sliviko, baigiamasis projektas tema „Vėjo jėgainių generacijos laiko eilučių analizė ir prognozavimas“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Turinys

Įvadas	13
1. Literatūros apžvalga	15
1.1. Prognozavimas versle	15
1.2. Prognozavimas energetikoje	17
1.3. Prognozavimo strategija	19
1.4. Prognozavimo metodai literatūroje	21
1.4.1. Išoriniai regresoriai.....	24
1.4.2. Kompleksinis sezoniškumas	26
1.4.3. Kolektyviniai modeliai	27
1.4.4. Prognozavimo problemos	28
2. Medžiagos ir tyrimų metodai	30
2.1. Laiko eilučių analizė	30
2.1.1. Stacionarumas	30
2.1.2. Dekompozicija	33
2.2. Laiko eilučių prognozavimas	35
2.2.1. Paprastieji metodai	37
2.2.2. Regresiniai metodai	38
2.2.3. ARMA metodai	39
2.2.4. Eksponentinio glodinimo metodai.....	41
2.2.5. Sudėtingi prognozavimo metodai.....	43
2.3. Modelių kolektyvai	44
2.4. Paklaidų įvertinimas	45
3. Tyrimo rezultatai ir jų aptarimas	49
3.1. Naudojami duomenys	49
3.2. Duomenų ir laiko eilučių analizė	51
3.3. Prognozavimo etapai	55
3.3.1. I etapas. Prognozavimas naudojant vienmačius metodus.....	56
3.3.2. II etapas. Prognozavimas naudojant daugiamačius metodus	58
3.3.3. Priežastingumo analizė	59
3.3.4. Prognozės sudarymo būdai	62
3.3.5. III etapas. Prognozavimas naudojant daugiamačius metodus, prie išorinių regresorių įtraukiant prognozavimo sistemų paruoštas prognozes	64
3.3.6. Prognozių kolektyvo kūrimas	65

3.4. VE generacijos prognozių modelių ekonominis įvertinimas	66
Išvados	70
Bibliografija	72
Priedai	77

Paveikslų sąrašas

1 pav. Straipsnių susijusių su prognozavimu energetikos sektoriuje publikavimas (1970-ieji – 2014 m.) (Esteves, 2015)	18
2 pav. (a) Google akcijų kaina 200 dienų; (b) Dienos pokytis Google akcijų kainai 200 dienų; (c) Kasmetinis žaibų skaičius JAV; (d) Mėnesiniai vienos šeimos tipo namų pardavimai JAV; (e) Kasmetė tuzino kiaušinių kaina JAV (Amerikos doleriais); (f) Mėnesinis skerdziamų kiaulių skaičius Australijoje (Victoria); (g) Kasmetinis į spąstus pakliuvusių lūšių skaičius Kanadoje (McKenzie upės regione); (h) Mėnesinė australiško alaus produkcija; (i) mėnesinė Australijos elektros energijos gamyba. (R.J Hyndman, 2014).....	31
3 pav. Google akcijų dienos akcijų kainos <i>ACF</i> (kairėje) ir Google akcijų kainos pokyčių <i>ACF</i> (dešinėje). (R.J Hyndman, 2014).....	33
4 pav. Multiplikatyviosios vėjo greičio laiko eilutės dekompozicija (2013 metų gegužės mėnuo). (Rao, 2015).....	34
5 pav. <i>ARMA</i> modelio operatorinė išraiška (R.J Hyndman, 2014) Taigi, <i>SARMA</i> modelį užrašius operatorine išraiška gauname tokią lygtį:.....	41
6 pav. VE elektrinių, kurių duomenys naudojami darbe, išsidėstymas Lietuvos teritorijoje.	49
7 pav. VE parkų pavadinimai ir instaliuotos galios.....	50
8 pav. VE parkų generacija (<i>MWh</i>) 2017 - 2018 metais.....	50
9 pav. Suminė VE generacija 2017, 2018 metais.....	51
10 pav. Mėnesinis suminės VE generacijos pasiskirstymas.	52
11 pav. VE generacijos pasiskirstymas pagal valandas.	52
12 pav. VE generacijos skirstinys pagal valandas.....	53
13 pav. VE generacijos laiko eilutės su dviem sezoniškumais (paros ir ketvirtiniu) dekompozicija.	53
14 pav. Liekanų grafikas atliekant dekompoziciją <i>TBATS</i> metodu (kairėje) ir liekanų <i>ACF</i> grafikas (dešinėje).	54
15 pav. <i>ACF</i> grafikas <i>STL</i> dekompozicijos liekanoms (laiko eilutės sezoniškumas – paros).	54
16 pav. VE generacijos laiko eilutės <i>ACF</i> ir <i>PACF</i> grafikai 48 atsilikimams.	55
17 pav. diferencijuotos VE generacijos laiko eilutės <i>ACF</i> ir <i>PACF</i> grafikai 48 atsilikimams.	55
18 pav. Prognozavimo tikslo iliustracija	56
19 pav. <i>ARIMA(2,1,1)</i> , <i>BATS</i> , <i>ETS</i> , <i>NNETAR</i> modelių paklaidų priklausomybė nuo laiko, atlikus kryžminį patikrinimą (vienos VE pavyzdys).....	57
20 pav. Vienmačio modelio <i>STL</i> prognozės pavyzdys, Vėjas 1 vėjo elektrinė.....	58
21 pav. Poveikio reikšmingumas Lietuvos VE generacijos laiko eilutei (jei $p < 0.05$ - poveikis statistiškai reikšmingas).	60

22 pav. Išorinių regresorių reikšmingumas (skaičius prie regresoriaus nurodo įtrauktą kintamojo atsilikimą).....	60
23 pav. Reikšmingiausi išoriniai kintamieji.....	61
24 pav. ARIMA(2,0,2), ARIMA(3,0,2), NNETAR ir STLM modelių paklaidų priklausomybė nuo laiko, atlikus kryžminį patikrinimą daugiamačiams modeliams (vienos VE pavyzdys).	61
25 pav. Modelių apmokymo trukmės priklausomybė, nuo apmokymo imties ilgio (1 būdas)	62
26 pav. Modelių apmokymo trukmės priklausomybė, nuo apmokymo imties ilgio (2 būdas)	63
27 pav. Daugiamačių modelių rezultatai (1 būdas).....	63
28 pav. Daugiamačių modelių rezultatai (2 būdas).....	64
29 pav. Daugiamačių modelių sukurtų naudojant sistemų prognozes rezultatai	65
30 pav. Vidutinės elektros energijos kainos 2018 m. ir 2019 m., ct/kWh (be PVM) šaltinis (https://www.regula.lt/Puslapiai/naujienos/2018-metai/2018-lapkritis/2018-11-30/komisija-patvirtino-2019-m-elektros-energijos-tarifus-buitiniams-vartotojams.aspx)	67
31 pav. Balansavimo kaštai pritaikius tyrime modeliuojamas VE generacijos prognozes	68
32 pav. Vienmačių modelių rezultatai.....	77
33 pav. Tyrime sukurtų prognozavimo modelių palyginimas tarpusavyje ir su 2 sistemų teikiamomis prognozėmis.....	78

Lentelių sąrašas

1 lentelė. Prognozavimo strategijų apibendrinimas.....	21
2 lentelė. Prognozavimo metodai suskirstyti į 5 grupes.	35
3 lentelė. Eksponentinio glodinimo klasifikavimas	43
4 lentelė. Į analizę įtraukiami regresoriai.....	59

Santrumpos

1. ARIMA – autoregresinis integruotas slenkantis vidurkis (*angl. Autoregressive integrated moving average*).
2. ADF – Augmented Dickey Fuller
3. ACF – Autokoreliacijos funkcija
4. PACF- Dalinės autokoreliacijos funkcija
5. AIC – Akaike informacijos kriterijus
6. BIC – Bayeso informacijos kriterijus
7. BSM – Pagrindinis Struktūrinis Modelis (*angl. Basic Structural Model*)
8. Bagging – savirankos agregavimas (*angl. Bootstrap aggregating*)
9. Bootstrap – savirankos imtis (imtis su pasikartojimu)
10. BATS - Ekspontinis glodinimas su Box-Cox transformacija, ARMA paklaidomis, Trendo ir Sezoniškumo komponentėmis
11. STL – Sezoninė laiko eilutės dekompozicija pagal Loess (*angl. Seasonal Decomposition of Time Series by Loess*)
12. LSTM - ilga trumpalaikė atmintis (*angl. Long Short Term Memory*)
13. Naive – Naivus (prognozavimo metodas)
14. BPNN – atgalinio sklidimo neuroninis tinklas (*angl. Back Propagation Neural Network*)
15. TES – Transformavimo išplėtimo imtis (Transform-Expand-Sample) tai stochastinio proceso klasė skirta laiko eilučių empirinei autokoreliacijai, dažnai ji naudojama Monte Carlo simuliacijose
16. XGBoost – Ekstremalus gradientų padidinimas (*angl. Extreme Gradient Boost*)
17. RNN – Rekurentinis neuroninis tinklas (*angl. Recurrent Neural Network*)
18. Ex-ante – prognozavimo būdas, kai laiko eilutės prognozei naudojama kitos laiko eilutės prognozė kaip išorinis kintamasis.
19. Ex-post – prognozavimo būdas, kai laiko eilutės prognozei naudojamos faktinės kitos laiko eilutės reikšmės.
20. NARX - netiesinis autoregresinis išorinių kintamųjų neuroninių tinklų modelis (*angl. nonlinear autoregressive exogenous artificial neural network*)
21. NNAR – neuroninio tinklo autoregresija (*angl. neural network autoregression*)
22. TBATS – Trigonometrinis Ekspontinis glodinimas su Box-Cox transformacija, ARMA paklaidomis, Trendo ir Sezoniškumo komponentėmis.
23. SMAPE – simetrinė vidutinė absoliučioji procentinė paklaida (*angl. symmetric mean absolute percentage error*)

24. MSAPE – medianos simetrinė absoliučioji procentinė paklaida (*angl. median symmetric absolute percentage error*)
25. SAPE – simetrinė absoliučioji procentinė paklaida (*angl. symmetric absolute percentage error*)
26. MAE – vidutinė absoliučioji paklaida (*angl. Mean Absolute Error*)
27. RMSE – vidutinė kvadratinė paklaida (*angl. Root Mean Squared Error*)
28. nMAE – normalizuota vidutinė absoliučioji paklaida (*angl. normalized Mean Absolute Error*)
29. VE – vėjo elektrinė
30. MW – Megavatas (Galios matavimo vienetas)
31. MWh – Megavatvalandė (Energijos matavimo vienetas)

Slivikas, Aivaras. Vėjo jėginių generacijos laiko eilučių analizė ir prognozavimas. Magistro baigiamasis projektas / vadovai doc. dr. Evaldas Vaičiukynas ir doc. dr. Vilda Gižienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypties grupė): Didžiųjų verslo duomenų analitika

Reikšminiai žodžiai: vėjas, generacija, prognozavimas, laiko, eilutės, vienmačiai, daugiamačiai, analizė.

Kaunas, 2019. 73 p.

Santrauka

Šio magistro projekto tikslas - sumažinti elektros energijos sistemos balansavimo kaštus, pasiekiant tikslesnes VE generacijos prognozes. Baigiamajame projekte iškelti uždaviniai: 1) Sudaryti VE generacijos prognozės modelius, kurie prognozuoja 38 val. horizontui į priekį. 2) Įvertinti skirtingų metodų modelių tikslumą. 3) Palyginti sukurtų modelių tikslumą su VE prognozavimo sistemų tikslumu. 4) Kombinuojant sistemų prognozes su išoriniais kintamaisiais, gauti geresnį rezultatą, nei bet kurios iš dviejų sistemų atskirai. 5) Panaudojus tiksliausių modelių prognozes atlikti ekonominį įvertinimą.

Magistro projekte aprašomi vėjo elektrinių generacijos laiko eilučių prognozės modeliavimai naudojant 17 skirtingų modelių. Išbandyta 17 vienmačių modelių, kai vėjo elektrinių generacija prognozuojama remiantis tik praeities duomenimis. Atrinkti stipriausiai laiko eilutę veikiantys išoriniai regresoriai. Išbandyti 4 daugiamačiai modeliai su išoriniais regresoriais. Daugiamačių modelių prognozių kūrimas atliktas dviem skirtingais būdais: 1) Fiksuojant apmokymo imtį ir 2) apmokymo imtį periodiškai didinant. Sudaryti trys skirtingi kolektyvų modeliai. Prognozės sudarytos trylikai skirtingų vėjo elektrinių ir suminei visų elektrinių generacijos laiko eilutei. Atlikus modeliavimą patikrinta kiekvieno modelio paklaida. Atrinkus geriausius modelius apskaičiuota finansinė nauda.

Tyrime gauti rezultatai atskleidžia, kad prognozuojant vėjo elektrinių generaciją daugiamačiai modeliai yra pranašesni už vienmačius. Modelių kolektyvų kūrimas praktiškai visuomet pasiteisina ir kolektyvo gaunamas rezultatas yra tikslesnis už pavienių modelių. Keičiant apmokymo imtį, gaunamas rezultatas šiek tiek tikslesnis, nei nekeičiant imties, tačiau skaičiavimo resursai ženkliai išauga. Tiksliausi vienmačiai modeliai – *ARIMA* ir *NNETAR*, daugiamačiai – *ARIMAX* ir *LOESS* regresija. Sistemų rezultatus pavyksta pranokti tik kai į modelį įtraukiamos šių sistemų prognozės.

Slivikas, Aivaras. Title of Final Degree Project. Master's Final Degree Project / supervisors assoc. prof. Evaldas Vaičiukynas and assoc. prof. Vilda Gižienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Business Big Data Analytics

Keywords: wind, generation, forecast, time, series, univariate, multivariate, analysis.

Kaunas, 2019. 73 pages.

Summary

The aim of this master thesis is to reduce electricity system imbalance prices with more accurate wind power plants generation forecasts. The main objectives were: 1) to develop wind turbines generation forecast models which are eligible to forecast up to 38 hours horizon ahead. 2) To evaluate accuracy of different forecast models. 3) To compare developed model errors with two systems forecasts output. 4) To combine systems forecasts with exogenous variables and achieve the better result than any of those systems separately. 5) To calculate financial benefit, using the most accurate models results.

This master thesis represents wind power turbine generation time series analysis and forecasting. 17 different models have been tested. 17 univariate models, with input of forecasted time series historical data have been developed. Selection of exogenous variables, which have strongest impact to time series values, has been performed. 4 multivariate with exogenous variables models have been tested. Multivariate forecast modelling has been performed in two following ways: 1) with constant training set and 2) with periodically growing training set. Three different ensemble models have been created. The forecasts have been produced for thirteen different wind turbine plants and sum generation time series. Each model has been evaluated with its performance error. Financial benefit has been introduced using the best performing models results.

The research shows that multivariate models are more advanced than univariate. Model ensembles give positive results and ensembles most of the times provides with more accurate result than any of the models individually. Train set re-evaluation gives slightly more positive result than no re-evaluation, although calculation resources are way higher. The most accurate of the univariate models are *ARIMA* and *NNETAR*. The most accurate of the multivariate models are *ARIMAX* and *LOESS* regression. Two system results have been outperformed only when their forecasts have been included in the models.

Ivadas

Temos aktualumas: Dabartinėje konkurencingoje ir dinaminėje aplinkoje vis daugiau sprendimo priėmimo procesų energetikos sektoriuje remiasi prognozėmis ir jų dinamikos modeliavimu. Energetikoje prognozavimo taikymas pasiskirstęs planavimo, rinkos analizės, rinkos plėtros ir sistemos valdymo segmentuose. Prognozuojama generacija, suvartojimas, elektros energijos kaina, dujų kaina, vandens tėkmė, ledo tirpsmas, saulės ir vėjo elektrinių generacija ir dar daugiau. Vėjo elektrinių (toliau - VE) generacijos prognozė yra viena iš sudėtingiausių, bet tuo pačiu ir didelę svarbą sistemai ir VE savininkams teikiančių prognozių. Sugebėdami tiksliai prognozuoti VE generaciją, elektros energijos sistemos operatoriai gali lengviau suvaldyti sistemą, pasiruošti būsimiems elektros energijos šuoliams ir taip būti vienu žingsniu priešaky prieš įvykstant dideliems elektros energijos sistemos balanso nuokrypiams. Turėdami tikslią prognozę VE savininkai gali optimaliai susiplanuoti savo elektrinių gamybą, taip išvengdami didesnių kaštų. Žinodami, kad VE generacijos suvaldymo kaštai yra įtraukiami į visuomeninę elektros energijos kainą kaip viešuosius interesus atitinkančios paslaugos (toliau – VIAP) galime išvelgti platesnę VE generacijos prognozavimo naudą, kuri svarbi visuomenei, o ne atskiroms interesų grupėms.

Problema: Baigiamajame darbe tiriama vėjo elektrinių generacijos prognozės tikslumo problema. Esant dideliems generacijos prognozių nuokrypiams, valstybė turi padengti balansavimo kaštus. Balansavimo kaštai tiesiogiai priklauso nuo nuokrypių dydžio, bei laiko, kada įvyko nuokrypis. Netikslios prognozės taip pat lemia sistemos nestabilumą. Kyla klausimas, kurie metodai prognozuoti VE generaciją yra tiksliausi ir ar įmanoma patobulinti dabartines prognozes?

Tyrimo objektas: Vėjo elektrinės.

Tyrimo tikslas: Tikslas – sumažinti elektros energijos sistemos balansavimo kaštus, pasiekiant tikslesnes VE generacijos prognozes.

Uždaviniai:

1. Sudaryti VE generacijos prognozės modelius, kurie prognozuoja 38 val. horizontui į priekį.
2. Įvertinti skirtingų metodų modelių tikslumą ir modelių kolektyvų tikslumą.
3. Palyginti sukurtų modelių tikslumą su VE prognozavimo sistemų tikslumu ir kombinuojant sistemų prognozes su išoriniais kintamaisiais, gauti geresnį rezultatą, nei bet kurios iš dviejų sistemų atskirai.
4. Panaudojus tiksliausių modelių prognozes atlikti ekonominį įvertinimą.

Darbo metodai: Užsienio mokslinės literatūros analizė, duomenų analizė naudojant „Gretl“ ir „R studio“ programinius paketus, modelių kūrimas naudojant „R studio“

programinį paketą, modelių įvertinimas naudojant „R studio“ programinį paketą, ekonominio įvertinimo parengimas.

Darbo struktūra: Pirmojoje darbo dalyje analizuojami moksliniai literatūriniai šaltiniai apžvelgiantys įvairius prognozavimo metodus ir modelius. Antroje dalyje pristatoma tyrimo metodologija, paaiškinami ir pagrindžiami metodai. Trečioje dalyje pristatomi VE generacijos laiko eilučių modeliavimo rezultatai. Siekiant gauti tiksliausius modelius ir apžvelgti daugiau metodų, modeliuojama trimis etapais: 1) vienmačių modelių kūrimas, 2) daugiamačių modelių kūrimas ir 3) daugiamačių modelių kūrimas pasitelkiant specialių prognozavimo sistemų sugeneruotas prognozes. Kiekvieno iš šių etapų rezultatai įvertinami ekonomiškai, patikrinant kokie VE balansavimo kaštai galėjo būti patiriami pritaikius skirtingus modelius. Trečioje dalyje taip pat įvertinama galutinė nauda elektros energijos vartotojui.

1. Literatūros apžvalga

Prognozuojant svarbu pasirinkti tinkamus regresorius, turėti pakankamai istorinių duomenų ir atlikti jų apžvalgą. Analogiškai – rašant tiriamąjį darbą svarbu išnagrinėti ir tinkamai pritaikyti literatūrinius šaltinius išleistus praeityje, juos palyginti, bei atsirinkti naudingą informaciją. Norint tai padaryti atliekama trumpa mokslinės literatūros apžvalga. Dauguma pasirinktų nagrinėti šaltinių yra koncentruoti į prognozavimą elektros energijos sektoriuje, laiko eilučių analizę ir vienmačių laiko eilučių prognozių sujungimą į kolektyvus. Taip pat apžvelgiami su laiko eilučių modeliavimu susiję žurnalai ir straipsniai, kuriuose pateikiama informacija apie prognozių naudą ir sėkmingą jų taikymą versle. Su energetika susijusi literatūra daugiausia dėmesio skiria vėjo greičio ar elektros energijos suvartojimo laiko eilučių analizei ir jų prognozėms. Nagrinėjami darbai pateikia svarbių įžvalgų tiek teorijoje, tiek praktikoje. Kai kurie jų yra istorinių žinių ar tyrimų apibendrinimai, kiti – paskirti naujų teorijų plėtojimui ir šių teorijų naudos įrodymui.

1.1. Prognozavimas versle

(Hyndman R. J., *Business Forecasting Methods*, 2010) teigia, kad organizacijos vadovybės sprendimų priėmimo prognozavimas turi būti esminė dalis ir gali būti pritaikomas daugelyje veiklos sričių. Savo knygoje profesorius pažymi, kad versle prognozavimas dažnai yra painiojamas su planavimu ir tikslais. Prognozavimas – tai bandymas nuspėti ateitį taip tiksliai, kaip tai yra įmanoma, naudojantis visa prieinama informacija, istoriniais duomenimis ir iš anksto apie ateitį turimomis žiniomis. Tuo tarpu planavimo etape numatomi veiksmai, kuriuos reikia atlikti, norint pasiekti išsikeltus tikslus, nustatomus pagal atliktas prognozes. Todėl reikia suprasti prognozės, plano ir tikslų išsikėlimo skirtumus.

Daugelio organizacijų strateginė sėkmė priklauso nuo efektyvios sąsajos tarp kiek įmanoma tikslesnės prognozės ir kiek įmanoma lankstesnio resursų paskirstymo plano (Wacker & R., 2002). Anot autoriaus, pagrindinė priežastis, dėl kurios verslo vadovai reikalauja kuo tikslesnių prognozių, yra jų pagrįstas įsitikinimas, kad tikslios prognozės padeda sumažinti kaštus. Daugiau nei 10% galutinio pelno bendrovėse yra prarandama dėl netikslių prognozių ar tinkamo inventoriaus bei duomenų neužtikrinimo kuriant prognozes. Prognozių horizonto ilgis taip pat turi didelę įtaką galutiniam finansiniam rezultatui. Prognozes atliekant dažniau ir trumpesniam periodui į ateitį, galima užtikrinti mažesnę paklaidą, kuri padės sutaupyti. Todėl prognozavimas turėtų tapti natūraliu verslo valdymo įpročiu ir būti neatsiejama procesų dalimi.

(Grushka-Cockayne, 2019) savo straipsnyje apie prognozių kokybės pagerinimą pamini, kad dauguma organizacijų išties rimtai žiūri į prognozavimą ir siekia pagerinti savo prognozes. Tiesa,

dažna šių organizacijų klaida, kad jos kiek atleidžia žiūri į prognozių sekimą. Nesunku surasti šaltinių, kurie pastoviai skelbia savo prognozes, tačiau kur kas sunkiau surasti tokius, kurie seka savo prognozes ir skelbia jų kokybę. „Vienintelis būdas tobulinti prognozes, jas sekti ir įvertinti“ sako autorė. Ji pamini meteorologinių prognozių teikėjus ir jų gerus rezultatus, patvirtindama, kad vienas iš faktorių, dėl ko jų prognozės dažnai būna labai tikslios, tai pastovus klaidų sekimas ir netgi visuomenės spaudimas. Visuomenė kasdien seka orų prognozes, taigi orų prognozių teikėjus tai verčia pasitempti. Jie sugeba nuspėti uraganus ir kitus gamtos reiškinius minučių tikslumu. Tokio akylo sekimo, autorės manymu, trūksta verslo sektoriuje. Verslo sektorius įgytą didelį pranašumą panaudodamas meteorologinių prognozių teikėjų praktiką, sekant prognozių tikslumą.

Prognozavimo procesas yra ypatingas laiko naudotojas verslo valdyme, taip pat prognozavimo procesas gali būti didelis laiko švaistytojas, jei pastangos pagerinti prognozę nueina perniek. Kai verslo pastangos koncentruojasi tik į tikslumo pagerinimą, dažnai pamirštama, kad yra kitų faktorių, tokių kaip prognozės šališkumas (prognozė visada per aukšta, ar per žema) arba neužtikrintumas (prognozės pasikliautinumo intervalas). Koncentravimasis tik į prognozės tikslumą ignoroja efektyvumą ir naudingumą (Gilliland, 2011).

Kaip vieną iš verslo sričių, kuriose taikomas prognozavimas, galima apžvelgti marketingą. Marketinge savo kliento pažinimas yra pagrindinis verslo žaidimo taisyklių variklis. Suprasdami kuo daugiau apie savo klientą ir tai ko klientai nori, verslo vadovai gali juos lengviau pasiekti naudodami tinkamus kanalus tinkamu laiku. Naudodami mašininio mokymosi prognozavimo technikas marketingo specialistai gali nuspėti tolimesnį savo klientų elgesį pasitelkdami praeities šablonus ar klientų panašumus. Pavyzdžiui regresinė analizė gali būti panaudota identifikuojant ryšį tarp klientų praeities pirkimo įpročių, nuspėjant jų ateities pirkinius. Tokie metodai gali puikiai pasitarnauti aprašant naują marketingo strategiją (O’Connell, 2019).

Apibendrinant aukščiau apžvelgtus mokslinius šaltinius, verta pastebėti, kad dažnai verslo atstovai supranta prognozavimo svarbą savo veiklos rezultatams. Prognozių taikymas versle nėra naujiena, jis plačiai taikomas jau kurį laiką. Tačiau dar geresnius įmonių rezultatus galėtų nulemti platesnis prognozių integravimas į kitas veiklos sritis ir kokybinių parametrų sekimas. Reguliarus kokybinių parametrų sekimas leistų užtikrinti mažesnius operacinius kaštus ir taip pateisinti prognozių modeliavimui skiriamus įmonės resursus.

1.2. Prognozavimas energetikoje

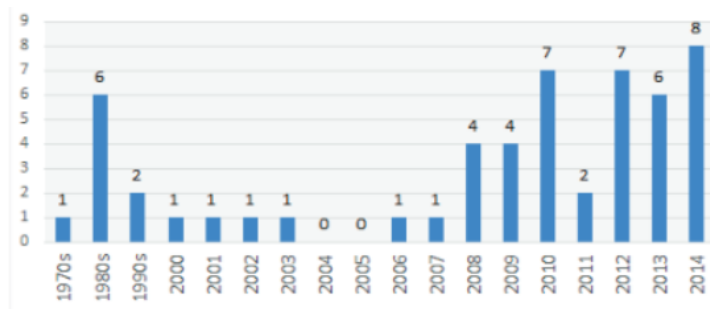
Energetikos industrija plačiai apima visas kitas industrijas ir dominuoja visose pasaulio valstybėse. Moderni visuomenė vartoja milžiniškus energijos kiekius ir energetikos sektorius yra kritinė infrastruktūros dalis kiekvienoje šalyje. Nuo energetikos išteklių kainos priklauso labai didelė dalis prekių ir paslaugų kainų visame pasaulyje. Taigi, energetika yra neatsiejama verslo dalis ir joje prognozavimas užima labai svarbią ir jautrią dalį. Ypatinga prognozavimo sparta energetikos sektoriuje pastebima paskutinius 20 metų, kai įprastinius energijos šaltinius pradėjo keisti alternatyvieji.

