

VILNIAUS UNIVERSITETAS

LAURA PAULIONIENĖ

ERDVĖS-LAIKO DUOMENŲ STATISTINIS MODELIAVIMAS,
PAGRĮSTAS LAIKO EILUČIŲ PARAMETRŲ ERDVINIŲ
INTERPOLIAVIMU

Daktaro disertacija

Fiziniai mokslai, Matematika (01P)

Vilnius, 2013

Disertacija rengta 2012 – 2013 metais Vilniaus universitete.

Disertacija ginama eksternu.

Mokslinis konsultantas: prof. dr. Kęstutis Dučinskas (Klaipėdos universitetas, fiziniai mokslai, matematika 01P).

Padėka

Dėkoju mokslinio darbo konsultantui prof. dr. Kęstučiui Dučinskui už skirtą laiką ir energiją bei visokeriopą pagalbą rengiant šį disertacinį darbą. Dėkoju recenzentams prof. habil. dr. Viliandui Bagdonavičiui bei prof. dr. Marijui Radavičiui už pastabas, padėjusias patobulinti disertaciją, sesei Jūratei už nukreipimus teisinga linkme. Taip pat dėkoju visiems už tikėjimą ir palaikymą.

Turinys

Žymėjimai.....	6
Santrumpos.....	7
ĮVADAS.....	8
1. ERDVĖS-LAIKO DUOMENYS.....	14
1.1. Erdvės-laiko kovariacinės funkcijos.....	14
1.1.1 Atskiriamumas ir simetriškumas.....	17
1.1.2 Erdvės-laiko variogramos.....	20
1.2. Erdvės-laiko kriginas.....	20
1.2.1. Paprastasis kriginas.....	21
1.2.2. Ordinariusis kriginas.....	23
1.3. Erdvinių laiko eilučių tipai bei jų statistinės savybės.....	24
1.3.1. Tolydaus erdvinio indekso procesai (Geostatistiniai procesai).....	25
1.3.2. Diskretaus erdvinio indekso procesai (Gardelės procesai).....	26
1.4. Erdvės-laiko duomenų modeliai.....	28
1.4.1. Vektorinis autoregresijos slenkamojo vidurkio modelis (VARMA).....	28
1.4.2. Erdvės-laiko autoregresijos slenkamojo vidurkio (STARMA) modelis.....	30
2. ERDVINIŲ LAIKO EILUČIŲ MODELIAVIMAS.....	33
2.1. Stochastinis erdvės-laiko modelis.....	33
2.2. Vienmačių laiko eilučių modeliai.....	34
2.2.1. Laiko eilučių modelių patikrinimas.....	40
2.3. Erdvinis laiko eilučių parametrų interpoliavimas.....	41
2.3.1. Semivariogramos modeliavimas.....	41
2.3.2. Empirinės semivariogramos įverčio tarp dviejų erdvės taškų savybės.....	43
2.3.3. Interpoliavimas, pagrįstas erdviniais svoriais (ES).....	44
2.3.4. Interpoliavimas, pagrįstas „formaliu“ ordinariuoju kriginu (OK).....	46
2.3.5. Interpoliavimas, pagrįstas trendo paviršiaus modeliu (TPM).....	47
2.3.6. Interpoliavimas, pagrįstas atvirkštinių atstumų svoriais (IDW).....	48
2.4. Kryžminio patikrinimo metodas.....	49
3. ERDVINIŲ LAIKO EILUČIŲ MODELIS VĖJO GREIČIO DUOMENIMS.....	51
3.1. Žvalgomoji duomenų analizė.....	52
3.2. Laiko eilučių modeliavimas.....	57
3.3. Erdvinis interpoliavimas.....	70

3.4. Erdvinių modelių patikrinimas	74
3.5. Erdvės-laiko modelio parinkimas kryžminio patikrinimo metodu	75
Išvados.....	77
Literatūra.....	79
Publikacijų sąrašas.....	86
Priedas: Erdvinių laiko eilučių modeliavimo algoritmas	87

Žymėjimai

$Z(\mathbf{s}; t)$	tolydaus laiko erdvės-laiko atsitiktinis procesas
$Z_t(\mathbf{s})$	diskretaus laiko erdvės-laiko atsitiktinis procesas
$D_S \times D_t$	erdvės-laiko indeksų aibė
$\delta(\mathbf{s}; t)$	nulinio vidurkio atsitiktinis efektas, aprašantis variacijas apie trendą laike ir erdvėje
$\varepsilon(\mathbf{s}; t)$	atsitiktinės stacionarios Gauso paklaidos su nuliniu vidurkiu
$Z_i(t)$	erdvės taškų $s_i \in D, i = 1, \dots, n, t = 1, 2, \dots, T$ laiko eilutė
ϕ_k	autoregresiniai parametrai
θ_l	slenkamojo vidurkio parametrai
$\hat{\eta}^i, \hat{\eta}^{*i}$	įvertintų laiko eilučių parametrų rinkiniai
Σ	kovariacijų matrica
W_l	svorių matrica
$\gamma(\mathbf{h}; \boldsymbol{\vartheta})$	parametrinis semivariogramos modelis
$\hat{\gamma}_{ij}$	semivariograma tarp taškų i ir j
$\hat{\delta}_{0j}$	erdviniai svoriai
\otimes	Kronekerio sandauga

Santrumpos

ACF – autokoreliacinė funkcija

AIC – Akaikės informacinis kriterijus

ARIMA – integruotas autoregresijos slenkamojo vidurkio modelis

ARMA – autoregresijos slenkamojo vidurkio modelis

BLUP – geriausia tiesinė nepaslinkta prognozė

ES – erdvinių svorių metodas

GLS – apibendrintas mažiausių kvadratų metodas

IDW – atvirkštinių atstumų svorių metodas

OK – formalaus ordinariojo kriginio metodas

OLS – paprastasis mažiausių kvadratų metodas

PACF – dalinės autokoreliacijos funkcija

PE – laiko eilučių prognozės liekanos

RMSE – šaknis iš vidutinės kvadratinės klaidos

STARMA – erdvės-laiko autoregresijos slenkamojo vidurkio modelis

STARMAG – apibendrintas erdvės-laiko autoregresijos slenkamojo vidurkio modelis

TP – trendo paviršiaus modelis

VARMA – vektorinis autoregresijos slenkamojo vidurkio modelis

WLS – svartinis mažiausių kvadratų metodas

ĮVADAS

Nagrinėjama problema ir darbo aktualumas. Darbe nagrinėjama erdvės-laiko duomenų modeliavimo problema nėra visiškai nauja, tačiau dėl besivystančių technologijų atsirandant vis naujų modeliavimo galimybių, ši tema vis dar lieka gana aktuali. Dažnai erdvinių duomenų rinkiniai yra gana nedideli, o taškai, kuriuose pasklidę stebėjimai, išsidėstę netaisyklingai. Renkant duomenis tam tikrą laiko periodą (paprastai vienodais laiko intervalais), jų gali būti labai daug. Tokių duomenų pavyzdžiai gali būti kasdien arba net kas valandą matuojama oro temperatūra, kritulių kiekis, vėjo greitis, matuojamas gamtos užterštumas, fiksuojami įvairūs ekonominiai rodikliai ir panašiai. Monitoringo sistemą tokiuose stebėjimuose paprastai sudaro ne daugiau kaip šimtas taškų, tačiau kiekviename taške gaunamos gana ilgos laiko eilutės. Sprendžiant „erdvinį“ uždavinį, paprastai siekiama interpoliuoti arba įvertinti erdvinį vidurkį. Laiko eilučių duomenys dažniausiai naudojami ateities reikšmėms prognozuoti. Tuo tarpu erdvės-laiko uždaviniai jungia abu uždavinių tipus [32]. Vienas iš akivaizdžių sprendimo būdų yra analizuoti erdvėje rinktus duomenis kiekvienu atskiru laiko momentu, t.y. ignoruoti reiškinių kitimą laike. Taip pat galima analizuoti laiko eilutes skirtinguose taškuose, t. y. daugiamates laiko eilutes. Tačiau tokiu atveju neįmanoma modeliuoti, prognozuoti bei įvertinti reikšmių taškuose, nesančiuose imtyje. Bendru atveju reikia atsižvelgti į koreliacijas ir erdvėje, ir laike bei nustatyti ryšius tarp jų.

Erdvės-laiko duomenų analize buvo susidomėta apie 1975 metus. Vieni iš pirmųjų autorių, pasiūliusių erdvės-laiko modelį buvo Cliff ir Ord [20]. Nuo šio laiko, sparčiai vystantis technologijoms, erdvės-laiko duomenų analizė stipriai pasistūmėjo į priekį. Pats bendriausias modelis, aprašantis kintamųjų tarpusavio priklausomybę erdvėje ir laike yra 1980 metais Pfeifer ir Deutsch

pasiūlytas erdvės-laiko autoregresijos slenkamojo vidurkio (*STARMA*) modelis. *STARMA* modelio sudarymo etapai, panašūs į vienmatės laiko eilutės modelio *ARMA* sudarymo etapus, buvo plačiai aprašyti minėtų autorių 1980–1981 metų darbuose [60], [61]. Taip pat galima rasti keletą darbų, analizuojančių ir aprašančių *STARMA* modelio taikymo galimybes, analizuojant konkrečius erdvės-laiko duomenis [42], [48], [57]. Kita modelių klasė, nagrinėjanti daugiamates laiko eilutes yra vektoriniai autoregresiniai slenkamojo vidurkio (*VARMA*) modeliai [40] arba dažniau ekonominiuose tyrimuose nagrinėjami jų atskiri atvejai (*VAR*) modeliai [5], [21], [33], [62]. Didžiausias šių modelių trūkumas – didelis vertinamų parametrų skaičius.

Didžioji dalis autorių, nagrinėjančių erdvės-laiko duomenis, siūlo savo metodus, kurie remiasi iteracinėmis procedūromis, taikydami įvairius erdvės ar laiko pjūvius.

Dėl dažniausiai ilgų laiko eilučių bei santykinai nedidelio erdvės taškų skaičiaus, nemažai autorių siūlo pirmiausia analizuoti vienmates laiko eilutes, aprašant jas trupmeniškai diferencijuotų autoregresinių slenkamojo vidurkio (*ARFIMA*), autoregresinių slenkamojo vidurkio (*ARMA*) ar tiesiog autoregresinių (*AR*) procesų pagalba [8]. Pašalinus priklausomybę laike, laiko eilučių liekanos modeliuojamos kaip nepriklausomi stebėjimai ar replikacijos erdvėje. Erdvinė priklausomybė nustatoma vertinant erdvines kovariacines funkcijas ar semivariogramas [5], [6], [26], [30], [31], [39], [42], [65], [66], [68], [77]. Erdviniam modeliavimui taip pat siūloma taikyti Bajeso hierarchinį modeliavimą [43], [71], [75], [76]. Taip pat analizuojami dinaminiai tiesiniai ar būsenų erdvės modeliai [33], [50]. Ekonominiuose taikymuose siūlomas dar vienas modeliavimo metodas – erdvinių laiko eilučių agregavimas [38].

Viena svarbiausių charakteristikų, apibūdinančių erdvės-laiko modelius yra *kovariacinė funkcija*, kuri, pagal analizuojamo erdvės-laiko proceso prigimtį, gali būti *atskiriama ir neatskiriama*. Atskiriama kovariacinė struktūra

reiškia, kad kovariacija gali būti išreikšta grynai erdvės ir grynai laiko kovariacijų sandauga. Pagrindinė atskiriamų kovariacinių funkcijų taikymo priežastis yra ženkliai mažesnis parametru skaičius kovariacijų matricoje, o tai lengvina didelių erdvės-laiko duomenų masyvų analizę [37]. Tačiau dalis autorių prieštarauja dėl atskiriamų kovariacinių funkcijų tinkamumo realių erdvės-laiko duomenų analizei, nes atskiriamos funkcijos neatsižvelgia į erdvės-laiko tarpusavio sąveikas. Todėl yra žinoma nemažai neatskiriamų kovariacinių funkcijų bei jų aproksimacijų [7], [13], [23], [25], [27], [28], [34], [35].

Darbe pasiūlyta keletas originalių erdvinių laiko eilučių modeliavimo metodų. Siūlomi metodai pirmiausia analizuoja vienmates laiko eilutes, o pašalinus jų laikinę priklausomybę, laiko eilučių liekanoms vertinama erdvinė priklausomybė. Mūsų tikslas – sudaryti modelį, leidžiantį prognozuoti požymio reikšmę naujame, nestebėtame taške, nauju laiko momentu. Tokio modelio sudarymas remiasi laiko eilučių parametru erdviniu interpoliavimu. Darbe analizuojami du vienmačių laiko eilučių modeliai: klasikinis *ARIMA* modelis bei modelis, susidedantis iš trendo, sezoninės komponentių bei *ARMA* proceso. Darbe remiamasi prielaida, kad kiekviename erdvės taške stebimą informaciją galima aprašyti tokiu pačiu (tos pačios eilės) laiko eilutės modeliu. Erdviniam parametru interpoliavimui siūlomi dviejų tipų metodai. Pirmo tipo metodai atsižvelgia į erdvinę koreliaciją, darbe nagrinėjami du šio tipo metodai: erdvinių svorių (ES), apskaičiuojamų semivariogramų pagalba, formulė bei pritaikytas formalus ordinarusis kriginas (OK). Antrajam tipui priklauso metodai, ignoruojantys erdvinę koreliaciją, t.y. siūlomas trendo paviršiaus metodas, kurio parametrai vertinami mažiausių kvadratų metodu (TPM) bei atvirkštinių atstumų svorių metodas (IDW). Šių metodų pranašumas yra tas, kad sumažėjus vertinamų parametru skaičiui, sumažėja ir nepageidaujamos vertinamos paklaidos, todėl neretai šie metodai konkuruoja su erdvinę koreliaciją inkorporuojančiais metodais. Trečiajame skyriuje visi šie

metodai aprašyti, interpoliuojant įvertintus laiko eilučių parametrus vėjų greičių duomenims Lietuvos respublikoje (LR), be to atliktas visų keturių metodų palyginimas.

Darbo tikslas

Sudaryti laiko eilutės modelį naujame, nestebėtame taške, erdvinę sąveiką (angl. spatial interacion) įtraukiant per įvertintų laiko eilučių parametrų erdvinį interpoliavimą.

Darbo uždaviniai:

Sudaryti erdvės-laiko modelių realizavimo algoritmus.

Patikrinti siūlomus modelius erdvės-laiko duomenų realizacijai.

Atlikti išsamų siūlomų modelių palyginimą su realiais duomenimis R programos pagalba.

Tyrimų metodika

Darbe naudojami šie tyrimų metodai: mokslinės literatūros apžvalga disertacijos tema, klasikiniai statistiniai metodai, erdvinės statistikos metodai. Siūlomi metodai realizuojami atvirojo kodo sistema R.

Darbo naujumas ir jo reikšmė

1. Pasiūlyta keletas originalių erdvės-laiko duomenų modeliavimo metodų, pagrįstų laiko eilučių parametrų interpoliavimu ir ištirtos jų savybės.
2. Išanalizuotos siūlomų erdvės-laiko modelių prognozės galimybės.
3. Atliktas erdvės-laiko modelių patikrinimas bei palyginimas, remiantis kryžminio patikrinimo metodu, modeliuojant LR vėjų greičių duomenis.

Darbo rezultatų praktinė reikšmė

1. Pateikta gana paprasta erdvės-laiko duomenų modeliavimo metodika, kurią nesudėtinga realizuoti kompiuterinių programų pagalba.
2. Sudarytas erdvinių laiko eilučių modelis, leidžiantis apskaičiuoti prognozę naujame erdvės taške s_0 , nauju laiko momentu $T + k$.

Ginamieji teiginiai

1. Dažnai erdvės-laiko atsitiktinį lauką $Z(\mathbf{s}; t)$ galima sumodeliuoti, atskiriant erdvės ir laiko komponentes.
2. Pasiūlyti erdvinių svorių bei ordinaraus kriginio metodai, pagrįsti semivariogramomis, leidžia pakankamai tiksliai aprašyti analizuojamus erdvės-laiko duomenis. Šie metodai tikslesni už trendo paviršiaus bei svertinių atvirkštinių atstumų metodus.
3. Pasiūlyta erdvės-laiko modeliavimo technika gali būti plačiai taikoma įvairiuose tyrimuose.

Darbo rezultatų aprobavimas

Disertacijos tema atspausdinti 5 straipsniai. Publikacijų sąrašas pateiktas disertacijos pabaigoje.

Disertacijos tema skaityti 2 pranešimai respublikinėse konferencijose bei 6 pranešimai tarptautinėse:

Tarptautinės konferencijos:

Modified STARIMA model. 9th International Vilnius Conference on Probability Theory and Mathematical Statistics, June 25-30, 2006, Vilnius, Lithuania.

Spatio-temporal modeling in case of anisotropic variogram structure. Nordic Conference on Mathematical Statistics 2008 (NORDSTAT), 16-19 June, 2008, Vilnius, Lithuania.

Adequacy Testing of Two Methods for Modeling Spatial Time Series.
The XIIIth International Conference “Applied Stochastic Models and Data Analysis” (ASMDA-2009), Vilnius, Lithuania.

Spatial-temporal modeling of Baltic Sea coastal zone parameters. 2nd Nordic-Baltic Biometric conference, 10-12 June, 2009, Tartu, Estonia.

Spatio-temporal model based on separable characteristics. 4th Nordic-Baltic Biometric Conference, June 10-12, 2013, Stockholm, Sweden.

Comparison of two Spatio-temporal modelling techniques for Wind Speed data. Spatial Statistics Conference, June 4-6, 2013, Columbus, Ohio, USA.

Respublikinės konferencijos:

Modifikuotas STARIMA modelis erdvės-laiko duomenims. 46 LMD konferencija, birželio 15–16 d., 2005, VU, Vilnius.

Erdvinis ARIMA modelis vėjų greičių duomenims LR. 53 LMD konferencija, 11-12 birželio, 2012, Klaipėda, KU.

Disertacijos struktūra.

Disertaciją sudaro įvadas, trys skyriai, išvados, literatūros sąrašas bei autorės publikacijų disertacijos tema sąrašas.

Bendra disertacijos apimtis 87 puslapiai, numeruotų formulių 66, 13 lentelių, 20 paveikslų. Literatūros sąrašą sudaro 79 šaltiniai.

1. ERDVĖS-LAIKO DUOMENYS

Šiame skyriuje aprašytos pagrindinės erdvės-laiko duomenų charakteristikos, pagrindiniai duomenų tipai bei modeliai.

Erdvės-laiko atsitiktinis procesas paprastai apibrėžiamas tokiu būdu:

$$\{Z(\mathbf{s}; t): \mathbf{s} \in D_s, t \in D_t\}, (\mathbf{s}; t) \in D_s \times D_t.$$

Erdvė $D_s \times D_t$ vadinama erdvės-laiko indeksų aibe.

Erdvė ir laikas, priklausomai nuo nagrinėjamų duomenų, gali būti tolydūs arba diskretūs. Erdviniam matavimui diskretumas (nevisada, bet dažnai) yra sąlygojamas tolydaus indekso sumavimo į tam tikrus regionus. Laikiniam matavimui diskretumas dažniausiai yra sąlygojamas, sujungiant tolydų laiką į vienodo ilgio intervalus [43].

1.1. Erdvės-laiko kovariacinės funkcijos

Erdvės-laiko kovariacinė funkcija glaustai bet informatyviai aprašo atsitiktinį procesą aibėje $D_s \times D_t$. Todėl gana įprasta kovariacinę funkciją laikyti erdvės-laiko proceso $Z(\mathbf{s}; t)$ pagrindine charakteristika.

Erdvės-laiko procesas $Z(\mathbf{s}; t)$ gali būti užrašytas tokia forma:

$$Z(\mathbf{s}; t) = \mu(\mathbf{s}; t) + \delta(\mathbf{s}; t), \quad (1.1)$$

čia $\mathbf{s} \in D_s, t \in D_t$; $\mu(\mathbf{s}; t)$ yra determinuota vidurkio funkcija; $\delta(\mathbf{s}; t)$ – nulinio vidurkio atsitiktinis efektas, aprašantis fluktuacijas apie trendą laike ir erdvėje, kurių neaprašo determinuota vidurkio funkcija $\mu(\mathbf{s}; t)$. Paprastai yra laikoma, kad (1.1) formulėje erdvės ir laiko procesai yra tolydūs. Antros eilės momentai erdvės-laiko procesui (1.1) gali būti užrašyti tokiu būdu:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) \equiv C_Z(\mathbf{s}, \mathbf{x}; t, r), \quad (1.2)$$

kuri vadinama erdvės-laiko *kovariacine funkcija*. Bet kuri erdvės-laiko kovariacinė funkcija yra neneigiamai apibrėžta funkcija ir atvirkščiai, todėl neneigiamai apibrėžtų funkcijų tyrimas erdvėje ir laike užima gana svarbią vietą erdvės-laiko statistikoje [52].

Apibrėžimas. Funkcija $\{f(\mathbf{u}, \mathbf{v}): \mathbf{u}, \mathbf{v} \in D\}$, apibrėžta $D \times D$ yra vadinama neneigiamai apibrėžta, jeigu bet kokiam skaičių rinkiniui $\{a_i: i = 1, \dots, m\}$, bet kokiam $\{\mathbf{u}_i: i = 1, \dots, m\}$ erdvėje D ir bet kokiam sveikam skaičiui m , galioja lygybė:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j f(\mathbf{u}_i, \mathbf{u}_j) \geq 0. \quad (1.3)$$

Apibrėžimas. Sakoma, kad f yra stacionari erdvės-laiko kovariacinė funkcija erdvėje $\mathbb{R}^d \times \mathbb{R}$, jeigu ji tenkina (1.3) nelygybę ir gali būti užrašyta tokia forma [54]:

$$f((\mathbf{s}; t), (\mathbf{x}; r)) = C(\mathbf{s} - \mathbf{x}; t - r),$$

čia $\mathbf{s}, \mathbf{x} \in \mathbb{R}^d$, $t, r \in \mathbb{R}$.

Jeigu atsitiktinis procesas $Z(\mathbf{s}; t)$ turi pastovų vidurkį ir stacionarią kovariacinę funkciją $C_Z(\mathbf{h}; \tau)$, jis yra vadinamas *antros eilės (plačiąja, silpnąja prasme) stacionariu*. *Stiprus (siaurąja arba griežtąja prasme) stacionarumas* reikalauja, kad atsitiktinių procesų $Z(\mathbf{s}; t)$ ir $Z(\mathbf{s} + \mathbf{h}; t + \tau)$ pasiskirstymai, visiems $\mathbf{h} \in \mathbb{R}^d$ ir $\tau \in \mathbb{R}$, sutaptų [4].

Paprastumo dėlei, stacionarios kovariacinės funkcijos žymėjime dažniausiai yra praleidžiama jos priklausomybė nuo Z ir rašoma paprasčiausiai $C(\mathbf{h}; \tau)$ [41]. Į (1.3) formulę įsistačius $C(\mathbf{h}; \tau)$ išraišką, gauname stacionarios erdvės-laiko kovariacinės funkcijos savybę:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j C(\mathbf{s}_i - \mathbf{s}_j; t_i - t_j) \geq 0, \quad (1.4)$$

visiems $\{a_i\}$, $\{(s_i; t_i)\}$ ir m . Galima pažymėti, kad $C(\mathbf{h}; 0)$, $\mathbf{h} \in \mathbb{R}^d$ yra stacionari erdvinė kovariacinė funkcija; o $C(\mathbf{0}; \tau)$, $\tau \in \mathbb{R}$ - stacionari laiko kovariacinė funkcija.

Turint erdvės-laiko kovariacinę funkciją, apibrėžiame stacionarią erdvės-laiko koreliacinę funkciją:

$$\rho(\mathbf{h}; \tau) \equiv \frac{C(\mathbf{h}; \tau)}{C(\mathbf{0}; 0)}, \mathbf{h} \in \mathbb{R}^d, \tau \in \mathbb{R}. \quad (1.5)$$

Intuityvus spėjimas, kad tolimos reikšmės turi mažesnę įtaką, gali būti patvirtintas remiantis kovariacine funkcija, kurios reikšmės artėja į nulį, kai erdvės-laiko koordinatės tolsta viena nuo kitos. Dažnai yra naudojami kovariaciniai modeliai, kurių reikšmės yra lygios nuliui, jeigu atstumas erdvėje arba laike pasidaro didesnis, nei tam tikras fiksuotas atstumas; tokie modeliai dažnai yra vadinami kompaktiškai apibrėžtais kovariaciniais modeliais [44].

Kovariacinės funkcijos stacionarumą galima apibrėžti atskirai erdvei ir laikui [64]. *Erdvinis* kovariacinės funkcijos *stacionarumas* apibrėžiamas tokiu būdu:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) \equiv C(\mathbf{s} - \mathbf{x}; t, r),$$

laikinis kovariacinės funkcijos *stacionarumas*:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) \equiv C(\mathbf{s}, \mathbf{x}; t - r),$$

o erdvės-laiko kovariacinės funkcijos stacionarumas:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) \equiv C(\mathbf{s} - \mathbf{x}; t - r).$$

Kita svarbi sąvoka, analizuojant erdvės-laiko kovariacinę funkciją, yra *izotropija* [29]. *Erdvinė izotropija* nusakoma tokia kovariacinės funkcijos savybe:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) \equiv C(\|\mathbf{s} - \mathbf{x}\|; t, r),$$

čia $\|\mathbf{s} - \mathbf{x}\|$ – Euklido atstumas, t. y. kovariacinė funkcija priklauso tik nuo vektoriaus ilgio ir nepriklauso nuo jo krypties. Priešingu atveju kovariacinė funkcija vadinama *anizotropine* [18].

Geometrinė erdvinė anizotropija nusakoma tokia kovariacinės funkcijos savybe:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) \equiv C\left(\{p_1(s_1 - x_1)^2 + \dots + p_d(s_d - x_d)^2\}^{1/2}; t, r\right),$$

čia $p_1 > 0, \dots, p_d > 0$.

Geometrinė erdvės-laiko anizotropija nusakoma savybe:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) \equiv C\left(\{p_1(s_1 - x_1)^2 + \dots + p_d(s_d - x_d)^2 + q(t - r)^2\}^{1/2}\right),$$

čia $p_1 > 0, \dots, p_d > 0, q > 0$.

1.1.1 Atskiriamumas ir simetriškumas

Apibrėžimas. Sakoma, kad atsitiktinis procesas $Z(\mathbf{s}; t)$ turi *atskiriamą* erdvės-laiko kovariacinę funkciją, jeigu visiems $\mathbf{s}, \mathbf{x} \in \mathbb{R}^d$ ir $t, r \in \mathbb{R}$, galioja lygybė [37]:

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) = C^{(s)}(\mathbf{s}, \mathbf{x}) \cdot C^{(t)}(t, r), \quad (1.6)$$

čia $C^{(s)}$ ir $C^{(t)}$ yra atitinkamai erdvės ir laiko kovariacinės funkcijos.

Iš (1.6) formulės seka, kad paprasta erdvės-laiko kovariacinių funkcijų klasė gali būti gaunama sudauginus individualias erdvės ir laiko kovariacines funkcijas. Tuomet (1.5) formulė įgyja tokį pavidalą:

$$C(\mathbf{h}; \tau) = C^{(s)}(\mathbf{h}) \cdot C^{(t)}(\tau). \quad (1.7)$$

Taip pat atskiriamumas reiškia, kad erdvės-laiko koreliacinė funkcija (1.5) gali būti užrašyta tokia forma:

$$\rho(\mathbf{h}; \tau) = \rho^{(s)}(\mathbf{h}; 0) \cdot \rho^{(t)}(\mathbf{0}; \tau). \quad (1.8)$$

Kita vertus, dažnai erdvės-laiko modeliai pasižymi erdvės-laiko tarpusavio sąveika - tokiu atveju atskiriami modeliai netinka. Iš kitos pusės, dėl savo paprastumo ir dėl to, kad matriciniai skaičiavimai erdvinėje statistikoje atskiriamumo atveju yra daug trumpesni, atskiriami kovariacijų modeliai iki XXI a. pradžios dominavo literatūroje. Tokiems modeliams būdingos savybės, kurios retai sutinkamos empiriniuose erdvės-laiko priklausomybės tyrimuose. Tarkime, iš atskiriamumo savybių seka, kad $C(\mathbf{h}_1; \tau) \propto C(\mathbf{h}_2; \tau)$; $\tau \in \mathbb{R}$, t. y. kiekviena laiko eilutė pasižymi lygiai tokiomis pačiomis kryžminės koreliacijos savybėmis kaip ir bet kuri kita laiko eilutė bet kuriame taške, nesvarbu kokiu atstumu nuo pirmosios.

Kita erdvės-laiko kovariacinės funkcijos savybė, kurios dažnai yra reikalaujama, tačiau dažniausiai netenkinama yra *visiškas simetriškumas*.

Apibrėžimas. Atsitiktinis procesas $Z(\mathbf{s}; t)$ turi *visiškai simetrišką* erdvės-laiko kovariacinę funkciją, jeigu visiems $\mathbf{s}, \mathbf{x} \in \mathbb{R}^d$ ir $t, r \in \mathbb{R}$

$$\text{cov}(Z(\mathbf{s}; t), Z(\mathbf{x}; r)) = \text{cov}(Z(\mathbf{s}; r), Z(\mathbf{x}; t)). \quad (1.9)$$

Pavyzdžiui, jeigu Z yra dienos temperatūra (vidutinė, minimali ar maksimali) (1.9) formulė teigia, kad kovariacija tarp vakar dienos temperatūros Klaipėdoje ir šios dienos temperatūros Nidoje yra lygiai tokia pati kaip kovariacija tarp vakar dienos Nidos ir šios dienos Klaipėdos temperatūros. Toks modelis netinka dienos temperatūrų aprašymui.

Kaip atskirą atvejį panagrinėkime tokias erdvės ir laiko koreliacines funkcijas:

$$\rho^s(h) = e^{-h(\alpha/\beta)^{1/2}} \quad h > 0$$

ir

$$\rho^t(\tau) = \text{Erfc}(\tau^{1/2} \alpha^{1/2}), \tau > 0,$$

čia $Erfc(z) \equiv \left(2/\pi^{1/2}\right) \int_z^\infty e^{-v^2} dv$, kai $z \geq 0$ ir $Erfc(z) = 2 - Erfc(-z)$,

kai $z < 0$. Tada bet kokiam $p > 0$

$$C(\mathbf{h}; \tau) = p \cdot \rho^s(\mathbf{h}) \cdot \rho^t(\tau) \quad (1.10)$$

apibrėžia atskiriamą kovariacinę funkciją. Taip pat bet kokiems $p > 0, q \geq 0$ ir $r \geq 0$

$$C(\mathbf{h}; \tau) = p \cdot \rho^s(\mathbf{h}) \cdot \rho^t(\tau) + q \cdot \rho^s(\mathbf{h}) + r \cdot \rho^t(\tau) \quad (1.11)$$

yra visiškai simetrinės kovariacinės funkcijos pavyzdys, kuris virsta į atskiriamos kovariacinės funkcijos modelį, kai $q = r = 0$. Tuo atveju, kai $p = 0$, (1.11) virsta į adityvų kovariacinės funkcijos modelį. Bendru atveju, procesų su atskiriamomis kovariacijomis klasė patenka į procesų, kurių kovariacijos yra visiškai simetriškos, klasę. Kadangi atskiriamumas įtakoja simetriškumą, atmetus prielaidą apie visiškai simetriškumą, prielaida apie atskiriamumą atmetama automatiškai. Remiantis žemiau pateikta teorema, suformuluota Cressie [24], atvirkštinė priklausomybė neegzistuoja.

Teorema 1. Tegu μ yra baigtinis neneigiamas ne tuščios aibės Θ matavimas. Tarkime, kad kiekvienam $\vartheta \in \Theta$, $C^{(s\vartheta)}$ ir $C^{(t\vartheta)}$ yra stacionarios kovariacinės funkcijos atitinkamai erdvėse \mathbb{R}^d ir \mathbb{R} ir tarkime, kad $\int_\Theta |C^{(s\vartheta)}(\mathbf{0}) \cdot C^{(s\vartheta)}(0)| d\mu(\vartheta) < \infty$, tada

$$C(\mathbf{h}; \tau) = \int_\Theta |C^{(s\vartheta)}(\mathbf{h}) \cdot C^{(t\vartheta)}(\tau)| d\mu(\vartheta) < \infty \quad (1.12)$$

yra stacionari erdvės-laiko kovariacinė funkcija.

Tuo atveju, kai (1.12) formulės dešinės pusės kiekvienas narys yra atskiriamas ir dėl to visiškai simetriškas, $C(\mathbf{h}; \tau)$, išreikšta (1.12) formule, nėra atskiriama, tačiau visiškai simetriška.

1.1.2 Erdvės-laiko variogramos

Erdvinėje statistikoje yra žinomas kitas būdas, susijęs su kovariacine funkcija ir padedantis nusakyti priklausomybę. Atsitiktinio proceso $Z(\mathbf{s}; t)$ erdvės-laiko variograma apibrėžiama tokiu būdu:

$$\text{var}(Z(\mathbf{s}; t) - Z(\mathbf{x}; r)) \equiv 2\gamma(\mathbf{s}, \mathbf{x}; t, r), \quad (1.13)$$

stacionarumo atveju $2\gamma(\mathbf{h}; \tau)$; $\mathbf{h} \in \mathbb{R}^d$, $\tau \in \mathbb{R}$. Dydis γ yra vadinamas *semivariograma*. Procesas $Z(\mathbf{s}; t)$ vadinamas *vidiniai (intrinsically) stacionariu*, jeigu jo vidurkis yra konstanta, o semivariograma stacionari. Jeigu $Z(\mathbf{s}; t)$ stacionarus plačiąja prasme su (stacionaria) kovariacine funkcija $C(\mathbf{h}; \tau)$, tada $Z(\mathbf{s}; t)$ yra vidiniai stacionarus su semivariograma [24]:

$$\gamma(\mathbf{h}; \tau) = C(0; 0) - C(\mathbf{h}; \tau), \quad \mathbf{h} \in \mathbb{R}^d, \tau \in \mathbb{R}. \quad (1.14)$$

1.2. Erdvės-laiko krigingas

Erdvės-laiko kovariacinė funkcija gali būti naudojama ne tik statistinių priklausomybių erdvėje ir laike charakterizavimui. Erdvinėje statistikoje gerai žinomas optimalus tiesinės erdvinės prognozės metodas – krigingas, kiekybiškai įvertinamas remiantis erdvine variograma arba erdvine kovariacine funkcija [64]. Erdvės-laiko kontekste, remiantis (1.13)

$$2\gamma(\mathbf{s}, \mathbf{x}; t, r) \equiv \text{var}(Z(\mathbf{s}; t) - Z(\mathbf{x}; r)), \quad \mathbf{s}, \mathbf{x} \in \mathbb{R}^d \text{ ir } t, r \in \mathbb{R}. \quad (1.15)$$

Krigingo tikslas yra prognozuoti $Z(\mathbf{s}_0; t_0)$ iš nepilnų ir daug triukšmo turinčių duomenų. Bet kokios nepaslinktos tiesinės $Z(\mathbf{s}_0; t_0)$ prognozės $Z^{\textcircled{a}}(\mathbf{s}_0; t_0)$, vidutinė kvadratinė prognozės paklaida (MSPE) gali būti išreikšta variogramos arba kovariacinės funkcijos pagalba [19]. Turint erdvės-laiko kovariacinę funkciją, semivariograma gali būti užrašyta taip:

$$\gamma(\mathbf{s}, \mathbf{x}; t, r) = \left(\frac{1}{2} \right) \{C(\mathbf{s}, \mathbf{s}; t, t) + C(\mathbf{x}, \mathbf{x}; r, r)\} - C(\mathbf{s}, \mathbf{x}; t, r). \quad (1.16)$$

Erdvės-laiko kovariacinės funkcijos arba semivariogramos stacionarumo sąlyga yra patogi prielaida vertinant parametrus, tačiau tokia prielaida yra per stipri didelio matavimo erdvės-laiko duomenų atveju. Tuo tarpu krigingas nereikalauja stacionarumo prielaidos, t. y. remiasi bendro pobūdžio erdvės-laiko semivariograma arba kovariacine funkcija. Paprastai krigingo lygtys rašomos darant prielaidą, kad laikas yra tolydus [24].

Tarkime, kad stebimų duomenų modelį galima užrašyti tokiu būdu:

$$Y(\mathbf{s}_i; t_{ij}) = Z(\mathbf{s}_i; t_{ij}) + \omega(\mathbf{s}_i; t_{ij}), j = 1, \dots, T_i, i = 1, \dots, m, \quad (1.17)$$

čia $\{\omega(\mathbf{s}_i; t_{ij})\}$ yra nepriklausomos nuo $Z(\cdot; \cdot)$ matavimo paklaidos su nuliniu vidurkiu ir dispersija σ_ω^2 . Pažymėkime $\mathbf{Y}^{(i)} \equiv (Y(\mathbf{s}_i; t_{ij}); j = 1, \dots, T_i)'; i = 1, \dots, m$. Mūsų tikslas yra prognozuoti reikšmę $Z(\mathbf{s}_0; t_0)$, remiantis visais duomenimis, apibrėžtais (1.17) išraiška.

1.2.1. Paprastasis krigingas

Paprastasis krigingas remiasi prielaida, kad vidurkis $\mu(\mathbf{s}; t) \equiv E(Z(\mathbf{s}; t))$, $\mathbf{s} \in D_s, t \in D_t$ erdvėje ir laike yra pastovus ir žinomas. Tokiu atveju prognozė $Z^*(\mathbf{s}_0; t_0)$ yra tiesinė duomenų kombinacija:

$$Z^*(\mathbf{s}_0; t_0) \equiv \sum_{i=1}^m \sum_{j=1}^{T_i} l_{ij} Y(\mathbf{s}_i; t_{ij}) + c \equiv \boldsymbol{\ell}' \mathbf{Y} + c, \quad (1.18)$$

čia $\mathbf{Y} \equiv (Y^{(1)'}, \dots, Y^{(m)'})'$, $\boldsymbol{\ell}$ ir c yra parinkti optimaliai, t. y. minimizuota vidutinė kvadratinė prognozės paklaida.

Lyginant su erdvine statistika, erdvinio krigingo atžvilgiu, laikas yra paprasčiausiai kita dimensija. Žinoma bet kuri erdvės-laiko kovariacinė funkcija (ar variograma) turi atsižvelgti į tai, kad atstumas laike turi būti interpretuojamas šiek tiek kitaip nei atstumas erdvėje [32]. Panagrinėsime šią situaciją paprastojo krigingo atveju. Pažymėkime $\mathbf{C}_Y \equiv var(\mathbf{Y})$, $\mathbf{c}_0 \equiv cov(Z(\mathbf{s}_0; t_0); \mathbf{Y})$ ir $C_{0,0} \equiv var(Z(\mathbf{s}_0; t_0))$. Tarkime, kad (1.17) formulėje tiek $Z(\cdot; \cdot)$, tiek $\omega(\cdot; \cdot)$ yra Gauso procesai, su kovariacijų matricomis $\boldsymbol{\Sigma}_Z$ ir $\boldsymbol{\Sigma}_\omega$

atitinkamai. Remiantis (1.17) išraiška bei erdviniu paprastuoju kriginu, gausime:

$$\begin{bmatrix} Z(\mathbf{s}_0; t_0) \\ \mathbf{Y} \end{bmatrix} \sim N \left(\begin{bmatrix} \mu(\mathbf{s}_0; t_0) \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} C_{0,0} & \mathbf{c}_0' \\ \mathbf{c}_0 & \mathbf{C}_Y \end{bmatrix} \right), \quad (1.19)$$

čia $\boldsymbol{\mu} \equiv (\mu(\mathbf{s}_i; t_{ij}) : j = 1, \dots, T_i, i = 1, \dots, m)'$, o $\mathbf{C}_Y = \boldsymbol{\Sigma}_Z + \boldsymbol{\Sigma}_\omega$. Tokiu atveju sąlyginis skirstinys bus:

$$Z(\mathbf{s}_0; t_0) | \mathbf{Y} \sim N(\mu(\mathbf{s}_0; t_0) + \mathbf{c}_0' \mathbf{C}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}), C_{0,0} - \mathbf{c}_0' \mathbf{C}_Y^{-1} \mathbf{c}_0). \quad (1.20)$$

Prisiminus faktą, kad paprastasis kriginas remiasi tiesine kombinacija $Z^*(\mathbf{s}_0; t_0) \equiv \boldsymbol{\ell}' \mathbf{Y} + c$, kur $\boldsymbol{\ell}$ ir c yra parenkami taip, kad minimizuotų vidutinę kvadratinę prognozės paklaidą $E(Z(\mathbf{s}_0; t_0) - \boldsymbol{\ell}' \mathbf{Y} - c)^2$ ir remiantis (1.19) prielaida, nesunku pastebėti, kad paprastojo kriginio išraiška gaunama skaičiuojant (1.20) išraiškos sąlyginį vidurkį:

$$Z^*(\mathbf{s}_0; t_0) = E(Z(\mathbf{s}_0; t_0) | \mathbf{Y}) = \mu(\mathbf{s}_0; t_0) + \mathbf{c}_0' \mathbf{C}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}). \quad (1.21)$$

Paprastojo kriginio dispersija σ_{pk}^2 yra minimizuota vidutinė kvadratinė prognozės paklaida. Remiantis (1.19) išraiška, tai yra tiesiog sąlyginė dispersija:

$$\sigma_{pk}^2 \equiv E(Z(\mathbf{s}_0; t_0) - Z^*(\mathbf{s}_0; t_0))^2 = C_{0,0} - \mathbf{c}_0' \mathbf{C}_Y^{-1} \mathbf{c}_0. \quad (1.22)$$

Reikėtų pažymėti, kad kriginio lygtyse yra įtraukta $T_+ \times T_+$ matavimo teigiamai apibrėžta kovariacijų matricos $\mathbf{C}_Y \equiv \text{var}(\mathbf{Y})$ atvirkštinė, čia $T_+ \equiv \sum_{i=1}^m T_i$. Pažymėtina, kad geofiziniuose bei aplinkos taikymuose dažnai rasti šią matricą yra gana sudėtinga dėl per didelio duomenų kiekio. Šis praktinis erdvės-laiko kriginio apribojimas yra viena iš priežasčių, dėl kurių priklausomybės tarp duomenų nustatymui, yra naudojamos atskiriamos kovariacinės funkcijos. Tokiu atveju matrica \mathbf{C}_Y įgyja blokinę struktūrą, dėl kurios atvirkštinės matricos \mathbf{C}_Y^{-1} radimas pasidaro daug paprastesnis.

Erdvės-laiko duomenų vektorius \mathbf{Y} užrašymui dažniausiai naudojami du būdai. (1.18) išraiškoje vektorius yra daugiamatis erdvinis procesas, kur

$\mathbf{Y}^{(i)}$ yra T_i dimensijos laiko duomenų vektorius i -ajame erdvės taške. Alternatyvus būdas yra surinkti visus duomenis laiko momentu t ir pažymėti juos \mathbf{Y}_t . Tokiu atveju reikia pertvarkyti vektoriaus \mathbf{Y} duomenis (1.21) išraiškoje.

Tarkime kiekviename erdvės taške $\{\mathbf{s}_i: i = 1, \dots, m\}$ ir kiekvienu laiko momentu $t = 1, \dots, T$ fiksuotas vienas stebėjimas. Tada $\mathbf{Y} \equiv (\mathbf{Y}^{(1)'}, \dots, \mathbf{Y}^{(m)'})'$ yra daugiamatis (laike) erdvinis procesas. Šiuo atveju erdvinėje statistikoje naudojamas plačiai žinomas daugiamatės erdvinės prognozės metodas – *kokrigingas*. Daugiamatei erdvinei prognozei ir kokrigingui yra skirta gana daug literatūros [24], [25]. Optimali prognozė yra gana paprasta. Tačiau kokrigingas remiasi kryžminėmis erdvės-laiko variogramomis, kurias įvertinti yra gana sudėtinga netgi naudojantis šiuolaikinėmis technologijomis. [24]

Duomenys taip pat gali būti užrašyti $(\mathbf{Y}'_1, \dots, \mathbf{Y}'_T)'$, t. y. daugiamatis (erdvėje) laiko eilučių procesas (pavyzdžiui $VAR(p)$ procesas).

Duomenų \mathbf{Y} formos pasirinkimas priklauso nuo nagrinėjamo proceso fizinių savybių bei nuo sudaromo modelio parametrizacijos galimybių.

1.2.2. Ordinarusis krigingas

Ordinarusis krigingas (OK) naudojamas tuo atveju, kai $Z(\cdot; \cdot)$ turi pastovų nežinomą vidurkį μ . Tokiu atveju $\boldsymbol{\mu} = \mu \mathbf{1}$, kur $\mathbf{1}$ yra T_+ dimensijos vienetų vektorius. Remiantis erdvinio ordinaraus krigingo apibendrintuoju mažiausių kvadratų (GLS) įverčiu

$$\hat{\mu}_{gls} \equiv (\mathbf{1}' \mathbf{C}_Y^{-1} \mathbf{1})^{-1} \mathbf{1}' \mathbf{C}_Y^{-1} \mathbf{Y}, \quad (1.23)$$

(1.21) formulėje įvedus pakeitimą $\mu(\cdot; \cdot) \equiv \mu$, gauname ordinariojo krigingo išraišką

$$Z^*(\mathbf{s}_0; t_0) = \hat{\mu}_{gls} + \mathbf{c}'_0 \mathbf{C}_Y^{-1} (\mathbf{Y} - \hat{\mu}_{gls} \mathbf{1}) \equiv \boldsymbol{\lambda}' \mathbf{Y}, \quad (1.24)$$

čia $\lambda' \equiv \{c_0 + \mathbf{1}(1 - \mathbf{1}'\mathbf{C}_Y^{-1}c_0)/\mathbf{1}'\mathbf{C}_Y^{-1}\mathbf{1}\}'\mathbf{C}_Y^{-1}$. (1.24) formulė duoda prognozę, kuri minimizuoja visų tiesinių nepaslinktų prognozių vidutinę kvadratinę paklaidą. Minimizuota vidutinė kvadratinė prognozės paklaida yra lygi ordinaraus kriginio dispersijai:

$$\sigma_{ok}^2 \equiv C_{0,0} - c_0' \mathbf{C}_Y^{-1} c_0 + (1 - \mathbf{1}' \mathbf{C}_Y^{-1} c_0)^2 / \mathbf{1}' \mathbf{C}_Y^{-1} \mathbf{1}. \quad (1.25)$$

Dydis \mathbf{C}_Y^{-1} figūruoja visose kriginio išraiškose. Dažnai, siekiant rasti matricos \mathbf{C}_Y atvirkštinę, daromos įvairios prielaidos. Tarkime, kad kiekvienas erdvės-laiko duomenų aibės stebėjimas $Y(\mathbf{s}_i; t)$ yra gautas kiekviename erdviniame taške \mathbf{s}_i , kiekvienu laiko momentu t ; $i = 1, \dots, m$, $t = 1, \dots, T$. Tokiu atveju viena iš prielaidų būtų, kad modeliuojant Y kovariaciją erdvė ir laikas yra atskiriami, t.y.:

$$\mathbf{C}_Y = \mathbf{C}_Y^{(s)} \otimes \mathbf{C}_Y^{(t)}, \quad (1.26)$$

čia \otimes yra Kronekerio sandauga, $\mathbf{C}_Y^{(s)}$ - $m \times m$ erdvinių kovariacijų matrica, $\mathbf{C}_Y^{(t)}$ - $T \times T$ laikinių kovariacijų matrica, $T_+ = mT$. Tokiu atveju, matricos \mathbf{C}_Y atvirkštinė matrica:

$$\mathbf{C}_Y^{-1} = (\mathbf{C}_Y^{(s)})^{-1} \otimes (\mathbf{C}_Y^{(t)})^{-1}. \quad (1.27)$$

Taigi vietoje $(mT) \times (mT)$ matavimo matricos, skaičiuojame matavimų $m \times m$ bei $T \times T$ matricų atvirkštines, tai daug mažiau resursų. Taigi, apibendrinant, be atskiriamų modelių arba dimensijų redukavimo, arba matricos \mathbf{C}_Y^{-1} apksimavimo, skaičiavimai pasidaro labai sudėtingi [24].