Prognozavimas energetikoje sparčiai keičiasi. Anksčiau pakakdavo numatyti kokiu pajėgumu dirbs dujomis varomas generatorius ir jis tokiu režimu dirbdavo visą suplanuotą laiką. Dabar, nustatyti koks energijos kiekis bus pagamintas ir perduotas į elektros energijos tinklą, tuo pat metu nustatant, koks šios energijos kiekis bus suvartotas iš tinklo, tampa vis sudėtingiau. Sparčiai augant atsinaujinančios energijos išteklių (AEI) panaudojimui ir krintant šių resursų panaudojimo kainai, prognozavimas tampa pagrindinių įrankiu bandant suplanuoti elektros energijos sistemos darbą ar siekiant pasipelnyti iš greitai kintančių kainų rinkoje. Visa AEI pagaminama energija visiškai priklauso nuo natūralių procesų tokių kaip vėjas, saulė ar audros. Šie procesai yra kintantys laike, tačiau juos galima prognozuoti. AEI – ne vienintelė priežastis dėl kurios energetikos sektorius užsiima prognozėmis. Praeityje tinklo operatoriai naudodavo meteorologines prognozes nustatant audras ir ekstremalius atvejus, siekiant pasiruošti tinklo remonto darbams ar apsaugoti savo infrastruktūrą. Dabar kaip pagrindines priežastis dėl kurių energetikoje ypatingai pasitarnauja prognozavimas galima įvardyti šias (Bax Energy, 2019):

- 1) Mažinamos sąnaudos tiesioginiams rinkos dalyviams. Rinkos dalyviai tiksliau suprognozavę savo elektrinių generaciją, energiją gali parduoti geresne kaina. Lygiai taip pat rinkos dalyviai, perkantys energiją rinkoje prognozių pagalba gali susiprognozuoti tikslų energijos poreikį ir nusipirkti optimalų energijos kiekį.
- 2) Padedama energijos prekeiviams (*angl. traders*) ir regioniniams energijos tiekėjams tiksliai apskaičiuoti energijos srautus ir susiprognozuoti būsimas kainas rinkoje.
- 3) Suteikiama daug informacijos elektros sistemos operatoriams, kurie gali pasiruošti remonto darbams (kai pvz. VE generacija sumažėja), ar numatyti kada reikės keisti elektrinių dalis.
- 4) Prognozės padeda nustatyti ir įvardyti faktorius, kurie nulemia energijos produkciją (pvz. sniego tirpsmas, saulės apšvieta)
- 5) Užtikrinamas reikiamas energijos kiekis rezervuaruose ar baterijose.

Energijos planavimas ir valdymas visuomet buvo iššūkis daugeliui valstybių pasaulyje. Nuo pat industrinės revoliucijos pradžios pagrindinis „butelio kakliukas“ ekonomikos augimui buvo energijos prieinamumas, ne tik dėl neišplėtotų elektros tinklų, bet ir dėl skirtingų energijos

šaltinių stygiaus. Taigi, sukurti metodikas ir procedūras skirtas elektros energijos vartojimo ir gamybos prognozavimui tapo pagrindinis akademikų, valdžios ir elektros kompanijų tikslas. (Esteves & T, 2015) savo straipsnyje apžvelgia šias metodikas, straipsnius ir procedūras kurios buvo publikuojamos nuo 1970 m.



1 pav. Straipsnių susijusių su prognozavimu energetikos sektoriuje publikavimas (1970-ieji – 2014 m.)

(Esteves & T, 2015)

Pasaulinės energetikos prognozavimo varžybos 2017 metais (GEFCom2017) pritraukė daugiau nei 300 studentų ir profesionalų iš daugiau nei 30 šalių. Komandos sprendė hierarchines ir tikimybinės suvartojimo prognozavimo problemas. Šios varžybos turėjo kvalifikacinę atranką, pirmosios kuriose buvo naudojama hierarchiniai duomenys su daugiau nei 2 lygiais, pirmosios kuriose buvo leidžiama naudotis išoriniais duomenų šaltiniais ir pirmosios, kuriose buvo prašoma realiu laiku pateikti ex-ante prognozes, taip pat jos buvo ilgiausios. Tai parodo, kad energetikos sektoriuje prognozavimas tobulėja, reikalauja iššūkių ir yra viena iš pirmaujančių šakų (Tao, Jingrui, & Jonathan, 2017).

VE generacijos prognozės suteikia elektros sistemai stabilumo. Išankstinis žinojimas koks energijos kiekis bus pagamintas, taip pat pasikliautinumo intervalo nustatymas yra labai naudingas veiksnys elektros energijos sistemos operatoriui. Vieną iš dažniausiai šiam tikslui naudojamų stochastinių modelių prognozuoti VE generacijai – *ARIMA* (Dumitru, 2018).

Kalbant apie VE generacijos prognozavimą, paskutinių metu daugėja literatūros susijusios su praeities duomenimis paremtų modelių prognozėmis. Įvairūs duomenų apdorojimo tipai sėkmingai taikomi padedant šiems modeliams gerinti prognozės tikslumą. (Hui & Chao, 2019) norėdamas akcentuoti žinių trūkumą duomenų apdorojime prieš prognozuojant VE generaciją, išleido straipsnį, kuriame pasakoja apie 7 duomenų apdorojimo technikos kategorijas: dekompoziciją, kintamųjų pasirinkimą, kintamųjų išgavimą, triukšmo panaikinimą, liekanų modeliavimą, išskirčių aptikimą ir filtravimu paremtą korekciją.

1.3. Prognozavimo strategija

Kaip ir bet kurioje kitoje srityje, prognozavime taip pat svarbu turėti strategiją. Pagal pasirinktą strategiją dažnai priklauso kokius programinius paketus naudosime, kiek laiko skirsime modeliavimui ir koks bus galutinis rezultatas.

Kelių žingsnių į priekį prognozavime galima taikyti keletą skirtingų strategijų:

- Rekursinė (*angl. Recursive*) strategija.
- Tiesioginė (*angl. Direct*) strategija.
- Statinė (*angl. Static*) strategija.
- Besiridenanti (*angl. Rolling*) strategija.
- Ištaisyta (*angl. Rectify*) strategija.

(Taieba, Bontempia, Atiyac, & Sorjamaab, 2012) savo straipsnyje apžvelgia tiesioginės kelių žingsnių ir rekursinės kelių žingsnių strategijų kombinacijas bandant išgauti abiejų strategijų geriausias savybes. Tiesioginė prognozavimo strategija, tai strategija, kai kiekvienam horizonto taškui sudaroma atskira prognozė ir vėliau jos kombinuojamos. Rekursinė prognozavimo strategija – tai kai praėjusio žingsnio prognozė į prognozės modelį įtraukiama kaip stebėjimas ir sudaroma nauja prognozė. Autoriai pamini, kad rekursinės strategijos trūkumas – šališkumas, o tiesioginės - didelė dispersija. Naudojant simuliacijas sukurta nauja strategija, kuri darbe aprašoma kaip ištaisyta strategija, ji sumažina rekursinės strategijos šališkumą ir tiesioginės strategijos dispersiją. Didesnis laiko eilutės triukšmas ir trumpesnės laiko eilutės, šios strategijos rezultatus paverčia netgi geresniais, lyginant su kitomis strategijomis, tačiau praktiškai ši strategija gali būti sunkiai pritaikoma.

Likusios dvi strategijos tai statinė ir besiridenanti. Statinė strategija naudoja žingsnių skaičiavimo metodą, kuris reiškia, kad apmokymo duomenys nekinta iki kito prognozavimo periodo. Besiridenanti strategija tai slenkantis praeties duomenų priėmimo/ išmetimo procesas nuspėjant ateitį nustatytam periodui į priekį. Ši strategija tai kontrastas rekursinei ir statinei strategijoms.

(Wang, Smith-Miles, & Hyndman, 2009) teigia, kad renkantis tinkamą prognozavimo strategiją ir metodus turi būti atsižvelgiama į keturis pagrindinius komponentus: 1) prognozavimo horizontą, 2) technikos specifiškumą, 3) analizuojamą sritį ir 4) praeties duomenų struktūrą. Realiame taikyme vartotojai turi pasirinkti jiems labiausiai tinkantį metodą, pagal prognozavimo užduotį ir atsižvelgti į šiuos faktorius. Autoriai kaip pagrindinius tikslus renkantis strategiją įvardija:

1. Vienmačių laiko eilučių charakteristikų identifikavimą ir metrikų nustatymą.
2. Pagrindinių ir populiariausių prognozavimo metodų efektyvumo įvertinimą.

3. Meta-mokymosi darbo aplinkos kūrimą, kuris apima 2 aukščiau paminėtus punktus ir juos apjungia pagal prognozavimo metodo ir duomenų charakteristikos ryšį. Šios taisyklės padeda prognozuotojui pasirinkti metodą (sprendimo priėmimo procesas).

(Hyndman & Athanasopoulos, 2014) taip pat aprašo pagrindinius penkis žingsnius kuriuos reikia atlikti norint išspręsti prognozavimo užduotį:

1. Problemos apibrėžimas – šis žingsnis dažnai įvardijamas kaip vienas iš sunkiausių, nes prognozę atliekantis analitikas turi gerai išsiaiškinti verslo poreikį. Reikia žinoti kaip prognozės bus naudojamos, kas turi gauti prognozes, kokių dažnumu ir kokiam periodui į priekį. Analitikas turi praleisti daug laiko rinkdamas informaciją iš įmonės apie duomenų bazes, gauti reikiamą istorinę informaciją ir pasakyti kaip sudaryta prognozė padės planavimui.
2. Informacijos surinkimas – informacija kurios prireiks yra dviejų tipų: a) istoriniai duomenys ir b) specifinės specialistų, kurie naudos prognozes, žinios ir įžvalgos. Dažnai gali kilti sunkumų surenkant pakankamai istorinių duomenų norint sukurti gerą statistinį modelį. Pasenę duomenys gali būti ne tokie informatyvūs ir sukelti problemų kuriant modelį, tačiau nereikia skubėti jų atsisakyti, nes geri statistiniai modeliai gali susitvarkyti su evoliuciniais pokyčiais.
3. Preliminari (tiriamoji) analizė – visuomet reikia atvaizduoti duomenis grafiškai. Ar yra atsikartojanti struktūra? Ar yra aiškus trendas? Ar svarbu sezoniškumas? Ar yra verslo ciklų apraiškų? Ar duomenyse yra išskirčių kurias galėtų paaiškinti specialistai ekspertai? Ar ryšys tarp kintamųjų naudojamų analizei yra stiprus? Padėti atsakyti į šiuos klausimus gali atsakyti įvairūs statistiniai paketai ir programos.
4. Modelio pasirinkimas ir apmokymas – geriausias pasirinktas modelis priklauso nuo istorinių duomenų prieinamumo, ryšio tarp prognozuojamo kintamojo ir paaiškinamųjų kintamųjų bei modelio panaudojimo atvejo. Natūralu išbandyti bent du ar tris potencialius modelius ir juos palyginti.
5. Modelio naudojimas ir įvertinimas – kai modelis pasirenkamas ir nustatomi jo parametrai modelis naudojamas atlikti prognozėms. Modelis gali būti įvertintas tik tuomet, kai turimi duomenys periodui, kuriam modelis prognozavo. Yra sukurta daugybė metodų padedančių įvertinti modelio tikslumą. Tačiau dažnai iškyla praktinių problemų prognozuojant, pavyzdžiui kaip elgtis su trūkstamomis istorinėmis reikšmėmis arba kaip elgtis su labai trumpo periodo laiko eilutėmis.

Mašininio mokymosi prognozavimo varžybų *NN5* apžvalga (Taieba, Bontempia, Atiyac, & Sorjamaab, 2012) atskleidė, kad kelių rezultatų išvedimo (*angl. Multiple-Output*) taikymas, kai vietoje vieno prognozės skaičiaus išvedamas vektorius, yra visais atvejais pranašesnis už vieno

rezultato išvedimo (*angl. Single-Output*) taikymą prognozuojant vienmates laiko eilutes kelis žingsnius į priekį. Taip pat varžybų išvadose atskleidžiama, kad sezoniskumo panaikinimas laiko eilučių prognozėms padarė teigiamą įtaką ir rezultatai dažnu atveju buvo geresni. Kryptį, tolimesniems kelių žingsnių į priekį laiko eilučių prognozavimo modeliams kurti, autoriai įvardija sezoniskumo naikinimą ir kelių rezultatu išvedimo strategijos tobulinimą.

1 lentelė. Prognozavimo strategijų apibendrinimas.

Nr.	Strategija	Panaudojimas	Trūkumas
1	Rekursinė	Naudojama kai aktualu turėti visas prognozės intervalo reikšmes, o ne vieną taškinę prognozę	Šališkumas
2	Tiesioginė	Koncentruota į vieną, konkretaus taško, prognozę	Didelė dispersija
3	Statinė	Naudojama kai taupomi skaičiavimo resursai	Rečiau atnaujinant istorinius duomenis rizikuojama sukurti mažiau tikslią prognozę
4	Besiridenanti	Naudojama kai modeliui sudaryti norima įtraukti naujesnius duomenis ir išmesti senesnius	Dideli skaičiavimo resursai
5	Ištaisyta	Tiesioginės ir rekursinės strategijų kombinacija	Sunkus praktinis pritaikymas

Apibendrinant prognozavimo strategijas galima pastebėti, kad strategijos pasirinkimas visuomet priklauso nuo prognozavimo tikslo ir turimų duomenų specifiškumo. Kartais galima bandyti taikyti kelias strategijas ir pasirinkti tinkamesnę. Greitesnis strategijos pasirinkimas taip pat priklauso nuo patirties sprendžiant įvairius prognozavimo uždavinius.

1.4. Prognozavimo metodai literatūroje

(Hyndman & Athanasopoulos, 2014) knygoje aprašomi įvairūs metodai padedantys sukurti tinkamą prognostinį modelį priklausomai nuo užduoties tikslo. Autorius taip pat pamini paprasčiausius metodus, kurie dažnai kelią nuostabą, dėl savo paprastumo ir efektyvumo. Tai tokie metodai kaip vidurkių, kai sekančio žingsnio prognozė yra visų turimų istorinių duomenų vidurkis. Naive metodas - kai sekančio žingsnio prognozė yra praėjusio žingsnio reikšmė ir Naive metodo atmaina – Seasonal Naive, kai sekančio žingsnio prognozei atkartojama praėjusio sezono periodo ta pati reikšmė. Drift metodas – kai paimama pirma ir paskutinė istorinių duomenų reikšmės, tarp jų

brėžiama tiesė ir ją bandoma ekstrapoliuoti į ateitį. Kartais vienas iš šių metodų gali būti geriausiai tinkantis prognozuoti, tačiau praktikoje dažniausiai šie metodai naudojami kaip palyginimas (*angl. benchmark*), o ne kaip prognostinis metodas. Autorius rekomenduoja savo sukurtą prognozės modelį visuomet palyginti su vienu iš paprastųjų metodų norint įsitikinti ar tai tikrai geriausia prognozė kokia galėjo būti.

Dekompozicija yra ne tik naudinga analizuojant laiko eilutes ir nagrinėjant jų istorinius pokyčius laike, ji taip pat gali būti puikiai pritaikoma prognozavime. Norint prognozuoti dekomponuotas laiko eilutes, reikia prognozuoti sezoninę ir sezoniškai pataisytą komponentes atskirai. Dažniausiai yra laikoma, kad sezoninė komponentė nekinta laike arba kinta labai lėtai, taigi jai prognozuoti imama praėjusio periodo reikšmė. Kitaip sakant naudojamas Seasonal Naive metodas. Norint prognozuoti sezoniškai pataisytą komponentę, naudojamas bet kuris, sezoniškumui nepritaikytas metodas. Pavyzdžiui atsitiktinio žingsnio su dreifu (*angl. random walk with drift*), Holt's arba nesezoninis *ARIMA* modelis (John Wiley & Sons, 2015).

(Rao & Uma, 2015) koncentruodamiesi į vėjo greičio prognozė panaudojant duomenis iš vėjo elektrinių, atliko tyrimą ir prognozavo vėjo greitį 2 žingsnius į priekį naudodami dekompozicijos metodą. Dekompozicijos metodas buvo lyginamas su kitais tradiciniais prognozavimo metodais tokiais kaip *ARIMA*, *BPNN*, *TES*. Modelių rezultatai buvo lyginami pagal *MAPE*, gautas rezultatas parodė, kad dekompozicijos modelis buvo tiksliausias apmokant modelius su įvairiomis imtimis: dviejų, keturių ir šešių mėnesių, taip pat keičiant apmokymo imties sezonus.

Dar vienas prognozavimo metodas yra eksponentinis glodinimas, kuris buvo pasiūlytas 1950 (Brown, Holt, Winters) ir tuomet jis buvo laikomas vienu iš sėkmingiausių prognozavimo metodų. Prognozės sukurtos naudojant eksponentinį glodinimą naudoja praeities reikšmių vidurkius su skirtingais svorio koeficientais. Svorio koeficientas eksponentiškai mažėja senesnių stebėjimų kryptimi. Kitaip sakant, kuo naujesnis yra stebėjimas, tuo didesnę svorį jis įgyja. Šis metodas gali greitai generuoti patikimas prognozes plataus spektro laiko eilutėms ir turi didelį pritaikymo potencialą industriniame sektoriuje. Skirtingi eksponentinio glodinimo būdai gali būti pritaikomi skirtingas charakteristikas turinčioms laiko eilutėms. Paprastas eksponentinis glodinimas naudojamas laiko eilutėms neturinčioms nei trendo nei sezoniškumo. Holtas išplėtojo paprastąjį eksponentinį glodinimą ir pritaikė jį laiko eilučių turinčių trendą prognozavimui, šis metodas vadinamas Holto tiesiniu trendo metodu arba tiesiog Holto eksponentiniu glodinimu. Trečias eksponentinis glodinimas taikomas laiko eilutėms turinčioms tiek trendą, tiek sezoniškumą, šis metodas vadinamas Holto Winterso eksponentiniu glodinimu arba tiesiog dvigubu eksponentiniu glodinimu. Kiekvienas iš šių metodų gali būti taikomas tiek adityvių tiek multiplikatyvių elementų turinčioms laiko eilutėms.

(Bergmeir, Hyndman, & Benítez, 2016) aprašo bagging naudojimą eksponentiniam glodinimui su Boxo-Coxo transformacijomis *STL* dekompozicija ir judančių blokų bootstrap. Šis metodas gali pranokti paprastą eksponentinio glodinimo metodą ir yra kur kas efektyvesnis nei tradiciniai eksponentinio glodinimo metodai. Bootstrap vyksta tokia tvarka: pirmiausia laiko eilutei atliekama Boxo-Coxo transformacija, tada atliekama laiko eilutės dekompozicija į tendą, sezoniškumą ir triukšmą naudojant *STL* metodą. Tuomet padaroma sumaišyta triukšmo komponentės versija, šiam tikslui naudojamas „blokų bootstrap“, kur artimos laiko eilutės triukšmo reikšmės sumaišomos ir atsitiktine tvarka sujungiamos atgal. Tada šios triukšmo reikšmės prijungiamos atgal prie sezoniškumo ir trendo komponentių ir atliekama atvirkštinė Boxo-Coxo transformacija.

ARIMA metodai dar ne kartą buvo paminėti literatūros apžvalgoje. Jie su eksponentinio glodinimo metodais yra plačiausiai taikomi laiko eilučių prognozavimo praktikoje. Tuo tarpu kai eksponentinio glodinimo metodai yra grįsti duomenų trendo ir sezoniškumo parametrais, *ARIMA* modelių tikslas - apibūdinti duomenų imties autokoreliaciją. (Shumway & Stoffer, 2000) savo knygoje aiškiai atskleidžia *AR*, *MA*, *ARMA* ir *ARIMA* procesų savybes ir jas apibūdina. Autoregresiniai modeliai (*AR*) yra grįsti idėja, kad laiko eilutės ateities reikšmės gali būti paaiškintos tos pačios laiko eilutės praeities reikšmių funkcija. Slenkantis vidurkis (*MA*) modelyje parenka praeities klaidas ir jas pritaiko tolimesnei prognozei. Dėl to prieš kuriant *ARMA* modelį paprastai atsižvelgiama į autokoreliacijos ir dalinės autokoreliacijos grafikus.

Dažnai praktikoje prognozėmis užsiimančios duomenų analitikai linkę eksperimentuoti ir kurti hibridinius modelius, kurie apima keletą metodų ir juos kombinuoja atrinkdami kiekvieno metodo geriausias savybes. Praktiškai tai yra sunkiau įgyvendinama dėl programinių paketų trūkumo kombinuojant šias prognozes, visą programinį kodą tenka kurti patiems arba reikia susirinkti jį iš fragmentų.

Teleskopas (*angl. Telescope*) - taip savo naująjį hibridinį prognozavimo modelį pavadino (Züfle & Bauer, 2017). Autorių sukurtas kelių žingsnių į priekį prognozavimo modelis naudoja klasterizavimą kategorinių kintamųjų kūrimui (pvz. savaitgaliai ir darbo dienos). Taip pat naudojama *STL* dekompozicija išskirti laiko eilutę į tendą, sezoniškumą ir triukšmą. *ARIMA* naudojamas trendo prognozei, o sezoniškumo struktūra gauta panaudojus *STL*, paprasčiausiai atkartojama. Galiausiai naudojamas *XGBoost* metodas naudojantis tendą, sezoniškumą ir klasterius. Sukurtas modelis rodo gana neblogus rezultatus palyginus su pavieniais modeliais, taip pat skaičiavimo laikas yra trumpas, teigia autoriai.

(Kline, 2004) nagrinėjo nepriklausomą ir jungtinių dirbtinių neuroninių tinklų metodus, bei lygina šiuos metodus su iteraciniu ir sezoniškai pataisytu Naive metodais. Nepriklausomas metodas tai kai kiekvienam prognozės horizontui sukuriamas atskiras neuroninis tinklas, jungtinis – kai visiems prognozės horizontams naudojamas tas pats neuroninis tinklas. Tyrimo rezultatas parodo, kad

nepriklausomas dirbtinių neuroninių tinklų metodas rodo geriausius rezultatus esant trumpam prognozės horizontui. Horizontui didėjant autoriai pataria taikyti jungtini dirbtinių neuroninių tinklų metodą.

(Masum, Liu, & Chiverton, 2018) pristatinėja kelių žingsnių į priekį elektros apkrovos prognozavimą. Eksperimentiniai rezultatai su elektros apkrovos laiko eilutėmis parodo, kad *LSTM* (angl. *Long Short Term Memory*) modelis yra pranašesnis nei *ARIMA*. *LSTM* – tai speciali *RNN* (angl. *Recurrent Neural Network*) rūšis, kuri sugeba išmokti ilgo laikotarpio priklausomybes. Šis modelis pirmą kartą buvo pristatytas (Schmidhuber & Hochreiter, 1997), vėliau daug tyrėjų plėtojo ir tobulino šį modelį. *LSTM* sprendžia daug problemų ir naudojimo paskirtis yra gana plati, dėl to paskutiniu metu yra plačiai naudojamas. Šis modelis išskirtinai sukurtas dėl ilgo laikotarpio priklausomybių problemos. Jis puikiai tinka ilgą periodą turinčioms laiko eilutėms ir naudodamas mažai resursų gali įsiminti didelės duomenų imties struktūrą.

Netikėtas analizės ir modeliavimo rezultatas buvo gautas (Bouktif, Fiaz, Ouni, & Serhani, 2018). Autoriai sukūrė optimizuotą *LSTM-RNN* vienmatį modelį, skirtą prognozuoti elektros energijos suvartojimą tiek trumpam periodui (1-3 paroms į priekį), tiek ilgam periodui (kelioms savaitėms į priekį). Palyginimui buvo sukurti kiti 7 plačiai naudojami mašininio mokymosi modeliai skirti elektros energijos suvartojimo prognozavimui. Geriausiai prognozuojantis modelis buvo naudojamas kaip lyginamasis koeficientas (angl. *benchmark*). Hyper parametrų parinkimas taip pat buvo naudojamas kolektyvinio mašininio mokymosi prognozės tobulinimui. Ši technika netikėtai panaikino temperatūrą ir ketvirčio metų kategorinius kintamuosius kaip svarbius kintamuosius, daugiausia dėmesio ji skyrė istoriniams elektros energijos suvartojimo laiko atsilikimams (angl. *lags*). Taip yra dėl to, kad laikini suvartojimo parametrai savyje ir taip įterpia bei užfiksuoja orų ir savaitės tendencijų parametrus. Tiesa, autoriai pastebi, kad dažna problema yra tinkamai parinkti atsilikimo dydį.

1.4.1. Išoriniai regresoriai

Ankstesniame skyriuje apžvelgta literatūra daugiausia užsimena apie laiko eilučių prognozes pasitelkiant jų istorinius duomenis, tačiau į modelį neįtraukiant kitos informacijos, kuri taip pat gali būti reikšminga. Pavyzdžiui, atostogų įtaką, konkurentų veikla ar įstatymų pokyčiai gali paaiškinti didelę dalį istorinės variacijos ir padėti sukurti tikslesnę prognozę. Toliau apžvelgiama literatūra atskleidžia prognozavimo metodus, kurie suteikia galimybę panaudoti šią informaciją.

Kaip vieną iš pagrindinių silpnųjų, norint tiksliai prognozuoti elektros energijos kainą trumpam periodui į priekį, (Weron, 2014) įvardija informacijos trūkumą. Prognozei reikalinga informacija dažnai būna paviešinama tik po kelių valandų ar net kitą parą. Tai yra reikšmingas trikdys,

neleidžiantis gauti reikiamos informacijos reikiamu metu, ne tik prognozuojant elektros energijos kainą, bet ir vėjo elektrinių generaciją. Turėdami išorinę informaciją (pavyzdžiui, vėjo elektrinių faktinė generacija kaimyninėse valstybėse realiu laiku) galėtume ją taikyti kaip patikimą išorinį regresorių, kuris padėtų pasiekti didesnę tikslumą.

(Hyndman & Fan, Short-term load forecasting based on a semi-parametric additive model, 2010) atliktame tyrime aprašo, kad trumpalaikės elektros suvartojimo prognozės (iki 7 dienų horizonto ilgio) yra kur kas efektyvesnės taikant ex-post, nei ex-ante išorinius regresorius. Tai paaiškina, kad prognozuojamas valandinis suvartojimas vienai dienai į priekį galėtų būti kur kas tikslesnis, jei išoriniai temperatūros kintamieji būtų suprognozuoti tiksliau. Šis teiginys dar kartą primena, jog, siekiant didesnio tikslumo, labai svarbu prognozėse naudoti kuo naujesnę ir esamam laikui artimesnę informaciją, bei iš daugelio išorinių regresorių atrinkti tinkamiausius.

(Cui & Peng, 2015) savo tyrime, kuriame yra prognozuojamas elektros energijos suvartojimas trumpam laiko horizontui (iki 3 dienų į priekį), palygina *ARIMAX*, *ARMA* ir *AR* modelius. Tyrimas atskleidžia, kad vienmačiame *ARIMAX* modelyje teisingai pritaikyti išoriniai regresoriai duoda geresnę rezultatą nei *AR* ir *ARMA* modeliai, su kuriais šis buvo lyginamas.

Į vėjo greičio laiko eilutės prognozes įtraukti kuo daugiau prieinamų išorinių kintamųjų rekomenduoja (Cadenas, Rivera, Campos-Amezcuca, & Heard, 2016). Autorių darbe lyginami vieno žingsnio į priekį vėjo greičio prognozavimo modeliai *ARIMA* ir *NARX* (*angl. nonlinear autoregressive exogenous artificial neural network*) netiesinis autoregresinis išorinių kintamųjų neuroninių tinklų modelis. Rezultatas parodo, kad neuroninių tinklų modelis yra šiek tiek tikslesnis nei *ARIMA*, vieną iš tai lemiančių veiksnių autoriai įvardija papildomus išorinius kintamuosius, kurie buvo naudojami *NARX* modelyje.