1.3. Erdvinių laiko eilučių tipai bei jų statistinės savybės

Erdvės-laiko atsitiktinis procesas $\{Z(\mathbf{s}, t): \mathbf{s} \in D_s, t \in D_t\}$ (tolydus laiko) arba $\{Z_t(\mathbf{s}): \mathbf{s} \in D_s, t = 0, 1, \dots\}$ (diskretaus laiko) yra charakterizuojamas baigtiniamatėmis pasiskirstymo funkcijomis. Jeigu procesas Z yra Gauso, tada pakanka išvesti tik pirmuosius du atsitiktinio

proceso momentus, t.y. vidurkio bei kovariacinę funkcijas, gaunant efektyvų statistinį modelį, leidžiantį aprašyti erdvės-laiko priklausomybę.

Toliau šiame darbe nagrinėjamas atvejis, kai $D_t = \{0,1,2, \dots\}$, t.y. diskretaus laiko modelis. Tuo atveju, kai laiko indeksas yra diskretus, erdvės-laiko procesas gali būti užrašytas tokiu būdu

$$\begin{aligned} Z_0(\cdot) &\equiv \{Z(\mathbf{s}; 0): \mathbf{s} \in D_s\}, \\ Z_1(\cdot) &\equiv \{Z(\mathbf{s}; 1): \mathbf{s} \in D_s\}, \\ &\dots \end{aligned} \tag{1.28}$$

kuris gali būti apibendrintas, darant prielaidą, kad erdvinių indeksų aibė D_s (1.28) formulėje priklauso nuo laiko t . Tai yra (1.28) lygtis yra aproksimuotas erdvės-laiko procesas $Z(\cdot; \cdot)$ vadinamas erdvinių procesų laiko eilute:

$$\{Z_t(\cdot): t = 0,1, \dots\}. \tag{1.29}$$

Taigi bet koks erdvės-laiko procesas gali būti analizuojamas kaip erdvinis laiko eilučių procesas, t.y. daugiamatis erdvinis procesas, kur daugiamatės priklausomybės yra paveldėtos iš laikinių priklausomybių, kurios gali priklausyti nuo erdvinių taškų.

1.3.1. Tolydaus erdvinio indekso procesai (Geostatistiniai procesai)

Formaliai, tolydaus erdvinio indekso atvejis yra apibrėžtas tik tuo atveju, kai $|D_s| > 0$, kur bet kokiam $A \subset \mathbb{R}^d$, $|A| \equiv \int_A ds$ yra A yra Lebego matas.

Vienmačių laiko eilučių atveju, atsitiktinių dydžių $\{Z_t: t = 0,1, \dots\}$ laiko eilutės dinaminis modelis gali būti užrašytas tokia forma

$$Z_t = \mathcal{M}_t(Z_{t-1}) + \delta_t, \tag{1.30}$$

čia $\mathcal{M}_t(\cdot)$ yra paprastai netiesinė funkcija, kuri gali priklausyti nuo t ; δ_t yra paklaida, nepriklausanti nuo Z_{t-1} , taip pat $\{\delta_t: t = 0,1, \dots\}$ yra tarpusavyje

nepriklausomi. Remiantis (1.30) išraiška, galima apibrėžti geostatistinių procesų laiko eilutę

$$Z_t(\mathbf{s}) = \mathcal{M}_t(\mathbf{s}, Z_{t-1}(\cdot)) + \delta_t(\mathbf{s}), \mathbf{s} \in D_s, t = 0, 1, \dots, \quad (1.30)$$

čia erdvinis procesas $\delta_t(\cdot)$ yra nepriklausomas nuo erdvinio proceso $Z_{t-1}(\cdot)$, o $\{\delta_t(\cdot): t = 0, 1, \dots\}$ yra tarpusavyje nepriklausomi. Vienas iš būdų yra daryti prielaidą, kad $\delta_t(\cdot)$ yra erdvėje koreliuotas paklaidų procesas su nuliniu vidurkiu, o \mathcal{M}_t laikyti nepriklausomą nuo t , t. y.

$$Z_t(\mathbf{s}) = \mathcal{M}(\mathbf{s}, Z_{t-1}(\cdot)) + \delta_t(\mathbf{s}),$$

kur

$$\mathcal{M}(\mathbf{s}, f(\cdot)) \equiv \int_{\mathbb{R}^d} m(\mathbf{s}, \mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \mathbf{s} \in D_s, \quad (1.31)$$

bet kokiai funkcijai $f(\cdot)$ erdvėje \mathbb{R}^d , kuriai egzistuoja dešinėje lygybės pusėje apibrėžtas integralas. Svorinė funkcija $m(\mathbf{s}, \mathbf{x})$ kontroliuoja proceso $Z_{t-1}(\cdot)$ dalių įtaką dabartinei reikšmei $Z_t(\mathbf{s})$. Pažymėtina, kad $\{\delta_t(\cdot): t = 0, 1, \dots\}$ yra nepriklausomų laike paklaidų procesas; tačiau kiekvienas $\delta_t(\cdot)$ gali išreikšti erdvinę priklausomybę.

Alternatyvus metodas erdvinį laiko eilučių procesų modeliavimui yra laikinės stacionarios kovariacinės funkcijos taikymas $C_s(\tau)$, $\tau = 0, \pm 1, \pm 2, \dots$, kur indeksas nurodo laiko eilutės vietą (tašką).

Šiame skyriuje aptarti dinaminiai geostatistinių procesų modeliai, kur erdvinį indeksų aibę D_s buvo tolydi. Dažnai šią aibę galima diskretizuoti, tokiu atveju gaunant gardelės proceso laiko eilutes. Apie diskretaus erdvinio indekso procesus kalbama sekančiame skyrelyje.

1.3.2. Diskretaus erdvinio indekso procesai (Gardelės procesai)

Kai D_s yra baigtinė arba suskaičiuojama, jos Lebegeo matas yra lygus nuliui; tai yra $|D_s| = 0$. Suskaičiuojamos aibės $D_s \in \mathbb{R}^d$ pavyzdys yra d -matė

sveikųjų skaičių gardelė $\{0,1, \dots\}^d$. Praktikoje tokios gardelės paprastai yra apribojamos, gaunant baigtines gardeles. Toliau šiame skyrelyje bus daroma prielaida, kad

$$D_s = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\} \quad (1.32)$$

yra baigtinis taškų rinkinys erdvėje. Tada erdvės-laiko procesas yra

$$\{Z_t(\mathbf{s}_i): i = 1, \dots, n; t = 0,1, \dots\}. \quad (1.33)$$

Taip pat tikėtina, kad $Z_t(\cdot)$ gali būti stebimas kiekviename D_s taške. Kadangi D_s yra baigtinė, (1.29) išraiška gali būti užrašyta tokiu būdu:

$$\mathbf{Z}_t \equiv (Z_t(\mathbf{s}_1), \dots, Z_t(\mathbf{s}_n))', \quad t = 1,2, \dots \quad (1.34)$$

t. y. bet kuriam laiko momentui (1.34) gali būti užrašyta n -matės laiko eilutės pavidalu

$$\mathbf{Z}_0, \mathbf{Z}_1, \dots \quad (1.35)$$

Dažnai daroma prielaida, kad modeliuoti laikinę priklausomybę yra daug svarbiau, nei erdvinę. Tarkime Markovo modelis (1.35) tipo laiko eilutei

$$[\mathbf{Z}_t | \mathbf{Z}_0, \dots, \mathbf{Z}_{t-1}] = [\mathbf{Z}_t | \mathbf{Z}_{t-1}], \quad t = 1,2, \dots, \quad (1.36)$$

neatskleidžia erdvinės priklausomybės. Daugiau \mathbf{Z}_t erdvinio modeliavimo yra naudojama bendrame autoregresiniame (SAR) bei sąlyginiame autoregresiniame (CAR) modeliuose.

Procesą $\{Z_t(\mathbf{s}_i)\}$ taip pat galima modeliuoti kaip erdvinį laiko eilučių procesą, tai yra $\{\mathbf{Z}(\mathbf{s}_i): i = 1, \dots, n; t = 0,1, \dots\}$, kur $\mathbf{Z}(\mathbf{s}_i) \equiv (Z_0(\mathbf{s}_i), \dots, Z_T(\mathbf{s}_i))'$. Taip pat galima erdvės-laiko priklausomybę nagrinėti, remiantis prielaida, kad tai yra $(T + 1)$ -matis erdvinis gardelės procesas.

1.4. Erdvės-laiko duomenų modeliai

Vieni populiariausių erdvės-laiko duomenų modelių yra vektorinis autoregresijos slenkamojo vidurkio modelis (VARMA) bei erdvės-laiko autoregresijos slenkamojo vidurkio modelis (STARMA) bei įvairios jų modifikacijos. Šiame skyriuje aprašyti šie modeliai, išanalizuotos kai kurios jų savybės, pranašumai bei trūkumai.

1.4.1. Vektorinis autoregresijos slenkamojo vidurkio modelis (VARMA)

n -matis procesas \mathbf{Z}_t yra vadinamas vektoriniu autoregresijos slenkamojo vidurkio procesu, jeigu $\{\mathbf{Z}_t\}$ yra stacionarus ir su visais t tenkina tokią lygtį:

$$\mathbf{Z}_t + \mathbf{M}_1\mathbf{Z}_{t-1} + \dots + \mathbf{M}_p\mathbf{Z}_{t-p} = \boldsymbol{\varepsilon}_t + \mathbf{A}_1\boldsymbol{\varepsilon}_{t-1} + \dots + \mathbf{A}_q\boldsymbol{\varepsilon}_{t-q}, \quad (1.37)$$

kur $\mathbf{Z}_t \equiv (Z_t(\mathbf{s}_1), \dots, Z_t(\mathbf{s}_n))'$, $t = 1, 2, \dots, T$; $\mathbf{A}_1, \dots, \mathbf{A}_p, \mathbf{M}_1, \dots, \mathbf{M}_q$ yra $n \times n$ eilės svorių matricos (kartais vadinamos būsenų perėjimo matricomis), o $\{\boldsymbol{\varepsilon}_t\}$ yra n -matė balto triukšmo seka su nuliniu vidurkiu ir kovariacijų matrica $\boldsymbol{\Sigma}$, t.y. $\boldsymbol{\varepsilon}_t \sim BT(\mathbf{0}, \boldsymbol{\Sigma})$. Tokia seka paprastai žymima $VARMA(p, q)$.

Lygybę (1.37) galima užrašyti operatorine forma:

$$\mathbf{M}(B)\mathbf{Z}_t = \mathbf{A}(B)\boldsymbol{\varepsilon}_t,$$

čia $\mathbf{M}(z) = \mathbf{I} + \mathbf{M}_1z + \dots + \mathbf{M}_pz^p$; $\mathbf{A}(z) = \mathbf{I} + \mathbf{A}_1z + \dots + \mathbf{A}_qz^q$ yra pilno rango (angl. full rank) matriciniai daugianariai, o B – postūmio atgal operatorius, t.y. $B^j\mathbf{Z}_t = \mathbf{Z}_{t-j}$.

Atskiras $VARMA(p, q)$ proceso atvejis yra $VAR(p)$ procesas, kuris susieja keleto laiko eilučių reikšmes laiko momentu t su tų laiko eilučių tiesinėmis kombinacijomis praėjusiais laiko momentais. n -matės laiko eilutės atveju, $VAR(p)$ procesas užrašomas tokiu būdu:

$$\mathbf{Z}_t = \mathbf{M}_1\mathbf{Z}_{t-1} - \mathbf{M}_2\mathbf{Z}_{t-2} - \dots - \mathbf{M}_p\mathbf{Z}_{t-p} + \boldsymbol{\varepsilon}_t. \quad (1.38)$$

Pavyzdžiui $VAR(1)$ procesas būtų užrašomas tokiu būdu:

$$\mathbf{Z}_t = \mathbf{M}\mathbf{Z}_{t-1} + \boldsymbol{\varepsilon}_t. \quad (1.39)$$

Šiuo atveju, perėjimų matricos \mathbf{M} i – toji eilutė sudaryta iš svorių, kurie parodo kaip keletas laiko eilučių laiko momentu $t - 1$ tiesiškai įtakoja i – tają laiko eilutę laiko momentu t :

$$Z_t(s_i) = m_{i,1}Z_{t-1}(s_1) + m_{i,2}Z_{t-1}(s_2) + \dots + m_{i,n}Z_{t-1}(s_n) + \varepsilon_t(s_i),$$

čia $m_{i,j}$ atitinka matricos \mathbf{M} elementą (i, j) , o $\varepsilon_t(s_i)$ yra vektoriaus $\boldsymbol{\varepsilon}_t$ i – tasis elementas; $i = 1, \dots, n$. Žinoma tokia sąveika tarp visų laiko eilučių bei galimai koreliuoto triukšmo gana stipriai apsunkina modeliavimą.

Be papildomų apribojimų ir prielaidų, pažymėtina, kad parametų, aprašančių (1.35) pasiskirstymą, skaičius gali būti labai didelis. (1.38) modelis atsižvelgia į laikinę priklausomybę, leisdamas sudaryti paprastesnį (parametų atžvilgiu) modelį. Siekiant nustatyti erdvinių „kaimynų“ priklausomybes, galima atsižvelgti tik į erdvinę priklausomybę, taip sudarant dar paprastesnį modelį. n -matis $VAR(1)$ modelis yra atskiras (1.36) modelio atvejis. Čia, bendru atveju, matrica \mathbf{M} turi n^2 parametų, o $\Sigma_\varepsilon = var(\boldsymbol{\varepsilon})$ turi $O(n^2)$ parametų. Parametų mažinimui yra siūloma nemažai būdų, dalis jų aprašyta Cressie ir Wickle [24].

Atskiras vektorinio autoregresinio slenkamojojo vidurkio ($VARMA$) modelio atvejis yra erdvės-laiko autoregresijos slenkamojojo vidurkio modelis ($STARMA$). $VARMA$ modeliai naudoja bendras $n \times n$ autoregresijos ir slenkamojojo vidurkio parametų matricas tam, kad nusakyti visas autokoreliacijas ir kryžmines koreliacijas viduje ir tarp n laiko eilučių. Jeigu daroma prielaida, kad diagonaliniai tų matricų elementai yra lygūs (tai yra atvejis, kai n eilučių aprašo tą patį atsitiktinį procesą, vykstantį skirtinguose taškuose), o nediagonaliniai elementai yra tiesinės svorių matricos \mathbf{W}_t kombinacijos, tai bendra $VARMA$ šeima virsta į $STARMA$ modelių klasę, žr.

(1.40). *STARMA* (arba *STARIMA*) metodologija leidžia žymiai sumažinti vertinamų parametru skaičių, lyginant su *VARMA* modeliu [45].

1.4.2. Erdvės-laiko autoregresijos slenkamojo vidurkio (*STARMA*)

modelis

STARMA modelyje priklausomybė tarp n laiko eilučių yra susijusi su taškų padėtimi. Modelio forma priklauso nuo $n \times n$ svorių matricos, kuri atsižvelgia į erdvinį taškų išsidėstymą. Tam, kad suformuluoti erdvės-laiko modelį, reikalingas erdvinio postūmio operatorius. Tegu $L^{(l)}$ yra erdvinis l -os eilės postūmio operatorius, toks kad:

$$L^{(l)}Z_t(\mathbf{s}_i) = \sum_{j=1}^n w_{ij}^{(l)} Z_t(\mathbf{s}_j),$$

kur $w_{ij}^{(l)}$ yra tokia svorių aibė, kad $\sum_{j=1}^n w_{ij}^{(l)} = 1$ visiems i ir $w_{ij}^{(l)}$ nenulinė tik tada, kai taškai i ir j yra l -os eilės kaimynai. $n \times n$ matavimo matrica, sudaryta iš svorių $w_{ij}^{(l)}$, kurios kiekvienos eilutės elementų suma yra lygi vienam, yra žymima \mathbf{W}_l .

Svorių $w_{ij}^{(l)}$ įvertinimas duotiems l yra paliekamas modelio sudarytojui: jie turi būti parinkti taip, kad atspindėtų esamą erdvinę situaciją. Pvz.: svoriai $w_{ij}^{(l)}$ gali būti parenkami taip, kad atspindėtų fizines nagrinėjamos sistemos savybes: ribos tarp apygardų (rajonų, miestų) i ir j ilgis, atstumas tarp apygardų (rajonų, miestų) centrų, natūralių barjerų (pvz.: kalnai ir upės) egzistavimas ir trumpiausias kelias tarp i ir j . Šis paskutinysis faktorius gali atsižvelgti į įvairias fizines savybes: kelių tarp i ir j skaičius, susisiekimo tarp i ir j galimybių kiekis ir panašiai [31].

Svoriai turi atspindėti hierarchinę erdvinių kaimynų eilę. Pirmos eilės kaimynai yra „artimiausi“ nagrinėjamai vietai taškai. Antros eilės kaimynai turi būti „toliau“ nei pirmos eilės, bet „arčiau“ nei trečios eilės kaimynai. Tuo atveju, kai taškai yra tvarkingai išsidėstę erdvėje (gardelės tipo duomenys),

žinoma veiksminga taisyklė, padedanti nustatyti erdvinių kaimynų eilę. Šis erdvinės eilės apibrėžimas atspindi rangavimą remiantis Euklido atstumais tarp nagrinėjamų taškų.

Kiekviena iš n laiko eilučių yra tuo pačiu tiesinė praėjusių stebėjimų ir klaidų gretimose stotyse kombinacija. Vienmačiai *ARMA* modeliai atspindi idėją, kad netolimos praeities įvykiai turi daugiau įtakos negu tolimesnės, atitinkamai *STARMA* modeliai atspindi (per svorinės matricos specifikaciją) idėją, kad artimesni taškai daug labiau įtakoja viena kitą nei tolimesni.

Taigi *STARMA* modelių klasėje kiekvienas stebėjimas laiko momentu t taške \mathbf{s}_i yra tiesinė prieš tai buvusių stebėjimų su vėlavimu erdvėje ir laike kombinacija. Šios išraiškos bazinis mechanizmas yra kiekvienos stoties kaimynų hierarchinė tvarka ir atitinkama svorinių matricų seka.

Daugiamatei laiko eilutei

$$\mathbf{Z}_t \equiv (Z_t(\mathbf{s}_1), \dots, Z_t(\mathbf{s}_n))', t = 1, 2, \dots$$

STARMA modelis yra užrašomas formule:

$$\mathbf{Z}_t = \sum_{k=1}^p \left(\sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}_l \right) \mathbf{Z}_{t-k} + \sum_{k=1}^q \left(\sum_{l=0}^{\mu_k} \theta_{kl} \mathbf{W}_l \right) \boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\varepsilon}_t, t = 0, 1, \dots, \quad (1.40)$$

čia $\{\mathbf{W}_l\}$ yra žinomos svorių matricos; p ir q yra atitinkamai autoregresijos ir slenkamojo vidurkio eilės; $\{\phi_{kl}\}$ ir $\{\theta_{kl}\}$ – modelio parametrai; λ_k – k -ojo autoregresijos elemento erdvinė eilė; μ_k – k -ojo slenkamojo vidurkio elemento erdvinė eilė; $\{\boldsymbol{\varepsilon}_t\}$ – nepriklausomi vienodai pasiskirstę atsitiktiniai vektoriai su nulniais vidurkiais ir kovariacijų matrica $\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \mathbf{I}$.

STARMA(p_λ, q_μ) modeliai yra identifikuojami remiantis erdvės-laiko autokoreliacinėmis (ACF) ir dalinės autokoreliacijos (PACF) funkcijomis bei Yule - Walker lygtimis. Plačiau apie *STARMA*(p_λ, q_μ) modelių identifikavimą vertinimą ir patikrinimą galima rasti Pfeifer ir Deutsch darbuose [60], [61].

Pastaruoju metu $STARMA(p_\lambda, q_\mu)$ modelis sulaukė nemažai kritikos dėl erdvinio heterogeniškumo nebuvimo, kaip pabrėžia, Cressie [24] ir kiti autoriai $STARMA$ modelių klasė, sudaryta griežtai remiantis vienmačių laiko eilučių ARMA modeliavimo technika, nesuteikia pilnai adekvataus erdvinio duomenų charakteristikų modeliavimo. Standartinė $STARMA$ modelio specifikacija netiesiogiai teigia, kad priklausomai nuo praeities stebėjimų, procesas pasidaro nebekoreliuotas erdvėje, t. y. erdvinę koreliaciją apriboja laikinė priklausomybė.

Alternatyva $STARMA(p_\lambda, q_\mu)$ modeliui yra apibendrintas erdvės-laiko autoregresijos slenkamojojo vidurkio modelis $STARMA_G$. Šiame modelyje į erdvinį heterogeniškumą atsižvelgiama per erdvėje kintančius parametrus, t. y. tiek laikiniai autoregresiniai efektai, tiek erdviniai autoregresiniai efektai yra indeksuojami erdvėje. $STARMA_G$ modelis yra užrašomas tokiu būdu:

$$\mathbf{Z}_t = \sum_{k=1}^p \left(\sum_{l=0}^{\lambda_k} \phi_{kl}^i \mathbf{W}_l \right) \mathbf{Z}_{t-k} + \sum_{k=1}^q \left(\sum_{l=0}^{\mu_k} \theta_{kl}^i \mathbf{W}_l \right) \boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\varepsilon}_t, \quad (1.41)$$

Šis modelis prie tam tikrų sąlygų vadinamas apribotu $VARMA$ modeliu, o parametrų vertinimui taikomas didžiausio tikėtimumo (ML) metodas.

Pagrindiniai $VARMA$ bei $STARMA$ (ar $STARMA_G$) modelių trūkumai yra didelis parametrų skaičius. Nors, lyginant, su $VARMA$ modeliu, kuriame vertinamų parametrų skaičius yra $(p + n^2) + n$, čia p – vėlavimų skaičius laike, o n – taškų skaičius, $STARMA$ modelyje parametrų skaičius yra mažesnis. Tačiau $STARMA$ modelio parametrų vertinimui reikalingos kryžminės erdvės-laiko koreliacijos, kurių skaičiavimas yra gana sudėtingas ir daug laiko reikalaujantis uždavinys. Šio darbo 2-ajame skyriuje aprašyti metodai neatsižvelgia į kryžmines erdvės-laiko koreliacijas - tai sumažina vertinamų parametrų skaičių ir palengvina darbą tyrėjui.

2. ERDVINIŲ LAIKO EILUČIŲ MODELIAVIMAS

Šiame skyriuje aprašyti siūlomi erdvinių laiko eilučių modeliavimo metodai. Dažniausiai analizuojamos gana ilgos laiko eilutės, o erdvės taškų skaičius yra žymiai mažesnis, dėl to laikinė priklausomybė yra svarbesnė nei erdvinė priklausomybė. Dėl šios priežasties kiekviename taške pirmiausia apibrėžiamas vienmatis laiko eilutės modelis, kuris nepriklauso nuo erdvinės informacijos. Išsamiai aprašomi du vienmačių laiko eilučių modeliavimo metodai: klasikinis *ARIMA* modelis bei modelis, sudarytas iš trendo, sezoninės komponentių bei *ARMA* proceso. Įvertinus vienmačių laiko eilučių parametrus, erdvinės priklausomybės įvertinama semivariogramos pagalba. Semivariograma vertinama, normuotoms liekanoms, interpretuojant jas kaip pakartotinius stebėjimus (replikacijas). Erdviniam laiko eilučių interpoliavimui siūloma erdvinių svorių apskaičiavimo formulė, „formalus“ ordinarusis kriginas bei du nestochastiniai - trendo paviršiaus bei svertinis atvirkštinių atstumų - metodai.

2.1. Stochastinis erdvės-laiko modelis

Darbe analizuojami erdvės-laiko duomenys yra atsitiktinio lauko $\{Z(\mathbf{s}; t): \mathbf{s} \in D_s, t \in D_t\}$ realizacija, kur $D_s \subset \mathbb{R}^2$, $D_t \subset [0, \infty)$. $Z(\mathbf{s}; t)$ gali būti apibrėžtas (1.1) modelio pagalba.

Proceso $Z(\mathbf{s}; t)$ paklaidos $\epsilon(\mathbf{s}; t)$ skaidomos į dviejų procesų sandaugą:

$$\epsilon(\mathbf{s}; t) = \sigma(\mathbf{s}; t)\varepsilon(\mathbf{s}; t), \quad (2.1)$$

čia $\sigma(\mathbf{s}; t)$ yra neatsitiktinė k -periodo funkcija, tenkinanti sąlygą

$$\sigma(\mathbf{s}; t) = \sigma(\mathbf{s}; t + k) \quad (2.2)$$

bet kuriam laiko momentui t ir periodui $k \in \mathbb{R}$. Taip pat daroma prielaida, kad

$$\varepsilon(\mathbf{s}; t): \mathbf{s} \in \mathbb{R}^2, t \in [0, \infty) \quad (2.3)$$

yra atsitiktinės stacionarios Gauso paklaidos su nuliniu vidurkiu, o erdvinė priklausomybė gali būti išreikšta parametrine koreliacine funkcija, nepriklausoma nuo t :

$$\text{corr}(\varepsilon(\mathbf{s}; t), \varepsilon(\mathbf{s} + \mathbf{h}; t)) = \rho(\mathbf{h}; \boldsymbol{\vartheta}), \quad (2.4)$$

visiems $\mathbf{s}, \mathbf{s} + \mathbf{h} \in D_s$, $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \dots)' \in \Theta \subset \mathbb{R}^l$ yra $l \times 1$ erdvės parametrų vektorius, arba semivariograma

$$\frac{1}{2} \text{var}(\varepsilon(\mathbf{s}; t) - \varepsilon(\mathbf{s} + \mathbf{h}; t)) = \gamma(\mathbf{h}; \boldsymbol{\vartheta}). \quad (2.5)$$

Taip pat daroma prielaida, kad kryžminės erdvės-laiko koreliacijos yra lygios nuliui, tai yra nagrinėjamas atskiriamas modelis, o paklaidos $\varepsilon(\mathbf{s}; t)$ yra nekoreliuotos laike bei koreliuotos erdvėje. Erdvės-laiko paklaidų, apibrėžtų (2.1) lygtimi, kovariacinė funkcija yra apibrėžiama tokiu būdu:

$$C(\mathbf{s}, \mathbf{s} + \mathbf{h}; \boldsymbol{\vartheta}_{st}) = \rho(\mathbf{h}_s; \boldsymbol{\vartheta}_s) \boldsymbol{\Sigma}_s(\boldsymbol{\vartheta}_t) \boldsymbol{\Sigma}_{s+h}(\boldsymbol{\vartheta}_s), \quad (2.6)$$

kai $\boldsymbol{\Sigma}_s(\boldsymbol{\vartheta}_t)$ yra diagonalinė kovariacijų matrica taške $\mathbf{s} \in D_s$, o $\boldsymbol{\vartheta}_{st} = (\boldsymbol{\vartheta}_s, \boldsymbol{\vartheta}_t)' \in \Theta \subset \mathbb{R}^l$ – erdvės ir laiko parametrų vektorius.

2.2. Vienmačių laiko eilučių modeliai

Tegu $Z_i(t)$ yra erdvės taškų $s_i \in D, i = 1, \dots, n, t = 1, 2, \dots, T$ laiko eilutė. Šiame disertaciniame darbe aprašomi du laiko eilučių tipai: pirmiausia laiko eilutė skaidoma į trendo komponentę, sezoninę komponentę bei *ARMA* procesą, alternatyva šiam modeliui yra klasikinis nestacionarių laiko eilučių *ARIMA* procesas.

Laiko eilutės $Z_i(t)$ modelis gali būti užrašytas tokiu būdu:

$$Z_i(t) = L_i(t) + S_i(t) + \sum_{k=1}^p \phi_k^i Y_i(t-k) + \sum_{l=1}^q \theta_l^i \varepsilon_i(t-l) + \varepsilon_i(t), \quad (2.7)$$

kur $L_i(t) = \sum_m a_m^i t^m$, $m = 0, 1, \dots$ yra m -os eilės polinomas, aprašantis laiko eilutės tendą; $X_i(t) = Z_i(t) - L_i(t)$; $Y_i(t) = X_i(t) - S_i(t)$; ϕ_k^i yra autoregresijos proceso $AR(p_k)$ parametrai, o θ_l^i – slenkamojo vidurkio proceso $MA(q_l)$ parametrai, $S_i(t)$ – sezoninė laiko eilutės komponentė, apibrėžiama tokiu būdu:

$$S_i(t) = b_0^i + \sum_{r=0}^R \left[b_{2r+1}^i \cos\left(\frac{(2r+2)\pi t}{365}\right) + b_{2r+2}^i \sin\left(\frac{(2r+2)\pi t}{365}\right) \right]. \quad (2.8)$$

Be to daroma prielaida, kad procesas $\epsilon_i(t)$ gali būti užrašytas tokia išraiška:

$$\epsilon_i(t) = \sigma_i(t) \varepsilon_i(t), \quad (2.9)$$

čia $\sigma_i(t)$ yra nuo laiko priklausanti standartinio nuokrypio funkcija, o $\varepsilon_i(t)$ – laike nepriklausomi Gauso atsitiktiniai dydžiai su nuliniu vidurkiu bei standartiniu nuokrypiu lygiu 1.

Vienas iš dispersijos modeliavimo būdų yra Furje eilučių

$$\sigma_i^2(t) = c_0^i + \sum_{j=1}^J \left[c_{2j}^i \cos\left(\frac{2j\pi t}{365}\right) + c_{2j+1}^i \sin\left(\frac{(2j+1)\pi t}{365}\right) \right] \quad (2.10)$$

taikymas empirinei dienos dispersijai. Alternatyvus būdas yra naudoti empirinę dienos dispersiją.

Kiekviename taške atlikus aukščiau aprašytą laiko eilutės dekompoziciją, vertinami nežinomi modelio parametrai. Nustačius ar laiko eilutei būdingas trendas ir jį pašalinus, vertinami (2.8) lygtimi aprašytos sezoniškumo funkcijos parametrai. Įvertinus geriausiai tinkantį modelį, laiko eilutės liekanoms gali būti būdingas triukšmas, tačiau sezoninė priklausomybė turėtų būti pašalinta. Liekanų triukšmui įvertinti paprastai taikomi klasikiniai stacionarių laiko eilučių modeliai: autoregresijos slenkamojo vidurkio (*ARMA*) modelis arba atskiri jo atvejai - autoregresijos (*AR*) arba slenkamojo vidurkio (*MA*) modeliai.

ARMA modelius laiko eilutės duomenims vertiname tokiais etapais: duomenų grafinis vaizdavimas, galimas duomenų transformavimas, modelio

tipo nustatymas, *AR* bei *MA* eilės parinkimas, parametru vertinimas bei diagnostinis patikrinimas. Nubrėžus laiko eilutės $Z_i(t)$ priklausomybės nuo laiko t grafiką, galime patikrinti ar nėra jokių anomalijų, be to analizuojamų duomenų pasiskirstymas turi būti artimas normaliajam, laiko eilutė stacionari. Priešingu atveju duomenis būtina transformuoti. Dažniausiai taikoma logaritminė, t.y. $Y_i(t) = \ln Z_i(t)$, kai $Z_i(t) > 0$ arba Box–Cox laipsninė transformacija:

$$Y_i(t) = \begin{cases} (Z_i^{\lambda_s}(t) - 1)/\lambda_s, & \text{kai } \lambda_s \neq 0, \\ \ln Z_i(t), & \text{kai } \lambda_s = 0, \end{cases} \quad (2.11)$$

kur parametro λ_s parinkimui yra naudojami įvairūs metodai. Analizuojant erdvės-laiko procesus, Box–Cox laipsninė transformacija nėra patogi, nes parametras λ_s priklauso nuo erdvės.

Sekantis žingsnis yra laiko eilutės modelio ir jo autoregresijos eilės p ir slenkamojo vidurkio eilės q nustatymas. Preliminarias p ir q reikšmes galima pasirinkti remiantis autokoreliacinės (ACF) ir dalinės autokoreliacinės (PACF) funkcijų elgesiu: jei $p = 0$ ir $q > 0$, ACF nutrūksta po vėlavimo q , o PACF gęsta; jei $q = 0$ ir $p > 0$, PACF nutrūksta po vėlavimo l , o ACF gęsta; jei $p > 0$ ir $q > 0$, tiek ACF, tiek PACF gęsta. Jeigu tiek autokoreliacinė, tiek dalinė autokoreliacinė funkcijos gęsta, tai rodo, jog laiko eilutės neįmanoma aprašyti *AR* arba *MA* modelio pagalba, o reikia bendro *ARMA* modelio. Kadangi visada analizuojame ne tikrąsias ACF ir PACF reikšmes, o tik jų įverčius, ne visada yra aišku kur ACF ir PACF nutrūksta ar gęsta. Dažnai tenka lygiagrečiai vertinti keletą skirtingų modelių ir tik vėliau apsispręsti kuris yra tinkamiausias [71].

Kitas modelio parinkimo žingsnis yra diagnostika. Šiame etape tiriamos liekanos $\hat{\epsilon}_i(t)$, pagal kurias sprendžiama apie modelio tinkamumą.

Jei modelis tinkamas, normuotos liekanos turi būti pasiskirsčiusios pagal normalųjį skirstinį su nuliniu vidurkiu ir dispersija lygia vienetui. Tačiau jei liekanos yra Gauso, dar nereiškia, kad jos nėra koreliuotos. Marginalinį

normalumą galima nustatyti liekanų histogramos pagalba. Šio metodo patobulinimas yra $Q-Q$ grafikas, kuris gali padėti nustatyti nukrypimus nuo normalumo. Tarkime, kad liekanos yra išrikiuotos didėjimo tvarka: $\hat{\epsilon}_i(1) \leq \hat{\epsilon}_i(2) \leq \dots \leq \hat{\epsilon}_i(T)$. Empiriškai, liekanų proporcija mažesnė arba lygi $\hat{\epsilon}_i(k)$ yra k/T . Tiriant ilgesnę laiko eilutę proporcija aproksimuojama $(k - 1/2)/T$. Jei liekanos yra artimos normaliajam atsitiktiniam dydžiui, $\hat{\epsilon}_i(k)$ apytiksliai turi būti lygi $(k - 1/2)/T$ standartinio normaliojo skirstinio (tarkime $\alpha_{(k)}$) kvantiliui. Vadinasi, jei $\hat{\epsilon}_i(t) \sim N(0,1)$, tada $P[\hat{\epsilon} \leq \alpha_k] = (k - 1/2)/T$. $Q-Q$ grafikas yra $(\alpha_{(k)}, \hat{\epsilon}_i(k))$ grafikas visiems $k = 1, \dots, T$. Jei liekanos tenkina hipotezę apie normalųjį pasiskirstymą, poros bus susijusios tiesiškai. Tiesiškumo testą galima atlikti skaičiuojant porų $(\alpha_{(k)}, \hat{\epsilon}_i(k))$ koreliacijas.

Taip pat galima ištirti liekanų autokoreliacijas $\hat{\rho}_\epsilon(h)$, čia h yra vėlavimas, tikrinant ar jos neturi jokių „pėdsakų“ ar didelių reikšmių. Pervardijant tai balto triukšmo sekai, imties ACF yra nepriklausoma ir normaliai pasiskirsčiusi su nuliniiais vidurkiais ir dispersijomis $1/T$. Taigi norint patikrinti liekanų koreliacijų struktūrą, galima nubrėžti $\hat{\rho}_\epsilon(h)$ priklausomybės nuo h grafiką su klaidų ribomis $\pm 2/\sqrt{T}$. Žinoma parinkto modelio liekanos niekada nebus visiškai baltas triukšmas ir $\hat{\rho}_\epsilon(h)$ dispersija gali būti daug didesnė už $1/T$ [71].

Papildant $\hat{\rho}_\epsilon(h)$ grafinį vaizdavimą, galima apibrėžti bendrą *Portmanteau* testą, kuris nagrinėja $\hat{\rho}_\epsilon(h)$ kvadratų sumą. T.y. gali būti, kad kiekviena atskira $\hat{\rho}_\epsilon(h)$ reikšmė yra nedidelė, tačiau bendrai reikšmės yra didelės. Tokią hipotezę galima patikrinti *Ljung – Box – Pierce* statistikos pagalba:

$$Q = T(T + 2) \sum_{h=1}^H \frac{\hat{\rho}_\epsilon^2(h)}{T-h},$$

čia H reikšmė yra pasirenkama savarankiškai. Esant išpildytai nulinei hipotezei, kad modelis yra adekvatus, asimptotiškai $Q \sim \chi_{H-p-q}^2$.

Taip pat reikia nepamiršti, kad modelio negalima „pervertinti“, tai reiškia, kad didesnis parametru kiekis ne visada yra geriau. Todėl praktikoje dažnai apsiribojama AR modelių sudarymu.

Kriterijai, padedantys nustatyti, kuris iš kelių modelių geriausiai aprašo analizuojamą laiko eilutę yra Akaike informacinis kriterijus (AIC), koreguotas Akaike informacinis kriterijus (AICc) ir Bayeso informacinis kriterijus (SIC). AIC kriterijus buvo pasiūlytas Akaike [71]:

$$AIC = \ln \hat{\sigma}_u^2 + \frac{2u}{T},$$

čia $\hat{\sigma}_u^2$ parinkto modelio su u parametru dispersija.

AICc kriterijus yra AIC patobulinimas, kurį išplėtojo Hurvich ir Tsai [71]. Jis išreiškiamas:

$$AICc = \ln \hat{\sigma}_u^2 + \frac{T+u}{T-u-2}$$

SIC yra korekcija, atlikta remiantis Bayeso argumentais:

$$SIC = \ln \hat{\sigma}_u^2 + \frac{u \cdot \log T}{T},$$

SIC tikslesnį rezultatą duoda, kai turime pakankamai didelį duomenų masyvą, o AICc yra pranašesnis, kai dirbame su trumpesnėmis laiko eilutėmis ir parametru skaičius yra didelis.

Remiantis aukščiau aprašytais kriterijais, kiekvienai vienmateriai laiko eilutei $Z_i(t)$ yra sudaromas (2.7) lygtimi užrašytas laiko eilutės modelis bei atliekamas modelio diagnostinis patikrinimas.

Alternatyvus būdas (2.7) lygtimi aprašytam laiko eilutės modeliui, kiekvienai laiko eilutei $Z_i(t)$ gali būti įvertintas integruotas autoregresijos slenkamojo vidurkio modelis *ARIMA*:

$$\phi^{*i}(B) \nabla^d Z_i(t) = \theta^{*i}(B) \epsilon_i(t), \quad (2.12)$$

kuriame $\phi^i(B) = (1 - \phi_1^* B - \dots - \phi_p^* B^p)$ ir $\theta^i(B) = (1 - \theta_1^* B - \dots - \theta_q^* B^q)$ atitinkamai autoregresijos bei slenkamojo vidurkio operatoriai; $\nabla^d = (1 - B)^d$ – d -os eilės skirtumo operatorius.

ARIMA proceso, aprašyto (2.12) lygtimi, sudarymas remiasi aukščiau išanalizuotais *ARMA* proceso vertinimo kriterijais, tačiau pirmas ir svarbiausias etapas yra laiko eilutės suglodinimas, pašalinant sezoniskumą. Dažniausiai tinkamiausia skirtumo d eilė yra mažiausia eilė, kuri lemia laiko eilutės reikšmių svyravimą apie apibrėžtą vidurkj, o autokoreliacinės funkcijos reikšmės (ACF) gana greitai gęsta tiek iš teigiamos, tiek iš neigiamos pusės. Remiantis ACF elgesiu galima nuspręsti ar reikalinga aukštesnė diferencijavimo eilė. Aukštesnė eilė reikalinga tuo atveju, kai laiko eilutei būdingas ilgalaikis trendas ar laiko eilutė nesvyruoja apie savo vidurkj, arba autokoreliacijos dar dideliame vėlavimų skaičiui yra reikšmingos. Originalios laiko eilutės stacionarumas reiškia, kad $d = 0$. Kai $d = 1$, originali eilutė turi trendą, kuris yra konstanta; kai $d = 2$, reiškia, kad originaliai laiko eilutei būdingas nuo laiko priklausantis trendas. Nustačius preliminarią d eilę, sekantis žingsnis yra eilutės $\nabla^d Z_i(t)$ ACF ir PACF analizavimas. Įvertinus nežinomus modelio parametrus sudaromas integruotas autoregresijos slenkamojo vidurkio procesas [9].

Žinoma, stiprus sezoniskumas, analizuojant įvairius meteorologinius parametrus, reikalauja į *ARIMA* modelį įtraukti sezoninę komponentę, sudarant sezoninį *ARIMA*, arba kitaip tariant *SARIMA* modelį. Tokiu atveju sezoninė priklausomybė egzistuoja tarp $Z_i(t)$ ir $Z_i(t - g)$, kur g yra sezoninis ciklas. Sezoninis ciklas galėtų būti taikomas tam, kad užfiksuoti savaitės, mėnesio, metų ar kitokią sezoninę priklausomybę laiko eilutėse. Bendru atveju toks multiplikatyvus sezoninis modeliavimas yra žymimas $SARIMA(p, d, q) \times (P, D, Q)_g$, čia P, D , ir Q yra sezoninė atitinkamai autoregresijos, skirtumo bei slenkamojo vidurkio eilės. Jeigu modeliuojame

laiko eilutes, kur stebėjimai gaunami kiekvieną dieną, o ne kas mėnesį ar metus, sezoninė komponentė *ARIMA* modelyje pasidaro netinkama. Tokiu atveju taikomas *ARIMA* modelis su aukštesnėmis *AR* ir *MA* eilėmis [17].

Siekiant supaprastinti skaičiavimus ir dėl to, kad vienmačių laiko eilučių elgesys skirtinguose erdvės taškuose yra gana panašus, visuose taškuose s_i vertinamas tos pačios eilės *ARIMA* modelis. Tačiau, koeficientai ϕ_k^i bei θ_l^i priklauso nuo erdvinės padėties. Taigi modelis yra panašus tačiau tuo pačiu individualus kiekviename erdvės taške.

(2.7) bei (2.12) modelių liekanos turėtų būti nepriklausomi ir vienodai pasiskirstę Gauso atsitiktiniai dydžiai. Tačiau, remiantis rezultatais, aprašytais [65], [66], [68] darbuose, kvadratinių liekanų autokoreliacinė funkcija ACF dažnai rodo, kad liekanos vis dar priklausomos laike. Šią priklausomybę galima įvertinti, modeliuojant liekanų dispersiją $\hat{\sigma}_i^2(t)$. Kaip jau minėta, liekanų dispersija gali būti aprašyta Furjė eilutės pagalba, (2.9) formule, arba empirinės dienos dispersijos pagalba (EDD). Jeigu stebėjimai atliekami kasdien, daugiau negu vienerius metus, empirinė dienos dispersija gaunama, skaičiuojant kiekvienos dienos vidutinę empirinę dispersiją ir gaunant 365 stebėjimus.

Taigi laiko eilučių $Z_i(t)$ liekanos yra normuojamos, dalijant jas iš vienos iš standartinio nuokrypio funkcijų.

2.2.1. Laiko eilučių modelių patikrinimas

Laiko eilučių modelių patikrinimui naudojami į imtį neįtraukti duomenys. Kiekviename taške generuojamos vieno žingsnio į priekį prognozės laiko eilutės duomenims, neįtrauktiems į analizę ir skaičiuojamos prognozės klaidos (PE), apibrėžiamos kaip skirtumas tarp stebėjimo ir prognozės. PE analizei taikomi klasikiniai statistiniai testai.

PE normalumui patikrinti apskaičiuojami kiekvieno taško laiko eilučių liekanų vidurkiai bei standartiniai nuokrypiai bei taikomi Kolmogorovo–Smirnovo arba Shapiro-Wilk testai.

Taip pat galima rasti prognozių rėžius (PI)

$$\hat{Z}_i(t) \pm z_{\alpha/2} \sqrt{\text{Var}(\hat{\epsilon}_i(t))}, \quad (2.13)$$

čia $\hat{Z}_i(t)$ yra laiko eilutės $Z_i(t)$ prognozė, $z_{\alpha/2}$ yra $\alpha/2$ lygmens standartinio normaliojo skirstinio kvantilio reikšmė; i – taško indeksas; $\text{Var}(\hat{\epsilon}_i(t))$ – liekanų dispersija.

2.3. Erdvinis laiko eilučių parametų interpoliavimas

2.3.1. Semivariogramos modeliavimas.

Siekiant sudaryti erdvės-laiko modelį naujame, nestebėtame erdvės taške s_0 reikia interpoliuoti įvertintus vienmačių laiko eilučių parametrus. Erdvinės sąveikos įtraukimas remiasi semivariogramos įvertiniu. Erdvinėje statistikoje stacionaraus izotropinio lauko $Z(s)$ empirinės semivariogramos įvertinys paprastai randamas pagal Matheron formulę:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{(s_i, s_j) \in N(\mathbf{h})} (Z_i(t) - Z_j(t))^2, \quad (2.14)$$

kur $N(\mathbf{h}) = \{(s_i, s_j): s_i - s_j \in T(\mathbf{h})\}$, o $T(\mathbf{h})$ yra vadinamoji tolerancijos sritis.