(Hyndman & Athanasopoulos, 2014) taip pat aprašo *ARIMA* modelių sudarymą su papildomais išoriniais regresoriais. Paprastas *ARIMA* modelis į prognozę įtraukia tik prognozuojamojo kintamojo praeitį, tuo tarpu regresija prognozuojant kintamąjį y naudojami tik x_n regresoriais ir neįtraukia praeities reikšmių su atsilikimu. Taigi autorius aprašo *ARIMA* panaudojimą su išoriniais regresoriais, tiesa, visi išoriniai regresoriai kaip ir prognozuojama laiko eilutė turi būti stacionarūs. Dažnu atveju efektas kurį sukelia išorinis regresorius gali būti ne toks paprastas ir staigus, pavyzdžiui, reklamos kompanija gali daryti įtaką pardavimams kur kas vėliau nei pirmoji jos paleidimo diena. Šiuo atveju reikėtų įtraukti išorinius kintamuosius su atsilikimu (*angl. lagged predictor*).

Rob J Hyndman įvardija dirbtinius neuroninius tinklus kaip dar labiau pažengusį prognozavimo metodą. Paprasčiausi neuroniniai tinklai neturi paslėptų sluoksnių ir yra tolygūs tiesinei regresijai, tokiu atveju norint sutaupyti skaičiavimo resursų kur kas naudingiau yra naudoti tiesiog tiesinę regresiją. Kai tik yra įterpiamas papildomas sluoksnis, neuronų tinklas tampa nebe tiesinis tai yra

vaidinama daugiasluoksniu perdavimo į priekį tinklu (*angl. multilayer feed-forward network*) kai kiekvienas neuronų sluoksniu gauna informaciją iš prieš jį buvusio sluoksniu ir perduoda informaciją sekančiam sluoksniu. Žvelgiant į laiko eilutes, praėjusių periodų reikšmės su tam tikru atsilikimu taip pat gali būti perduodamos į neuroninių tinklų modelį taip atsiranda *NNAR* modelis (*angl. neural network autoregression*). Naudojant šį modelį prognozavimui procesas vyksta su daug iteracijų. Pirmajam žingsniui prognozuoti paprasčiausiai naudojami istoriniai duomenys, antrajam pirmojo žingsnio prognozė ir istoriniai duomenys, procesas vyksta ir toliau kol yra suskaičiuojami visi apibrėžti žingsniai į priekį. Tokiam neuroniniam tinklui papildomai galima įtraukti ir išorinius kintamuosius.

1.4.2. Kompleksinis sezoniškumas

Iki šiol apžvelgėme palyginti paprastas sezonines struktūras. Tačiau didesnio dažnio duomenys atskleidžia kur kas sudėtingesnę sezoninę struktūrą. Pavyzdžiui, paros duomenys gali turėti savaitinę ir metinę struktūrą. Valandiniai duomenys dažniausiai turi trijų tipų sezoniškumą: paros, savaitinį ir metinį. Tokia multisezoninė struktūra yra vis labiau pastebima didėjant duomenų dažniui. Multisezoniškumą turinčių didelio dažnio laiko eilučių pavyzdžiai galėtų būti skambučių kiekis skambučių centruose, vizitų į ligoninės priimamąjį skaičius, grynujų pinigų išėmimo operacijos, elektros energijos ir vandens vartojimas, prisijungimas į internetines svetaines.

(Livera, Hyndman, & Snyder, 2011) savo straipsnyje nagrinėja kompleksinį sezoniškumą ir aprašo trigonometrinių Boxo-Coxo transformacijų *ARMA* modelį *TBATS*. Šis modelis yra pranašesnis nagrinėjant tiesines ir netiesines laiko eilutes, kuriose susiduriama su multisezoniškumu. Remiantis *AIC* kriterijumi, taikant *TBATS* gaunami kur kas geresni rezultatai, nei naudojant tradicinius eksponentinio glodinimo metodus. Taip pat, autoriai pabrėžia, kad, norint gauti geresnį rezultatą, vertėtų pamėginti susikurti papildomus kintamuosius, žyminčius sezoną ar netipines dienas (pvz. mėnesių, ketvirčių, savaitgalių kintamieji).

(Gould, et al., 2008) savo straipsnyje pateikia analizę, kurioje yra lyginami trys metodai: Holto Winterso, Taylora dvigubo sezoniškumo ir Multisezoninis modelis, kurį autoriai patys išplėtojo. Autorių modeliuojamas Multisezoninis metodas nuo pirmųjų dviejų skiriasi tuo, kad sezoniniai reiškiniai, kurie atvaizduoja sezono ciklą, proceso metu yra atnaujinami daugiau nei vieną kartą. Autoriai savo straipsnyje visus modelius išbandė prognozuodami elektros energijos suvartojimą. Šiuo atveju, multisezoninis metodas leidžia pirmadienio sezoninius reiškinius atnaujinti antradienį. Kadangi savaitės darbo dienų suvartojimas yra labai panašus, tai leidžia dažniau atnaujinti sezoninius parametrus ir faktiniai duomenys, gauti už pirmadienį, suteikia galimybę koreguoti parametrus prognozuojant antradienio suvartojimą. Tyrimo metu modeliai buvo lyginami ne tik

prognozuojant elektros energijos suvartojimą, bet ir prognozuojant automobilių srautą. Duomenys turėjo aiškius paros ir savaitinius sezoniškumus. Multisezoniškumo modelis buvo kur kas tikslesnis nei Holto Winterso ir Taylor'o dvigubo sezoniškumo modeliai dėl savo lankstumo. Modelis kiekvienai dienai leidžia turėti savo valandinę struktūrą, arba turėti dienų su ta pačia struktūra. Taip pat, modelis suteikia galimybę skirtingų dienų struktūrai paveikti vienai kitą.

Vėjo greičio laiko eilučių analizei ir prognozei *TBATS* modelį taip pat pritaikė ir (Condeixa, Bastos, Oliveira, & Barbosa, 2017). Autoriai savo darbe mini, kad vėjo greičio laiko eilutės dekompozicijai *TBATS* modelis yra puikiai pritaikomas, todėl, kad tai yra vienas iš nedaugelio modelio galinčių laiko eilutę išskirti į kelis sezoniškumus.

1.4.3. Kolektyviniai modeliai

M – varžybos (arba Makridakis competitions) yra serija prognozavimo varžybų, kurias organizuoja prognozavimo tyrinėtojas Spyros Makridakis. Šių varžybų tikslas yra įvertinti ir palyginti įvairių prognozavimo metodų tikslumą. Varžybos prasidėjo 1982 metais, jos apima 24 skirtingas prognozavimo procedūras ir 3003 laiko eilutes. Laiko eilutės klasifikuojamos į mikroekonomikos (828), industrijos (519), makro ekonomikos (731), finansų (308), demografinės (413) ir kitas (204) kategorijas. Yra 5 skirtingos tikslumo (klaidos) metrikos: *SMAPE* (symmetric mean absolute percentage error), *MSAPE* (median symmetric absolute percentage error), vidutinis reitingavimas, *SAPE* (symmetric absolute percentage error) ir santykinė paklaida. Pastebima tendencija, kad paskutinių varžybų apžvalgoje ir analizėse ypatingas dėmesys skiriamas kolektyviniams modeliams.

Žvelgiant į literatūrą, detaliau aprašančią prognozavimo metodus, gausu informacijos apie kasmet organizuojamas prognozavimo varžybas ir jų rezultatų išvadas (Makridakis & Hibon, *The M3-Competition: results, conclusions and implications*, 2000). Tarptautiniame prognozavimo žurnale viešinamas šių autorių straipsnis, kuriame jie teigia, kad M3 varžybų metu prognozės, gautos naudojant vienmačius laiko eilučių modelius, beveik visada, t.y. išskyrus metinio dažnumo duomenis, yra tikslesnės nei naudojant daugiamačius, kada į modelį įtraukiamos papildomos laiko eilutės kaip nepriklausomi kintamieji. Lyginant vienmačius modelius tarpusavyje, sudėtingesnių metodų taikymas nebūtinai duoda tikslesnes prognozes. Kolektyvo, susidedančio iš vienmačių laiko eilučių modelių, tikslumas dažnai yra geresnis nei kiekvieno modelio atskirai. Prognozavimo tikslumas priklauso tiek nuo prognozuojamo horizonto ilgio, tiek nuo naudojamos klaidos metrikos. Pirmiausia yra prognozuojami ketvirtinio arba mėnesinio periodiškumo duomenys, vėliau prognozės yra agreguojamos iki metinio periodiškumo. Taip yra gaunamos tikslesnės prognozės nei iškart prognozuojant metinio dažnumo duomenis.

Kombinuojant keleto modelių prognozes, paprasto vidurkio strategija, kai modeliams skiriami lygiaverčiai svoriai, dažnai pranoksta sudėtingesnes strategijas, kai parenkami unikalūs svoriai, priklausomai nuo modelio tikslumo testuojant ant paskutinės dinamikos. Tuo tarpu praktikoje norint, kad vienas modelis nesugadintų visos kolektyvo prognozės, vertėtų taikyti išskirtims atsparius vidurkio variantus arba medianą, teigia (Hendry & Clements, 2004).

(Kolassa, 2011) savo publikacijoje pritaria anksčiau minėtų autorių nuomonei, kad, dažnais atvejais, naudojant modelių kolektyvus yra pasiekiamas geresnis rezultatas nei naudojant atskirus pagal *AIC* ar *BIC* kriterijus išrinkus pavienius geriausius modelius. Autorius savo straipsnyje teigia, kad atmetant tuos modelius, kuriuos pranoko geriausias modelis, yra prarandama didelė dalis aiškinamąją galią turinčios informacijos. Taip gali nutikti dėl to, kad vienas modelis yra per paprastas ir nesugeba užfiksuoti proceso esmės, arba, priešingai, vienas modelis yra labai sudėtingas ir linkęs persimokyti.

Paskutinėse M4 prognozavimo varžybose puikiai pasirodęs ir pirmąją vietą užėmęs „Uber“ atstovas, kurio prognozės buvo 9 % tikslesnės nei artimiausių varžovų, savo prognozei naudojo modelių kolektyvus (kaip ir kiti 12 iš 17 tiksliausių prognozavimo sprendimų). Šis modelis apjungė statistinius ir mašininio mokymosi metodus, taip pat pasiteisino statistinių modelių kolektyvo svorių derinimas, pasitelkiant mašininį mokymą (Makridakis, Spiliotis, & Assimakopoulos, The M4 Competition: Results, findings, conclusion and way forward, 2018). Kolektyvinio modelio rezultatai dar kartą patvirtina, kad tai yra šiuolaikinis sprendimas ir tendencija modelius sudarinėti iš kolektyvų auga.

1.4.4. Prognozavimo problemos

Natūralu, kad prognozuojant iškyla įvairių, nestandartinių problemų, keletą iš jų išskiria (Hyndman & Athanasopoulos, 2014). Autoriai mini, kad dažnai koreliacija yra maišoma su veikimu (*angl. causation*) ar veikimas su prognozavimu. Kintamasis x gali būti naudingas prognozuojant kintamąjį y , tačiau tai nereiškia, kad x veikia y . Žinoma, yra realu, kad x veikia y , bet galbūt y veikia x arba ryšys tarp šių kintamųjų yra kur kas sudėtingesnis nei paprastas veikimas. Svarbu suprasti, kad koreliacija yra dažnai naudinga prognozavimui, net tuomet, kai tarp dviejų kintamųjų nėra veikimo ryšio, ar koreliacija veikia priešinga kryptimi nei modelis. Kad ir kaip bebūtų, dažnu atveju gaunamas tikslesnis modelis, kai veikimo principai gali būti aiškiai nustatyti.

Taip pat prognozavimo uždaviniuose pasitaiko sunkūs (*angl. confounded*) kintamieji. Kaip pavyzdį sunkiems kintamiesiems galima paimti atvejį kai prognozuojami pardavimai vienai kompanijai. Nagrinėjant išorinių kintamųjų istorinius duomenis matoma, kad tam tikru metu praeityje bendrovė įgijo konkurentus, lygiai tuo pat metu prasidėjo ekonominis nuosmukis. Sudarant prognozę ir

įtraukiant šiuos kintamuosius į pardavimų prognozavimo modelį tampa labai sunku atskirti šių dviejų veiksnių poveikius kompanijos pardavimams. Prognozavimo modeliui sunkūs kintamieji didelės įtakos nedaro, nebent yra prognozuojami scenarijai, nes prognozuojant scenarijus turi būti atsižvelgiama į regresorių sąveiką, taip pat tai sukelia sunkumų atliekant duomenų analizę.

Atlikta mokslinės literatūros analizė leidžia teigti, kad laiko eilučių analizei ir prognozėms yra paskirta nemažai svarbių akademinų darbų. Senesni straipsniai ir knygos linkę apibrėžti klasikinę prognozavimo teoriją ar strategijas, tuo tarpu naujesnė literatūra dažnai analizuoja galimybes kombinuoti įvairius metodus ir kurti naujus modelius. Didelė dalis naujesnių mokslinių leidinių skiria dėmesio kolektyvinių modelių kūrimui, kuriuose atskleidžiama, kad modelių kolektyvas yra gana paprastas būdas siekiant didesnio tikslumo. Apžvalga taip pat atskleidžia, kad multisezoniškumas – dažnas didelio tankio duomenų reiškiny. Prognozuojant visuomet verta patikrinti kuo daugiau prieinamų išorinių regresorių, kurie dažnu atveju padeda pagerinti rezultatą. Pasitaiko, kad panašūs tyrimai ar apžvalgos prieina prie išvalgų, kurios turi skirtumų. Kartais autoriai bando įrodyti anksčiau iškeltas išvalgas ar teorijas, tačiau rezultatas būna geresnis nei tikėtasi. Tarp nagrinėtos literatūros jaučiamas medžiagos, skirtos daug žingsnių į priekį prognozavimui, kai yra didelis duomenų dažnumas ir yra multisezoniškumas, trūkumas. Taigi naudojantis galimybe galima papildyti šio tipo analizes ir plačiau išnagrinėti šią sritį. Analizuota akademinė literatūra turi didelę vertę ir yra svarbus pagrindas, kuris bus naudojamas šiame magistrinio darbo tyrime.

2. Medžiagos ir tyrimų metodai

Šiame skyriuje aprašomi tyrimų metodai, kurie buvo naudojami ištirti VE generacijos laiko eilutės ir atlikti jų prognozavimą. Aprašomi metodai yra atrinkti apžvelgus literatūrinius šaltinius ir ekspertiškai įvertinus (pagal laiko eilutės ir prognozavimo uždavinio tikslo specifiškumą) galimai tinkamiausius metodus.

2.1. Laiko eilučių analizė

(Tuftė, 1983), „Laiko eilutės grafikas – dažniausiai naudojama grafinio dizaino forma. Viena dimensija, žengianti paskui reguliarių sekundžių, minučių, valandų, dienų, savaitių mėnesių ar metų natūralų išsidėstymą laike, duoda tokios projekcijos interpretavimui stiprumo ir naudingumo daugiau nei bet kuris kitas grafinis metodas.“ Taigi pradedant laiko eilutės analizę svarbiausia dalis yra grafiškai atvaizduoti istorinius duomenis, kurie duos stiprų pagrindą tolimesnei laiko eilutės analizei.

Tinkamas laiko eilutės modelis gali būti paruoštas:

- 1) Suprantant slypinčius veiksnius ir struktūras, kurios nulėmė tokį nagrinėjamos laiko eilutės išsidėstymą istoriškai.
- 2) Pradedant apmokytą modelį ir remiantis jo prognozėmis, prognozių analizėmis, stebėjimu ir klaidų matavimu pakartotinai sudarant modelį.

Toliau apžvelgiamas stacionarumas, nestacionarumas, kaip identifikuojant laiko eilutės autokoreliaciją galima nuspręsti ar laiko eilutė yra stacionari. Kaip nestacionarią laiko eilutę, ją diferencijuojant, galima paversti stacionaria. Apžvelgiama laiko eilutės dekompozicija ir jos specifiškumai. Žinant laiko eilutę apibūdinantį procesą, vėliau pagal tai galima pritaikyti atitinkamą prognozavimo modelį.

2.1.1. Stacionarumas

Laiko eilučių analizėje dažniausiai naudojami 2 apibrėžimai apibūdinant ir tikrinant proceso stacionarumą - stacionarumas siaurąją ir plačiąją prasme. Antrasis apibrėžimas – stacionarumas plačiąją prasme, praktikoje naudojamas dažniau, nes apima ir pirmąjį, tačiau apžvelkime juos abu:

1. Procesas ξ_t vadinamas stacionariu siaurąja prasme, jei jo daugiamačiai pasiskirstymai nepriklauso nuo postūmio laike, t. y.:

$$F_{t_1, \dots, t_k}(\cdot) = F_{t_1 + \tau, \dots, t_k + \tau}(\cdot), \forall t_1, \dots, t_k \in T, k = 1, 2, \dots, \text{jei } t_i + \tau \in T, \quad (1)$$

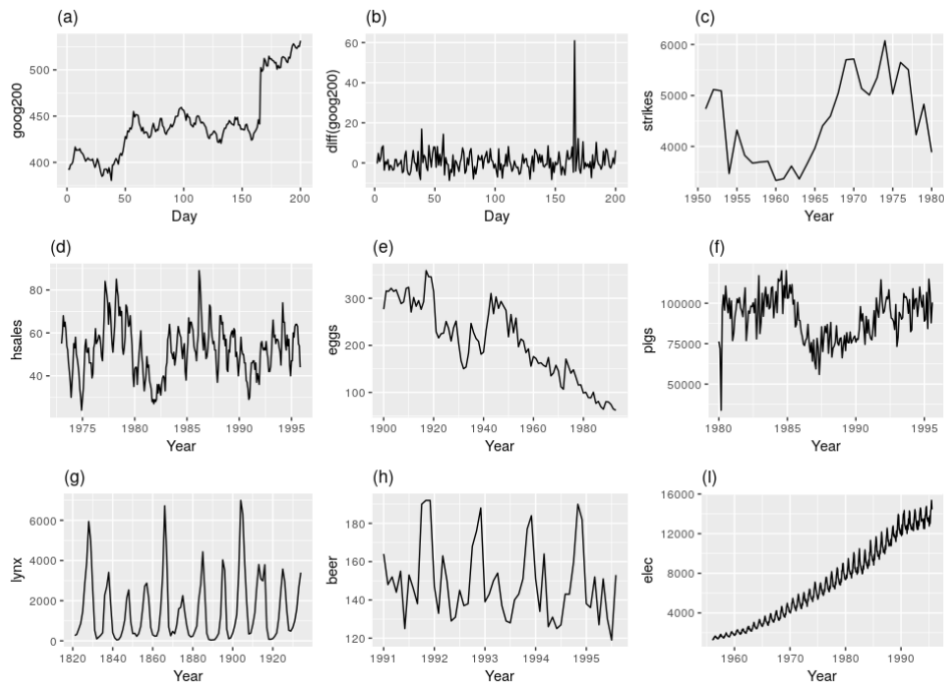
čia $F_{t_1, \dots, t_k}(\cdot)$ yra atsitiktinių vektorių $(\xi_{t_1}, \dots, \xi_{t_k})$ pasiskirstymo funkcijos.

2. Procesas ξ_t vadinamas stacionariu plačiaja prasme, jei jo matematinis vidurkis ir kovariacinė funkcija nepriklauso nuo poslinkio laike, t. y. jei:

$$m(t) = m(0), R(t, s) = R(t - s, 0), \forall t, s \in T \quad (2)$$

Stacionarios laiko eilutės parametrai (vidurkis, dispersija ir kovariacinė funkcija) nepriklauso nuo laiko. Taigi laiko eilutės turinčios tendrą nėra stacionarios. Trendas paveiks laiko eilutės reikšmes vienu ar kitu laiko momentu. Balto triukšmo laiko eilutės yra stacionarios – nesvarbu kuriuo laiko metu atlikti stebėjimai, jie bus labai panašūs bet kuriuo laiko momentu.

Apibendrinant, stacionarios laiko eilutės neturi nuspėjamo šablono ilgame laiko periode. Grafiškai atvaizduotos laiko eilutės gali būti daugiau mažiau horizontalios (vis dėlto cikliškumas gali būti pastebėtas) su pastovia dispersija. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Vienas iš paprasčiausių būdų nustatyti stacionarumui yra laiko eilutės vizualizacija.



2 pav. (a) Google akcijų kaina 200 dienų; (b) Dienos pokytis Google akcijų kainai 200 dienų; (c) Kasmetinis žaibų skaičius JAV; (d) Mėnesiniai vienos šeimos tipo namų pardavimai JAV; (e) Kasmetė tuzino kiaušinių kaina JAV (Amerikos doleriais); (f) Mėnesinis skerdziamų kiaulių skaičius Australijoje (Victoria); (g) Kasmetinis į spąstus pakliuvusių lūšių skaičius Kanadoje (McKenzie upės regione); (h) Mėnesinė australiško alaus produkcija; (i) mėnesinė Australijos elektros energijos gamyba.

(Hyndman & Athanasopoulos, 2014)

Iš 2 pateiktų laiko eilučių vizualizacijos galime matyti, kad stacionarios yra tik (b) ir (g). Norint patikrinti laiko eilutės stacionarumą detaliau, atliekami statistiniai vienietinės šaknies testai (*angl. Unit Root Test*). Matematinis šių testų paaiškinimas:

Tarkime, kad turime laiko eilutę:

$$y_t = a \cdot y_{t-1} + \varepsilon_t \quad (3)$$

kur y_t yra reikšmė laiko momentu t ir ε_t yra paklaida (arba triukšmas). Norint apskaičiuoti y_t mums reikia y_{t-1} reikšmės, kuri yra lygi:

$$y_{t-1} = a \cdot y_{t-2} + \varepsilon_{t-1} \quad (4)$$

Jei skaičiuosime visiems stebėjimams tai y_t reikšmė bus:

$$y_t = a^n \cdot y_{t-n} + \sum_{i=1}^n \varepsilon_{t-i} \cdot a^i \quad (5)$$

Jei a reikšmės (5) lygtyje bus lygios vienetui, tuomet prognozės bus lygios y_{t-n} ir visų klaidų nuo $t-n$ iki t sumai. Tai reikštų, kad laikui bėgant dispersija augtų. Tai apibrėžiama kaip vienetinė šaknis laiko eilutėse. Žinome, kad stacionariose laiko eilutėse dispersija negali būti funkcija nuo laiko. Vienetinės šaknies testai patikrina ar $a=1$. Praktikoje naudojami 2 pagrindiniai vienetinės šaknies testai stacionarumo nustatymui:

ADF (Augmented Dickey Fuller) testas (Fuller, 1976)

Šis testas tikrina ar laiko eilutėje yra vienetinių šaknų išskeldamas hipotezę: Laiko eilutė turi vienetinių šaknų ($a=1$). Alternatyva šiai hipotezei - laiko eilutė vienetinių šaknų neturi.

Jei hipotezės neatmetame, galime teigti, kad laiko eilutė yra nestacionari.

KPSS (Kwiatkowski-Phillips-Schmidt-Shin) testas (Kwiatkowski, Phillips, Schmidt, & Shin, 1992)

KPSS testo pagrindinė ir alternatyvioji hipotezės priešingos ADF testui. Taigi, pagrindinė hipotezė - laiko eilutė yra stacionari ir neturi vienetinių šaknų, o jos alternatyva, kad laiko eilutė nestacionari ir egzistuoja vienetinės šaknys.

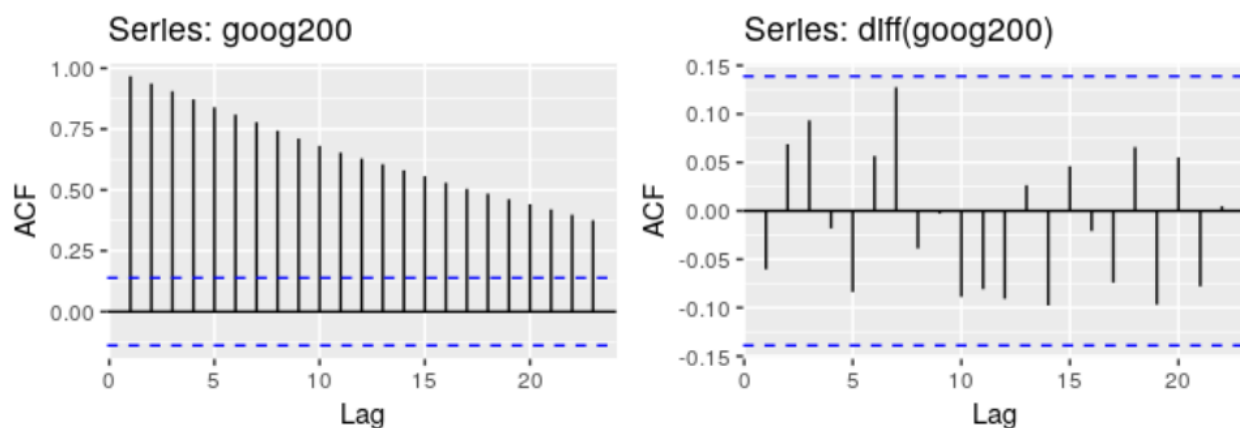
Literatūroje rekomenduojama atlikti abu šiuos testus norint įsitikinti ar laiko eilutės yra stacionarios. Gavus skirtingus rezultatus rekomenduojama diferencijuoti naikinant stacionarumą, kol abiejų testų rezultatai sutampa.

Diferencijavimas yra nesudėtingas būdas naudojamas laiko eilutes iš nestacionarių paversti į stacionarias. Matematiškai jis užrašomas taip:

$$y_t' = y_t - y_{(t-1)} \quad (6)$$

Gaunamas skirtumas tarp dviejų laike gretimų reikšmių. Transformacijos, pavyzdžiui, logaritminės, gali padėti stabilizuoti laiko eilutės dispersiją. Diferencijavimas šiuo atveju padeda stabilizuoti laiko eilutės vidurkį panaikinant (ar sumažinant) laiko eilutės tendą ir tam tikrais atvejais - sezoniškumą.

Nustatinėjant stacionarumą, autokoreliacijos funkcijos grafikai (toliau – *ACF*) gali būti nemažiau naudingi, nei laiko eilutės istorinių duomenų atvaizdavimas. Stacionariose laiko eilutėse *ACF* grafikai greitai nukrenta iki nulio, kai, tuo tarpu, nestacionarių laiko eilučių *ACF* grafikai mažėja lėčiau ir palaipsniui. Taip pat, nestacionarių laiko eilučių *ACF* grafiko reikšmės dažnai didelės ir teigiamos (Hyndman & Athanasopoulos, 2014).