Remiantis įvertinta empirine semivariograma yra parenkama geriausiai tinkanti parametrinė semivariograma $\gamma(\mathbf{h}; \boldsymbol{\vartheta})$. Parametrinės semivariogramos vertinimui gali būti taikomi trys gerai žinomi netiesinio mažiausių kvadratų metodo įverčiai: paprastasis, apibendrintasis bei svertinis.

Paprastasis (OLS) įvertis, kuriame nežinomų parametų rinkinys $\boldsymbol{\vartheta}$ vertinamas minimizuojant tokią išraišką:

$$\{\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}; \boldsymbol{\vartheta})\}'\{\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}; \boldsymbol{\vartheta})\}.$$

Apibendrintasis (GLS) įvertis, kuriame nežinomų parametru rinkinys $\boldsymbol{\vartheta}$ vertinamas minimizuojant tokią išraišką:

$$\{\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}; \boldsymbol{\vartheta})\}'\mathbf{V}(\boldsymbol{\vartheta})^{-1}\{\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}; \boldsymbol{\vartheta})\},$$

čia $\mathbf{V}(\boldsymbol{\vartheta}_s)$ yra $\hat{\gamma}(\mathbf{h})$ kovariacijų matrica ir, dėl netiesinio uždavinio, priklauso nuo $\boldsymbol{\vartheta}$.

Svertinis (WLS) įvertis

$$\{\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}; \boldsymbol{\vartheta})\}'\boldsymbol{\Sigma}(\boldsymbol{\vartheta})^{-1}\{\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}; \boldsymbol{\vartheta})\},$$

čia $\boldsymbol{\Sigma}(\boldsymbol{\vartheta})$ yra $\hat{\gamma}(\mathbf{h})$ diagonalinė dispersijų matrica.

Efektyviausias yra GLS įvertis, tačiau WLS bei GLS įverčiai reikalauja matricių $\mathbf{V}(\boldsymbol{\vartheta})$ bei $\boldsymbol{\Sigma}(\boldsymbol{\vartheta})$ specifikacijų.

Cressie [22] pasiūlė WLS kriterijaus aproksimaciją: jeigu $\hat{\gamma}(\mathbf{h})$ yra įvertinta baigtinei aibei $\{\mathbf{h}\}$, nežinomų parametru rinkinys $\boldsymbol{\vartheta}$ vertinamas minimizuojant tokią išraišką:

$$\sum_s |N(\mathbf{h})| \left\{ \frac{\hat{\gamma}(\mathbf{h})}{\gamma(\mathbf{h}; \boldsymbol{\vartheta})} - 1 \right\}^2.$$

Ši išraiška taip pat yra WLS, remiantis aproksimacija

$$\text{var}\{\hat{\gamma}(\mathbf{h})\} \approx \frac{2\gamma^2(\mathbf{h})}{|N(\mathbf{h})|}.$$

Disertaciniame darbe empirinės semivariogramos parametrai vertinami remiantis WLS kriterijaus aproksimacija.

Remiantis Lahiri S. N., Lee Y., Cressie N. [22], [47] rezultatais, esant išpildytoms gana bendroms reguliarumo sąlygoms, WLS įvertis $\hat{\boldsymbol{\vartheta}}$ yra suderintas, asimptotiškai normalus ir efektyvus.

Populiariausi parametriniai semivariogramų modeliai yra Matern tipo modeliai [25]. Tai yra eksponentinė semivariograma:

$$\gamma(\mathbf{h}; \boldsymbol{\vartheta}) = \begin{cases} 0, & \text{kai } \mathbf{h} = 0 \\ \vartheta_0 + \vartheta_1 \left(1 - \exp\left(-\frac{1}{\vartheta_2} \sqrt{\alpha^2 h_1^2 + h_2^2}\right) \right), & \text{kai } \mathbf{h} > 0 \end{cases} \quad (2.15)$$

Ornstein-Uhlenbeck'o semivariograma:

$$\gamma(\mathbf{h}; \boldsymbol{\vartheta}) = \begin{cases} 0, & \text{kai } \mathbf{h} = 0 \\ \vartheta_0 + \vartheta_1 \left(1 - \exp\left(-\frac{1}{\vartheta_2} (\alpha^2 h_1^2 + h_2^2)\right) \right), & \text{kai } \mathbf{h} > 0 \end{cases} \quad (2.16)$$

bei sferinė semivariograma

$$\gamma(|\mathbf{h}|; \boldsymbol{\vartheta}) = \begin{cases} 0, & \text{kai } h = 0 \\ \vartheta_0 + \vartheta_1 \left(\frac{3}{2} \frac{|\mathbf{h}|}{\vartheta_2} - \frac{1}{2} \left(\frac{|\mathbf{h}|}{\vartheta_2} \right)^3 \right), & \text{kai } 0 < |\mathbf{h}| < \vartheta_2. \\ \vartheta_0 + \vartheta_1, & \text{kai } |\mathbf{h}| > \vartheta_2 \end{cases} \quad (2.17)$$

Dydis α yra anizotropiškumo parametras: kai $\alpha = 1$, eksponentinė semivariograma tampa izotropine. Eksponentinė semivariograma yra tinkama visoms l reikšmėms, tačiau jos plotis yra begalinis. Dėl pastarosios savybės jos reikšmės sparčiai didėja augant \mathbf{h} . Be abejo, tai priklauso nuo parametro ϑ_2 . Kai $\alpha = 1$, Ornstein-Uhlenbeck'o semivariograma tampa izotropine Gauso semivariograma. Neneigiamiems parametrams $\vartheta_0, \vartheta_1, \vartheta_2$ sferinė semivariograma tinka semivariogramų modelių nagrinėjimui erdvėje \mathbb{R}^l , kai $l = 1, 2$ arba 3 , tačiau, kai $l > 3$ ji nėra neteigiamai apibrėžta.

2.3.2. Empirinės semivariogramos įverčio tarp dviejų erdvės taškų savybės

Paprastai semivariograma vertinama požymių stebiniams erdvėje, kiekviename taške turint tik vieną stebėjimą. Tokių įverčių savybės yra gerai žinomos, plačiai išnagrinėtos ir aprašytos daugelio autorių. Šiame darbe empirinė semivariograma vertinama ne požymių stebiniams, o normuotoms vienmačių laiko eilučių liekanoms, kurių kiekviename taške yra T , ir kurios, adekvataus modelio atveju, turi būti nepriklausomos ir vienodai

pasiskirsčiusios pagal normalųjį dėsnį, t. y. $\varepsilon_i(t) \sim N(0,1)$, čia $\hat{\varepsilon}_i(t)$ – normuotos laiko eilučių modelių liekanos. (2.7) modelio atveju $\hat{\varepsilon}_i(t) = \frac{\hat{Z}_i(t) - \hat{L}_i(t) - \hat{S}_i(t) - \sum_{k=1}^p \hat{\phi}_k^i(Y_i(t-k)) - \sum_{l=1}^q \hat{\theta}_l^i(\varepsilon_i(t-l))}{\hat{\sigma}_i(t)}$. Analogiškai liekanos yra apibrėžiamos ir (2.12) modeliui.

Matheron įvertinys erdvės taškuose i ir j apskaičiuojamas tokiu būdu:

$$\hat{\gamma}_{ij} = \sum_{t=1}^T (\hat{\varepsilon}_i(t) - \hat{\varepsilon}_j(t))^2 / (2T),$$

čia $T(\mathbf{h}) = 1$, nes įvertinys skaičiuojamas kiekvienai taškų i ir j porai, kurių yra T .

Remiantis prielaida, kad $cov(\hat{\varepsilon}_i(t), \hat{\varepsilon}_j(t')) = 0$, kai $t \neq t'$,

$cov(\hat{\varepsilon}_i(t), \hat{\varepsilon}_j(t)) = r_{ij}(\theta)$, kai $i \neq j$, $t = 1, \dots, T$, $i, j = 1, \dots, n$, bei esant išpildytoms bendroms reguliarumo sąlygoms, $\hat{\gamma}_{ij}$ yra nepaslinktas, pagrįstas bei asimptotiškai normalus įvertis.

2.3.3. Interpoliavimas, pagrįstas erdviniais svoriais (ES)

Kiekviename taške įvertinus laiko eilučių modelį (2.7), gauname parametrų rinkinius $\hat{\eta}^i = \{\hat{a}_m^i, \hat{b}_r^i, \hat{\phi}_k^i, \hat{\theta}_l^i : m = 0, 1, 2, \dots; r = 0, 1, 2, \dots; k = 0, 1, 2, \dots; l = 0, 1, 2, \dots\}$, kur \hat{a}_m^i yra polinominio trendo parametrai; \hat{b}_r^i - (2.12) lygtimi aprašyto modelio parametrai, o $\hat{\phi}_k^i, \hat{\theta}_l^i$ – autoregresijos ir slenkamojo vidurkio parametrai aprašyti (2.7) lygtyje. Analogiškai, įvertinus (2.12) modelį, gaunamas parametrų rinkinys $\hat{\eta}^{*i} = \{\hat{\phi}_k^{*i}, \hat{\theta}_l^{*i} : k = 0, 1, 2, \dots; l = 0, 1, 2, \dots\}$, kur $\hat{\phi}_k^{*i}, \hat{\theta}_l^{*i}$ yra autoregresijos ir slenkamojo vidurkio parametrai aprašyti (2.12) lygtyje. Siekiant sudaryti erdvės-laiko duomenų $Z(\mathbf{s}; t)$ modelį, kiekvienam parametrai reikia sudaryti erdvinį modelį. Tam taikoma keletas erdvinio interpoliavimo metodų, vienas iš jų – *erdvinių svorių* (ES) metodas.

Įvertinus empirinį parametrinį semivariogramos modelį, galima apskaičiuoti erdvinius svorius tarp dviejų erdvės taškų. *Erdviniai svoriai* tarp fiksuoto erdvės taško s_0 , bei bet kurio kito erdvės taško s_j , apskaičiuojami formule

$$\hat{\delta}_{0j} = \frac{1/\gamma(h_{0j}; \hat{\boldsymbol{\vartheta}})}{\sum_{i=1}^n 1/\gamma(h_{0i}; \hat{\boldsymbol{\vartheta}})}, \quad (2.18)$$

čia $\gamma(h_{ij}; \hat{\boldsymbol{\vartheta}})$ yra parametrinės semivariogramos įvertis tarp i -ojo ir j -ojo taškų; $h_{ij} = s_i - s_j$.

Erdvinių svorių formulėje (2.18) atvirkštinis dydis imamas dėl to, kad semivariograma yra atstumo matas, o atvirkštinis dydis leidžia daryti prielaidą, kad stipresnė priklausomybė bus tarp artimesniuose taškuose esančių parametų.

Kadangi $\hat{\boldsymbol{\vartheta}}$ yra suderintas, asimptotiškai normalus ir efektyvus, remiantis (2.18) išraiškai galima parodyti, kad erdvinių svorių įverčiai, kaip tiesinės suderintų atsitiktinių dydžių funkcijos, taip pat yra suderinti, t.y.

$$\hat{\delta}_{0j} \xrightarrow{p} \delta_{0j}. \quad (2.19)$$

Erdvinių svorių įverčių matrica erdviniams taškams $s_i, i = 1, \dots, n$ gali būti užrašyta tokiu būdu:

$$\hat{\Delta} = \begin{pmatrix} 0 & \cdots & \hat{\delta}_{1j} & \cdots & \hat{\delta}_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{\delta}_{i1} & \vdots & 0 & \vdots & \hat{\delta}_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{\delta}_{n1} & \cdots & \hat{\delta}_{nj} & \cdots & 0 \end{pmatrix}.$$

Parametrų rinkinys naujam taškui s_0 apskaičiuojamas taip:

$$\hat{\boldsymbol{\eta}}_{0(ES)} = \hat{\boldsymbol{\eta}} \cdot \hat{\boldsymbol{\delta}}_0', \quad (2.20)$$

čia $\hat{\eta} = (\hat{\eta}^1, \dots, \hat{\eta}^n)$, kur $\hat{\eta}^i$ yra (2.7) bei (2.12) lygtyse aprašytų įvertintų parametru vektorius i -ajame taške arba $\hat{\eta} = (\hat{\eta}^{*1}, \dots, \hat{\eta}^{*n})$, kur $\hat{\eta}^{*i}$ yra (2.12) lygtyje aprašytų įvertintų parametru vektorius, o $\hat{\delta}_0 = (\hat{\delta}_{10}, \dots, \hat{\delta}_{n0})$ – svorių vektorius.

Iš (2.7) formulės gauname laiko eilutės modelį taške s_0

$$\hat{Z}_0(t) = \hat{L}_0(t) + \hat{S}_0(t) + \sum_{k=1}^p \hat{\phi}_k^0(Y_0(t-k)) + \sum_{l=1}^q \hat{\theta}_l^0(\epsilon_0(t-l)) + \epsilon_i(t).$$

Erdvinio sujungimo metodą pritaikius (2.12) formule užrašytam vienmatės laiko eilutės *ARIMA* procesui, gauname modifikuotą *STARIMA* modelį, aprašytą 1.4.2. skyrelyje.

(2.20) formule apskaičiuotus parametru įverčius, *ARIMA* modelis taške s_0 gali būti užrašytas taip:

$$\hat{Z}_0(t) = \sum_{k=1}^p \hat{\phi}_{0k}^* Z_0(t-k) + \epsilon_0(t) + \sum_{l=1}^q \hat{\theta}_{0l}^* \epsilon_0(t-l). \quad (2.21)$$

(2.21) lygtis tam tikra prasme yra modifikuotas *STARMA* modelis. Šiame modelyje erdvinė priklausomybė išreikšta per parametru $\hat{\phi}_{0k}^*$ bei $\hat{\theta}_{0k}^*$ įverčius, kurie apskaičiuoti remiantis erdvine priklausomybe tarp stebėjimų, tačiau, priešingai nei *STARMA* modelyje, neatsižvelgiant į kryžmines erdvės-laiko kovariacijas.

2.3.4. Interpoliavimas, pagrįstas „formaliu“ ordinariuoju krigingu (OK)

Krigingas yra gerai žinomas erdvinėje statistikoje prognozavimo (interpoliavimo) metodas, užtikrinantis geriausią tiesinę nepaslinktą prognozę (BLUP). Šis metodas taip pat glaudžiai susijęs su semivariograma, kuri naudojama tam, kad imtyje esantiems stebėjimams būtų priskirti svoriai, priklausantys nuo taškų, kuriuose jie buvo atlikti bei taško, kuriame reikšmė prognozuojama vietos erdvėje [32]. Erdvės-laiko krigingas buvo aprašytas 1.2 skyriuje, tačiau kaip jau minėta, skaičiavimai reikalauja labai daug resursų bei

atskiriųjų kovariacinių funkcijų (ar semivariogramų) naudojimo. Šiame skyrelyje aprašomas metodas remiasi „formaliu“ ordinariuoju kriginu, nes tiek semivariograma vertinama ne stebiniams, o normuotoms laiko eilučių liekanoms, tiek pats kriginas taikomas laiko eilučių parametru, o ne pradinių stebinių interpoliavimui.

Ordinariojo kriginio įverčiai gali būti apskaičiuoti taip:

$$\hat{\eta}_{0(OK)} = \hat{\eta} \cdot \hat{\lambda}'_0. \quad (2.22)$$

Svoriai $\hat{\lambda}'_0$ apibrėžti Le ir Zidek [50]:

$$\hat{\lambda}'_0 = \frac{1'_n \Gamma^{-1}}{\Gamma} + \gamma'_0 \Gamma^{-1} \left(I - \frac{J \cdot 1'_n \Gamma^{-1}}{\Gamma} \right), \quad (2.23)$$

čia $\Gamma = 1'_n \Gamma^{-1} 1_n$, $J = 1_n 1'_n$; $\gamma_0 = \gamma(h_{0j}; \hat{\vartheta})$, o Γ – semivariogramų matrica.

2.3.5. Interpoliavimas, pagrįstas trendo paviršiaus modeliu (TPM)

Aukščiau aprašyti interpoliavimo metodai atsižvelgia į erdvines koreliacijas tarp taškų. Tačiau yra žinoma ir nemažai interpoliavimo metodų, kurie erdvines koreliacijas ignoruoja ir interpoliuoja tik atsižvelgdami į taškų geometrinį išsidėstymą. Tikslėsius rezultatus turėtų duoti metodai, įtraukiantys erdvinę koreliaciją, tačiau tokiu atveju vertinamų parametru skaičius, o taip pat ir vertinimo paklaidos auga.

Vienas iš tokių prognozavimo (interpoliavimo) metodų yra mažiausių kvadratų metodu įvertintas trendo paviršiaus modelis (TPM). Tai yra modelis, priklausantis nuo stebėjimo taško koordinatų. Taškui $s_i = (x_i, y_i) \in D$, $i = 1, 2, \dots, n$, šis modelis užrašomas taip:

$$p^{\hat{\beta}_i} = \sum_{\substack{f+u \leq d \\ 0 \leq f, u \leq d}} \lambda_{fu}^{\hat{\beta}_i} x_i^f y_i^u, \quad (2.24)$$

čia $\hat{\beta}_i$ – atitinkama parametru vektorių $\hat{\eta}^i$ ir $\hat{\eta}^{*i}$ komponentė taške s_i ; $\lambda_{fu}^{\hat{\beta}_i}$ yra siūlomo modelio parametras, d – modelio (daugianario), arba kitaip – trendo

paviršiaus eilė, kai $d = 0$, paviršius atitinka pastovaus vidurkio modelį, kai $d = 1$ yra vadinamas tiesiniu, arba pirmos eilės modeliu, kuris generuoja plokštumą, kvadratinis, arba antros eilės, atitinka $d = 2$, kubinis, arba trečios eilės - $k = 3$ ir t.t. Taigi aukštesnė eilė generuoja sudėtingesnį paviršių [42].

Taikant trendo paviršiaus modelį n stebėjimų atvejui, parametru vektorių galima išreikšti lygybe:

$$\mathbf{P}_n = \mathbf{A}\boldsymbol{\lambda},$$

kur \mathbf{A} yra taškų koordinatų atitinkamų sandaugų matrica

$$\mathbf{A} = \begin{pmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1 y_1 & \dots & x_1^t y_1^t & \dots \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & x_2 y_2 & \dots & x_2^t y_2^t & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & y_n & x_n^2 & y_n^2 & x_n y_n & \dots & x_n^t y_n^t & \dots \end{pmatrix},$$

o $\boldsymbol{\lambda} = (\lambda_{00}, \lambda_{10}, \lambda_{01}, \lambda_{20}, \lambda_{02}, \lambda_{11}, \dots, \lambda_{fu})^T$ – trendo paviršiaus parametru vektorius.

2.3.6. Interpoliavimas, pagrįstas atvirkštinių atstumų svoriais (IDW)

Erdvinių svorių modeliavimas yra atvirkštinių atstumų svorių (Inverse distance weighting) analogas. Tai yra vienas iš paprasčiausių interpoliavimo metodų kai apie kiekvieną tašką, kuriame siekiama interpoliuoti yra nustatoma tam tikra „kaimyninė“ riba ir skaičiuojamas visų kaimyninėje aplinkoje stebėtų reikšmių svertinis vidurkis. Svoriai yra mažėjanti atstumo funkcija. Paprasčiausia atvirkštinio atstumo svertinės interpoliacijos forma dar vadinama „Shepardo metodu“. Šiuo atveju taške s_0 , parametru įverčiai yra apskaičiuojami remiantis (2.20) formule, kurioje svoriai apskaičiuojami formule:

$$\hat{w}_{0j} = \frac{h_{0j}^{-p}}{\sum_{i=1}^n h_{0i}^{-p}}, \quad (2.25)$$

čia $p > 0, p \in \mathbb{R}$; h_{0i} yra atstumas nuo kiekvieno taško iki interpoliuojamo taško, t.y. $h_{0i} = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2}$. Sviurių funkcijos reikšmės svyruoja nuo vieneto iki nulio bei yra normalizuojama taip, kad sviurių suma būtų lygi vienam.

2.4. Kryžminio patikrinimo metodas

Aprašyta siūlomų erdvinių laiko eilučių modeliavimo metodika leidžia sudaryti keletą skirtingų modelių naujoje stotyje s_0 . Kadangi modeliavimo tikslas yra sudaryti modelį, leidžiantį prognozuoti naujoje stotyje s_0 ir naujais laiko momentais $T + k$, todėl reikia nustatyti, kuris iš siūlomų modelių geriausiai aprašo analizuojamas erdvines laiko eilutes.

Kryžminio patikrinimo metodas, dar vadinamas rotaciniu įvertinimu, yra technika, leidžianti nustatyti kaip statistinės analizės rezultatai gali būti apibendrinti nepriklausomai duomenų aibei. Dažniausiai šis metodas taikomas tada, kai pagrindinis tikslas yra prognozė ir siekiama nustatyti kaip gerai siūlomas modelis aprašo analizuojamus duomenis [1].

Dažniausiai yra taikomi trys kryžminio patikrinimo tipai: „ k – imčių“ (k - fold) metodas; pakartotinis atsitiktinis sub-imčių patikrinimas bei „palikti vieną“ (leave-one-out) metodas. „ k -imčių“ metodas dalija pradinę imtį į k sub-imčių iš kurių viena yra paliekama patikrinimui, o visos kitos laikomos mokymo imtimis. Pritaikius metodą k kartų, ieškomasis įvertis gaunamas suskaičiavus gautų rezultatų vidurkj. „Palikti vieną“ metodas yra atskiras „ k - imčių“ metodo atvejis, kai $k = 1$. Pakartotinis atsitiktinis sub-imčių patikrinimas atsitiktinai padalija pradinę imtį į mokymo ir patikrinimo imtis, metodą pakartojus keletą kartų, skaičiuojamas gautų rezultatų vidurkis. Pastarojo metodo pranašumas prieš „ k -imčių“ metodą yra tas, kad šiuo atveju mokymo/tikrinimo imčių padalijimas nepriklauso nuo iteracijų numerio. Paprastai, pritaikius kryžminio patikrinimo metodą, skaičiuojamos

vidutinė kvadratinė klaida (MSE) arba šaknis iš vidutinės kvadratinės klaidos (RMSE).

Šiame skyriuje aprašyti erdvinių laiko eilučių modeliavimo bei interpoliavimo metodai pritaikyti vėjų greičio duomenims, rinktiems 20-yje Lietuvos Respublikos miestų nuo 1977 m. sausio 1 dienos iki 2007 metų gruodžio 31 d. Modeliavimo bei kryžminio patikrinimo rezultatai pateikti 3 skyriuje.

3. ERDVINIŲ LAIKO EILUČIŲ MODELIS VĖJO GREIČIO DUOMENIMS

Vėjo energija yra vienas greičiausiai pasaulyje augančių energijos šaltinių per pastaruosius 10 metų. Dažnai vėjo jėgainių parkai yra statomi ten, kur atlikti vėjo greičio matavimus yra neįmanoma. Viena iš išeičių yra apskaičiuoti prognozuojamą vėjo greitį, remiantis rezultatais surinktais aplink esančiose hidrometeorologijos stotyse. Paprastai hidrometeorologinės tarnybos vėjo greičius bei kitus hidrometeorologinius parametrus fiksuoja nuo vieno iki kelių kartų per dieną. Taip sudaromos gana ilgos laiko eilutės, priklausomos erdvėje. Tai viena iš pagrindinių priežasčių kodėl tinkamo erdvių laiko eilučių modelio sudarymas yra pakankamai aktuali tema.

Gana daug autorių nagrinėjo laiko eilutes, aprašančias vėjo greitį. Klasikiniai statistiniai modeliai, aprašantys vėjo greitį, remiasi Weibull'io, Rayleigh'o ar lognormaliuoju pasiskirstymais ([36], [53], [15]). Šie modeliai remiasi prielaida, kad vėjo greičio duomenys yra nepriklausomi, kas praktikoje negalioja. Kitas, tinkamesnis, būdas aprašyti vėjo greičio dinamiką laike yra autoregresijos (AR) arba autoregresijos slenkamojo vidurkio (ARMA) laiko eilučių modeliai ([3], [10], [11], [12], [14], [15], [62], [14], [55], [56], [67], [70], [72]). Bouette ir kt. [8] modeliavo vienmates laiko eilutes, remiantis apibendrintu ARMA procesu.

Sahin ir Sen [66] pasiūlė trigonometrines taškinės kumuliatyvinės semivariogramos koncepciją, pagal kurią galima parinkti erdvių priklausomybių funkcijas ir jas taikyti prognozėms. Vienas iš pirmųjų darbų, aprašančių tam tikro regiono vėjo greičio erdvės-laiko struktūrą, buvo Haslett ir Raftery [40] straipsnis apie Airijos vėjo greitį. Autorių sudarytas modelis rėmėsi tokiomis klasikinės statistikos idėjomis: sezoniškumo iš laiko eilučių

pašalinimas, kringas, trupmeniškai diferencijuotas ARMA modelis. Yan ir kt. [30] taiko apibendrintą tiesinį modelį vėjo greičio Šiaurės vakarų Europoje modeliavimui. Cripps ir kt. [26] remiasi Bayes'o hierarchiniu modeliu, prognozuodamas vėjo lauko paviršių vienai dienai į priekį Sidnėjaus uoste. Alliot ir kt. [3] siūlo Markovo autoregresinį modelį erdvės-laiko vėjo greičio evoliucijos modeliavimui. Panašūs metodai taikomi ir kitų meteorologinių erdvės-laiko parametrų (pavyzdžiui, kritulių kiekio, oro užterštumo) modeliavimui. Erdvės-laiko meteorologinių parametrų modeliavimo uždaviniai nagrinėti [49], [50], [51], [59], [62], [65], [68], [73] [74], [78], [79] darbuose.

Siūlomas modeliavimo algoritmas pateiktas 1 priede.

3.1. Žvalgomoji duomenų analizė

Analizuojami duomenys buvo gauti iš Lietuvos hidrometeorologijos tarnybos. Pateikta informacija apie vidutinius dienos vėjo greičius, matuojamus metrais per sekundę (m/s) 20 Lietuvos miestų (1 pav.). Duomenys Lietuvoje rinkti nuo 1976 m.



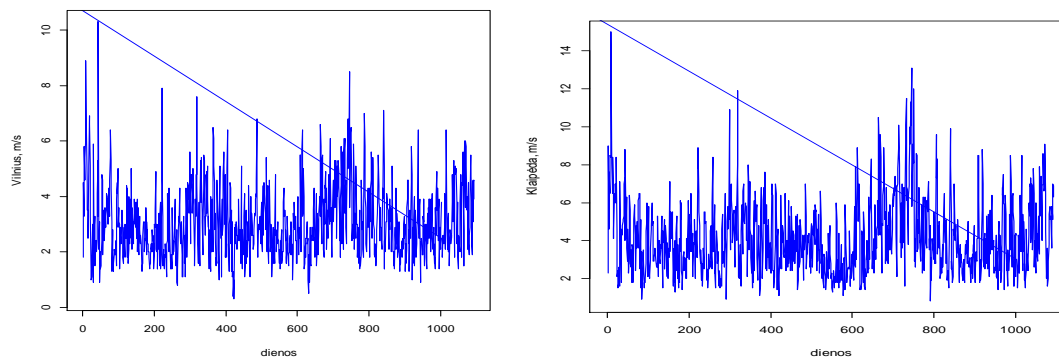
1 pav. Lietuvos miestai, kuriuose buvo renkama informacija apie vėjo greitį.

Dalyje miestų, elektroninėse laikmenose, duomenys kaupiami nuo 1976 m. sausio 1 dienos. Siekiant, kad analizuojamos laiko eilutės būtų vieno ilgio, analizei buvo pasirinkti duomenys nuo 1977 m. sausio 1 dienos iki 2007 metų gruodžio 31 d., taip kiekviename mieste gaunant 11315 stebėjimų. Dvi stotys, Palanga ir Trakų Vokė, buvo pašalintos iš analizės, nes Palangoje duomenys buvo pradėti fiksuoti tik 1992 m., o Trakų Vokėje trūko paskutinių 92 stebėjimų, kas sudaro 0,8%, taip pat Trakų Vokė yra netoli Vilniaus, todėl ji buvo pašalinta iš tyrimo. Šios dvi stotys vėliau bus panaudotos modelio validumui (tinkamumui) patikrinti. Iš likusių 18 stočių trijose taip pat buvo keletas trūkstamų stebėjimų. Kiekvienoje stotyje trūkstamos reikšmės buvo pakeistos vidutine atitinkamos dienos temperatūra per visą analizuojamą laikotarpį. Trūkstamų duomenų kiekis (%) buvo nedidelis (žr. 1 lentelę), dėl to toks duomenų atstatymo būdas jokios reikšmingos įtakos duomenų masyvui neturėjo.

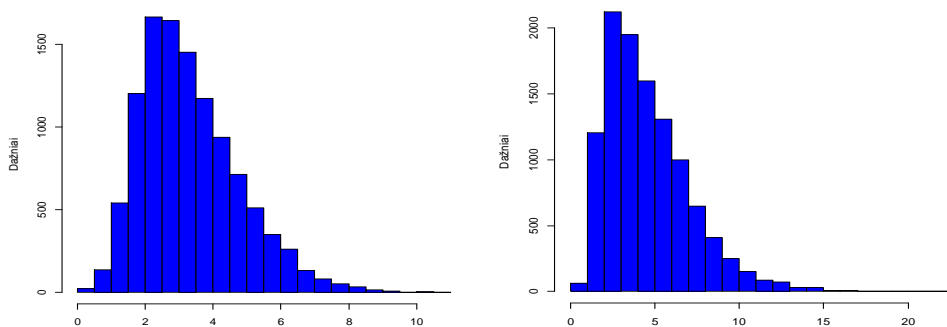
2 pav. pavaizduoti Klaipėdoje ir Vilniuje fiksuoti vėjo greičiai per paskutinius trejus metus. Visam grafiniam vaizdavimui pasirinkti šie du miestai, nes Klaipėdos duomenys atspindi vėjo elgesį Baltijos jūros pakrantėje, o Vilniaus duomenys – vėjo elgesį žemyninėje šalies dalyje. Modeliuojant laiko eilutes, buvo remiamasi prielaida, kad visose stotyse vėjo greičio elgesys yra panašus, šią prielaidą galima pagrįsti tuo, kad Lietuvos paviršius yra gana vienodas. 2 pav. matome, kad vėjo greitis yra didesnis ir labiau varijuojantis šaltuoju metų laikotarpiu. Išanalizavus visas stotis taip pat buvo pastebėta, kad daug silpnesnis sezoniškumas fiksuojamas žemyninėje dalyje esančiose stotyse.

Histogramos, pateiktos 3 paveiksle rodo, kad vėjo greičio duomenims būdinga gana ryški asimetrija. Duomenų simetrizavimui buvo pritaikyta logaritminė bei Box-Cox transformacija, tačiau paaiškėjus, kad Box-Cox transformacija neduoda ženkliai geresnių rezultatų, dėl mažesnio vertinamų parametrų kiekio, taip pat dėl galimų transformacijos parametro variacijų,

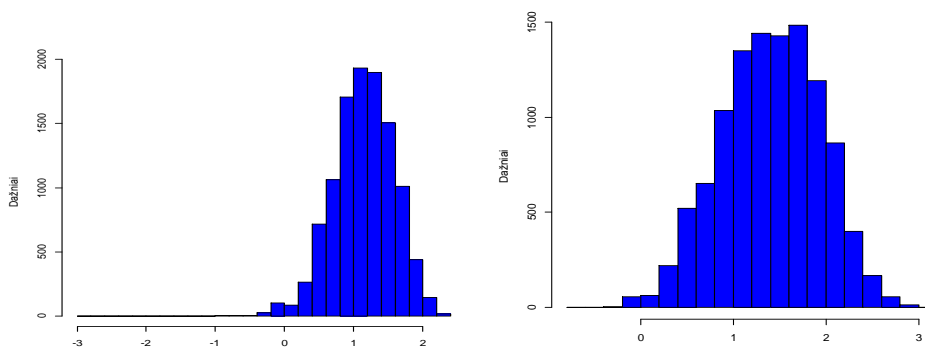
priklausomai nuo stoties, pasirinkta logaritminė transformacija. Logaritmuotų duomenų histogramos pateiktos 4 paveiksle.



2 pav. Vėjo greičio duomenys Vilniaus ir Klaipėdos miestuose

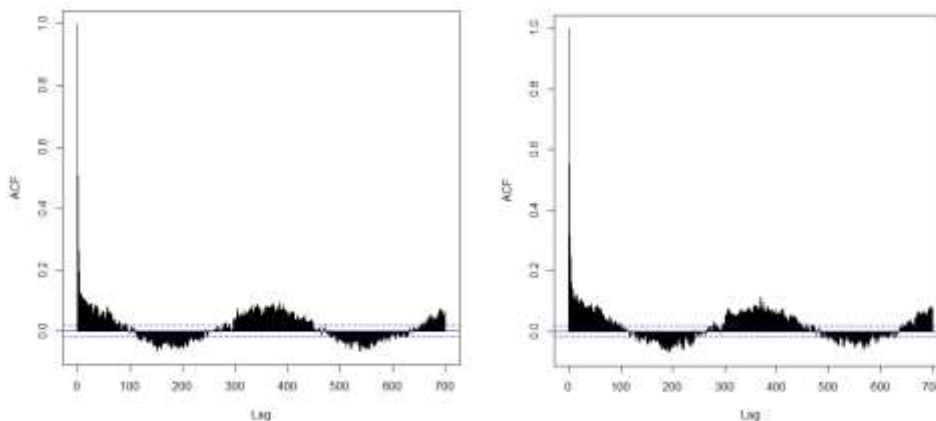


3 pav. Vėjo greičio duomenų histogramos (Vilniaus duomenys kairėje, Klaipėdos – dešinėje)



4 pav. Vėjo greičio logaritmuotų duomenų histogramos (Vilniaus duomenys kairėje, Klaipėdos – dešinėje)

Logaritmuotiems duomenims apskaičiuotos aprašomosios statistikos charakteristikos pateikiamos 1 lentelėje. Vidutinis logaritmuotas vėjo greitis svyruoja nuo 0,72 m/s iki 1,40 m/s su gana stabiliu standartiniu nuokrypiu visame analizuojamame regione. Nors po logaritminės transformacijos duomenys išsidėstę simetriškiau, asimetrijos bei eksceso koeficientai visose stotyse, išskyrus asimetrijos koeficientą Dotnuvoje, statistiškai reikšmingai skiriasi nuo 0. Logaritmuotų duomenų autokoreliacinė funkcija (5 pav.) rodo, kad analizuojamiems duomenims būdingas ryškus sezoniškumas bei reikšmingos autokoreliacijos.



5 pav. Logaritmuotų duomenų autokoreliacinės funkcijos (Vilnius – kairėje, Klaipėda – dešinėje).

Erdvinių savybių nustatymui, visoms stočių poroms buvo apskaičiuoti koreliacijos koeficientai. Visi koeficientai buvo didesni už 0,5 – tai yra natūralu analizuojant Lietuvos Respublikos skirtinguose miestuose užfiksuotus vėjo greičius, nes atstumai tarp miestų nėra dideli, ilgiausias atstumas tarp dviejų stočių yra apytiksliai 292 km. Kaip ir tikėtasi, koreliacijos priklauso nuo atstumų tarp stočių. Stebėjimai tarp trumpesnių atstumų buvo stipriau koreliuoti.

	Trūkstantos reikšmės (%)	Vidurkis	SN	Asimetrijos koef.	Eksceso koef.
Biržai	0(0)	1,14	0,45	-0,48	0,61
Dotnuva	0(0)	0,86	0,52	0,003*	-0,18
Dūkštas	0(0)	0,81	0,56	-0,20	0,15
Kaunas	0(0)	1,25	0,46	-0,43	0,32
Kybartai	1(9x10 ⁻⁴)	1,03	0,54	-0,36	0,70
Klaipėda	0(0)	1,40	0,54	-0,15	-0,36
Laukuva	61 (0,5)	1,12	0,55	-0,11	0,47
Lazdijai	0(0)	1,07	0,50	-0,48	0,61
Nida	0(0)	1,38	0,53	-0,44	0,81
Panevėžys	0(0)	1,09	0,48	-0,70	1,21
Raseiniai	0(0)	1,25	0,41	-0,43	1,09
Šiauliai	0(0)	0,94	0,48	-0,99	2,74
Šilutė	0(0)	1,22	0,50	-0,42	1,03
Telšiai	0(0)	1,04	0,42	-0,49	1,16
Ukmergė	0(0)	1,16	0,54	-0,65	1,56
Utena	31 (0,3)	0,81	0,55	-0,78	1,52
Varėna	0(0)	0,72	0,56	-0,83	1,54
Vilnius	0(0)	1,13	0,44	-0,43	0,81

1 lentelė. Aprašomosios statistikos charakteristikos logaritmuotiems vėjo greičio duomenims (n=11315). Žvaigždute pažymėti parametrai, kurių reikšmės statistiškai reikšmingai nesiskiria nuo 0.

Laiko eilučių modeliai buvo įvertinti duomenims, kuriuos sudarė 10950 stebėjimų laike ir 18 stočių erdvėje.

Pirmiausia buvo vertinamas (2.7) lygtimi aprašytas modelis, kai remiantis laiko eilučių dekompozicija, pažingsniui iš duomenų buvo eliminuotos skirtingos komponentės. Pirmiausia buvo patikrinta ar duomenims būdingas tiesinis trendas ir, remiantis tiesinės regresinės analizės rezultatais nustatyta, kad analizuojamiems vėjo greičio duomenims trendas nebūdingas. Toliau kiekvienoje stotyje buvo eliminuota laiko eilučių sezoninė komponentė. Sekančiame etape, kiekvienoje meteorologinėje stotyje, pasirinkta AR modelio eilė ir įvertinti nežinomi modelio parametrai, bei ištirtos laiko eilučių liekanos. Modelio pasirinkimui buvo naudoti Akaike informacinis kriterijus (AIC), ištirtos liekanų ACF (Ljung-Box testas) bei histogramos. Liekanos pasirodė nekoreliuotos ir pasiskirsčiusios pagal normalųjį dėsnį. Tačiau kvadratinių liekanų analizė parodė, kad dar liko neeliminuo-

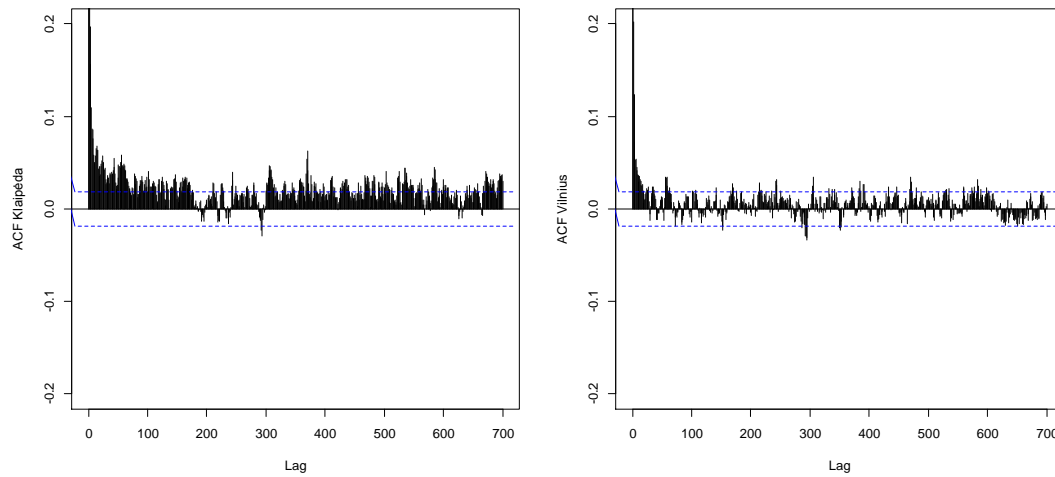
sezoniškumo, kuris buvo modeliuojamas remiantis dviem skirtingomis variacijos (dispersijos) funkcijomis. Vertinant (2.12) lygtimi aprašytą modelį, pirmiausia buvo nustatyta reikalinga skirtumo eilė, vėliau, remiantis tais pačiais kriterijais, kiekvienoje stotyje įvertinti *ARIMA* modeliai su tokiu pačiu parametru skaičiumi.

Kiekvienoje stotyje atskirai įvertinus laiko eilučių modelį, parametrai buvo sujungti į erdvinį modelį, kiekvienai parametru aibeį įvertinant antros eilės trendo paviršiaus modelį, pritaikant atvirkštinius svertinius atstumus bei įvertinant erdvinius svorius. Įvertinti modeliai buvo patikrinti remiantis duomenimis, neįtrauktais į imtį (2007 metų duomenys). Laike nepriklausomoms ir normaliai pasiskirsčiusioms liekanoms buvo įvertinta semivariograma. Erdvės laiko modelio patikrinimas buvo atliktas remiantis kryžminio patikrinimo metodu.

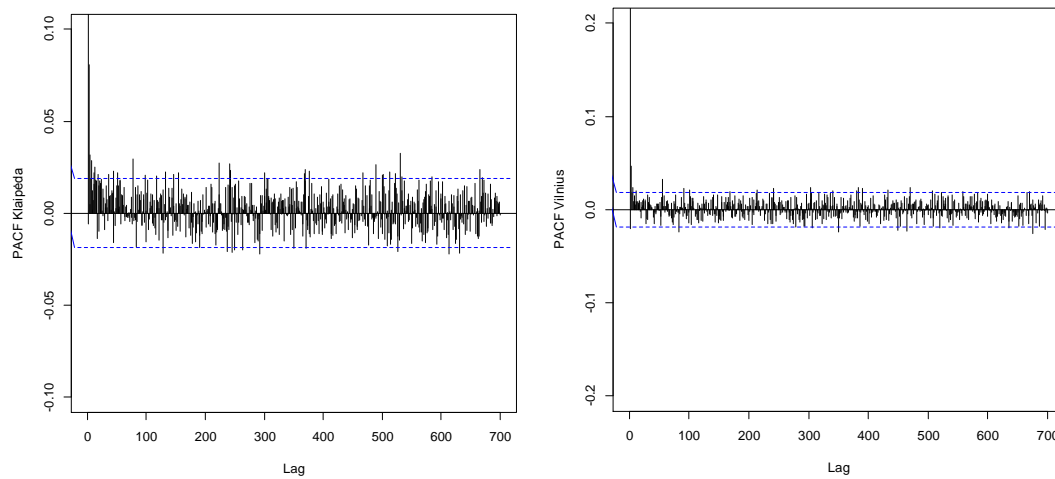
3.2. Laiko eilučių modeliavimas

Modelių įvertinimui buvo pasirinkti duomenys nuo 1977 01 01 iki 2006 12 31, kurie sudarė 10950 stebėjimų 18 stočių. Sudarant (2.7) lygtimi aprašytą modelį, pirmiausia kiekvienoje stotyje buvo įvertinta sezoninė funkcija, aprašyta (2.8) lygtimi. Sezoninių svyravimų pašalinimui pakako funkcijos su trimis parametrais. Šių parametru įverčiai yra pateikti 2 lentelės pirmuosiuose trijuose stulpeliuose. 6 ir 7 paveiksluose pavaizduotos Klaipėdos bei Vilniaus miestų ACF bei PACF funkcijų reikšmės, gautos laiko eilutėse eliminavus sezoniškumą. Nors sezoniniai svyravimai buvo pašalinti, matome, kad liekanos vis dar pasižymi stipria autokoreliacija. Remiantis PACF grafikais, šiam triukšmui įvertinti reikalingi gana aukštos eilės *AR* procesai. *AR* proceso eilės nustatymui buvo taikytas AIC kriterijus, liekanų ACF tyrimas bei liekanų histogramos. Pradėjus nuo *AR*(1) modelio, eilė buvo didinama. *AR* modelio eilė p_i buvo apribota iki 10. Daugumai stočių pilnai pakako *AR*(3) ar *AR*(4) modelio, tačiau keturiose stotyse prireikė aukštesnės eilės modelių. Įvertintos

AR proceso parametrų reikšmės pateiktos 2 lentelės 4 – 11 stulpeliuose. Taip pat buvo patikrintas ir autoregresijos slenkamojo vidurkio modelio (*ARMA*) tinkamumas, tačiau šis modelis nedavė tikslesnių rezultatų.



6 pav. Laiko eilučių liekanų, gautų eliminavus sezoninę komponentę, ACF su 95% pasikliautinių intervalų ribomis.

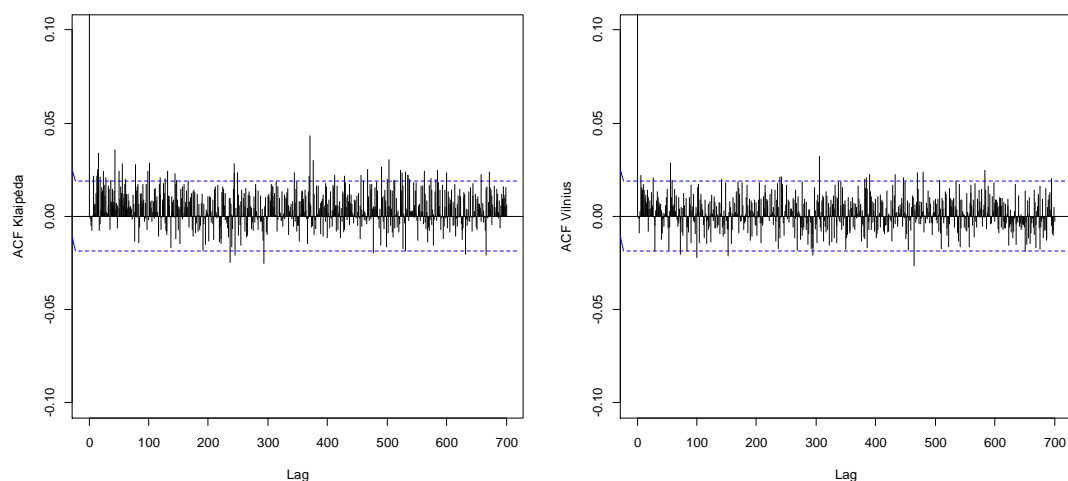


7 pav. Laiko eilučių liekanų, gautų eliminavus sezoninę komponentę, PACF su 95% pasikliautinių intervalų ribomis.