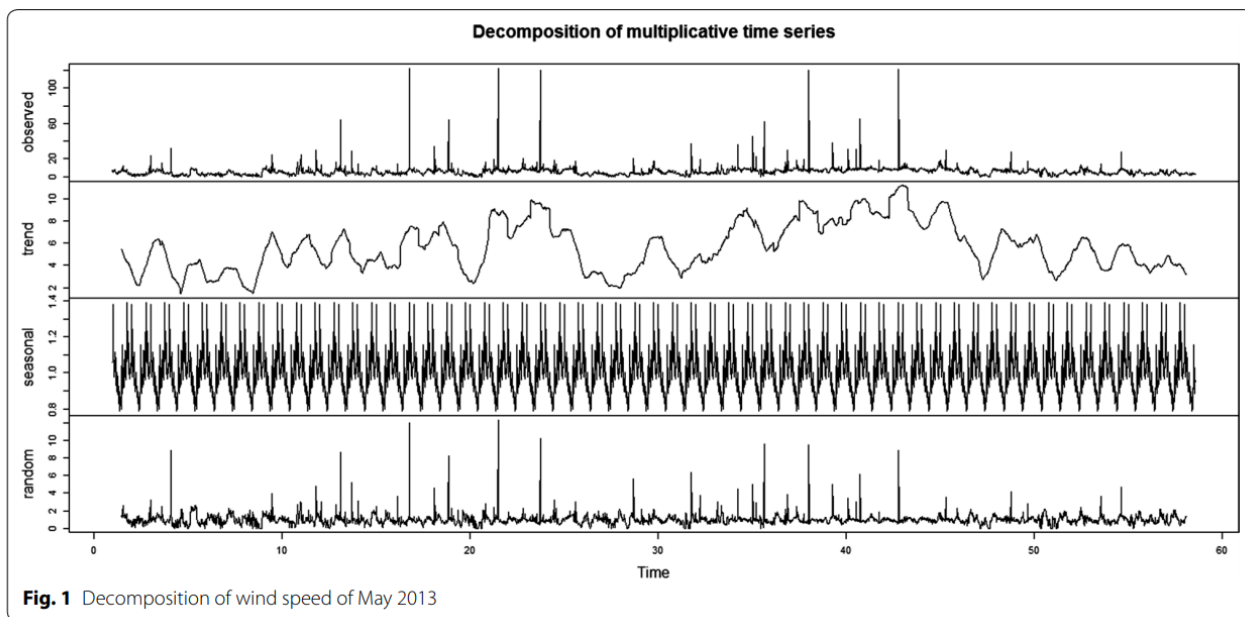


3 pav. Google akcijų dienos akcijų kainos *ACF* (kairėje) ir Google akcijų kainos pokyčių *ACF* (dešinėje).
(Hyndman & Athanasopoulos, 2014)

Taip pat verta paminėti, kad *ACF* ir *PACF* (dalinė autokoreliacijos funkcija) taip pat plačiai naudojamos parenkant *ARMA* modelį, apie kurį aprašoma laiko eilučių prognozavimo skyrelyje.

2.1.2. Dekompozicija

Laiko eilutės turi tris pagrindines komponentes – trendo, sezoninę ir atsitiktinio triukšmo. Tiesinis ar netiesinis trendas atspindi ilgalaikes raidos tendencijas ir gali būti kylantis arba besileidžiantis. Sezoniškumas parodo sisteminių duomenų atsikartojimą keičiantis ir besikartojant sezonams. Atsitiktinio triukšmo komponentė – stacionarus procesas, kuris vyksta atsitiktinai ir pažymi atsitiktinius svyravimus, kurie negali būti paaiškinti ankstesnėmis dvejomis komponentėmis.



4 pav. Multiplikatyviosios vėjo greičio laiko eilutės dekompozicija (2013 metų gegužės mėnuo).

(Rao & Uma, 2015)

4 pavaizduota originalios vėjo greičio eilutės (viršutinis grafikas) dekompozicija. Ši laiko eilutė išskaidyta į trendo (antras grafikas) sezoniškumo (trečias grafikas) ir atsitiktinio triukšmo (ketvirtas grafikas) komponentes (Rao & Uma, 2015).

Praktikoje dažnai naudojamas matematinis modelis išreiškiantis adityvias laiko eilutės komponentes:

$$X_t = \mu_t + s_t + \xi_t, \quad E \xi_t = 0 \quad (7)$$

Čia X_t — stebimas atsitiktinis procesas, μ_t — trendas, t. y. glodi determinuota funkcija, atspindinti X_t kitimo ilgalaikes tendencijas, s_t — sezoniškumas, ξ_t — stacionarus procesas, parodantis atsitiktinius svyravimus. Sezoniškumas s_t yra determinuota periodinė funkcija.

Tuo tarpu multiplikatyvusis modelis atrodo taip:

$$X_t = \mu_t \cdot s_t \cdot \xi_t \quad (8)$$

Adityvusis modelis naudojamas kai sezoniniai svyravimai kinta skirtingu lygiu, nepriklausomai nuo laiko eilutės ir jos vidurkio. Priešingu atveju, jei sezoniniai svyravimai proporcingi vidurkio kitimui multiplikatyvaus modelio taikymas yra tinkamesnis (Chatfield, 2000).

2.2. Laiko eilučių prognozavimas

Prieš pradėdant prognozuoti VE generaciją svarbu pasirinkti tinkamiausius modelius, kuriuos vėliau galima apjungti ir bandyti atlikti dar tikslesnę prognozę. Praktikoje nėra vieno modelio, kuris yra tiksliausiai prognozuojantis visais atvejais t.y. pritaikytas visoms laiko eilutėms. Tą sako ir teorema „Nemokamų pietų nebūna“ (*angl. No Free Lunch Theorem*), kurią išskėlė (Macready & G., 1997). Teorema sako, kad jei yra bent vienas algoritmas A, kuris veikia geriau, nei algoritmas B, tuomet turi egzistuoti kitos funkcijos, kur algoritmas B pralenkia algoritmą A. Jei ši teorema pritaikoma renkantis geriausią modelį prognozei, tai reiškia, kad kiekvienas modelis yra pats tinkamiausias bent vienam duomenų rinkiniui iš visų duomenų rinkinių aibės. Atvirkščiai, nėra vienintelio modelio, kuris būtų tinkamiausias visiems duomenų rinkiniams.

Darbe naudojama 17 modelių, su kuriais sudaromos prognozės ir bandoma jas sujungti į kolektyvą. Pagal metodų sudėtingumą ir prognozės sudarymo principą modeliai suskirstyti į 5 pagrindines grupes:

Paprastieji, Regresiniai, *AR(I)MA*, Eksponentinio glodinimo, Sudėtingieji.

2 lentelė. Prognozavimo metodai suskirstyti į 5 grupes.

Metodo grupė	Metodo pavadinimas lietuvių k.	Metodo pavadinimas anglų k.
Paprastieji metodai	Vidurkio	Mean
	Atsitiktinio žingsnio (Naivusis)	Random Walk (Naïve)
	Atsitiktinio žingsnio su dreifu (Naivusis)	Random Walk with Drift (Naïve)
	Sezoninis naivusis	Seasonal Naïve
Regresiniai metodai	Tiesinė regresija	Linear Regression
	Tiesinė regresija su sezoniniais kategoriniais kintamaisiais	Linear Regression with Seasonal Dummies
	Kvadratinio trendo	Quadratic Trend

	Kvadratinio trendo su sezoniniais kategoriniais kintamaisiais	Quadratic trend with Seasonal Dummies
<i>AR(I)MA</i> metodai	<i>ARMA</i>	<i>ARMA</i>
	<i>ARIMA</i>	<i>ARIMA</i>
	<i>ARIMAX</i>	<i>ARIMAX</i>
	<i>SARIMA</i>	<i>SARIMA</i>
	<i>SARIMAX</i>	<i>SARIMAX</i>
Eksponentinio glodinimo metodai	Paprastasis eksponentinis glodinimas	Simple Exponential Smoothing
	Dvigubas eksponentinis glodinimas (Holt-Winters)	Double Exponential Smoothing (Holt-Winters)
	Trigubas eksponentinis glodinimas (Holt-Winters Seasonal)	Triple Exponential Smoothing (Holt-Winters Seasonal)
Sudėtingi metodai	<i>BATS</i> (Eksponentinio glodinimo su Boxo-Coxo transformacija, <i>ARMA</i> paklaidomis, Trendo ir Sezoniškumo komponentėmis)	<i>BATS</i> (Exponential smoothing state space model with Box-Cox transformation, <i>ARMA</i> errors, Trend and Seasonal components)
	<i>TBATS</i> (Trigonometrinis Eksponentinio glodinimo su Boxo-Coxo	<i>TBATS</i> (Trigonometric Exponential

	transformacija, <i>ARMA</i> paklaidomis, Tendo ir Sezoniskumo komponentėmis)	smoothing state space model with Box-Cox transformation, <i>ARMA</i> errors, Trend and Seasonal components)
	Neuroniniai tinklai	Neural Network

2.2.1. Paprastieji metodai

Paprastieji prognozavimo metodai yra stebėtinai paprasti, tačiau kai kuriais atvejais gali būti labai veiksmingi. Toliau apžvelgsime juos po vieną.

- 1) **Vidurkio** metodas ateities reikšmes prognozuoja pagal visų praeities duomenų vidurkį:

$$Y_{t+1} = \frac{Y_1 + \dots + Y_t}{T} \quad (9)$$

- 2) **Naivusis** arba **atsitiktinio žingsnio** metodas remiasi paprasta idėja – kito žingsnio prognozė yra paskutinis užfiksuotas stebėjimas. Šis metodas dažnai naudojamas kaip lyginamasis indeksas (*angl. benchmark*) su kitais, sudėtingesniais ir labiau pažengusiais modeliais. Metodo formulė, kur ε_t yra baltas triukšmas:

$$Y_t = Y_{t+1} + \varepsilon_t \quad (10)$$

- 3) **Naivusis su dreifu** arba **atsitiktinio žingsnio su dreifu** metodas - tai naiviojo metodo variacija. Dreifo didinimas arba mažinimas susijęs su laiko eilutės trendu. Atliekant prognozavimą šiuo metodu atsižvelgiama į istorinį duomenų vidurkio didėjimą ir mažėjimą. Formulė išreiškianti prognozę:

$$Y_{T+h|T} = Y_t + \frac{h}{T-1} \sum_{t=2}^T (Y_t - Y_{t-1}) = Y_t + h \left(\frac{Y_t - Y_1}{T-1} \right) \quad (11)$$

- 4) **Sezoninis naivusis** metodas yra naudojamas duomenims, turintiems aiškų sezoniskumą. Kito žingsnio reikšmė yra prognozuojama pagal analogišką praėjusio sezono žingsnį. Matematinė išraiška, kur m nusako sezoną, atrodo taip:

$$Y_{T+h|T} = Y_{T+h-m(k+1)} \quad (12)$$

2.2.2. Regresiniai metodai

Pagrindinė regresinių metodų idėja yra prognozuojamojo kintamojo y ryšys su kitais kintamaisiais x arba su savo paties atsilikimu y_{t-n} laike. Savo paties atsilikimas plačiau taikomas ARMA procesuose. Detalesnė informacija pateikiama skyrelyje 2.2.3. Dažniausiai regresiniame modelyje ryšys yra tiesinis, tačiau priklausomybių gali pasitaikyti įvairių. Norint, kad šis modelis būtų efektyvus, visuomet reikia patikrinti, ar modelio liekanos (paklaidos) autokoreliuoja. Paprastai liekanų autokoreliacijos neturėtų būti (Hyndman & Athanasopoulos, 2014). Toliau apžvelgiami darbe taikomi regresijos metodai.

- 1) **Tiesinė regresija**, kaip nurodo jos pavadinimas, apibūdina tiesinį ryšį tarp prognozuojamojo kintamojo ir regresoriaus. Jos matematinė išraiška:

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (13)$$

čia, koeficientai β_0, β_1 yra apskaičiuojamos konstantos, x_t – regresorius ir ε_t atsitiktinė paklaida.

- 2) **Tiesinė regresija su sezoniniais kategoriniais kintamaisiais** - tai tas pats **tiesinės regresijos** metodas, tik kiekvienam sezonui sukuriame kategorinį kintamąjį s , kuris turi tam tikrą koeficientą. Šiuo atveju, kai turime daugiau kintamųjų, kuriems priskiriami koeficientai, regresiją galima vadinti daugiamate. Jos išraiška:

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 s_{1,t} + \dots + \beta_k s_{k,t} + \varepsilon_t \quad (14)$$

- 3) **Kvadratinio trendo** regresija gaunama į formulę įtraukiant regresoriaus kvadratą, paliekant ir pirmojo laipsnio regresorių:

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{1,t}^2 + \varepsilon_t \quad (15)$$

- 4) **Kvadratinio trendo regresija su sezoniniais kategoriniais kintamaisiais** - tai padidintos regresoriaus įtakos metodas, kuris jungia sezoniškumui atspindėti sukurtus kategorinius kintamuosius.

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{1,t}^2 + \beta_2 s_{1,t} + \dots + \beta_k s_{k,t} + \varepsilon_t \quad (16)$$

- 5) **Loess regresija** tai neparametrinė technika, kuri naudoja vietinę pasvertą regresiją norint gauti sklaidos diagramos taškus apibūdinančią suglodytą kreivę. Loess kreivės gali atskleisti trendą ir ciklus duomenyse, kai juos aptikti parametriniam modeliui gali būti sudėtinga. Loess kreivėms gauti naudojamas glodinimo parametras, kuris nusako Loess regresijos funkcijos lankstumą. Kuo didesnis glodinimo parametras tuo Loess regresija mažiau jautri staigiems duomenų kitimams. Loess regresijos trūkumas – ji neturi lengvai matematine forma išreiškiamos regresinės funkcijos. Todėl rezultatą tyrimuose interpretuoti gali būti sunku. Šios regresijos privalumas – ji reikalauja vienintelio parametro, glodinimo koeficiento. Kitas privalumas tai, kad ji puikiai tinka didelio dažnio duomenimis, ir kuo duomenų daugiau, tuo geresnis rezultatas yra gaunamas (Fox, 2010).

2.2.3. ARMA metodai

ARMA modeliai yra kuriami duomenų autokoreliacijos pagrindu. Klasikinio leidinio, išpopuliarinusio *ARMA* metodą, autoriai yra Box, G. E. P., ir Jenkins, G. M. (Box & Jenkins, 1970). *ARMA* metodai yra paremti dviem atskirais procesais: *AR* – autoregresijos procesu ir *MA* – slenkančio vidurkio procesu. *AR* procesas primena daugiamatės regresijos modelį, tik vietoje išorinių regresorių yra naudojamos praeities reikšmės su skirtingu atsilikimo žingsniu:

$$Y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (17)$$

čia p yra atsilikimo eilė.

Vietoj praeities reikšmių, kurios įtraukiamos į regresiją, *MA* modelis naudoja praeities prognozių klaidas išraiškoje, kuri yra panaši į regresiją, tačiau tai nėra regresija:

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (18)$$

čia, ε_t yra baltas triukšmas. Galima pastebėti, kad kiekviena Y_t reikšmė yra svertinis slenkantis praeities klaidų vidurkis.

Jeigu sujungsime diferencijavimą, aprašytą 2.1.1 skyrelyje, su *AR* ir *MA* modeliais, gausime *ARIMA* modelį, kurio akronimas yra auto regresinis integruotas slenkantis vidurkis (*angl. Autoregressive Integrated Moving Average*). Šiame kontekste integravimas reiškia priešingą veiksmą diferencijavimui. Jei diferencijavimo atlikti nereikėjo, nes laiko eilutė jau buvo stacionari, gaunamas tiesiog *ARMA* modelis, kurio išraiška:

$$Y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (19)$$

ARIMA modeliai turi p , d ir q parametrus, kurie atitinkamai nusako kiekvienos modelio dalies *AR*, I ir *MA* eiles. Autoregresinei ir slenkančio vidurkio daliai eilės gali būti nustatytos iš *ACF* ir *PACF* grafikų. Kaip nustatoma I yra aprašyta 2.1.1 skyrelyje.

Kaip jau buvo minėta, *ARMA* modelius išpopuliarino Box ir Jenkins. Šių autorių vardu yra pavadinta metodologija, skirta tirti laiko eilutėms. Boxo-Jenkinso metodologiją sudaro keli etapai:

- 1) integruotumo eilės nustatymas ir diferencijavimas arba trendo pašalinimas;
- 2) stacionaraus proceso modelio eilės nustatymas;
- 3) modelio parametrų įvertinimas. *AR* parametrus galime įvertinti mažiausių kvadratų metodu, *MA* parametrus įvertinti yra kiek sudėtingiau. Tam atlikti dažniausiai yra naudojamas didžiausio tikėtimumo metodas. Tuo tarpu *ARMA* modeliui programiniuose paketuose dažnai naudojamas *AIC* kriterijus (*angl. Akaike's Information Criterion*) arba pataisytas *AICc* kriterijus:

$$AIC = -2 \log(L) + 2(p + q + k + 1) \quad (20)$$

čia, L – duomenų tikėtimumas, $k = 1$ jei $c \neq 0$ ir $k = 0$, jei $c = 0$.

$$AICc = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2} \quad (21)$$

- 4) modelio patvirtinimas (*angl. validation*). Prognozės liekanos ε_t turi būti nekoreliuotos. Patikrinimui yra naudojamas Ljungo – Boxo testas, kuris tikrina nulinę hipotezę, kad liekanos yra nekoreliuotos.
- 5) modelis taikomas prognozei.

SARMA modelis - tai sezoninis *ARIMA* modelis, kuriam įvedami papildomi sezoniniai parametrai, kad šis galėtų tvarkytis su sezoniškumu. Gaunamas modelis $ARMA(p, d, q) (P, D, Q)_m$. Verta paminėti, kad *ARMA* procesų išraiškų suvedimas į operatorines palengvina darbą:

$$\begin{array}{ccccc} (1 - \phi_1 B - \dots - \phi_p B^p) & (1 - B)^d y_t & = & c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \\ \uparrow & \uparrow & & \uparrow \\ AR(p) & d \text{ diferencijavimas} & & MA(q) \end{array}$$

pav. 5 *ARMA* modelio operatorinė išraiška (Hyndman & Athanasopoulos, 2014) Taigi, *SARMA* modelį užrašius operatorine išraiška gauname tokią lygtį:

$$(1 - \phi_1 B)(1 - \phi_1 B^m)(1 - B)(1 - B^m)y_t = (1 + \theta_1 B)(1 + \theta_1 B^m)\varepsilon_t \quad (22)$$

čia, m – sezono parametras.

2.2.4. Eksponentinio glodinimo metodai

Holt, C. E. (1957) vienas pirmųjų pasiūlė eksponentinio glodinimo metodus. Šie metodai remiasi tuo, kad ateities reikšmės yra svertinis praeities stebėjimų vidurkis, kai link tolimesnės praeities stebėjimų koeficientai mažėja eksponentiškai. Kitaip sakant, kuo stebėjimas yra naujesnis, tuo didesnę svorį jis įgauna. Toliau aptarsime tris eksponentinio glodinimo metodus.

- 1) **Paprastasis eksponentinis glodinimas** naudojamas laiko eilutėms, neturinčioms trendo ir sezoniškumo. Jo matematinė išraiška:

$$Y_{T+1|T} = \alpha Y_T + \alpha(1 - \alpha)Y_{T-1} + \alpha(1 - \alpha)^2 Y_{T-2} + \dots \quad (23)$$

čia, $0 \leq \alpha \leq 1$ yra glodinimo parametras, arba koeficientas, kuris, kaip matome iš formulės, toldamas nuo paskutinio atlikto stebėjimo mažėja.

Parametrai šiam modeliui parenkami minimizuojant kvadratinių paklaidų sumą *SSE* (*angl. Sum of Squared Errors*). Tačiau tai nėra tiesinis metodas, todėl papildomai reikia atlikti optimizaciją.

- 2) **Dvigubas eksponentinis glodinimas** arba **Holt-Winters** metodas yra skirtas laiko eilutėms, turinčioms trendą. (Holt, 1957) patobulino paprastąjį eksponentinį metodą. Šį metodą apima trys lygtys: prognozės ir dvi glodinimo lygtys (trendo ir lygio). Prognozės lygtis:

$$Y_{t+h|t} = l_t + hb_t \quad (24)$$

Lygio lygtis:

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (25)$$

Trendo lygtis:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (26)$$

čia, l_t yra lygis laiko momentu t , b_t yra trendas laiko momentu t , α yra lygio glodinimo parametras, β trendo glodinimo parametras, h – horizontas, kuriam prognozuojame. Taigi, pritaikius Holto-Winterso metodą, prognozės nebėra tolygios, bet turi tendą. Prognozė, turinti h žingsnių į priekį, yra lygi paskutinio įvertinto lygio ir žingsnių skaičiaus, padauginto iš paskutinės nustatytos trendo reikšmės, sumai. Rezultate gauname, kad prognozė yra tiesinė h (horizonto) funkcija.

- 3) **Trigubas eksponentinis glodinimas** arba **sezoninis Holto-Winterso** metodas prie dvigubo eksponentinio glodinimo prideda ir sezoniškumo lygtį, taip modelį apibūdina jau nebe trys, o keturios lygtys:

prognozės lygtis:

$$Y_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \quad (27)$$

lygio lygtis:

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (28)$$

trendo lygtis:

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (29)$$

sezoniškumo lygtis:

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (30)$$

čia k parametras, kuris užtikrina, kad sezoniniai indeksai ateina iš paskutinių laiko eilutės metų, m šiuo atveju yra periodų (sezonų) skaičius.

Apibendrinant eksponentinio glodinimo metodus galima pastebėti, kad literatūroje randama įvairių variacijų kaip šiuos metodus galima klasifikuoti priklausomai nuo laiko eilutės trendo ir

sezoniškumo, bei jų adityvumo ar multiplikatyvumo. Pirmasis šį grupavimą pasiūlė (Pegels, 1969) grupavimas pavaizduotas 3 lentelėje, žemiau:

3 lentelė. Eksponentinio glodinimo klasifikavimas

Trendo komponentė	Sezoninė komponentė		
	N (nėra)	A (adityvioji)	M (multiplikatyvioji)
N (nėra)	(N,N)	(N,A)	(N,M)
A (adityvioji)	(A,N)	(A,A)	(A,M)
Ad (adityvioji prislopinta)	(Ad,N)	(Ad,A)	(Ad,M)

Šioje lentelėje trūksta multiplikatyviosios, bei multiplikatyviosios prislopintos trendo komponentių, tačiau (Hyndman & Athanasopoulos, 2014) teigia, kad prognozės naudojant šias komponentės retai kada būna tikslios, dėl to darbe jos nebus naudojamos.

2.2.5. Sudėtingi prognozavimo metodai

Pagrindinis struktūrinis modelis (BSM), buvo aprašytas (Peters & Harvey, 1997). Tai modelis, kuris naudojami trendu ir sezoniškumu generuojant prognozes. Jo matematinė išraiška:

$$Y_t = \mu_t + \gamma_t + \varepsilon_t \quad (31)$$

čia μ_t , γ_t , ε_t yra trendas, sezoniškumas ir paklaidos atitinkamai.

Procesas generuojanti trendą gali būti apibūdintas kaip:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad (32)$$

$$\beta_t = \beta_{t-1} + \zeta_t \quad (33)$$

čia η_t ir ζ_t yra nepriklausomai pasiskirstę su vidurkiu lygiu nuliui ir dispersijomis σ_η^2 ir σ_ζ^2 .

Procesas generuojantis sezonines komponentes atrodo taip:

$$\gamma_t = \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \quad (34)$$

čia ω_t nepriklausomai pasiskirstęs triukšmo kintamasis su vidurkiu lygiu nuliui ir dispersija σ_ω^2 , o s žymi sezonų skaičių metuose.

BATS modelis tai eksponentinio glodinimo su Boxo-Coxo transformacija, *ARMA* paklaidomis, trendo ir sezoniškumo komponentėmis modelis. Šis modelis - tai apibendrinimas visų metodų tinkančių sezoniškumu pasižyminčioms laiko eilutėms. Jis leidžia turėti kelis sezoniškumus, tačiau nėra tinkamas jei sezono reikšmė nėra sveikasis skaičius.

TBATS tai alternatyvus *BATS* modelio taikymas, kurį pasiūlė (Livera, Hyndman, & Snyder, 2011). Modelis naudoja Furjė sąlygas su eksponentiniu glodiniu ir Boxo-Coxo transformacijomis ir visa tai vyksta visiškai automatizuotai. Dėl automatizavimo priežasties rezultatai kartais gali būti itin prasti, tačiau kai kuriais atvejais šio modelio taikymas gali būti labai naudingas. *TBATS* modelio privalumas yra tai, kad jis leidžia nestipriai kisti sezoniškumui laike, minusas – modelio parametru įvertinimas gali būti labai lėtas, ypač taikant modelį ant ilgų laiko eilučių.

Neuroninio tinklo modelis remiasi sluoksnių, turinčių daug mazgų principu, kai pradiniai duomenys praleidžiami pro daug mazgų turinčius sluoksnius, kiekvieno mazgo išeinantys duomenys tai - į kitą mazgą įeinantys duomenys padauginami iš koeficiento. Pagal šį principą gaunamas **neuroninio tinklo autoregresinis** modelis, kai pirminiams duomenims imamos prognozuojamos laiko eilutės atsilikusios reikšmės. Pavyzdžiui *NNAR* (p, k) modelis indikuoja, kad modeliui bus naudojama p atsilikusių reikšmių ir k mazgų paslėptame sluoksnyje.

2.3. Modelių kolektyvai

(Grimit & Mass, 2002) aprašė, kad modelių vidurkių kolektyvas prognozuoja atmosferinius reiškinius tiksliau nei bet kuris vienas modelis atskirai. (Raftery, Gneiting, Balabdaoui, & Polakowski, 2005) savo darbe apžvelgia galimybę individualiems modeliams suteikti skirtingus svorius, nustatytus pagal modelių praeities rezultatus. Darbe lyginamas geriausias modelis, kai ateities prognozei parenkamas vienintelis modelis, praėjusiame laiko periode prognozavęs tiksliausiai. Šis modelis lyginamas su skirtingų svorių modeliu. Priimtos išvados ir rezultatai paraiškina, kad įtraukiant tik geriausia modelį prarandama daug papildomos informacijos, todėl naudingiau yra naudoti daug modelių su skirtingais svoriais. Tačiau, kartais kai kurie modeliai yra beverčiai, todėl verta pasirinkti tik tikslesnius modelius ir prasčiausių modelių neįjungti į kolektyvą. Taip sutaupomi ne tik skaičiavimo resursai, bet ir dažnu atveju padidinamas kolektyvo tikslumas.

Darbe bus lyginami trys modelių kolektyvai: vidurkio, geriausio modelio ir skirtingų svorių kombinacijos. Pirmo kolektyvo sudarymas yra paprastas ir detalesnio paaiškinimo nereikalauja. Geriausio modelio kolektyvo sudarymą galima išreikšti taip:

$$F_{Best(t-24)} = \min(F_1; F_2; F_n)_{t-24} \quad (35)$$

$$F_{Best} = F_{Best(t-24)}_{t+24} \quad (36)$$

Čia, $F_{Best(t-24)}$ praėjusios paros tiksliausias modelis, F_{Best} – kitai parai naudojamą modelio prognozė, F_n – prognozių modeliai.

Skirtingų svorių kombinacijos modelis remiasi keliais tiksliausiais praėjusios paros modeliais ir pagal jų tikslumą priskiria atitinkamus koeficientus. Geriausias modelis gauna didžiausią koeficientą, antras pagal tikslumą šiek tiek mažesnę ir t.t. šį metodą aprašė (Aiolfia & Timmermannb, 2005). Sviurių kombinacijos modelį galima išreikšti taip:

$$F_{Komb(t-24)} = \text{rank}(F_1; F_2; F_n)_{t-24} \quad (37)$$

$$F_{Komb} = \sum_{k=1}^k F_{Komb(t-24)} * W_k \quad (37)$$

Čia, $F_{Komb(t-24)}$ praėjusios paros pagal tikslumą išrikiuoti modeliai, tiksliausiam modeliui priskiriamas 1 rangas, W_k – kiekvienam rangui parinktas koeficientas, F_{Komb} – kitos paros prognozei naudojamas kolektyvas. Tyrime svorių koeficientas parenkami ekspertiškai, t.y. testavimo imtyje, cikle, bandomi įvairūs svoriai ir parenkami tiksliausių rezultatų sugeneravę svoriai.

2.4. Paklaidų įvertinimas

Norint įvertinti prognozės tikslumą svarbu naudoti originalią prognozę. Prognozės sukurtos ant duomenų ir vėliau patikrintos ant tos pačios duomenų imties skirtumų dydis nėra patikimas indikatorius parodyti kokios iš tikrųjų bus prognozės paklaidos ateityje. Prognozės tikslumas gali būti teisingai įvertintas tik tikrinant kaip gerai modelis prognozuos ateityje, naudodamas duomenis, kurie nebuvo panaudoti apmokant modelį. Renkantis geresnį modelį praktikoje įprasta duomenis atskirti į dvi imtis, apmokymo ir testavimo. Ap mokymo imtis skirta nustatyti prognozės modelio parametrus, o testavimo imtis - įvertinti modelių tikslumui. Kadangi testavimo imtis nėra naudojama modelio sudarymui, ji yra objektyvus ir patikimas indikatorius galintis parodyti kaip modelis prognozuos turėdamas naujus duomenis. Paprastai testavimo imtis sudaro apie 20 % visos imties ir ji turi būti bent vieno prognozės horizonto ilgio (Hyndman & Koehler, 2006). Verta

pažymėti, kad modelis, kuris puikiai pasirodo su apmokymo imtimi nebūtinai prognozuos gerai. Modelio persimokymas yra taip pat blogai kaip ir nesugebėjimas aptikti besikartojančios struktūros. Kai kurie šaltiniai testavimo imtį įvardija kaip “atskirtąją dalį” (*angl. “hold-out set”*), todėl, kad ši duomenų dalis atskiriama apmokant modelį.

Verta atkreipti dėmesį, kad prognozės paklaidos nuo skirtumų skiriasi dėl dviejų dalykų. Pirmia, skirtumai yra skaičiuojami kai naudojami apmokymo imtis, o paklaidos kai naudojama testavimo imtis. Antra, skirtumai yra paremti prognozavimu vieną žingsnį į priekį, o paklaidos gali būti skaičiuojamos kai yra daugiažingsnis (*angl. multi-step*) prognozavimas į priekį.

Nuo skalės priklausančios paklaidos

Paklaidos yra toje pačioje skalėje kaip ir duomenys. Tikslumo matavimo būdai, kurie priklauso nuo skalės negali būti naudojami kelių laiko eilučių paklaidų palyginimui kai laiko eilučių skalės skiriasi. Du pagrindiniai paklaidų matavimo būdai priklausantys nuo skalės yra *MAE* (*angl. Mean Absolute Error*) ir *RMSE* (*angl. Root Mean Squared Error*). *MAE* yra populiarus nes yra lengvai suprantamas ir lengvai suskaičiuojamas. Prognozės metodas, kuris minimizuoja *MAE*, naudojamas medianos prognozei, *RMSE* minimizavimas yra nuorodą į vidurkio prognozę. *RMSE* taip pat plačiai naudojamas, tiesa, jį interpretuoti yra šiek tiek sunkiau.

Procentinės paklaidos

Procentinės paklaidos turi privalumą, nes jos nepriklauso nuo matavimo vienetų, dėl to yra labai lengva palyginti prognozių naudojančių skirtingas laiko eilutes paklaidas. Dažniausiai naudojamas yra *MAPE* (*angl. Mean Absolute Percentage Error*). Matavimai grįsti procentinėmis paklaidomis turi trūkumą dažnai būti begaliniais arba neapibrėžtais. Pavyzdžiui, procentinė paklaida neturi prasmės kai matuojamas temperatūros prognozės tikslumas, nes temperatūra turi nustatytą 0 lygmenį. Taip pat yra išvesta simetrinė *MAPE* paklaida (*sMAPE*), kurią pasiūlė (Armstrong, 1978). Tačiau (Hyndman & Koehler, 2006) nerekomenduoja naudoti *sMAPE* apskritai. Simetrinės procentinės paklaidos (*sMAPE*) turi trūkumą dėl to, kad skiria didesnę baudą neigiamoms paklaidoms nei teigiamoms.

Skalei pritaikytos paklaidos

Skalei pritaikytas paklaidas pasiūlė (Hyndman & Koehler, 2006) kaip alternatyvą procentinėms paklaidoms norint palyginti paklaidas skirtingoms laiko eilutėms. Jie pasiūlė paklaidas skaičiuoti pagal paprastųjų prognozavimo metodų *MAE*. Nesezoninėms laiko eilutėms naudinga naudoti skalei pritaikytas paklaidas, kurios lyginamos su Naive metodu. Jei skalei pritaikyta paklaida yra mažesnė nei 1, tai reiškia, kad sukurtas modelis yra geresnis nei Naive metodas. Kontraversiškai, jei paklaida

didesnė nei 1 paklaida didesnė nei Naive metodo. Taip pat sezoniškumui pritaikytų modelių paklaidas galima lyginti su Naive metodo paklaidomis.

Tyrimė naudojamos keturių tipų paklaidų metrikos: *MAE*, *RMSE*, *MASE* ir *nMAE*. *MAE* ir *RMSE* yra nuo laiko eilutės skalės priklausančios paklaidos, kurios naudojamos modelių palyginimui prognozuojant tos pačios laiko eilutės reikšmes. Nuo skalės priklausančių paklaidų negalima lyginti su kitomis laiko eilutėmis. Pavyzdžiui tyrimo atveju vieno vėjo parko *RMSE* paklaidos negalima lyginti su kito, nes jų skalės skiriasi.

$$MAE = E(|e_t|) \quad (38)$$

$$RMSE = \sqrt{E(e_t^2)} \quad (39)$$

čia E yra vidurkis, o e_t paklaida.

MASE paklaida puikiai tinka palyginti įvairių laiko eilučių paklaidas. Šias paklaidas pasiūlė (Hyndman & Koehler, 2006) kaip alternatyvą procentinėms paklaidoms lyginant prognozes tarp skirtingų laiko eilučių.

$$q_j = \frac{q_j}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|} \quad (40)$$

Iš formulės matosi, kad prognozių paklaidos lyginamos su vidutine naiviosios prognozės paklaida tai pačiai laiko eilutei prognozuoti. Jei $MASE < 1$ tai reiškia, kad prognozė buvo tikslesnė nei naivusis prognozavimas, kai sekančios reikšmės prognozė yra paskutinė užfiksuota reikšmė. Jei $MASE > 1$ vadinasi, naivusis modelis buvo tikslesnis.

Paskutinis paklaidos įvertinimo metodas kuris yra naudojamas tai *nMAE* – normalizuota vidutinė absoliutinė paklaida. Ši paklaida tinkama įvertinti, kai laiko eilutės skalė kinta laike. Pavyzdžiui jei suminė VE instaliuota galia didėja kasmet. Šis metodas paprasčiausiai vidutinę absoliutinę paklaidą padalina iš instaliuotos galios.

$$nMAE = \frac{MAE}{P} \quad (41)$$

čia P yra instaliuota galia, kuri yra skirtinga kiekvienai elektrinei ir visą analizuojamą laikotarpį vienoda suminei instaliuotai galiai.

Apibendrinant prognozavimo tyrimo metodus pateiktus mokslinėje literatūroje galima išskirti pagrindinius 5 metodų grupes: paprastųjų, regresinių, *ARIMA*, eksponentinio glodinimo ir sudėtingųjų. Kiekvienos iš šių grupių metodas naudojamas atliktame tyrimė. Naudojant skirtingų metodų modelius taip pat kuriami 3 skirtingi kolektyvai: modelių vidurkių, geriausio modelio

(BEST) ir svertinių koeficientų (KOMB.). Sukūrus modelius, kiekvieno modelio parametrų tinkamumas įvertinamas pagal *RMSE* paklaidą, o skirtingų modelių skirtingoms laiko eilutėms palyginimas atliekamas naudojant *nMAE* paklaidas.

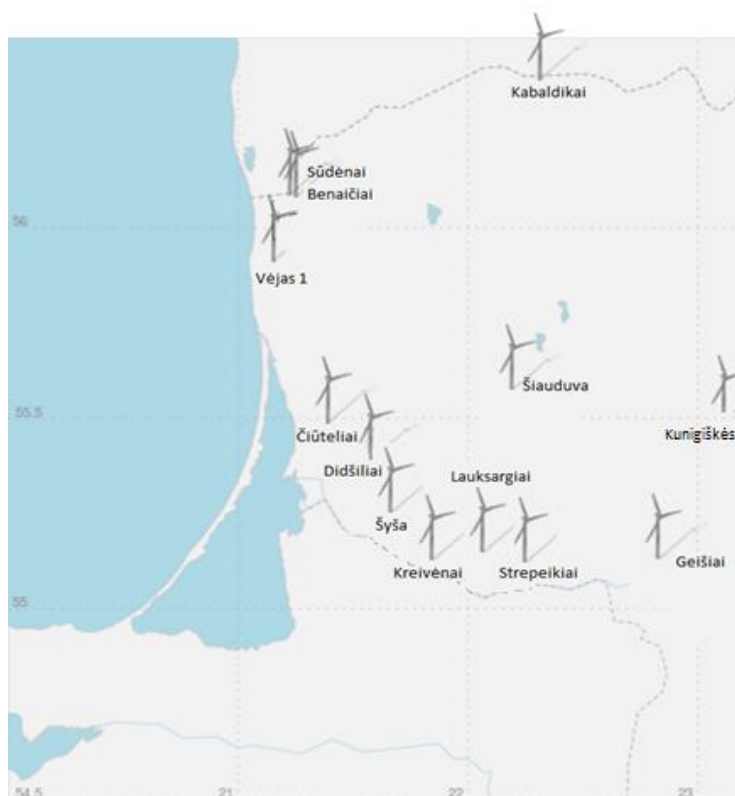
3. Tyrimo rezultatai ir jų aptarimas

Tyrimui atlikti naudoti „R“, „R Studio“ ir „Gretl“ programiniai paketai. „R“ pagalba buvo kuriami prognozavimo modeliai ir nustatinėjami jų parametrai. „Gretl“ naudotas išorinių regresorių priežastingumo analizei atlikti ir regresorių reikšmingumui patikrinti.

3.1. Naudojami duomenys

Tyrimui buvo naudojama 13 Lietuvos VE, prijungtų prie PT (elektros sistemos perdavimo tinklas), faktiniai valandiniai generacijos duomenys (2017 – 2018 metai). Elektrinės išsidėsčiusios įvairiuose Lietuvos regionuose, tačiau didžiausia jų dalis yra Lietuvos vakaruose, kur vidutiniai vėjo greičiai yra statistiškai didžiausi.

Papildomai naudojami šie valandiniai parametrai: 5 skirtingų šaltinių vėjo greitis, vėjo kryptis, temperatūra, kitų valstybių (Latvijos, Lenkijos, Švedijos ir Danijos) VE valandinė generacija, 2 skirtingų sistemų jau paruošta VE generacijos prognozė.



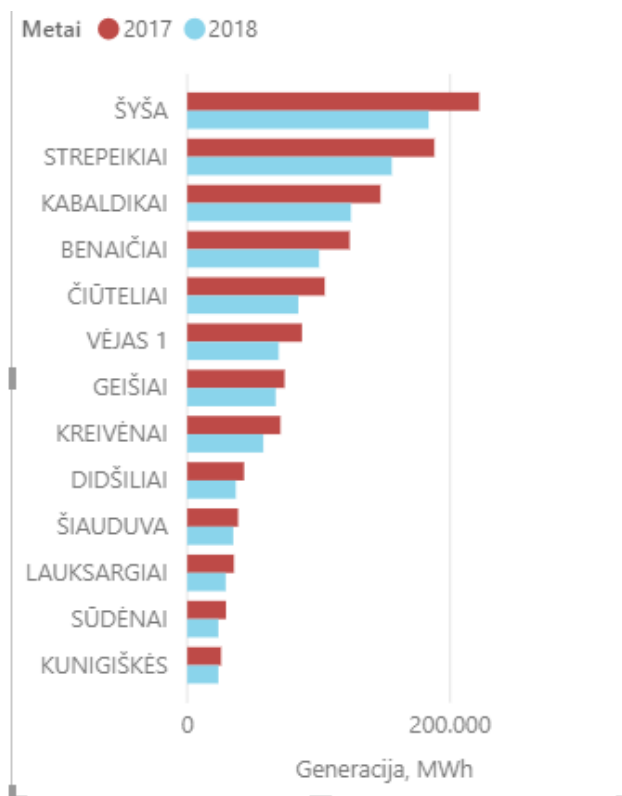
6 pav. VE elektrinių, kurių duomenys naudojami darbe, išsidėstymas Lietuvos teritorijoje.

Dauguma VE prijungtų prie PT išsidėsčiusios vakarinėje Lietuvos dalyje, kur vidutinis vėjo greitis yra didžiausias.

VE Parkas	Instaliuota galia, MW
STREPEIKIAI	73,50
ŠYŠA	60,00
BENAIČIAI	50,50
KABALDIKAI	45,00
VĖJAS 1	39,13
ČIŪTELIAI	39,10
KREIVĖNAI	30,00
GEIŠIAI	24,00
DIDŠILIAI	21,50
LAUKSARGIAI	14,90
SŪDĖNAI	14,00
ŠIAUDUVA	13,80
KUNIGIŠKĖS	7,50
Total	432,93

7 pav. VE parkai ir jų instaliuotos galios

Kiekvieno parko instaliuota galia yra skirtinga, instaliuota galia varijuoja nuo 7,5 iki 73,5 MW. Instaliuota galia apibūdina visos VE atskirų turbinų suminę galią. Vienos turbinos galia gali svyruoti nuo 1,8 iki 3,5 MW instaliuotos galios.

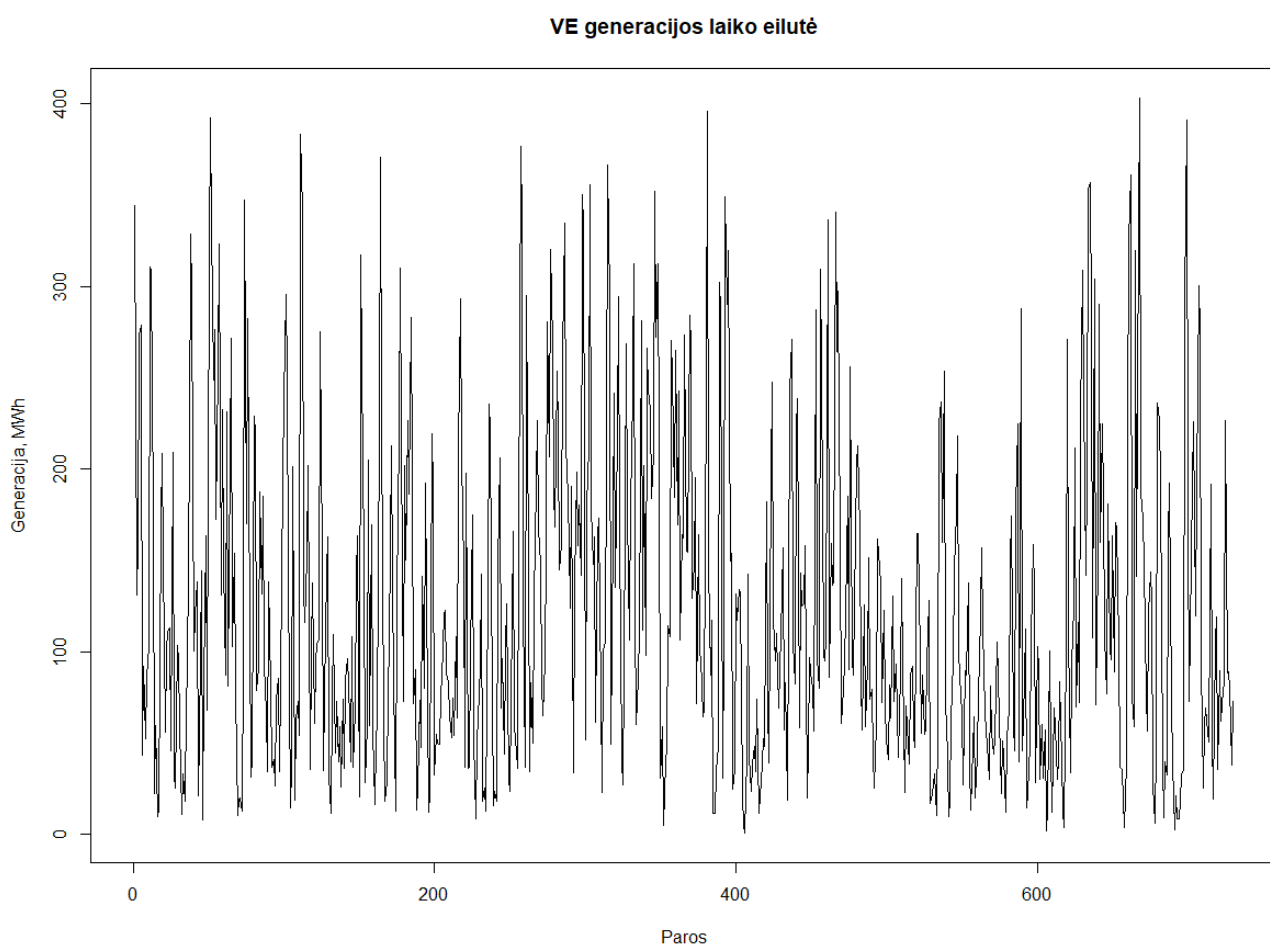


8 pav. VE parkų generacija (MWh) 2017 - 2018 metais.

Apžvelgus generacijos duomenis (**8 pav.**) matoma, kad didesnės instaliuotos galios VE nebūtinai gamina daugiau energijos. Prie perdavimo tinklo prijungtų VE generacija 2018 m. sudarė 31% visos Lietuvos generacijos. Toliau tiriamos atskirų VE generacijos laiko eilutės kuriose užfiksuota **8 pav.** išvardintų VE informacija.

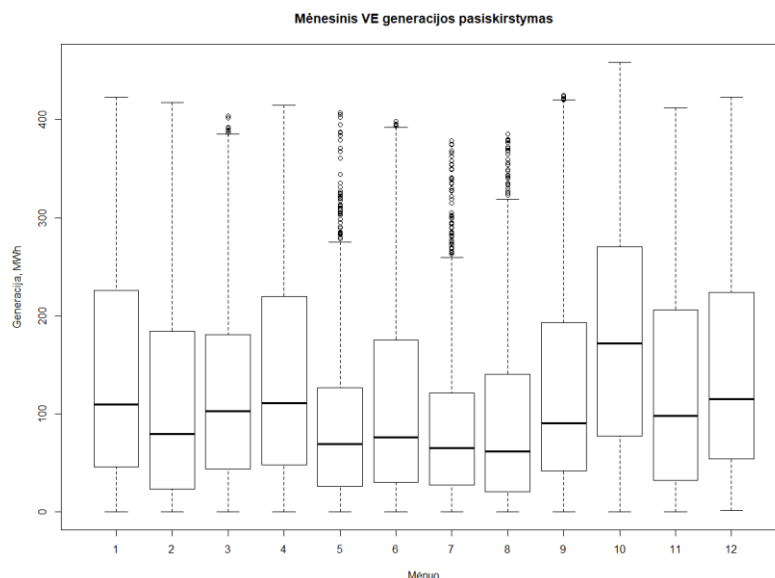
3.2. Duomenų ir laiko eilučių analizė

Toliau apžvelgiamas laiko eilutės grafikas, atvaizduojama suminė VE generacijos laiko eilutė, kuri atspindi visų VE atitinkamų valandų reikšmes.



9 pav. Vidutinė paros VE generacija 2017, 2018 metais.

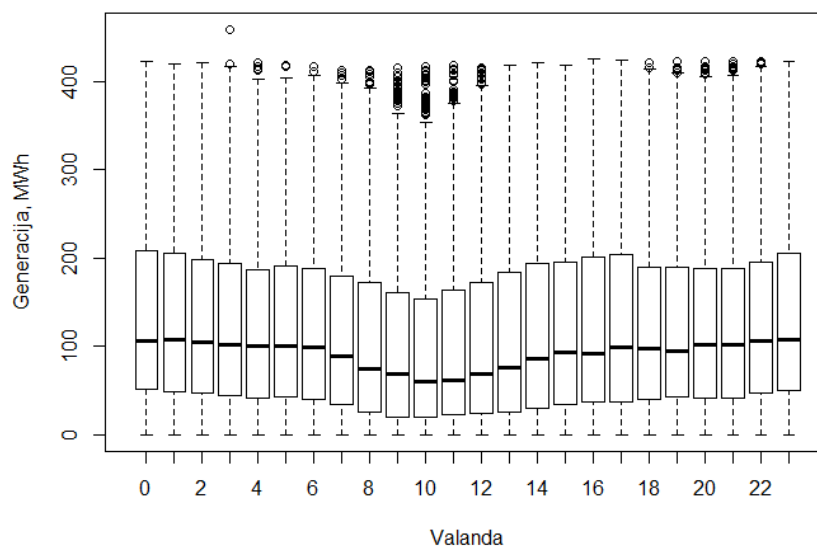
Kitame grafike apžvelgiama kaip atrodo VE generacija kiekvieną mėnesį. Pastebima, kad šiltuoju periodu generacija linkus sumažėti, tuo tarpu rudenį ir pavasarį, pučiant smarkesniai vėjui – generacija išauga. Taip pat matosi, kad vasarą padidėja išskirčių, tai lemia netipinės dienos, kuriomis atplūstantys vėjo gūšiai smarkiai pakelia generaciją nuo vidurkio reikšmės.



10 pav. Mėnesinis suminės VE generacijos pasiskirstymas.

Žvelgiant į valandinį VE generacijos pasiskirstymą galima pastebėti, kad generacija linkusi sumažėti šviesiuoju paros metu, 8-12 valandomis ir pakilti likusiomis.

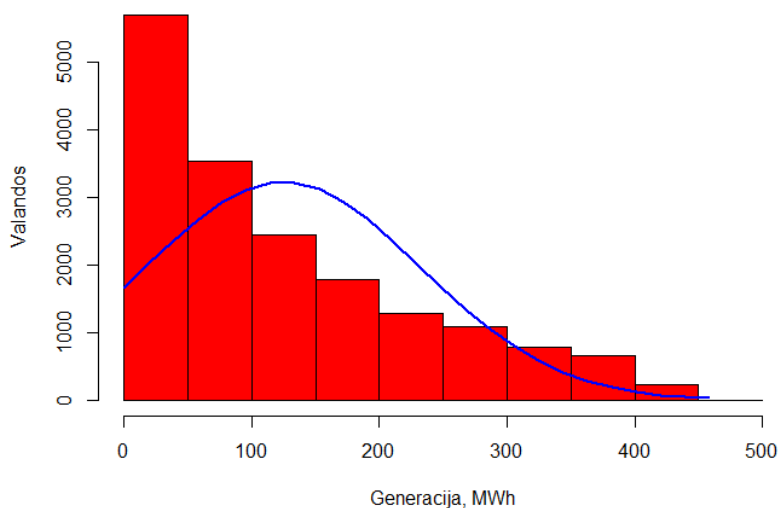
Valandinis VE generacijos pasiskirstymas



11 pav. VE generacijos pasiskirstymas pagal valandas.

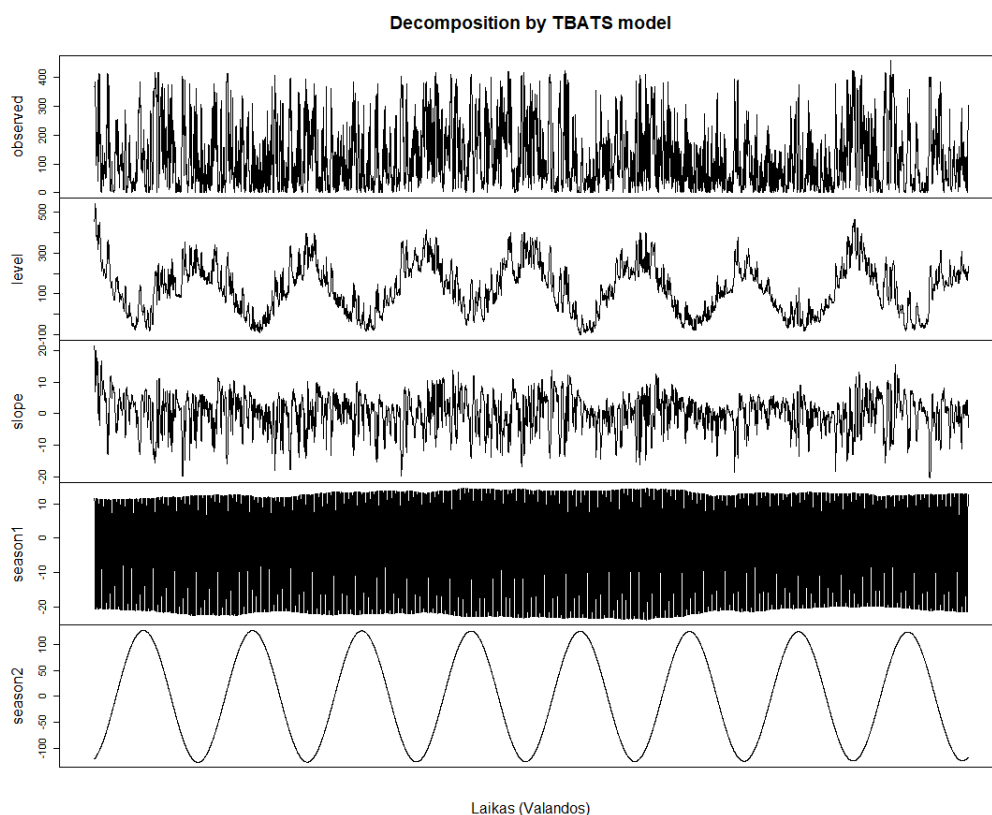
Toliau vaizduojamas VE generacijos skirstinys pagal valandas. Skirstinys parodo, kad faktinė generacija didžiąją laiko dalį kur kas mažesnė, nei potenciali (instaliuota galia = 433 MW).

Histograma su normaliojo skirstinio kreive



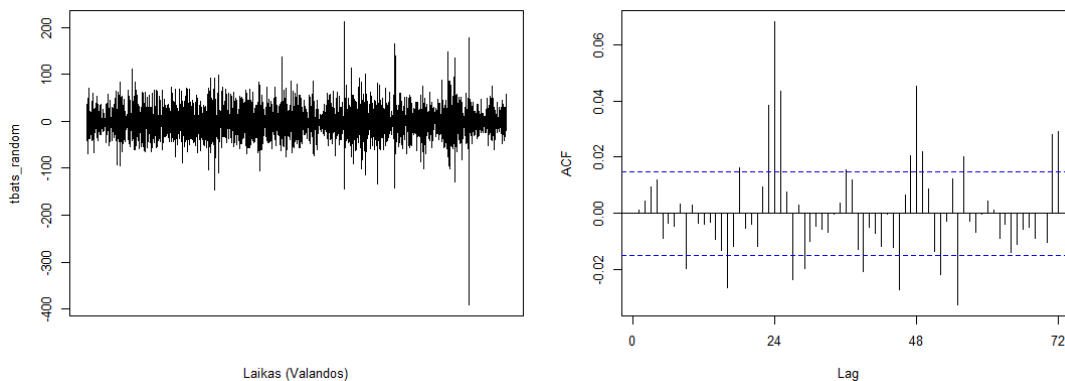
12 pav. VE generacijos skirstinys pagal valandas.

Žvelgiant į laiko eilutės grafiką sunku nustatyti koks sezoniškumas vyrauja. Iš VE generacijos pasiskirstymo pagal valandas grafiko galima numanyti, kad yra paros sezoniškumas, tačiau panašu, kad yra dar vienas sezoniškumas priklausantis nuo metų laiko. Turimai laiko eilutei sukuriame 2 sezoniškumus ir atliekame laiko eilutės dekompoziciją naudodami *TBATS*.