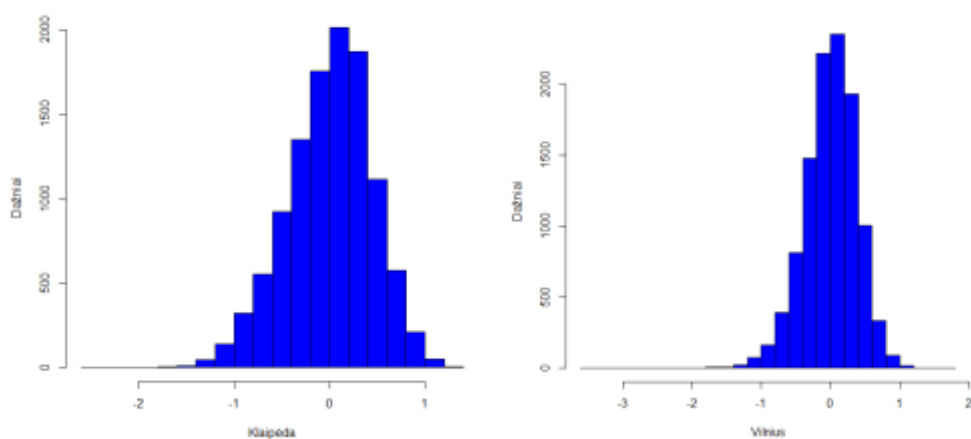
	b_0^i	b_1^i	b_2^i	ϕ_1^i	ϕ_2^i	ϕ_3^i	ϕ_4^i	ϕ_5^i	ϕ_6^i	ϕ_7^i	ϕ_8^i
Biržai	1,14	0,15	0,04	0,46	-0,05	0,05	0,02				
Dotnuva	0,86	0,14	0,08	0,49	-0,04	0,08	0,09	0,03	0,06		
Dūkštas	0,81	0,14	0,04	0,46	-0,04	0,07	0,06				
Kaunas	1,25	0,20	0,03	0,47	-0,05	0,05					
Kybartai	1,03	0,19	0,03	0,48	-0,02*	0,06					
Klaipėda	1,40	0,18	-0,06	0,52	-0,05	0,06	0,03				
Laukuva	1,12	0,17	0,03	0,48	-0,04	0,05	0,04				
Lazdijai	1,07	0,17	0,02	0,49	-0,03	0,04	0,04				
Nida	1,38	0,18	-0,05	0,47	0,02*	0,07	0,04	0,04	0,03	0,05	0,07
Panevėžys	1,09	0,14	0,05	0,48	-0,03	0,05	0,04				
Raseiniai	1,25	0,14	0,06	0,43	-0,04	0,05	0,02				
Šiauliai	0,94	0,16	0,06	0,47	-0,03	0,05	0,02				
Šilutė	1,22	0,11	0,06	0,47	-0,04	0,07					
Telšiai	1,04	0,12	0,03	0,48	-0,05	0,06					
Ukmergė	1,16	0,17	0,04	0,46	-0,04	0,07	0,02				
Utena	0,81	0,18	0,06	0,45	-0,01*	0,06	0,02	0,03	0,05		
Varėna	0,72	0,15	0,07	0,47	-0,03	0,04	0,03	-0,00*	0,05		
Vilnius	1,13	0,16	0,03	0,47	-0,03	0,06					

2 lentelė. 2.7 lygtimi aprašyto modelio įverčiai. Parametrai, su 5% reikšmingumo lygmeniu statistiškai reikšmingai nesiskiriantys nuo nulio pažymėti žvaigždute.

Determinacijos koeficiento R^2 reikšmės, po sezoninės funkcijos bei AR proceso įvertinimo svyravo nuo 23,4% Raseiniuose iki 40,5% Nidoje. Žinoma gana didelė dalis sklaidos dar lieka nepaaiškinta. Remiantis Ljung-Box testo rezultatais (8 pav.) laiko eilučių liekanos, po AR proceso įvertinimo nepasižymėjo autokoreliacija. Remiantis histogramomis, visose stotyse liekanos pasižymėjo neigiamu asimetrijos ir teigiamu eksceso koeficientu, tačiau pasiskirstymas atrodė beveik simetriškas (9 pav.). Liekanų vidurkis iš esmės buvo lygus nuliui, standartiniai nuokrypiai svyravo nuo 0,36 iki 0,48.

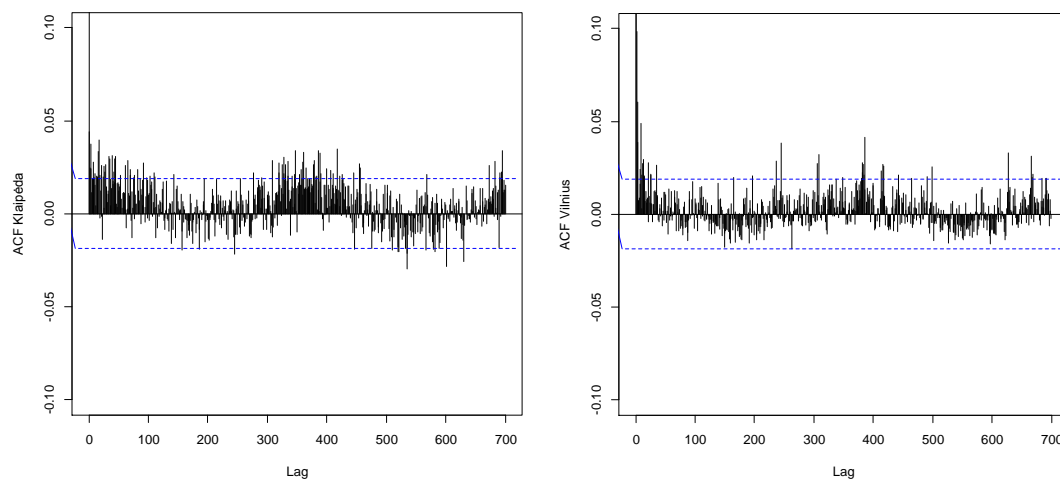


8 pav. Liekanų, gautų eliminavus sezoninę bei AR komponentes ACF su 95% pasikliautinųjų intervalų ribomis.

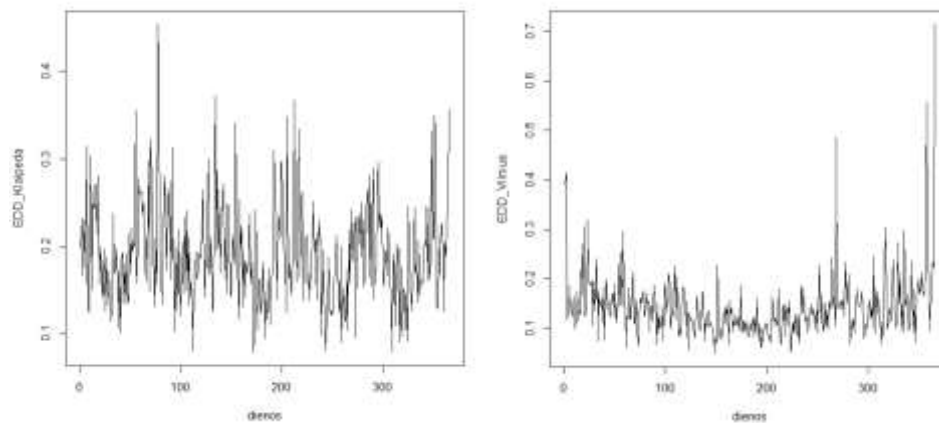


9 pav. Liekanų, gautų eliminavus sezoninę bei AR komponentes, histogramos.

Nors gautų liekanų autokoreliacijos svyravo apie nulį, tačiau ištirta kvadratinė liekanų ACF atskleidė dispersijos priklausomybę laike (10 pav.). Siekiant pašalinti šią priklausomybę, buvo apskaičiuota empirinė dienos dispersija, skaičiuojant kiekvienos dienos kvadratinė liekanų vidurkius. Platesni svyravimai pastebimos šaltuoju sezonu bei mažesni vasarą (11 pav.). Mažiau ryškus sezoninis elgesys užfiksuotas Vilniaus duomenis, būdingas ir kelioms kitoms žemyninėms stotims. Tačiau aiškūs sezoniniai svyravimai, užfiksuoti Klaipėdoje, būdingi daugeliui analizuojamų stočių.



10 pav. Kvadratinų liekanų, gautų eliminavus sezoninę bei AR komponentes ACF su 95% pasikliautinių intervalų ribomis.



11 pav. Liekanų empirinė dienos dispersija (EDD) po sezoninės komponentės bei AR proceso eliminavimo.

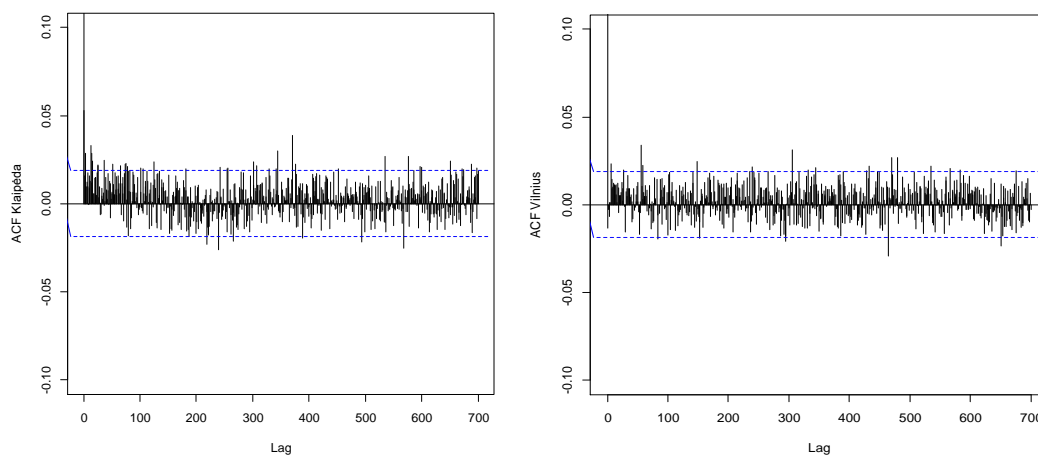
Tęsiant empirinės dispersijos modeliavimą, įvertinta (2.9) lygtimi apibrėžta funkcija su parametru $J = 1$. Sezoniškumo modeliavimui taip pat taikyta empirinė dienos dispersija. Dispersijos funkcijos $\sigma_i^2(t)$ įverčiai pateikti 3 lentelėje. Pirmas šios lentelės stulpelis taip pat rodo empirinę kiekvienos stoties dispersiją.

Liekanos buvo normuotos, dalijant jas iš vienos iš dviejų dispersijos funkcijų kvadratinės šaknies. Taip transformuotos liekanos pasirodė artimiausios baltajam triukšmui tuo atveju, kai buvo pasinaudota empirine dienos dispersija. Empirinės dienos dispersijos pagalba normuotų liekanų ACF

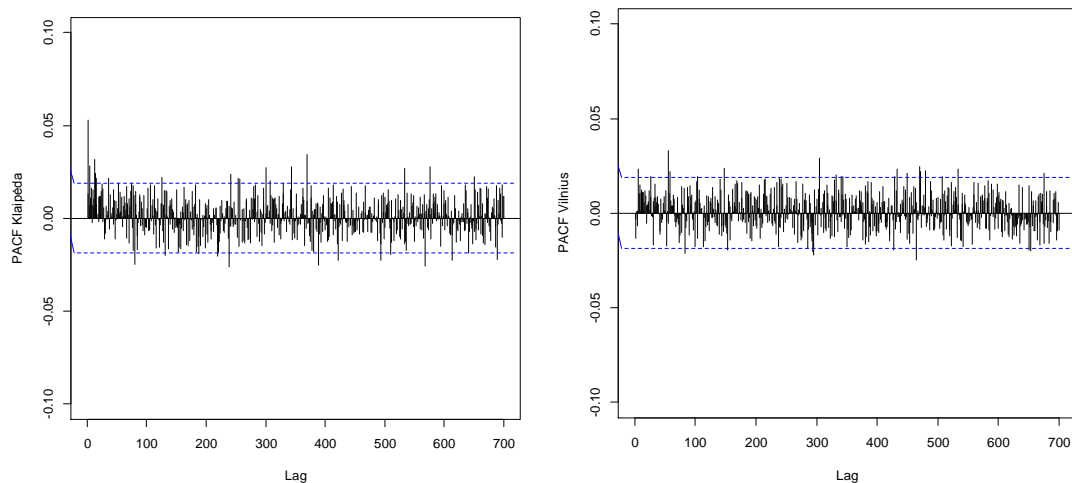
bei PACF pavaizduotos atitinkamai 12 ir 13 paveiksluose. Remiantis ACF galima daryti išvadą, kad liekanos yra nekoreliuotos ir su nuliniu vidurkiu, t. y. baltas triukšmas. Taigi, dispersijos sezoniškumas buvo eliminuotas. Vietoje empirinės dispersijos, liekanų normavimui pritaikius įvertintą dispersiją gauti identiški rezultatai, tačiau dėl mažesnio parametrų skaičiaus buvo pasirinkta empirinė dienos dispersija.

	c_1^i	c_2^i	c_3^i
Biržai	0,151	0,012	-0,004*
Dotnuva	0,177	0,030	-0,004
Dūkštas	0,228	0,037	-0,015*
Kaunas	0,148	0,011	-0,003*
Kybartai	0,213	0,048	-0,002*
Klaipėda	0,198	0,035	-0,017
Laukuva	0,219	0,064	-0,029
Lazdijai	0,182	0,022	-0,008*
Nida	0,169	0,022	-0,018
Panevėžys	0,165	0,007*	-0,013
Raseiniai	0,132	0,021	-0,001*
Šiauliai	0,166	0,014	-0,005*
Šilutė	0,189	0,052	-0,027
Telšiai	0,133	0,023	-0,005*
Ukmergė	0,221	0,014	-0,017*
Utena	0,210	0,027	-0,009*
Varėna	0,229	0,043	-0,019
Vilnius	0,143	0,030	-0,001*

3 lentelė. Funkcijos $\sigma_i^2(t)$ įverčiai. Statistiškai nereikšmingi parametrai pažymėti žvaigždute.

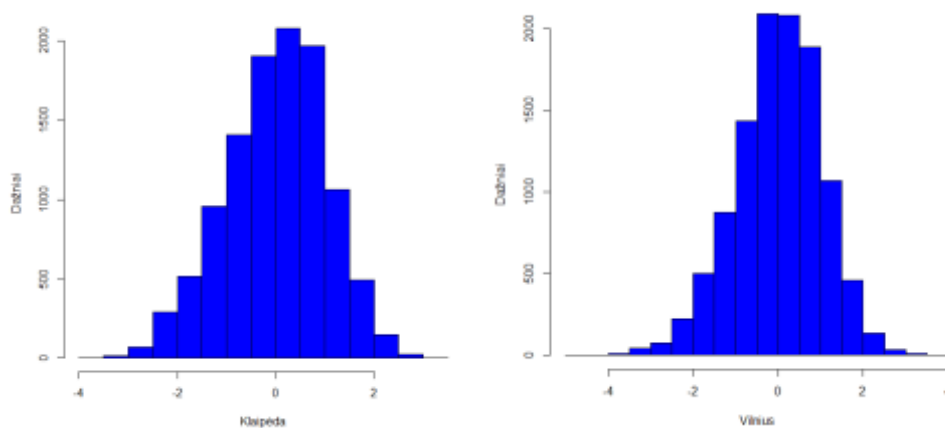


12 pav. Normuotų liekanų ACF su 95% pasikliautinųjų intervalų ribomis.



13 pav. Normuotų liekanų PACF su 95% pasikliautinųjų intervalų ribomis.

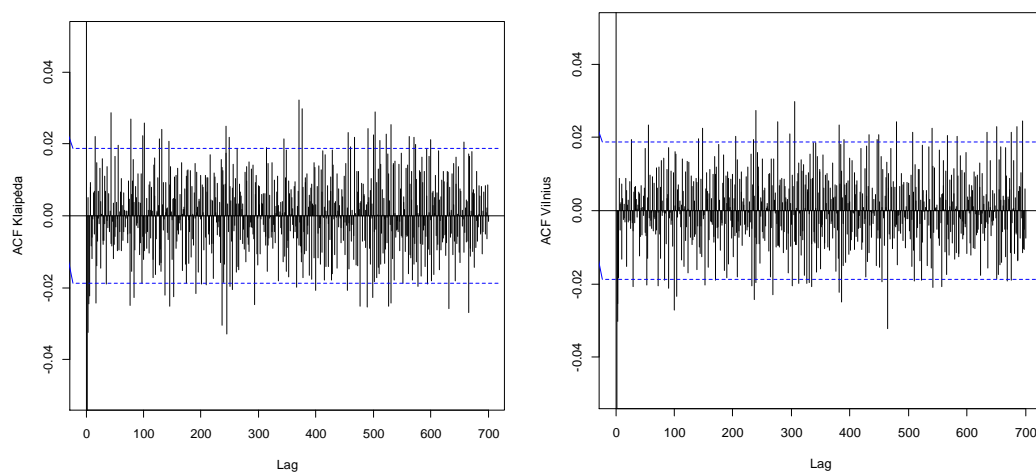
Visų trijų dispersijų atveju, normuotoms paklaidoms būdingas nežymus asimetriškumas. Teigiamas eksceso koeficientas rodo, kad pasiskirstymas smalesnis negu normalusis, tačiau pasiskirstymas yra artimas normaliajam su vidurkiu, praktiškai lygiu nuliui ir standartiniu nuokrypiu lygiu vienam (14 pav.).



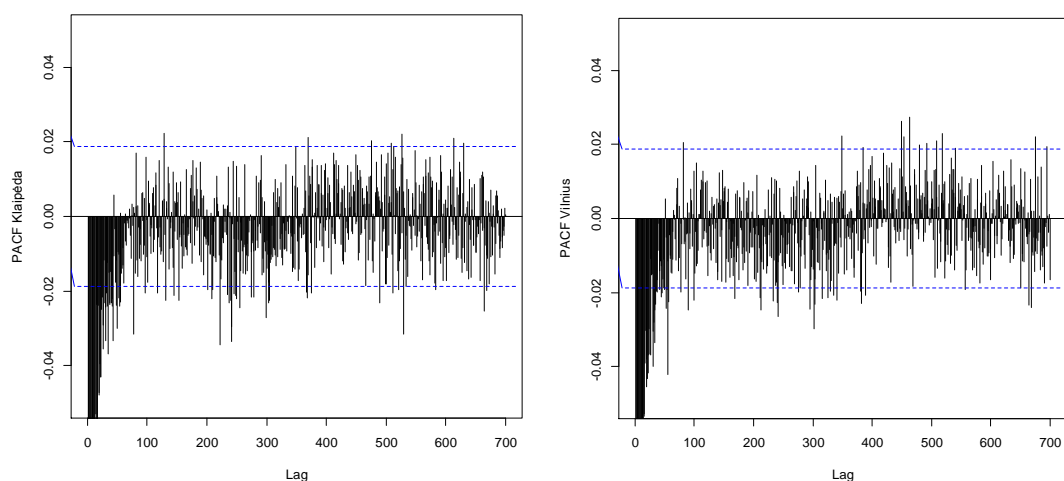
14pav. Normuotų Liekanų histogramos (Klaipėdos vėjo greičio duomenys kairėje ir Vilniaus – dešinėje).

Remiantis (2.12) formule, kiekvienoje stotyje taip pat buvo įvertinti *ARIMA* modeliai. Vertinant šiuos modelius buvo remiamasi prielaida, kad kiekvienoje stotyje vėjo greičio duomenis galima aprašyti tos pačios eilės

modeliu. Analizuojant jau logaritmuotas laiko eilutes, nustatyta, kad siekiant transformuoti į stacionarią laiko eilutę, pakanka apskaičiuoti 1 eilės skirtumus. Kadangi vėjo greitis pasižymi gana dideliu triukšmu, pagal diferencijuotų duomenų *ACF* bei *PACF* elgesį gana sudėtinga pasirinkti tinkamiausią modelio eilę (15 ir 16 pav.). Todėl, remiantis liekanų autokoreliacinės funkcijos elgesiu, Ljung-Box testo rezultatais bei AIC kriterijais buvo renkama iš keleto modelių, pradedant nuo *ARIMA(1,1,0)* ir apsiribojant *ARIMA(3,1,3)* daugelyje stočių tinkamiausias pasirodė *ARIMA(3,1,1)* modelis. Tokį pasirinkimą patvirtino ir projekto R versijos i386 2.15.2 funkcija `auto.arima()`, nustatanti geriausiai tinkančią modelio eilę. Tik 4 stotyse buvo siūlomas šiek tiek aukštesnės eilės modelis, tačiau ištyrus liekanų *ACF* bei histogramas, buvo nustatyta, kad ir šiose stotyse *ARIMA(3,1,1)* modelis pakankamai gerai aprašo laiko eilutės elgesį. 4 lentelėje pateikiami visose stotyse gauti *ARIMA(3,1,1)* parametrų įverčiai.

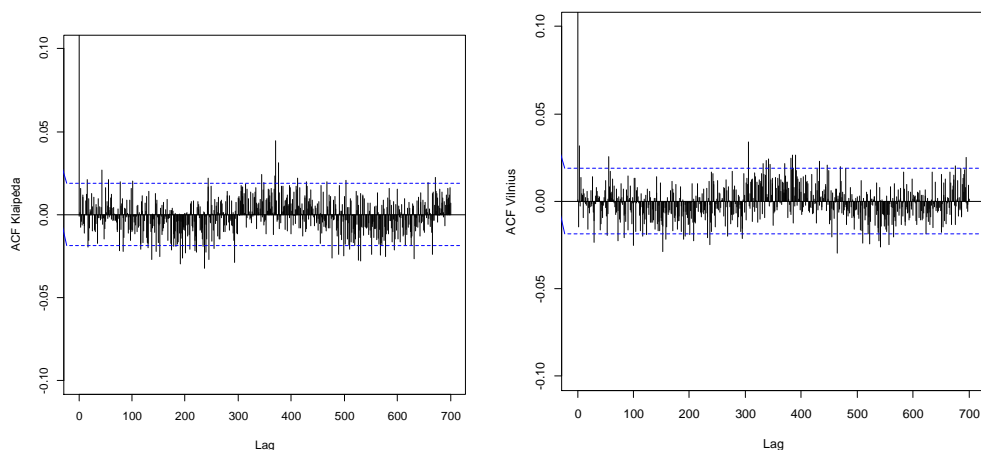


15 pav. ACF su 95% pasikliautinųjų intervalų ribomis, apskaičiuavus pirmos eilės skirtumus



16 pav. PACF su 95% pasikliautinųjų intervalų ribomis, apskaičiuavus pirmos eilės skirtumus

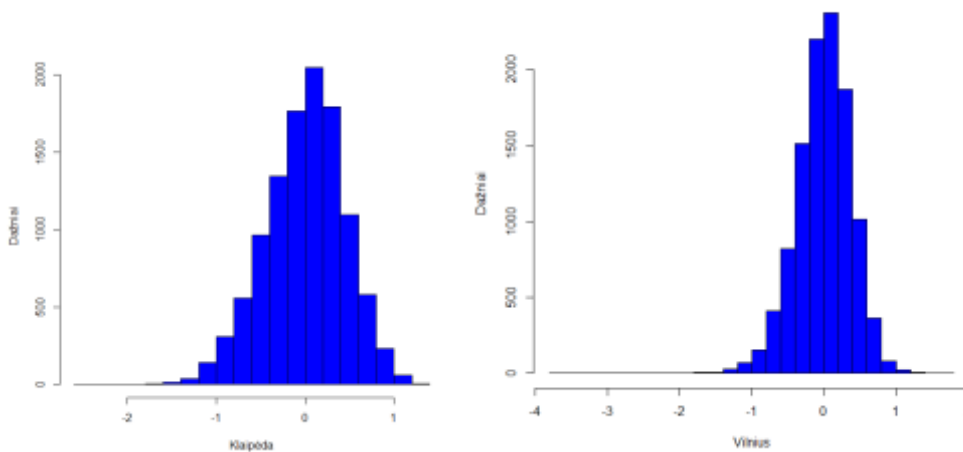
$ARIMA(3,1,1)$ modelių liekanų ACF dar rodo priklausomybę laike (17 pav.). Remiantis histogramomis, visose stotyse liekanos pasižymėjo neigiamu asimetrijos ir teigiamu eksceso koeficientu, tačiau pasiskirstymas atrodė beveik simetriškas (18 pav.). Liekanų vidurkis tenkino hipotezę apie lygybę nuliui, standartiniai nuokrypiai svyravo nuo 0,36 iki 0,49. Remiantis Ljung-Box testo rezultatais liekanų autokoreliacija buvo statistiškai nereikšminga.