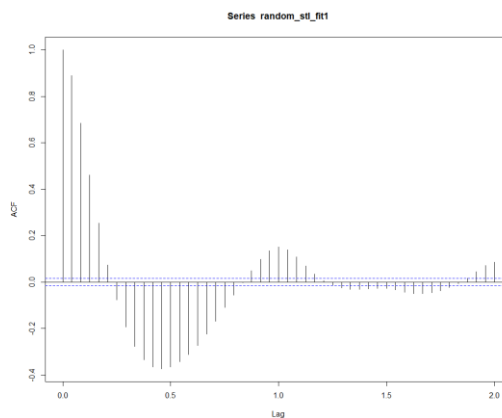
13 pav. VE generacijos laiko eilutės su dviem sezoniškumais (paros ir ketvirtiniu) dekompozicija.

Pirmasis sezoniškumas turi labai didelį dažnį dėl to grafiškai sunkiai pastebimas. Tačiau iš paros sezoniškumo grafiko y ašies galime pastebėti, kad paros laikotarpyje svyravimo amplitudė yra 30 MWh (nuo -20 iki 10). Tuo tarpu ketvirtiniame sezoniškume svyravimų amplitudė yra kur kas didesnė ir siekia 200 MWh. Atlikus analizuojamos laiko eilutės dekompoziciją, pritaikius skirtingo periodo antrąjį sezoniškumą (savaitinis, ketvirtinis, pusmečio), ketvirtinis sezoniškumas buvo tinkamiausias, nors liekanos po dekompozicijos turi autokoreliaciją:



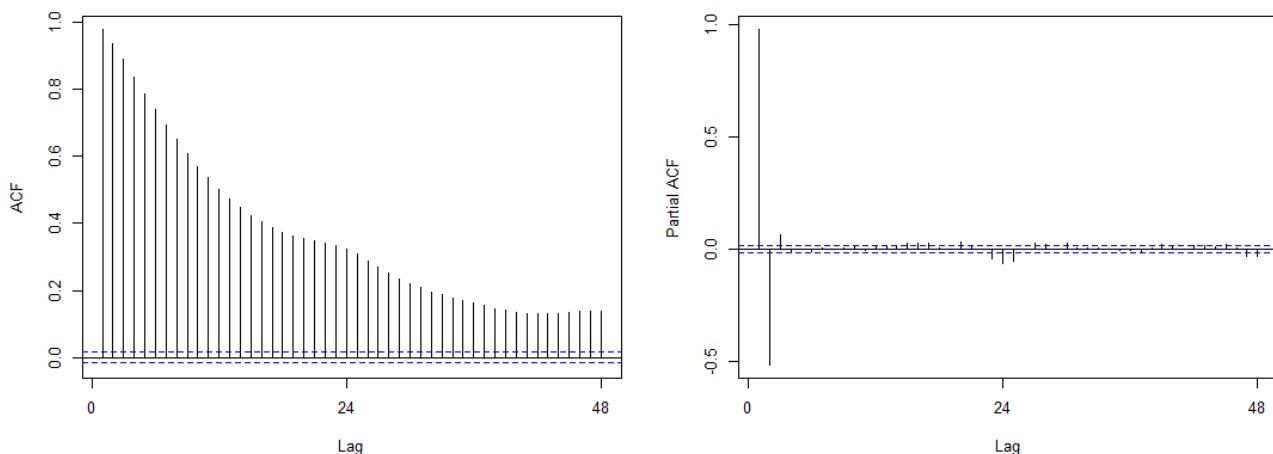
14 pav. Liekanų grafikas atliekant dekompoziciją *TBATS* metodu (kairėje) ir liekanų *ACF* grafikas (dešinėje).

Galima palyginti liekanų *ACF* grafiką pritaikius vieną sezoniškumą ir *STL* dekompozicijos metodą:



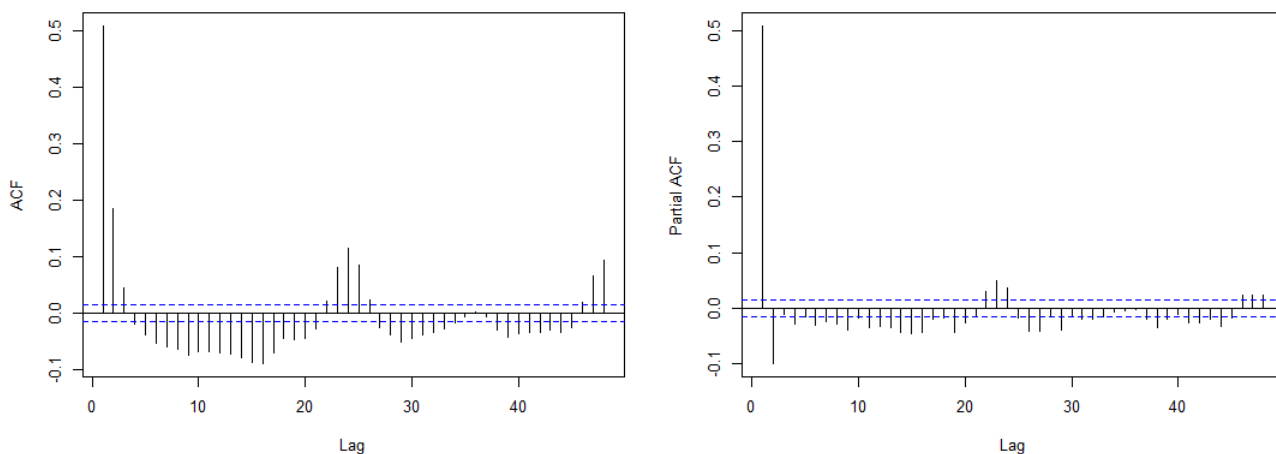
15 pav. *ACF* grafikas *STL* dekompozicijos liekanoms (laiko eilutės sezoniškumas – paros).

Toliau atvaizduojami VE generacijos laiko eilutės *ACF* ir *PACF* grafikai skirti stacionarumo nustatymui.



16 pav. VE generacijos laiko eilutės *ACF* ir *PACF* grafikai 48 atsilikimams.

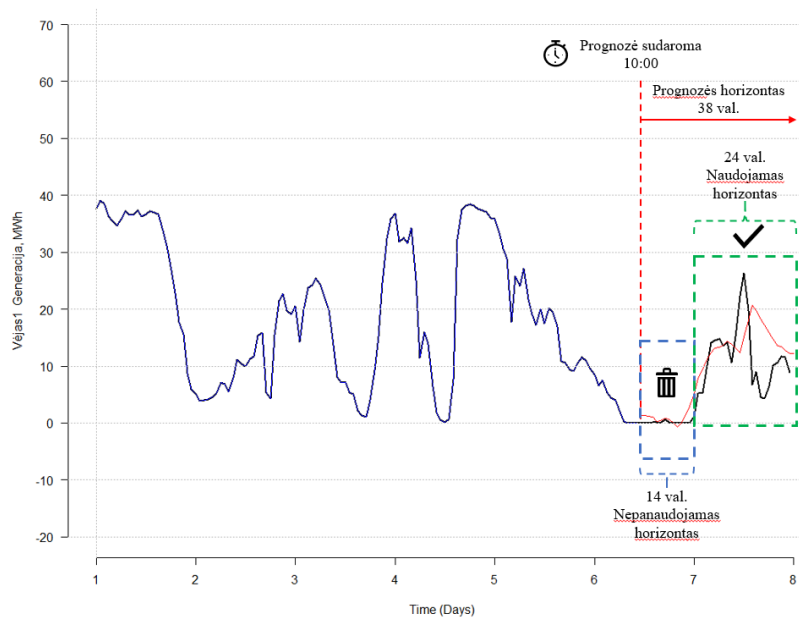
Literatūroje aprašoma (Hyndman & Athanasopoulos, 2014), jei *ACF* reikšmės krinta lėtai ir palaipsniui – laiko eilutė nėra stacionari. *ACF* turėtų nukristi staigiai ir per kelis žingsnius, tuomet laiko eilutę galime laikyti stacionaria. Atlikę vienietinės šaknies testus gauname 2 skirtingus rezultatus. *ADF* testo su 72 atsilikimais *p* reikšmė 0,01, tai sako, kad laiko eilutės nestacionarumo hipotezė atmetama ir priimama alternatyva, kad laiko eilutė yra stacionari. Tuo tarpu *KPSS* testo statistikos reikšmė lygi 1,29 (kritinė 5 % statistikos reikšmingumo riba 0,463), tai pažymi, kad hipotezė apie laiko eilutės stacionarumą atmetama ir priimama alternatyva – laiko eilutė nėra stacionari. Pagal literatūros rekomendacijas atliekame pirmos eilės diferencijavimą.



17 pav. diferencijuotos VE generacijos laiko eilutės *ACF* ir *PACF* grafikai 48 atsilikimams.

3.3. Prognozavimo etapai

Prognozavimo tikslas yra sugeneruoti kuo tikslesnę prognozę ateinančiai parai, kai prognozės generavimo terminas yra einamos paros 10 val. Vadinasi, gali būti panaudoti visi istoriniai duomenys, kurie yra prieinami iki einamos paros 10 valandos. Tikslu iliustracija pateikta žemiau.



18 pav. Prognozavimo tikslo iliustracija

Pagal pateiktą tikslo iliustraciją galima matyti, kad sudarant prognozę ateinančiai parai, 10 val. ryto bus prognozuojama 38 val. horizontui. Iš šių 38 val., pirmosios 14 valandų bus nepanaudojamos (mėlynas stačiakampis), tikslumo įvertinimui bus naudojamos 24 val. iš panaudojamo horizonto (žalias stačiakampis).

Tyrimas atliekamas trimis etapais:

I etapas - prognozavimas naudojant vienmačius metodus

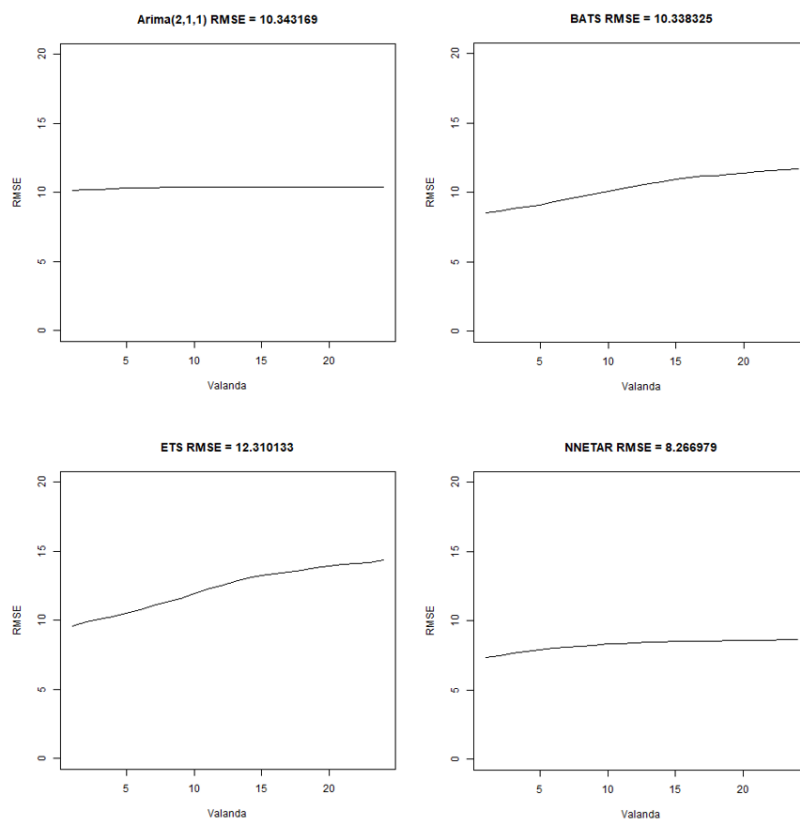
II etapas – prognozavimas naudojant 4 daugiamačius metodus.

III etapas – prognozavimas naudojant 4 daugiamačius metodus, prie išorinių regresorių įtraukiant prognozavimo sistemų paruoštas prognozes.

3.3.1. I etapas. Prognozavimas naudojant vienmačius metodus

Prognozavimas naudojant vienmačius metodus tyrime paremtas VE generacijos prognozavimu, naudojant tik kiekvieno VE parko generacijos faktinius duomenis, sudarant prognozę į ateitį. Kitaip tariant, prognozei sudaryti nenaudojami jokie kiti išoriniai kintamieji. Praktikoje šis metodas dažniau pasiteisina sudarant prognozes trumpam horizontui į priekį.

Atliekant I etapo tyrimą, atrenkant kiekvienam modeliui naudojamus parametrus buvo taikomas kryžminis patikrinimas (*angl. cross-validation*). Kryžminis patikrinimas buvo atliekamas tik suminei VE generacijos eilutei, remiantis prielaida, kad visų likusių VE generacijos eilutės turi panašias charakteristikas. Keletas kryžminio patikrinimo pavyzdžių pateikta apačioje.

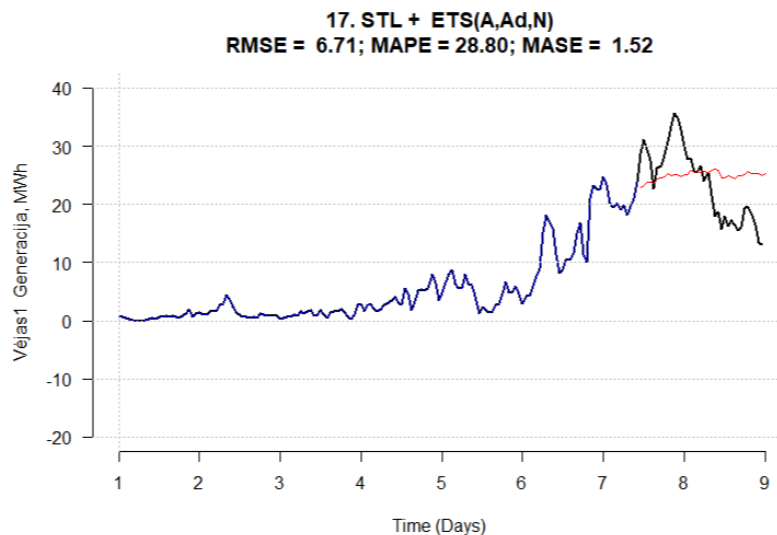


19 pav. *ARIMA(2,1,1)*, *BATS*, *ETS*, *NNETAR* modelių paklaidų priklausomybė nuo laiko, atlikus kryžminį patikrinimą (suminės VE generacijos pavyzdys).

Pagal pavyzdžio iliustraciją galima aiškiai matyti, kad paklaida didėjant horizontui auga, tačiau vieniems modeliams tai pastebima aiškiai, kitiems tokio stipraus ryšio tarp horizonto ir paklaidos nėra.

Sudarant *ARMA/ARIMA* modelius ir renkant *AR* ir *MA* eiles buvo naudojami *ACF* ir *PACF* grafikai iš 17. Taigi, *AR* modelio p eilė iš *PACF* grafiko parenkama 2, *MA* modelio q eilė – 3. Tinkamiausias modelis sprendžiant iš grafikų – *ARIMA* (2, 1, 3), kur $AICc = 152525$. Taip pat buvo atlikta automatinio eilės parinkimo funkcija, kuri pagal $AICc$ kriterijų geriausiu modeliu išrinko *ARIMA* (2,1,1) su AIC 152520, todėl nuspręsta naudoti šį modelį.

Atrinkus parametrus kiekvienam modeliui, buvo generuojamos prognozės. Prognozės generuojamos prieš tai apmokant modelį su vienu metų imtimi (2017 m.). Pagal apmokymo metu nustatytus parametrus modelis apskaičiuoja prognozę 38 val. horizontui ir išmeta pirmąsias 14 val.



20 pav. Vienmačio modelio STL prognozės pavyzdys, Vėjas 1 vėjo elektrinė.

Visos prognozės išsaugomos į bendrą failą ir lyginamos su atitinkamos VE faktine generacija. 1 priede **32 pav.** Vienmačių modelių rezultatai, pateikta lentelė su kiekvieno modelio paklaida, kiekvienai VE, taip pat ir VE suminio kiekio prognozės paklaida. Žvelgiant į vienmačių modelių paklaidų lentelę išsiskiria keli modeliai, kurių tikslumas buvo geriausias. *TBATS* modelis buvo tiksliausias 6 VE, *BATS* ir *TSLM* su trendu ir sezoniškumu kiekvienas, parodė geriausią rezultatą su 3 VE. *NNETAR* modelis buvo geriausias tik 1 VE, tačiau jis taip pat parodė geriausią rezultatą prognozuodamas suminę VE generaciją. Paskutinis stulpelis „Prognozių vidurkis“ parodo paklaidą, kuri yra gaunama skaičiuojant kiekvienos valandos visų modelių prognozių vidurkį. Vertėtų atkreipti dėmesį, kad imant visų modelių vidurkį, gaunamas geresnis rezultatas nei bet kurio modelio atskirai, tai pastebima su visomis VE. Paskutinė eilutė „VE suma iš atskirų VE prognozių“ parodo kokia paklaida yra gaunama sumuojant kiekvieno modelio prognozę, kiekvienam parkui, kiekvieną valandą. Gaunamas rezultatas - suminė VE prognozė, kuri gali būti palyginta su atskira suminės VE generacijos laiko eilutės prognozės paklaida, ši yra viena eilutė aukščiau – „Bendra VE prognozė“. Pastebimas įdomus rezultatas, *ARIMA (2,1,1)* modelis, suminę VE generaciją prognozuoja didžiausiu tikslumu, kai kiekviena VE sumuojama atskirai, nors iki šiol šis modelis nebuvo geriausias nei su viena VE. Taip pat matoma aiški tendencija, kad sumuotos prognozės dažniausia lenkia atskiros laiko eilutės prognozę savo tikslumu.

3.3.2. II etapas. Prognozavimas naudojant daugiamačius metodus

Prognozavimas naudojant daugiamačius metodus tyrime paremtas VE generacijos prognozavimu, naudojant tiek kiekvieno VE parko generacijos faktinius duomenis, tiek išorinius kintamuosius paminėtus šio skyriaus pradžioje. Kitaip tariant, prognozei sudaryti naudojama visa prieinama

informacija ir pasitelkiamos ex-ante prognozės. Kaip buvo atrinkti tik naudingi išoriniai kintamieji paašškinta kitame poskyryje. II etape prognozių modeliams kurti naudojama nediferencijuota laiko eilutė ir nediferencijuotos išorinių kintamųjų laiko eilutės.

3.3.3. Priežastingumo analizė

Norint patikrinti kurie išoriniai regresoriai veikia prognozuojamas laiko eilutes iš pradžių atliekama priežastingumo analizė naudojant „Gretl“ programinį paketą. Taip galima atrinkti kuriuos kintamuosius verta įtraukti į modelį, taip pat patikrinti kokią įtaką turi jų atsilikimas.

Buvo tikrinama šių regresorių įtaka priklausomam VE generacijos kintamajam:

4 lentelė. Į analizę įtraukiami regresoriai

Regresorius	Žymėjimas
Temperatūra, °C	Temp
Vėjo kryptis, nuo 0° iki 360 °	Kryptis
Vėjo greičio prognozė iš skirtingų prognozės tiekėjų, m/s	DMUVHR
	FORECA
	FORECA Wind
	YR
	FMI Pure
VE prieinamumas, %	SUM_VE_Unavailability
Sezonas, šaltasis - 0, šiltasis - 1	Sezonas
Kitų valstybių VE generacija, MWh	LV
	PL
	SE
	DK

Atlikus priežastingumo analizę su visais išoriniais kintamaisiais, gaunamas rezultatas parodo, kad Švedijos VE generacija ir viena iš vėjo greičio prognozių kiek mažiau veikia Lietuvos VE generaciją. VE generacijai Lietuvoje poveikį daro visi atrinkti kintamieji pasirinkus 3 atsilikimų periodą (*angl. lags*), kuris buvo rekomenduotas pagal *AIC* ir *HQC* kriterijus:

Mean dependent var	124,3037	S.D. dependent var	108,1680
Sum squared resid	5275910	S.E. of regression	17,37511
R-squared	0,974257	Adjusted R-squared	0,974198
F(40, 17476)	16534,48	P-value(F)	0,000000
rho	-0,001290	Durbin-Watson	2,002556

F-tests of zero restrictions:

All lags of SE	F(3, 17476) =	4,4562	[0,0039]
All lags of LV	F(3, 17476) =	130,33	[0,0000]
All lags of PL	F(3, 17476) =	76,342	[0,0000]
All lags of DK	F(3, 17476) =	9,8369	[0,0000]
All lags of SUM_VE_Unavailab-	F(3, 17476) =	5,8887	[0,0005]
All lags of DMUVHRWind	F(3, 17476) =	14,822	[0,0000]
All lags of FORECAWind	F(3, 17476) =	59,466	[0,0000]
All lags of FORECA_WindWind	F(3, 17476) =	11,749	[0,0000]
All lags of YRWind	F(3, 17476) =	31,067	[0,0000]
All lags of FMIPureWind	F(3, 17476) =	2,9679	[0,0306]
All lags of Temp	F(3, 17476) =	10,462	[0,0000]
All lags of Kryptis	F(3, 17476) =	5,9201	[0,0005]
All lags of SUM_VE	F(3, 17476) =	31112,	[0,0000]
All vars, lag 3	F(13, 17476) =	16,181	[0,0000]

21 pav. Poveikio reikšmingumas Lietuvos VE generacijos laiko eilutei (jei $p < 0.05$ - poveikis statistiškai reikšmingas).

Tuomet įtraukus visus lentelėje paminėtus regresorius ir kai kurių iš jų atsilikimus, atrinkti reikšmingiausi regresoriai:

	coefficient	std. error	t-ratio	p-value	
const	-88,7173	1,37412	-64,56	0,0000	***
Temp	-0,985191	0,666201	-1,479	0,1392	
Temp_1	-3,38552	1,23110	-2,750	0,0060	***
Temp_2	4,65299	0,685499	6,788	1,18e-011	***
Kryptis	-0,0212717	0,00381383	-5,578	2,49e-08	***
Kryptis_1	-0,000481206	0,000429126	-1,121	0,2621	
Kryptis_2	-0,000560171	0,000428859	-1,306	0,1915	
Kryptis_3	-0,000435868	0,000428671	-1,017	0,3093	
Kryptis_4	-0,000330740	0,000428504	-0,7718	0,4402	
DMUVHRWind	6,40916	0,306897	20,88	1,11e-095	***
FORECAWind	9,04078	0,249096	36,29	3,80e-278	***
FORECA_WindWind	-0,334578	0,453183	-0,7383	0,4604	
YRWind	11,2919	0,403375	27,99	1,01e-168	***
FMIPureWind	-1,59571	0,149999	-10,64	2,39e-026	***
SUM_VE_Unavailab-	-0,711222	0,131634	-5,403	6,64e-08	***
Sezonas	-4,14813	1,16192	-3,570	0,0004	***
SE	0,00444427	0,00266831	1,666	0,0958	*
SE_1	-0,000651617	0,00452602	-0,1440	0,8855	
SE_2	-1,97192e-05	0,00452738	-0,004356	0,9965	
SE_3	-0,00128675	0,00267318	-0,4814	0,6303	
LV	1,37857	0,105369	13,08	6,26e-039	***
LV_1	0,497547	0,162079	3,070	0,0021	***
LV_2	0,315239	0,162209	1,943	0,0520	*
LV_3	-0,229739	0,104955	-2,189	0,0286	**
PL	0,00253336	0,00293077	0,8644	0,3874	
PL_1	0,0134860	0,00588723	2,291	0,0220	**
PL_2	-4,65475e-05	0,00588524	-0,007909	0,9937	
PL_3	-0,00359869	0,00292550	-1,230	0,2187	
DK	0,00294096	0,00276941	1,062	0,2883	
DK_1	-0,00132064	0,00507915	-0,2600	0,7949	
DK_2	-0,000419023	0,00528200	-0,07933	0,9368	
DK_3	-0,00217972	0,00527607	-0,4131	0,6795	
DK_4	1,38020e-05	0,00527162	0,002618	0,9979	
DK_5	0,000457491	0,00507076	0,09022	0,9281	
DK_6	0,000145571	0,00278007	0,05236	0,9582	

Mean dependent var	124,2622	S.D. dependent var	108,1306
Sum squared resid	32159763	S.E. of regression	42,89416
R-squared	0,842944	Adjusted R-squared	0,842638
F(34, 17479)	2759,185	P-value(F)	0,000000
Log-likelihood	-90664,27	Akaike criterion	181398,5
Schwarz criterion	181670,5	Hannan-Quinn	181488,1
rho	0,903306	Durbin-Watson	0,193511

Excluding the constant, p-value was highest for variable 43 (DK_4)

22 pav. Išorinių regresorių reikšmingumas (skaičius prie regresoriaus nurodo įtrauktą kintamojo atsilikimą).

Gautas AIC kriterijus 181398, kuris įgis prasmę tik kai jį palyginsime su kitu modeliu. Taip pat gautas 84,26 % R^2 . Toliau paliekame tik reikšmingus kintamuosius, kurių $p < 0,01$.

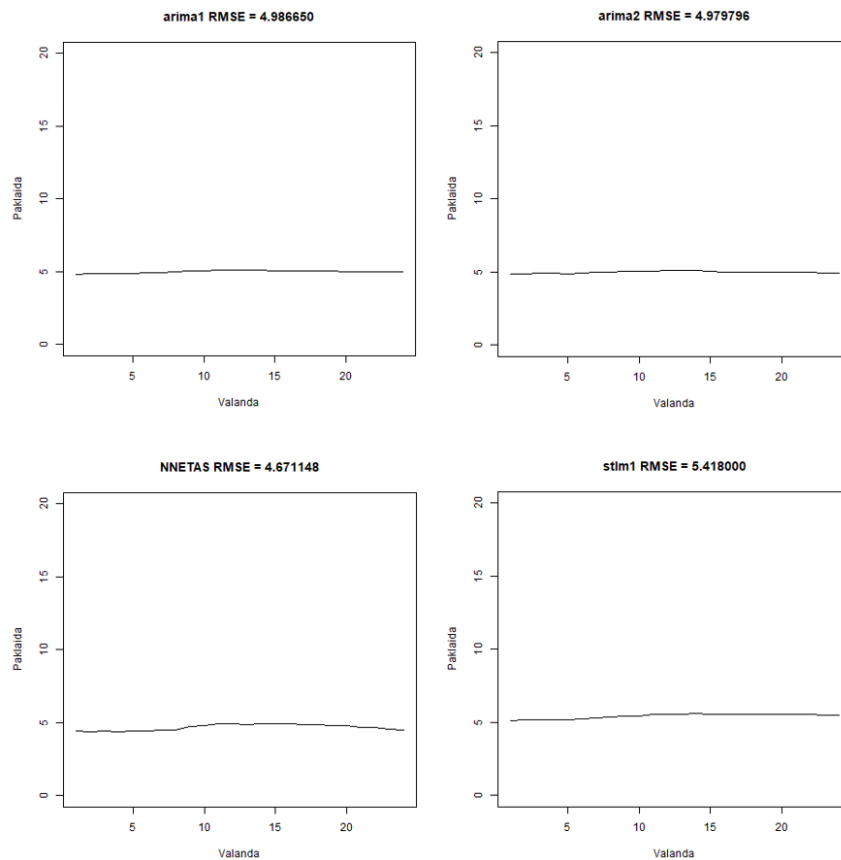
Model 13: OLS, using observations 1:07-730:24 (T = 17514)
 Dependent variable: SUM_VE

	coefficient	std. error	t-ratio	p-value	
const	-88,7635	1,34796	-65,85	0,0000	***
Temp_1	-5,01771	0,496687	-10,10	6,27e-024	***
Temp_2	5,32904	0,493119	10,81	3,89e-027	***
Kryptis	-0,0231459	0,00373837	-6,191	6,09e-010	***
DMUVRWind	6,30156	0,260120	24,23	1,50e-127	***
FORECAWind	8,99190	0,245855	36,57	2,84e-282	***
YRWind	11,1288	0,323685	34,38	9,75e-251	***
FMPureWind	-1,62131	0,148152	-10,94	8,78e-028	***
SUM_VE_Unavailab~	-0,776015	0,129047	-6,013	1,85e-09	***
Sezonas	-4,61795	1,14570	-4,031	5,58e-05	***
SE	0,00241388	0,000294435	8,198	2,60e-016	***
LV	1,38647	0,100810	13,75	8,13e-043	***
LV_1	0,578014	0,101119	5,716	1,11e-08	***
PL_1	0,0174049	0,00113130	15,38	4,60e-053	***
PL_3	-0,00532106	0,00113472	-4,689	2,76e-06	***
Mean dependent var	124,2622	S.D. dependent var	108,1306		
Sum squared resid	32201718	S.E. of regression	42,89759		
R-squared	0,842739	Adjusted R-squared	0,842613		
F(14, 17499)	6698,177	P-value (F)	0,000000		
Log-likelihood	-90675,68	Akaike criterion	181381,4		
Schwarz criterion	181497,9	Hannan-Quinn	181419,7		
rho	0,902393	Durbin-Watson	0,195327		

23 pav. Reikšmingiausi išoriniai kintamieji.

Pašalinus išorinius kintamuosius ir kintamųjų atsilikimus gaunamas $AIC = 181381$, tai yra geresnis rezultatas nei pirmojo modelio (naudojant visus kintamuosius). Todėl toliau darbe, kaip išoriniai regresoriai, bus naudojami tik šie kintamieji ir jų atsilikimai.

Atliekant II etapo tyrimą, atrenkant kiekvienam modeliui naudojamus parametrus kaip ir pirmajame etape buvo taikomas kryžminis patikrinimas. Keletas kryžminio patikrinimo pavyzdžių pateikta apačioje.



24 pav. ARIMAX(2,0,2), ARIMAX(3,0,2), NNETAR ir STLM modelių paklaidų priklausomybė nuo laiko, atlikus kryžminį patikrinimą daugiamačiams modeliams (vienos VE pavyzdys).

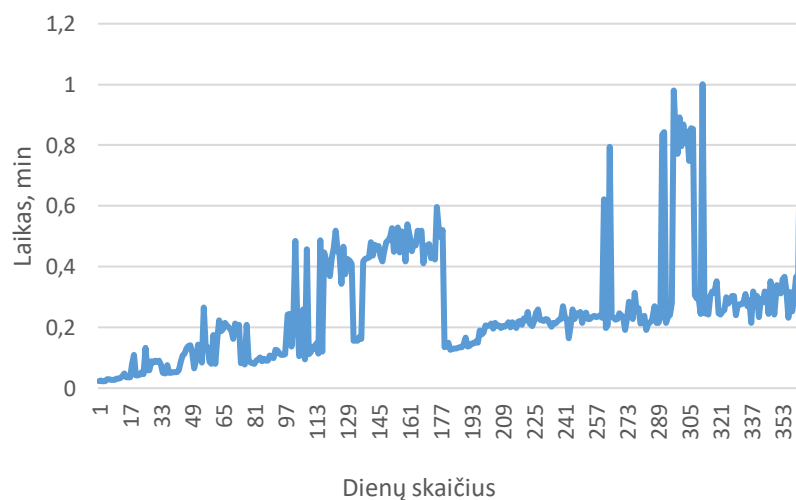
Šį kartą *ARIMAX* modelio *AR*, *I* ir *MA* reikšmės buvo pasirinktos atlikus automatizuotą modelio apmokymą ir pasirinkus 2 modelius su mažiausiais *AIC*, taigi buvo pasirinkti *ARIMAX(2,0,2)*, *ARIMAX(3,0,2)* modeliai. *STLM* modelyje metodas buvo pasirinktas - *ARIMA*, nes iš visų *STLM* galimų metodų jis vienintelis sugeba tvarkytis su išoriniais kintamaisiais. Paskutinis tikrinamas metodas buvo *NNETAR*. Visi šie 4 metodai buvo pasirinkti atsižvelgiant į jų galimybę įtraukti išorinius kintamuosius kuriant modelį.

3.3.4. Prognozės sudarymo būdai

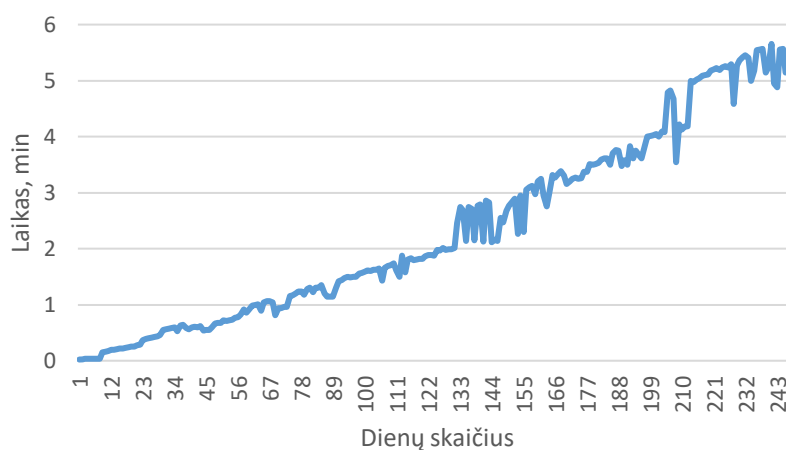
II etapo prognozavimas buvo atliekamas 2 būdais:

- 1) Kiekvieną modelį apmokant su 1 metų duomenimis ir vėliau, periodiškai, pagal parinktus parametrus generuojant prognozę 38 val. į priekį išmetant pirmąsias 14 val. Šiuo būdu parametrai nustatomi tik 1 kartą.
- 2) Kiekvieną modelį apmokant kaskart gavus papildomas 24 val. duomenų. Šis būdas skirtas kaskart, turint didesnę imtį istorijos, perskaičiuoti parametrus ir generuoti prognozę su naujaisiais parametrais. Tuo tarpu pirmajame būde generuojamos prognozės visuomet naudoja tuos pačius parametrus gautus apmokus modelius su 1 metų imtimi.

Antrasis būdas reikalauja kur kas daugiau resursų, nes, ypač apmokant *NNETAR* modelį, skaičiavimai užtrunka, o palaipsniui didėjant duomenų kiekiui jie tampa vis ilgesni. Pavyzdžiui, sudarant prognozę 1 metams, kai kaskart prognozuojama 38 val. periodui į priekį, generuojant 364 dienos prognozę reikia apmokyti modelį turint 363 dienų istorinius duomenis. O sudarant 365 dienos prognozę, modelio apmokymas vyksta iš naujo ir į modelio apmokymą įtraukiamos papildomos 24 val.



25 pav. Modelių apmokymo trukmės priklausomybė, nuo apmokymo imties ilgio (1 būdas)



26 pav. Modelių apmokymo trukmės priklausomybė, nuo apmokymo imties ilgio (2 būdas)

Prognozavimo modelių kūrimui galiausiai nuspręsta naudoti tik 3 vėjo greičio prognozes, su kuriomis apmokant modelius buvo gauti geriausi rezultatai. Apačioje pavaizduoti II etapo abejais būdais gauti rezultatai.

Vėjo elektrinė	Paklaida, nMAE				Prognozių Vidurkis
	ARIMAX 1	ARIMAX 2	STLM	NNETAR	
Benaičiai	9,6%	9,7%	9,9%	9,3%	9,3%
Kreivėnai	9,1%	9,1%	9,2%	9,8%	8,9%
Lauksargiai	9,5%	9,5%	10,0%	9,9%	9,3%
Šiauduva	13,0%	13,0%	14,5%	13,8%	12,8%
Sūdėnai	8,6%	8,6%	10,0%	8,3%	8,4%
Vėjas_1	9,5%	9,5%	9,9%	9,1%	9,0%
Čiuteliai	10,6%	10,6%	10,6%	10,2%	9,9%
Didšiliai	9,1%	9,1%	10,3%	9,3%	8,9%
Geišiai	12,4%	12,4%	13,6%	12,8%	12,1%
Strepeikiai	7,9%	8,0%	9,7%	20,3%	9,5%
Kabaldikai	12,3%	12,4%	13,4%	12,4%	11,9%
Šyša	14,2%	14,2%	15,9%	14,5%	13,2%
Kunigiškiai	15,7%	15,8%	16,0%	16,7%	15,2%
VE prognozė	7,9%	7,8%	8,8%	9,1%	7,6%
VE suma iš atskirų VE prognozių	7,9%	8,0%	8,5%	8,9%	7,7%

27 pav. Daugiamačių modelių rezultatai (1 būdas)

Vėjo elektrinė	Paklaida, nMAE				Prognozių Vidurkis
	ARIMAX 1	ARIMAX 2	STLM	NNETAR	
Benaičiai	9,1%	9,1%	9,3%	9,9%	8,7%
Kreivėnai	9,0%	9,0%	9,2%	9,5%	8,6%
Lauksargiai	9,5%	9,5%	10,0%	10,2%	9,2%
Šiauduva	13,4%	13,4%	15,7%	15,0%	13,2%
Sūdėnai	7,9%	7,9%	8,6%	8,6%	7,7%
Vėjas_1	8,5%	8,5%	9,0%	8,8%	8,3%
Čiuteliai	10,2%	10,3%	10,3%	11,0%	9,6%
Didšiliai	9,0%	8,9%	9,9%	9,7%	8,7%
Geišiai	12,5%	12,5%	13,8%	13,1%	12,2%
Strepeikiai	8,1%	8,2%	9,0%	20,5%	9,6%
Kabaldikai	12,2%	12,2%	13,1%	12,1%	11,7%
Šyša	14,4%	14,4%	15,4%	14,8%	13,4%
Kunižiškiai	15,7%	15,9%	15,8%	16,6%	15,2%
VE prognozė	8,0%	8,0%	8,8%	9,1%	7,6%
VE suma iš atskirų VE prognozių	7,7%	7,8%	8,1%	8,7%	7,5%

28 pav. Daugiamačių modelių rezultatai (2 būdas)

Pagal daugiamačių modelių rezultatų lenteles pastebima, kad prognozuojant tiek 1, tiek 2 būdais *STLM* modelis nei karto nebuvo pranašesnis už kitus. Aiškiai matoma, kad *NNETAR* modelis prognozuojant 1 būdu sugebėdavo dažniau aplenkti kitus modelius. Prognozių kombinavimas vidurkinant (t.y. naudojant vienodus svorius) kaip ir I etape pranoko individualius modelius. Geriausias pasiektas rezultatas buvo visų modelių vidurkis, kai kiekviena VE prognozė buvo sumuojama sudarant suminę VE prognozę, paklaida – 7,5 % *nMAE*.

3.3.5. III etapas. Prognozavimas naudojant daugiamačius metodus, prie išorinių regresorių įtraukiant prognozavimo sistemų paruoštas prognozes

Šiame etape siekiama dar labiau sumažinti prognozės paklaidas, panaudojus 2 skirtingų sistemų teikiamas suminės VE generacijos prognozes. Tikslas - panaudoti visą prieinamą informaciją ir kurti kombinacijas pasitelkiant turimas prognozes, išorinius kintamuosius, bei istorinius duomenis taip minimizuojant paklaidą. Žingsniai šiame etape analogiški II etapo žingsniams, išskyrus tai, kad naudojamas tik II etapo 1 būdas, atsisakant 2 būdo dėl didelių skaičiavimo resursų sąnaudų. Šiame etape kaip išoriniai regresoriai įtraukiamos 2 nepriklausomų sistemų prognozės ir 3 skirtingos vėjo greičio prognozės. Išorinių sistemų pavieniui paklaidos 6,04 % ir 6,95 % *nMAE*. Gautas rezultatas pateikiamas žemiau.

Vėjo elektrinė	Paklaida, nMAE							
	ARIMAX 1	ARIMAX 2	STLM	NNETAR	Prognозиų Vidurkis	KOMB.	BEST	LOESS
Benaičiai	7,9%	8,0%	9,3%	8,4%	7,80%	7,79%		
Kreivėnai	7,8%	7,8%	10,5%	8,5%	7,92%	7,81%		
Lauksargiai	8,3%	8,2%	9,2%	8,9%	8,18%	8,23%		
Siauduva	11,1%	11,0%	11,9%	12,3%	11,18%	11,04%		
Sudėnai	7,5%	7,5%	9,1%	8,3%	7,53%	7,56%		
Vėjas_1	7,7%	7,7%	9,0%	8,6%	7,61%	7,56%		
Čiuteliai	8,3%	8,3%	8,5%	9,6%	8,30%	8,25%		
Didšiliai	7,6%	7,6%	10,0%	8,7%	7,81%	7,66%		
Geišiai	10,7%	10,8%	12,0%	11,5%	10,79%	10,76%		
Strepeikiai	7,6%	7,6%	10,2%	13,7%	8,35%	7,90%		
Kabaldikai	11,0%	11,1%	12,1%	12,4%	10,98%	10,94%		
Šyša	11,8%	11,8%	12,5%	13,0%	11,56%	11,61%		
Kunigiškiai	13,8%	13,9%	15,0%	14,4%	13,54%	13,48%		
VE prognozė	6,24%	6,20%	9,20%	7,17%	6,45%	5,90%	6,18%	✓ 5,84%
VE suma iš atskirų VE prognozių	6,03%	6,03%	6,65%	7,69%	6,00%	5,96%		

29 pav. Daugiamačių modelių sukurtų naudojamų sistemų prognozes rezultatai

Trečiame etape paklaidos yra dar mažesnės, *ARIMAX 1* ir *ARIMAX 2* modeliams pavyko aplenkti kiekvieną sistemą pavieniui, kai kiekvieno parko prognozės buvo agreguojama į vieną. Pastebima, kad 3 etape prognozių vidurkis rečiau pranoksta pavienes modelių prognozes. Šiame etape taip pat buvo atliekamas prognozių kombinavimas, kitaip tariant, kolektyvo kūrimas.

3.3.6. Prognozių kolektyvo kūrimas

Prognozių kolektyvas buvo kuriamas 2 būdais:

- 1) Renkamas geriausias praėjusios paros modelis ir sekančiai parai naudojamas tik jis.
- 2) Modeliai reitinguojami pagal praėjusios paros rezultatus ir sekančiai parai parenkami atitinkami koeficientai.

Pirmasis kolektyvo kūrimas buvo panaudotas tik suminei VE generacijos laiko eilutei. Jo paklaida buvo 6,18 % *nMAE* (29 – BEST), šis kolektyvas pralenkė prognozių vidurkį 0,27 procentinio punkto. Kuriant antrąjį kolektyvinį modelį buvo vertinamas kiekvienos prognozės tikslumas praėjusiai parai. Modelis rodęs mažiausią paklaidą praėjusią parą dauginamas iš koeficiento – 0,4, antros vietos modelis dauginamas iš koeficiento – 0,3, trečią ir ketvirtą vietas užėmę modeliai dauginami iš 0,2 ir 0,1 atitinkamai. Šis modelių kolektyvas buvo pranašesnis už geriausio modelio kolektyvą, jo paklaida – 5,9 % *nMAE* (29 – KOMB.) - iki šiol geriausias rezultatas.

Nors kombinuotas modelių kolektyvas, su skirtingais svoriais pasirodė tikrai gerai. Toliau buvo siekiama tikslesnės prognozės. Tam buvo panaudota Loess regresija. Loess regresijai, keičiant parametrus su vieno mėnesio duomenimis buvo atrinktas optimaliausias *span* koeficientas, kuris lygus 0,35. Loess regresija turi galimybę naudoti tik 4 išorinius kintamuosius, taigi atrinkinėjant kintamuosius, taip pat buvo naudojama 1 mėn. duomenų imtis. Išorinių kintamųjų atranka vyko pagal mažiausią paklaidą. Buvo atrinkti šie išoriniai kintamieji, kurie buvo panaudoti suminei VE generacijos laiko eilutei prognozuoti: 1 ir 2 sistemos prognozės, vėjo greitis ir neprieinamumas. Žvelgiant į paklaidų lentelę (29 pav. – LOESS.) pastebima, kad pritaikius LOESS regresiją buvo gautas absoliučiai geriausias rezultatas ir paklaida minimizuota iki 5,84 % *nMAE*.

3.4. VE generacijos prognozių modelių ekonominis įvertinimas

Finansinei naudai nustatyti atliekamas ekonominis įvertinimas, kurio tikslas prognozių modelių tikslumą išreikšti finansine išraiška (EUR). Ekonominis įvertinimas darbe atliekamas naudojantis atsinaujinančių išteklių energetikos įstatyme aprašyta tvarka (AEI įstatymas, 2011), Valstybinės kainų ir energetikos kontrolės komisijos nustatytais tarifais (Tarifai, 2017), AEI balansavimo biudžetu 2018 metams (Biudžetas, 2017), Elektros energijos disbalanso kainomis 2018 metais (Baltic CoBA Dashboard, 2018) ir Elektros energijos biržos „NordPool“ kainomis (NordPool, 2019).

Norint apskaičiuoti VE balansavimo kaštus, reikia įvertinti skirtumą tarp prognozės ir fakto. Skirtumas apskaičiuotas kiekvienai valandai dauginamas iš balansavimo kainos ir vidutinės mėnesio elektros energijos biržos kainos skirtumo, atitinkamai valandai. Skaičiuojant VE balansavimo kaštus gali būti scenarijų kai balansavimo kaštai yra neigiami, kitaip sakant VE nuokrypis generuoja pajamas. AEI balansavimo kaštų formulė, kai *n* yra valandų skaičius:

$$P_{bal.} = \sum_{i=1}^n (|E_{fakt.} - E_{progn.}|) \cdot P_d. \quad (42)$$

čia $P_{bal.}$ - balansavimo kaštai, $E_{fakt.}$ - faktiškai VE pagaminta energija, $E_{progn.}$ - VE gamybos prognozė, $P_d.$ - kainos dedamoji, kurios apskaičiavimas parodytas žemiau.

Jei $E_{fakt.} - E_{progn.} > 0$:

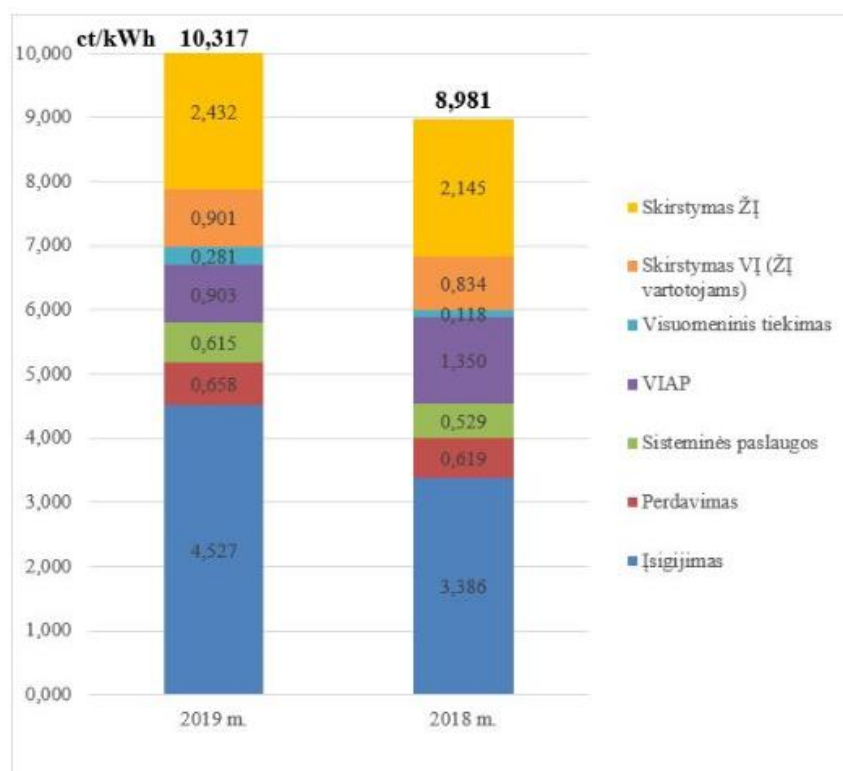
$$P_d. = P_{birž.} - P_{disbal.} \quad (43)$$

Jei $E_{fakt.} - E_{progn.} < 0$:

$$P_d. = P_{disbal.} - P_{birž.} \quad (44)$$

Vietoje $E_{progn.}$ į (43) formulę įstačius modeliuojamas prognozes, galima patikrinti kokie bus balansavimo kaštai taikant vieną ar kitą modelį. Čia, $P_{disbal.}$ - disbalanso kaina, kuri kiekvienai valandai yra skirtinga, $P_{birž.}$ – elektros energijos biržoje susiformavusi kaina, kuri taip pat kiekvienai valandai yra skirtinga.

Skaičiuojant įtaką galutinei elektros energijos kainai naudojamas 2018 metų viešuosius interesus atitinkančių paslaugų biudžetas (VIAP) iš VKEKK.



30 pav. Vidutinės elektros energijos kainos 2018 m. ir 2019 m., ct/kWh (be PVM)

(<https://www.regula.lt/Puslapiai/naujienos/2018-metai/2018-lapkritis/2018-11-30/komisija-patvirtino-2019-m-elektros-energijos-tarifus-buitiniams-vartotojams.aspx>)

Pagal 30, 2019 metams VIAP dedamosios kaina visiems asmenims – 0,903 ct/kWh be PVM.

Bendros lėšos 2019 metams elektros energetikos įmonėms, teikiančioms VIAP - 99,576 mln. Eur.

Iš šios sumos, perdavimo sistemos operatoriui už elektros energijos, pagamintos naudojant atsinaujinančius energijos išteklius, balansavimą skirta 4,769 mln. EUR. Pritaikius proporciją apskaičiuojame, kad 0,043 ct/kWh elektros energijos kainos dedamosios sudaro VE generacijos balansavimas.

Toliau pateikiami sumodeliuoti scenarijai, kuriuose suskaičiuojama kokie būtų buvę AEI balansavimo kaštai 2018 m., jei vietoje faktiškai naudotos VE generacijos prognozės būtų naudojamos tyrime sumodeliuotos VE generacijos prognozės. Palyginimui modeliuojami scenarijai pasitelkiant šias prognozes iš kiekvieno etapo:

I etapas: *TBATS* ir modelių vidurkis

II etapas: *ARIMAX (2,0,2)*, *ARIMA (3,0,2)*, *STLM*, *NNETAR* ir modelių vidurkis

III etapas: *ARIMAX (2,0,2)*, *BEST* kolektyvas, *KOMBINACIJA* kolektyvas ir *LOESS* regresija

Prognozės modelis	Rodiklis	Viso
Sistema 1	Balansavimo sąnaudos, tūkst. EUR	4510,82
	Paklaida, nMAE %	6,95%
Sistema 2	Balansavimo sąnaudos, tūkst. EUR	3370,14
	Paklaida, nMAE %	6,04%
Galutinė prognozė	Balansavimo sąnaudos, tūkst. EUR	4085,24
	Paklaida, nMAE %	6,12%
TBATS	Balansavimo sąnaudos, tūkst. EUR	4935,56
	Paklaida, nMAE %	17,82%
AVG	Balansavimo sąnaudos, tūkst. EUR	5339,83
	Paklaida, nMAE %	17,53%
ARIMAX 1	Balansavimo sąnaudos, tūkst. EUR	4843,25
	Paklaida, nMAE %	7,93%
ARIMAX 2	Balansavimo sąnaudos, tūkst. EUR	4863,07
	Paklaida, nMAE %	7,97%
STLM	Balansavimo sąnaudos, tūkst. EUR	4263,84
	Paklaida, nMAE %	8,48%
NNETAR	Balansavimo sąnaudos, tūkst. EUR	4906,54
	Paklaida, nMAE %	7,70%
AVG	Balansavimo sąnaudos, tūkst. EUR	4798,33
	Paklaida, nMAE %	7,70%
ARIMAX 1	Balansavimo sąnaudos, tūkst. EUR	4179,64
	Paklaida, nMAE %	6,03%
BEST	Balansavimo sąnaudos, tūkst. EUR	4102,58
	Paklaida, nMAE %	6,18%
KOMBINACIJA	Balansavimo sąnaudos, tūkst. EUR	4103,92
	Paklaida, nMAE %	5,90%
LOESS	Balansavimo sąnaudos, tūkst. EUR	3222,89
	Paklaida, nMAE %	5,84%

31 pav. Balansavimo kaštai pritaikius tyrime modeliuojamas VE generacijos prognozes

31 pav. lentelėje „Galutinė prognozė“ eilutėje pažymėtas rezultatas, kuris buvo gautas 2018 m. kombinuojant prognozes iš 2 skirtingų sistemų. Žvelgiant į rezultatus galima pastebėti, kad VE generacijos prognozės paklaidos yra susiję su balansavimo sąnaudomis, tačiau yra atvejų, kai esant mažesnei paklaidai balansavimo kaštai didesni. Taip nutinka dėl nevienodai pasiskirsčiusių

energijos disbalanso kainų skirtingoms valandoms. Pavyzdžiui prognozei nukrypus 5 MWh, 09:00 val. ryto, už disbalansą gali tekt susimokėti 70 EUR/MWh kainą, o nukrypus tą patį kiekį vakare, 22:00 val., už disbalansą gali tekt susimokėti 30 EUR/MWh kainą. Šios kainos kinta kas valandą ir priklauso nuo balanso elektros energijos sistemoje. Taigi, žinoma, kad nukrypus apskritai - balansavimo kaštų nebus. Kuo didesnis nuokrypis, tuo didesnė tikimybė, kad balansavimo kaštai išaugs. Toliau analizuojant lentelę pastebima, kad pritaikius *STLM* modelį, 2018 m. balansavimo kaštai būtų apie 247 tūkst. EUR mažesni. Mažiausi kaštai, kaip ir mažiausia paklaida gaunama pritaikius *LOESS* modelį, su kuriuo sutaupoma 863 tūkst. EUR.

Toliau apskaičiuojama galutinės elektros energijos kaina padarius prielaidą, kad vietoje 4,769 mln. EUR suplanuoto AEI balansavimo biudžeto 2019 metams, bus naudojami AEI balansavimo kaštai gauti pritaikius *LOESS* modelį. Šie kaštai padidinami 5 %, padarius dar vieną prielaidą dėl galimai išaugsiančių disbalanso kainų 2019 m. Taigi gaunamas perskaičiuotas 2019 m. biudžetas – 3,383 mln. EUR. Taikant tą pačią proporciją aprašytą 3.4 poskyrio pradžioje galime apskaičiuoti, kokia būtų galutinė elektros energijos kaina vartotojui su nauju biudžetu. Gaunamas rezultatas – AEI balansavimui skirta tarifo dedamoji sumažėtų nuo 0,043 ct/kWh iki 0,031 ct/kWh, vadinasi bendra elektros energijos kaina sumažėtų 0,012 ct/ kWh. 2018 m. Lietuvos gyventojai suvartojo 2,894 TWh elektros energijos, tai yra $2,894 \times 10^9$ kWh. Sudauginus elektros energijos suvartojimą ir elektros energijos kainos sutaupymą, gaunamas rezultatas – Lietuvos namų ūkių savininkai būtų sumokėję 347 tūkst. EUR mažiau. Jei vidutinis Lietuvos namų ūkis suvartoja vidutiniškai 200 kWh elektros energijos per mėnesį, tai per metus šis namų ūkis būtų sutaupęs 28,8 ct. Žvelgiant į individualius gyventojus šis sutaupymas nėra pastebimas, tačiau bendru mastu sutaupomi pinigai gali būti reikšmingi panaudojant juos energetikos plėtrai ar kitose srityse.

Apibendrinant tyrimą verta paminėti, kad prognozuojant VE generacijos laiko eilutes daugiamačiai modeliai daug pranašesni, nei vienmačiai. Modelių kolektyvų kūrimas suteikia pozityvių rezultatų ir padeda pasiekti mažesnę paklaidą. Apmokymo imties koregavimas neįdomus, tačiau prisideda prie geresnių rezultatų, tačiau laiko atžvilgiu nėra naudingas. Prognozavimo sistemų tikslumą pavyksta pagerinti tik įtraukus pačių sistemų prognozės rezultatus

Apibendrinant ekonominį įvertinimą verta paminėti, kad per 2018 metus potencialiai sutaupomi 863 tūkst. EUR balansavimo kaštai yra pakankamai didelė suma, dėl kurios verta plėtoti prognozavimo modelius. Siekiant dar didesnės ekonominės naudos, kuriant prognozių modelius vertėtų atkreipti dėmesį į balansavimo rinkos situaciją. Teisingai pritaikius VE generacijos prognozės modelį ir jį sujungus su prekiavimo elektros energija balansavimo rinkoje strategija, būtų galima siekti dar didesnio sutaupymo.

Išvados

1. Tyrimo metu sukurta 17 skirtingų modelių, kurie prognozuoja 38 val. periodui į priekį. Atliktas tyrimas atskleidė, kad taikant vienmačius metodus VE generacijos prognozių *nMAE* paklaida kur kas aukštesnė nei taikant daugiamačius metodus. Prognozuojant suminę VE generaciją tiksliausias vienmačių metodų modelis buvo *NNETAR* su 17,8% *nMAE*. Atskirai VE generacijai prognozuoti tinkamiausias iš vienmačių modelių pasirodė *TBATS*, kuris sudarė tiksliausias prognozes šešioms iš trylikos VE. Agreguojant atskiras to paties modelio VE prognozes į bendrą, 12-ai iš 18-os vienmačių modelių agregavimas pasiteisina - gaunama paklaida yra žemesnė nei prognozuojant suminę VE generaciją. Agregavimas labiausiai pasiteisino su *ARIMA (2,1,1)* modeliu, kur paklaida nukrito iš 20,2% į 17,44% *nMAE*. Pastebima, kad sudarius skirtingų modelių kolektyvą su vienodais svorių koeficientais kiekvienos VE generacijai ir suminei VE generacijai, paklaida yra žemesnė nei bet kurio modelio atskirai. Tiksliausia vienmačio metodo prognozė gauta sudarius modelių kolektyvą su vienodais svorio koeficientais suminei VE generacijos prognozei ir yra lygi 17,41 % *nMAE*.
2. Atliekant prognozių modeliavimą su keturiais daugiamačiais modeliais, modeliavimui taikyti 2 skirtingi būdai parodė panašius rezultatus. Pirmasis būdas, kai modelis apmokomas su vienerių metų imtimi ir pagal nustatytus parametrus prognozuoja kiekvienai kitų metų parai, šiek tiek nusileido antrajam būdai, *nMAE* – 7,7% ir 7,5% atitinkamai. Antrasis būdas, kai daugiamačis modelis apmokomas kiekvieną parą, buvo šiek tiek tikslesnis, tačiau reikalavo 25 kartus daugiau resursų. Daugiamačių modelių kolektyvų modeliavime pastebima ta pati tendencija – visų modelių vidurkis kiekvienu atveju tikslesnis nei bet kurio atskiro modelio. Žemiausia II etape pasiekta paklaida - 7,5% *nMAE* - 2,3 karto mažesnė nei geriausia vienmačių modelių paklaida. Tai patvirtina, kad išorinių kintamųjų, ypač vėjo greičio prognozės, įtraukimas daro teigiamą įtaką prognozės tikslumui.
3. Trečiajame etape atliktas modeliavimas atskleidė, kad pranokti kiekvienos iš sistemų prognozių tikslumą nėra paprastas, tačiau įmanomas uždavinys. *ARIMA (3,0,2)* ir *ARIMA (2,0,2)* modeliai aplenkė tikslesnę sistemą viena šimtają procentinio punkto ir parodė 6,03% *nMAE* rezultata. *STLM* ir *NNETAR* modeliai buvo mažiau tikslūs, 6,65% ir 7,69% *nMAE* atitinkamai. III etape atliktas modelių kolektyvų kūrimas davė teigiamą rezultatą. Kolektyvas, sukurtas kiekvienai modelio prognozei pritaikant svorius pagal šio modelio praėjusios paros rezultatus, buvo sėkmingas ir sumažino paklaidą iki 5,9% *nMAE*. Šio kolektyvo rezultatas 10-čiai VE iš 13-os parodė geresnį rezultatą nei kolektyvo, sudaryto iš

vienodų svorių, rezultatas. Kolektyvas, sudarytas pagal praėjusios paros tiksliausią modelį prognozuojant ateinančiai parai, nebuvo toks tikslus, jo paklaida – 6,18% *nMAE*. Loess regresija šiame tyrime buvo pranašiausia. Loess regresijos modelis su 0,35 glodinimo koeficientu, įtraukus 2-iejų sistemų prognozes, neprieinamumą ir vėjo greičio prognozę, buvo tiksliausias su 5,84% *nMAE* paklaida.

4. Pritaikius geriausią prognozės modelį ir sumodeliavus situaciją 2018 metams buvo apskaičiuoti VE balansavimo kaštai, kurie siekė 3,383 mln. EUR. Lyginant su faktiniu 2018 metų rezultatu šie kaštai buvo 863 tūkst. EUR mažesni. Vidutiniam elektros energijos vartotojui, suvartojančiam 200 kWh per mėnesį, apskaičiuotas 28 ct sutaupymas per metus.

Tolimesniam VE generacijos prognozės plėtojimui būtų galima atlikti platesnę kolektyvinių modelių analizę ir kurti modelių kolektyvą pagal ilgesnio nei vienos paros periodo rezultatus. Taip pat, tyrimą galima papildyti apmokymo imties keitimu slenkančio lango principu, kai kaskart apmokant modelį iš apmokymo imties išmetami senesni stebėjimai ir pridedami naujaisi, lango periodą paliekant tokį pat. Hibridinių modelių taikymas galėtų būti dar vienas tikslesnės prognozės ieškojimo sprendimas.

Norint sumažinti paklaidą ir VE balansavimo kaštus vertėtų prognozuoti trumpesnio horizonto periodui, galbūt panaudojant realaus laiko duomenis. Šį darbą taip pat galima pritaikyti siekiant prognozuoti disbalanso kainą elektros energijos balansavimo rinkoje, taip siekiant dar labiau sumažinti vėjo elektrinių generacijos balansavimo kaštus.

Bibliografija

1. AEI įstatymas, L. (2011). Atsinaujinančių išteklių energetikos įstatymas. Vilnius. Nuskaityta iš <https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.398874>
2. Aiolfia, M., & Timmermann, A. (2005). Persistence in forecasting performance and conditional combination strategies. *135*(1-2), 31-53. Nuskaityta iš <https://doi.org/10.1016/j.jeconom.2005.07.015>
3. Armstrong, J. S. (1978). *Long-range forecasting: From crystal ball to computer*. John Wiley & Sons.
4. Baltic CoBA Dashboard. (2018). Disbalanso kainos. Nuskaityta iš <https://dashboard-baltic.electricity-balancing.eu/en/imbalance/prices>
5. Bax Energy. (2019). Nuskaityta iš <https://www.baxenergy.com/power-forecasting-energy-trading/>
6. Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *32*(2)(2), 303-312. Nuskaityta iš <https://doi.org/10.1016/j.ijforecast.2015.07.002>
7. Biudžetas. (2017). Dėl viešuosius interesus atitinkančių paslaugų lėšų ir kainos 2018 metams nustatymo. Nuskaityta iš <https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/cbfd6eb2be7e11e7af36e75c0ac79247?jfwid=-1fhjp5q94>
8. Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches. *11*(7). Nuskaityta iš <https://doi.org/10.3390/en11071636>
9. Box, G. E., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
10. Cadenas, E., Rivera, W., Campos-Amezcu, R., & Heard, C. (2016). Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model. *9*(2), 109. Nuskaityta iš <https://doi.org/10.3390/en9020109>
11. Chatfield, C. (2000). Time-Series Forecasting. Nuskaityta iš <https://doi.org/10.1201/9781420036206>
12. Condeixa, L. D., Bastos, L. d., Oliveira, F. L., & Barbosa, S. D. (2017). Wind speed time series analysis using TBATS decomposition and moving blocks bootstrap. *International Journal of Energy and Statistics*. Nuskaityta iš <https://www.researchgate.net/publication/318465237>

13. Cui, H., & Peng, X. (2015). Short-Term City Electric Load Forecasting with Considering Temperature Effects: An Improved ARIMAX Model. *Mathematical Problems in Engineering*, 2015. Nuskaityta iš <http://dx.doi.org/10.1155/2015/589374>
14. Dumitru, C.-D. (2018). Wind energy forecasting: a comparative study between stochastic model ARIMA and model based on Neural Network. 32(1), 410-417. Nuskaityta iš <https://doi.org/10.1016/j.promfg.2019.02.234>
15. Esteves, & T, G. R. (2015). Long Term Electricity Forecast: A Systematic Review. Nuskaityta iš <https://reader.elsevier.com/reader/sd/pii/S1877050915015161?token=806ED4620BE9A36BA288FBBA1EB0F09F5A7DD9B921B5D0F5CEF6E999D036A54C6B127F3F6AD2217D3EFA5575997EB22>
16. Fox, J. (2010). Nonparametric Regression in R: An Appendix to An R Companion to Applied Regression, 2nd edition. Nuskaityta iš <https://socialsciences.mcmaster.ca/jfox/Books/Companion/appendices/Appendix-Nonparametric-Regression.pdf>
17. Fuller, W. A. (1976). Introduction to Statistical Time Series. New York. doi:ISBN 0-471-28715-6
18. Gilliland, M. (2011). Value Added Analysis: Business forecasting effectiveness. Nuskaityta iš <http://analytics-magazine.org/value-added-analysis-business-forecasting-effectiveness/>
19. Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., & Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, 191. Nuskaityta iš <https://www.sciencedirect.com/science/article/abs/pii/S0377221707008740?via%3Dihub>
20. Gritmit, E. P., & Mass, C. F. (2002). Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. Nuskaityta iš <https://journals.ametsoc.org/doi/full/10.1175/1520-0434%282002%29017%3C0192%3AIROAMS%3E2.0.CO%3B2>
21. Grushka-Cockayne, Y. (2019). The Better Way to Forecast the Future. Nuskaityta iš <https://hbswk.hbs.edu/item/the-better-way-to-forecast-the-future>
22. Hendry, D. F., & Clements, M. P. (2004). Pooling of forecasts. 7(1), 1–31. Nuskaityta iš <https://doi.org/10.1111/j.1368-423X.2004.00119.x>
23. Hyndman, R. J. (2010). Business Forecasting Methods. *International Encyclopedia of statistical Science*. Nuskaityta iš https://doi.org/10.1007/978-3-642-04898-2_156
24. Hyndman, R. J., & Athanasopoulos, G. (2014). Forecasting: principles and practice. Nuskaityta iš <https://otexts.com/fpp2/>

25. Hyndman, R. J., & Fan, S. (2010). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on power systems*. Nuskaityta iš <https://robjhyndman.com/papers/2010STLF-FinalR1.pdf>
26. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22. Nuskaityta iš <https://robjhyndman.com/publications/automatic-forecasting/>
27. Hyndman, R., & Athanasopoulos, G. (2014). Forecasting: principles and practice. Nuskaityta iš <https://otexts.com/fpp2/>
28. Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages. 20(1), 5-10. Nuskaityta iš <https://doi.org/10.1016/j.ijforecast.2003.09.015>
29. Hui, C., & Chao, L. (2019). Data processing strategies in wind energy forecasting models and applications: A comprehensive review. 249(1), 392-408. Nuskaityta iš <https://doi.org/10.1016/j.apenergy.2019.04.188>
30. John Wiley & Sons. (2015). Statistics for Big Data for Dummies. Nuskaityta iš <http://index-of.co.uk/Big-Data-Technologies/Statistics%20for%20Big%20Data%20For%20Dummies%20-%20Alan%20Anderson.pdf>
31. Kline, D. (2004). Methods for Multi-Step Time Series Forecasting with Neural Networks. Nuskaityta iš <https://doi.org/10.4018/978-1-59140-176-6.ch012>
32. Kolassa, S. (2011). Combining exponential smoothing forecasts using Akaike weights. 27(2), 238-251. Nuskaityta iš <https://doi.org/10.1016/j.ijforecast.2010.04.006>
33. Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? 54(1-3), 159-178. Nuskaityta iš [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
34. Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of American Statistics Association*, 106(496), 1513-1527. Nuskaityta iš <http://dx.doi.org/10.1198/jasa.2011.tm09771>
35. Macready, D. H., & G., W. (1997). Transactions on Evolutionary Computation. *No Free Lunch Theorems for Optimization*. Nuskaityta iš <https://ti.arc.nasa.gov/m/profile/dhw/papers/78.pdf>
36. Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. 16(4), 451-476. Nuskaityta iš [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1)

37. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *34*(4), 802-808. Nuskaityta iš <https://doi.org/10.1016/j.ijforecast.2018.06.001>
38. Masum, S., Liu, Y., & Chiverton, J. (2018). Multi-step Time Series Forecasting of Electric Load using Machine Learning Models. Nuskaityta iš https://doi.org/10.1007/978-3-319-91253-0_15
39. NordPool. (2019). Elektros energijos biržos kainos. Oslo. Nuskaityta iš <https://www.nordpoolgroup.com/Market-data1/Dayahead/Area-Prices/ALL1/Hourly/?view=table>
40. O'Connell, M. (2019). Using Predictive Analytics To Forecast Business Performance. Nuskaityta iš <https://digitalstrategy.ie/benefits-of-using-predictive-analytics-to-forecast-business-performance/>
41. Pegels, C. C. (1969). Exponential forecasting: Some new variations. *15*(5). Nuskaityta iš <https://doi.org/10.1287/mnsc.15.5.311>
42. Peters, A. C., & Harvey, S. (1997). Estimation Procedures for Structural Time Series Models. *Journal of Forecasting*. Nuskaityta iš <https://pdfs.semanticscholar.org/2bd2/6de225003be0c281e1373614b75f6991cc54.pdf>
43. Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Nuskaityta iš <https://doi.org/10.1175/MWR2906.1>
44. Rao, V. P., & Uma, K. (2015). Time series decomposition model for accurate wind speed forecast. Nuskaityta iš <https://doi.org/10.1186/s40807-015-0018-9>
45. Schmidhuber, & Hochreiter. (1997). Long Short-Term Memory. *9*(8), 1735-1780. Nuskaityta iš <https://doi.org/10.1162/neco.1997.9.8.1735>
46. Shumway, R. H., & Stoffer, D. S. (2000). Time Series Analysis and Its Applications. Nuskaityta iš https://doi.org/10.1007/978-1-4757-3261-0_2
47. Taieba, S. B., Bontempia, G., Atiyac, A. F., & Sorjamaab, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *39*(8), 7067-7083. Nuskaityta iš <https://doi.org/10.1016/j.eswa.2012.01.039>
48. Tao, B., Jingrui, H., & Jonathan, X. (2017). Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. Nuskaityta iš <https://doi.org/10.1016/j.ijforecast.2019.02.006>

49. Tarifai, V. (2017). Elektros energijos gamintojams, naudojantiems atsinaujinančius energijos išteklius, fiksuoti tarifai, Eur/kWh (be PVM). Vilnius. Nuskaityta iš <https://www.regula.lt/atsinaujinantys-istekliai/Puslapiai/tarifai.aspx>
50. Tufte, E. (1983). *The Visual Display of Quantitative Information*. Graphics Press. Cheshire, CT. Nuskaityta iš <https://pdfs.semanticscholar.org/c042/2a29881fc1fdc7021f5e8f9f042e5dee2c57.pdf>
51. Wacker, J. G., & R., L. R. (2002). Sales forecasting for strategic resource planning. 22(9), 1014-1031. Nuskaityta iš <https://doi.org/10.1108/01443570210440519>
52. Wang, X., Smith-Miles, K., & Hyndman, R. (2009). Rule induction for forecasting method selection: meta-learning the characteristics of univariate time series. 72(10-12), 2581-2594. Nuskaityta iš <https://doi.org/10.1016/j.neucom.2008.10.017>
53. Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. 30(4), 16-21. Nuskaityta iš <http://dx.doi.org/10.1016/j.ijforecast.2014.08.008> 0169-2070
54. Züfle, M., & Bauer, A. (2017). Telescope: A Hybrid Forecast Method for Univariate Time Series. Nuskaityta iš [https://www.researchgate.net/publication/318653962_Telescope_A_Hybrid_Forecast_Meth
od_for_Univariate_Time_Series](https://www.researchgate.net/publication/318653962_Telescope_A_Hybrid_Forecast_Method_for_Univariate_Time_Series)

Priedai

1 priedas

Vienmačių modelių rezultatai

Šiame priede pateikiama 18 vienmačių modelių prognozių, skirtingoms VE, paklaidų įvertinimo lentelė.

Vėjo elektrinė	Paklaida, nMAE																	Prognozių Vidurkis
	Mean	Naïve	RWF drift	Snaive	TSLM Trend	TSLM Trend + Season	TSLM Quad. Trend	TSLM Quad. Trend + Season	ARIMA (2, 1, 1)	STLM (ARIMA)	BATS	TBATS	Thetaf	Holt Winters	ETS	STLM (ETS)	NNETAR	
Benaičiai	19,0%	19,6%	19,7%	21,4%	17,8%	17,8%	19,9%	19,9%	19,6%	19,4%	17,3%	17,6%	22,4%	20,3%	19,6%	20,1%	20,7%	17,3%
Kreivėnai	18,5%	19,7%	19,8%	21,2%	17,4%	17,3%	18,9%	18,8%	17,3%	19,4%	17,4%	17,6%	21,7%	20,2%	19,8%	19,8%	19,2%	17,2%
Lauksargiai	19,1%	19,8%	19,9%	21,8%	18,0%	17,8%	19,5%	19,3%	17,9%	19,5%	18,1%	18,3%	21,9%	20,2%	19,9%	19,5%	20,2%	17,6%
Siauduva	23,9%	24,8%	25,0%	26,9%	23,1%	22,7%	24,5%	24,1%	22,8%	24,4%	22,6%	22,4%	27,3%	25,5%	25,1%	24,3%	24,7%	22,0%
Sudėnai	16,5%	17,4%	17,5%	19,0%	15,8%	15,8%	17,5%	17,5%	19,1%	17,2%	15,7%	15,4%	19,8%	19,0%	17,6%	17,7%	17,7%	15,0%
Vėjas_1	17,6%	17,7%	17,8%	19,7%	16,3%	16,3%	18,6%	18,6%	16,1%	17,6%	15,8%	16,1%	20,0%	18,5%	18,1%	18,1%	16,7%	15,8%
Čiuteliai	21,2%	21,3%	21,5%	23,9%	19,8%	19,6%	22,0%	21,8%	21,3%	21,1%	19,8%	19,6%	23,6%	22,0%	21,3%	21,1%	19,5%	19,1%
Didšiliai	17,0%	17,5%	17,7%	19,4%	16,3%	16,1%	17,8%	17,7%	16,2%	17,2%	16,0%	16,1%	19,9%	18,1%	17,6%	17,2%	17,3%	15,8%
Geišiai	25,7%	26,7%	26,9%	28,9%	24,9%	24,5%	26,3%	25,8%	24,3%	25,9%	24,1%	24,0%	28,4%	26,5%	26,7%	25,9%	25,0%	23,5%
Strepeikiai	20,4%	21,0%	21,1%	23,0%	19,2%	19,0%	20,9%	20,7%	19,1%	20,6%	19,1%	19,1%	23,1%	21,5%	21,7%	21,2%	21,0%	18,6%
Kabaldikai	25,0%	24,7%	24,8%	27,5%	24,5%	24,2%	25,6%	25,3%	24,8%	24,1%	23,3%	22,7%	27,5%	24,8%	24,1%	24,0%	25,8%	22,2%
Šyša	27,0%	26,9%	27,0%	29,7%	25,8%	25,4%	26,9%	26,5%	25,0%	26,0%	25,2%	24,3%	28,8%	29,2%	27,2%	26,5%	27,0%	24,0%
Kunigiškiai	28,3%	28,7%	28,8%	30,6%	27,7%	27,1%	28,7%	28,0%	26,9%	27,7%	26,3%	25,7%	31,3%	28,1%	27,9%	27,7%	28,1%	25,5%
Bendra VE prognozė	19,7%	20,0%	20,3%	22,0%	18,7%	18,5%	20,3%	20,1%	20,2%	19,6%	17,8%	17,9%	22,5%	20,4%	22,0%	21,4%	17,8%	17,4%
VE suma iš atskirų VE prognozių	19,1%	20,1%	19,9%	21,1%	18,7%	18,6%	20,3%	19,4%	17,4%	18,8%	17,9%	17,8%	22,3%	20,1%	20,0%	19,0%	18,0%	17,5%

32 pav. Vienmačių modelių rezultatai

Prognozavimo modelių tikslumo ir AEI balansavimo sąnaudų palyginimas kas mėnesį

Pilkai pažymėti modeliai sukurti I etape. Melsvai pažymėti modeliai sukurti II etape. Rusvai pažymėti modeliai sukurti III etape. Galutinė prognozė – dabartinis Sistema 1 ir Sistema 2 sistemų kombinacijos rezultatas.

Prognozės	Rodiklis	Sausis	Vasaris	Kovas	Balandis	Gegužė	Birželis	Liepa	Rugpjūtis	Rugsėjis	Spalis	Lapkritis	Gruodis	Viso
Sistema 1	Balansavimo sąnaudos, tūkst. EUR	589,44	192,44	155,46	435,22	567,15	413,36	260,63	290,46	373,53	560,28	198,06	474,80	4510,82
	Paklaida, nMAE %	7,14%	7,67%	5,93%	9,75%	7,70%	6,71%	6,49%	5,72%	6,00%	8,43%	5,10%	6,70%	6,95%
Sistema 2	Balansavimo sąnaudos, tūkst. EUR	198,32	273,60	55,90	236,39	349,01	334,40	86,08	376,13	426,89	471,87	188,62	372,93	3370,14
	Paklaida, nMAE %	6,34%	6,85%	6,01%	7,64%	7,01%	5,63%	6,32%	5,19%	5,18%	6,64%	4,39%	5,33%	6,04%
Galutinė prognozė	Balansavimo sąnaudos, tūkst. EUR	515,57	245,65	139,79	363,74	423,07	349,01	211,80	351,78	411,40	483,75	194,92	394,75	4085,24
	Paklaida, nMAE %	7,17%	6,67%	5,53%	8,56%	6,85%	5,85%	6,25%	5,17%	5,44%	6,55%	3,96%	5,47%	6,12%
TBATS	Balansavimo sąnaudos, tūkst. EUR	483,98	113,08	229,60	667,31	632,03	400,07	614,99	436,61	733,51	-58,94	542,94	140,38	4935,56
	Paklaida, nMAE %	28,53%	15,99%	15,96%	23,21%	12,08%	15,34%	10,46%	16,37%	18,19%	23,75%	19,99%	16,12%	17,82%
AVG	Balansavimo sąnaudos, tūkst. EUR	429,70	170,80	362,63	761,86	587,14	391,50	573,59	450,63	772,63	31,61	587,61	220,14	5339,83
	Paklaida, nMAE %	26,04%	14,77%	16,71%	23,57%	11,85%	14,69%	9,98%	16,59%	19,66%	23,32%	18,26%	16,57%	17,53%
ARIMAX 1	Balansavimo sąnaudos, tūkst. EUR	491,72	169,86	270,03	526,82	550,12	354,63	566,40	403,36	486,47	305,40	305,25	413,20	4843,25
	Paklaida, nMAE %	8,44%	7,90%	6,96%	11,10%	7,39%	7,19%	6,63%	6,38%	8,32%	9,36%	7,49%	8,18%	7,93%
ARIMAX 2	Balansavimo sąnaudos, tūkst. EUR	499,44	169,09	270,46	532,16	548,78	355,63	569,29	400,86	487,10	303,79	312,32	414,17	4863,07
	Paklaida, nMAE %	8,63%	7,97%	7,00%	11,15%	7,40%	7,23%	6,63%	6,38%	8,32%	9,37%	7,55%	8,20%	7,97%
STLM	Balansavimo sąnaudos, tūkst. EUR	483,85	181,43	197,47	525,50	403,82	283,22	450,42	348,56	488,45	308,62	279,88	312,62	4263,84
	Paklaida, nMAE %	10,20%	7,37%	7,14%	11,65%	7,03%	8,02%	6,23%	7,21%	9,05%	10,76%	8,79%	8,68%	8,48%
NNETAR	Balansavimo sąnaudos, tūkst. EUR	559,97	204,05	188,93	418,42	520,52	362,76	539,64	311,01	434,01	561,89	322,41	482,92	4906,54
	Paklaida, nMAE %	8,91%	8,02%	8,65%	12,22%	9,19%	8,11%	8,15%	7,91%	9,15%	9,72%	6,89%	9,93%	7,70%
AVG	Balansavimo sąnaudos, tūkst. EUR	507,17	197,31	237,13	498,50	492,84	364,02	534,70	366,98	481,50	415,70	283,70	418,77	4798,33
	Paklaida, nMAE %	8,50%	7,43%	6,80%	10,86%	7,28%	7,05%	6,51%	6,28%	7,86%	9,04%	7,00%	8,04%	7,70%
ARIMAX 1	Balansavimo sąnaudos, tūkst. EUR	513,69	211,29	179,89	359,47	416,15	333,63	309,96	365,90	420,00	455,11	222,60	391,95	4179,64
	Paklaida, nMAE %	7,11%	6,43%	5,44%	8,13%	6,40%	5,46%	5,77%	4,98%	5,54%	6,61%	4,70%	6,06%	6,03%
BEST	Balansavimo sąnaudos, tūkst. EUR	501,50	237,29	145,61	356,74	378,76	352,99	287,88	350,30	400,17	483,45	226,87	381,02	4102,58
	Paklaida, nMAE %	7,21%	6,20%	5,94%	8,52%	6,54%	5,65%	5,52%	5,54%	5,99%	6,52%	4,57%	6,21%	6,18%
KOMBINACIJA	Balansavimo sąnaudos, tūkst. EUR	503,11	240,59	143,19	356,06	378,82	353,45	296,25	350,80	400,88	474,55	225,19	381,04	4103,92
	Paklaida, nMAE %	6,89%	6,14%	5,54%	8,04%	6,16%	5,51%	5,31%	4,95%	5,50%	6,51%	4,63%	5,95%	5,90%
LOESS	Balansavimo sąnaudos, tūkst. EUR	198,38	203,42	46,51	273,13	311,18	305,00	192,79	347,88	377,96	423,13	146,95	396,57	3222,89
	Paklaida, nMAE %	8,18%	6,12%	5,43%	7,40%	6,04%	5,34%	5,73%	4,81%	4,95%	6,43%	4,14%	5,84%	5,84%

33 pav. Tyrimo sukurtų prognozavimo modelių palyginimas tarpusavyje ir su 2 sistemų teikiamomis prognozėmis.