17 pav. Liekanų, gautų įvertinus $ARIMA(3,1,1)$ modelį, ACF su 95% pasikliautinųjų intervalų ribomis

	ϕ_1^{*i}	ϕ_2^{*i}	ϕ_3^{*i}	θ_1^{*i}
Biržai	0,4342	-0,0629	0,033	-0,9701
Dotnuva	0,4178	-0,0817	0,0507	-0,9673
Dūkštas	0,4346	-0,0621	0,0673	-0,9785
Kaunas	0,4456	-0,0490	0,0281	-0,9641
Kybartai	0,452	-0,0203	0,0363	-0,9724
Klaipėda	0,5064	-0,0576	0,0614	-0,9740
Laukuva	0,4537	-0,0581	0,0386	-0,9713
Lazdijai	0,4661	-0,0407	0,035	-0,9713
Nida	0,4044	-0,0219	0,0241	-0,9665
Panevėžys	0,4430	-0,0465	0,0337	-0,9722
Raseiniai	0,3964	-0,0570	0,0248	-0,9696
Šiauliai	0,4398	-0,0441	0,0382	-0,9701
Šilutė	0,4517	-0,0444	0,0501	-0,9794
Telšiai	0,4530	-0,0465	0,0395	-0,9746
Ukmergė	0,4371	-0,0543	0,0565	-0,9747
Utena	0,4105	-0,0329	0,0284	-0,9689
Varėna	0,4389	-0,0476	0,0178	-0,9725
Vilnius	0,4635	-0,0489	0,0348	-0,9713

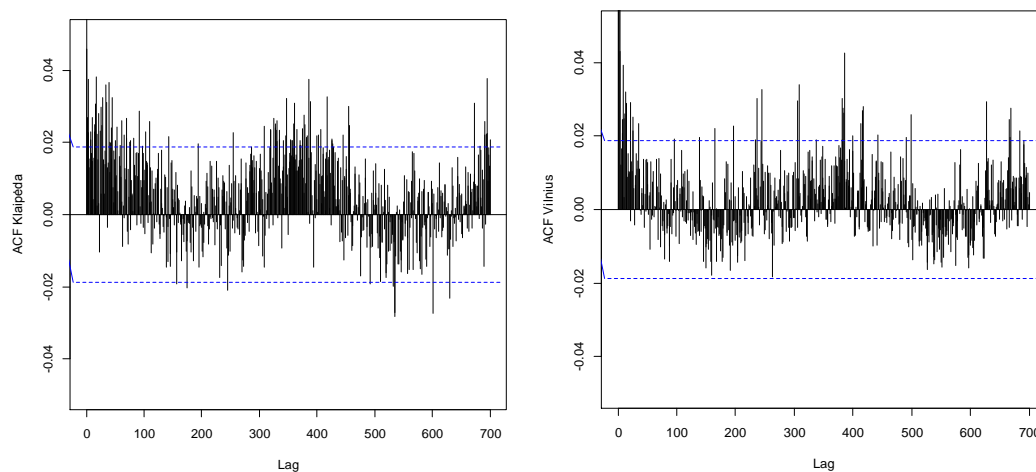
4 lentelė. *ARIMA(3,1,1)* modelio parametrai.



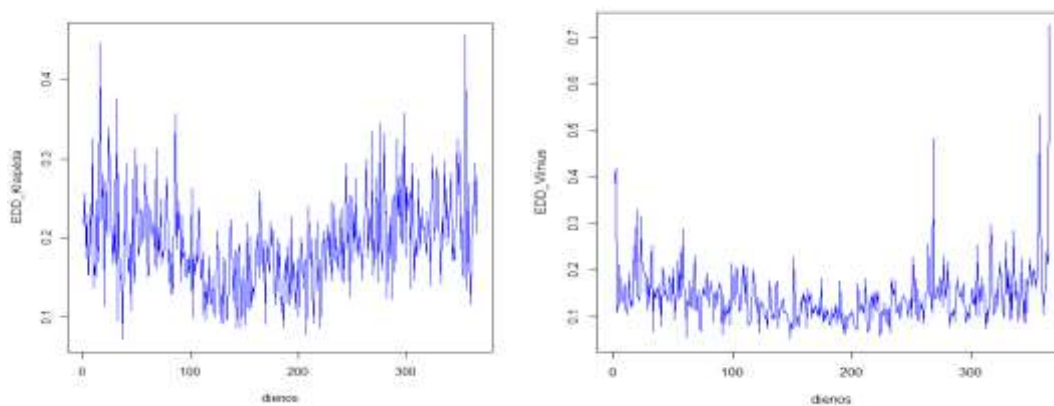
18 pav. Liekanų, gautų įvertinus *ARIMA(3,1,1)* modelį, histogramos

Išanalizavus kvadratinių liekanų ACF, nustatyta, kad liekanų dispersija vis dar priklausoma laike. Taip pat buvo apskaičiuotos empirinės dienos dispersijos (EDD) (20 pav.) bei įvertintas (2.9) lygtimi aprašytas modelis (5 lentelė). Kaip ir (2.7) modelio atveju, palyginus liekanų, normuotų empirinės

dienos dispersijos pagalba bei įvertintos dispersijos funkcijos (2.10) pagalba, ACF, nustatyta, jog abiem atvejais rezultatai yra gana panašūs. Dėl mažesnio vertinamų parametų skaičiaus pasirinktas normavimas, remiantis empirine dienos dispersija (EDD). EDD pagalba normuotų kvadratinių liekanų ACF, pavaizduota 21 paveiksle, rodo, kad sezoniškumas iš liekanų dispersijos buvo pašalintas.



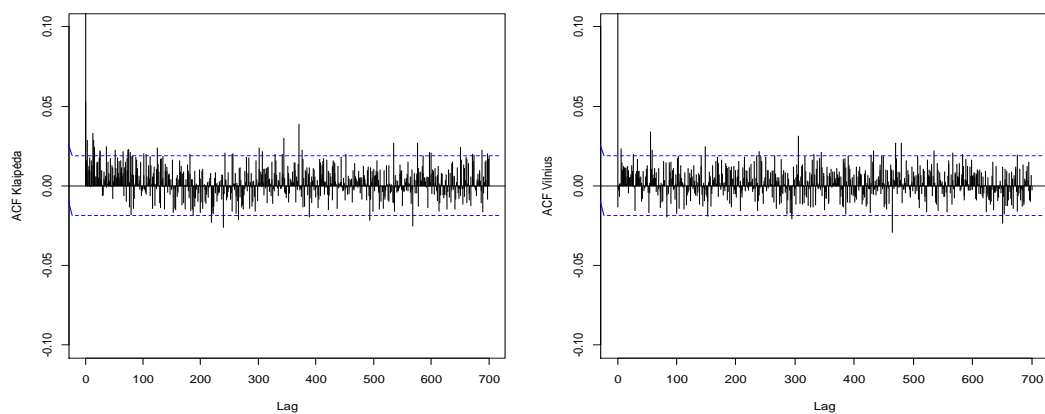
19 pav. Kvadratinių liekanų ACF su 95% pasikliautinųjų intervalų ribomis, įvertinus $ARIMA(3,1,1)$ modelį.



20 pav. Liekanų empirinė dienos dispersija.

	c_1^{*i}	c_2^{*i}	c_3^{*i}
Biržai	0,152	0,011	-0,003*
Dotnuva	0,173	0,031	-0,005*
Dūkštas	0,23	0,042	-0,016
Kaunas	0,151	0,011	-0,003*
Kybartai	0,221	0,054	-0,007*
Klaipėda	0,2	0,36	-0,017
Laukuva	0,221	0,065	-0,03
Lazdijai	0,185	0,023	0,011
Nida	0,164	0,023	-0,017
Panevėžys	0,169	0,014	-0,011
Raseiniai	0,133	0,022	-0,001*
Šiauliai	0,171	0,019	-0,008*
Šilutė	0,195	0,059	-0,031
Telšiai	0,139	0,032	-0,007*
Ukmergė	0,228	0,021	-0,019
Utena	0,211	0,034	-0,010*
Varėna	0,236	0,057	-0,022
Vilnius	0,149	0,037	-0,001*

5 lentelė. Funkcijos $\sigma_k^2(t)$ įverčiai. Statistiškai nereikšmingi parametrai pažymėti žvaigždute.



21 pav. Normuotų liekanų ACF su 95% pasikliautinių intervalų ribomis.

Prieš pereinant prie erdvinio modeliavimo, svarbu patikrinti laiko eilučių modelių adekvatumą.

Kaip jau minėta aukščiau, laiko eilučių modelių patikrinimui naudojami į imtį neįtraukti duomenys, tai yra 2007 metų vėjų greičiai, fiksuoti visose 18 stočių. Apskaičiavus vieno žingsnio į priekį prognozes, apskaičiuotos prognozės klaidos (PE). 6 lentelėje pateikti rezultatai rodo, kad (2.7) lygtimi

aprašyto modelio atveju, remiantis Kolmogorovo – Smirnovu testu, liekanos pasiskirsčiusios pagal normalųjį dėsnį visose išskyrus dvi (Šiaulių ir Utenos) stotis, tačiau ir šiose dviejose stotyse histogramos buvo simetriškos.

	Vidurkis	SN	KS reikšmė	p	EDD	FIT
Biržai	-0,03	0,34	0,243		2,24	1,96
Dotnuva	-0,01	0,34	0,701		1,96	1,12
Dūkštas	-0,13	0,42	0,987		2,24	1,96
Kaunas	-0,01	0,35	0,350		3,64	2,52
Kybartai	0,01	0,38	0,105		1,96	1,12
Klaipėda	-0,02	0,38	0,617		2,52	2,24
Laukuva	0,04	0,40	0,471		2,80	2,80
Lazdijai	0,01	0,37	0,160		2,80	2,52
Nida	-0,05	0,38	0,481		5,32	3,92
Panevėžys	-0,01	0,36	0,087		5,32	3,36
Raseiniai	-0,03	0,34	0,218		5,32	5,04
Šiauliai	-0,04	0,37	0,015		4,76	4,48
Šilutė	0,06	0,38	0,863		4,48	3,64
Telšiai	0,04	0,30	0,293		3,08	1,68
Ukmergė	-0,04	0,40	0,265		3,36	2,52
Utena	-0,08	0,55	0,006		10,36	8,12
Varėna	-0,08	0,45	0,064		5,60	5,32
Vilnius	-0,02	0,34	0,810		3,64	1,96

6 lentelė. (2.7) modelio PE aprašomosios statistikos charakteristikos, Kolmogorovo Smirnovu p reikšmės bei prognozės intervalai, dienos empirinei (EDD) bei įvertintai (FIT) dispersijos funkcijoms.

Analogiški (2.12) lygtimi aprašyto modelio vieno žingsnio į priekį prognozės rezultatai pateikti 7 lentelėje.

	Vidurkis	SN	KS reikšmė	p	EDD	FIT
Biržai	-0,02	0,29	0,251		2,24	1,96
Dotnuva	-0,03	0,31	0,691		1,96	1,12
Dūkštas	-0,11	0,38	0,871		2,24	1,96
Kaunas	-0,02	0,34	0,431		3,64	2,52
Kybartai	0,02	0,39	0,301		1,96	1,12
Klaipėda	-0,01	0,41	0,589		2,52	2,24
Laukuva	0,05	0,37	0,318		2,80	2,80
Lazdijai	0,02	0,35	0,160		2,80	2,52
Nida	-0,03	0,36	0,481		5,32	3,92
Panevėžys	-0,02	0,39	0,087		5,32	3,36
Raseiniai	-0,01	0,35	0,003		5,32	5,04
Šiauliai	-0,05	0,35	0,219		4,76	4,48
Šilutė	0,07	0,39	0,754		4,48	3,64
Telšiai	0,05	0,28	0,293		3,08	1,68
Ukmergė	-0,04	0,41	0,231		3,36	2,52
Utena	-0,04	0,59	0,003		10,36	8,12
Varėna	-0,03	0,48	0,064		5,60	5,32
Vilnius	-0,04	0,33	0,741		3,64	1,96

7 lentelė. (2.12) modelio PE aprašomosios statistikos charakteristikos, Kolmogorovo Smirnovo p reikšmės bei prognozės intervalai, dienos empirinei (EDD) bei įvertintai (FIT) dispersijos funkcijoms.

3.3. Erdvinis interpoliavimas

Įvertintame laiko eilučių modelyje, aprašytame (2.7) lygtimi, kiekvienoje stotyje yra po 12 parametrų. Tam, kad įvertinti duomenų, gautų pašalinus sezoninę komponentę, autokoreliacijas, daugelyje stočių pakako $AR(3)$ ar $AR(4)$ procesų. Tačiau keturiose stotyse prireikė aukštesnės eilės modelių (2 lentelė). Erdvinis modeliavimas atliekamas remiantis prielaida, kad kiekvienoje stotyje buvo įvertintas $AR(8)$ procesas, o neįvertinti parametrai laikomi lygiais nuliui. Analogiškai (2.12) modelio atveju, kiekvienoje stotyje turime po 5 parametrus (4 lentelė), t.y. $ARIMA(3,1,1)$ modelio parametrai ir liekanų dispersija.

Semivariogramos modeliavimas

Kiekvienoje stotyje įvertinus laiko eilučių modelius, gautos liekanos tenkina hipotezę apie nepriklausomumą ir vienodą pasiskirstymą. Remiantis šia prielaida, galima įvertinti erdvinę semivariogramą. Siekiant sumažinti

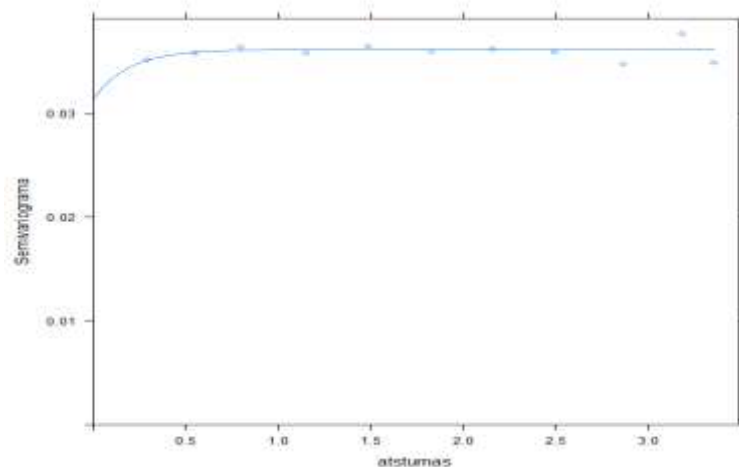
nepriklausomų stebėjimų skaičių kiekvienoje stotyje, abiejų laiko eilučių modelių atveju, buvo apskaičiuoti kiekvienos dienos standartizuotų liekanų vidurkiai, gaunant 365 stebėjimus 18 stočių. Remiantis (2.15)–(2.17) formulėmis, įvertinti eksponentinis, Ornstein-Uhlenbecko bei sferinis semivariogramų modeliai. Remiantis vidutine kvadratine paklaida (MSE), abiejų laiko eilučių modelių atveju, tiksliausiai erdvinę priklausomybę tarp stočių aprašo eksponentinė semivariograma.

(2.7) modelio atveju gauta semivariograma pavaizduota 22 pav. Įvertintas modelis ($MSE = 0,013$):

$$\gamma(h; \hat{\boldsymbol{\theta}}) = \begin{cases} 0, & \text{kai } h = 0 \\ 0,031 + 0,002 \left(1 - \exp \left(-\frac{1}{38472,13} \sqrt{1,15h_1^2 + h_2^2} \right) \right), & \text{kai } h > 0. \end{cases}$$

(2.12) modelio atveju gautas eksponentinės semivariogramos modelis su gana panašiais parametru įverčiais ($MSE = 0,015$):

$$\gamma(h; \hat{\boldsymbol{\theta}}) = \begin{cases} 0, & \text{kai } h = 0 \\ 0,034 + 0,002 \left(1 - \exp \left(-\frac{1}{38513,41} \sqrt{1,18h_1^2 + h_2^2} \right) \right), & \text{kai } h > 0. \end{cases}$$



22 pav. Eksponentinė semivariograma, įvertinta (2.7) modelio liekanoms. Atstumas metrais padaugintas iš 10^{-5} .

Erdviniai svoriai

Erdvinių modelių patikrinimui bei palyginimui pasinaudota dviem įimtj neįtrauktomis stotimis – Palangos ir Trakų Vokės. Remiantis įvertintomis semivariogramomis, (2.20) formulės pagalba Palangos ir Trakų Vokės stotims įvertinti erdviniai svoriai. (2.7) bei (2.12) modeliams gauti svorių įverčiai pateikti 8 lentelėje, kurioje $\hat{\delta}_P$ – svorių vektorius tarp Palangos ir likusių 18 stočių, įvertintas (2.7) modelio atveju; $\hat{\delta}_P^*$ - svorių vektorius tarp Palangos ir likusių 18 stočių, įvertintas (2.12) modelio atveju; $\hat{\delta}_{TV}$ - svorių vektorius tarp Trakų Vokės ir likusių 18 stočių, įvertintas (2.7) modelio atveju; $\hat{\delta}_{TV}^*$ - svorių vektorius tarp Trakų Vokės ir likusių 18 stočių, įvertintas (2.12) modelio atveju.

	$\hat{\delta}_P$	$\hat{\delta}_{TV}$	$\hat{\delta}_P^*$	$\hat{\delta}_{TV}^*$
Biržai	0,005	0,014	0,003	0,008
Dotnuva	0,015	0,041	0,017	0,063
Dūkštas	0,009	0,022	0,009	0,023
Kaunas	0,015	0,263	0,009	0,237
Kybartai	0,011	0,021	0,011	0,028
Klaipėda	0,257	0,012	0,247	0,005
Laukuva	0,114	0,009	0,098	0,007
Lazdijai	0,005	0,015	0,005	0,011
Nida	0,168	0,005	0,173	0,009
Panevėžys	0,004	0,057	0,004	0,053
Raseiniai	0,089	0,014	0,098	0,019
Šiauliai	0,093	0,035	0,097	0,035
Šilutė	0,119	0,013	0,123	0,013
Telšiai	0,078	0,013	0,078	0,013
Ukmergė	0,007	0,021	0,007	0,021
Utena	0,008	0,031	0,008	0,073
Varėna	0,005	0,022	0,006	0,031
Vilnius	0,007	0,391	0,007	0,351

8 lentelė. Svorių įverčiai Palangos ir Trakų Vokės stotims

Trendo paviršiaus modelis

Kiekvienam įvertintam laiko eilutės parametrai, mažiausių kvadratų metodu buvo įvertintas antros eilės trendo paviršiaus modelis apibrėžtas (2.24) lygtimi, t. y. tokio pavidalo modelis:

$$P^{\hat{\beta}} = \lambda_0^{\hat{\beta}} + \lambda_1^{\hat{\beta}}x + \lambda_2^{\hat{\beta}}y + \lambda_3^{\hat{\beta}}x^2 + \lambda_4^{\hat{\beta}}y^2.$$

Čia x ir y yra 18×1 matavimo vektoriai, atspindintys metrinės stočių koordinatas. Įvertinto antros eilės trendo paviršiaus modelio parametrai, (2.7) modelio atveju, pateikti 9 lentelėje kartu su atitinkamomis determinacijos koeficiento R^2 reikšmėmis bei vidutinėmis kvadratinėmis klaidomis (MSE). Kaip rodo determinacijos koeficientas R^2 , remiantis antros eilės trendo paviršiaus modeliu, didelė dalis parametru klaidos lieka nepaaiškinta.

	$\lambda_0^{\hat{\beta}}$	$\lambda_1^{\hat{\beta}}$	$\lambda_2^{\hat{\beta}}$	$\lambda_3^{\hat{\beta}}$	$\lambda_4^{\hat{\beta}}$	$R^2(\%)$	MSE
b_0	1,220	-0,159	0,180	0,013	-0,078	44,9	0,0263
b_1	0,171	-0,002	0,002	0,002	-0,007	12,0	0,0007
b_2	-0,055	0,090	0,043	-0,022	-0,017	58,8	0,0007
ϕ_1	0,507	-0,019	-0,033	0,003	0,0137	29,6	0,0003
ϕ_2	-0,013	-0,020	0,001	0,006	-0,005	20,6	0,0003
ϕ_3	0,048	-0,010	0,036	0,003	-0,015	55,5	0,0001
ϕ_4	0,013	-0,002	0,015	0,002	-0,005	5,8	0,0007
ϕ_5	0,004	-0,009	0,020	0,003	-0,009	17,3	0,0002
ϕ_6	0,011	0,001	0,004	0,001	0,005	6,8	0,0005
ϕ_7	0,020	-0,020	0,001	0,005	-0,002	34,4	0,0001
ϕ_8	0,028	-0,029	0,002	0,007	-0,003	34,4	0,0002
σ^2	0,217	-0,044	-0,013	0,015	-0,000	23,8	0,0011

9 lentelė. Antros eilės trendo paviršiaus modelio parametru įverčiai (2.7) modelio parametrams, R^2 ir MSE

$ARIMA(3,1,1)$ modelio atveju įvertinti 2-os eilės trendo paviršiaus modelio parametrai pateikti 10 lentelėje. Remiantis determinacijos koeficientais (R^2) taip pat galima padaryti išvadą, kad trendo paviršiaus modelis nepaaiškina didelės dalies sklaidos. Tokius rezultatus abiejų modelių

atveju galima pagrįsti tuo, kad trendo paviršiaus modelio parametrai neatsižvelgia į erdvines koreliacijas tarp stočių.

	$\lambda_0^{\hat{\beta}}$	$\lambda_1^{\hat{\beta}}$	$\lambda_2^{\hat{\beta}}$	$\lambda_3^{\hat{\beta}}$	λ_4^{η}	R^2 (%)	MSE
ϕ_1^*	0,492	-0,028	-0,061	0,006	0,025	24,87	0,0013
ϕ_2^*	-0,022	-0,015	-0,016	0,003	0,003	23,5	0,0023
ϕ_3^*	0,034	-0,013	0,018	0,005	-0,006	13,12	0,0018
θ_1^*	-0,976	0,006	0,004	-0,002	-0,002	18,84	0,0027
σ^{*2}	0,215	-0,035	-0,012	0,013	-0,002	22,86	0,0011

10 lentelė. *ARIMA*(3,1,1) modelio antros eilės trendo paviršiaus modelio parametru įverčiai, R^2 ir MSE

3.4. Erdvinių modelių patikrinimas

11 ir 12 lentelėse pateikiami atitinkamai (2.7) bei (2.12) lygtimis aprašytų vienmačių laiko eilučių modelių, įvertinų Palangos bei Trakų Vokės duomenims, parametrai su standartinėmis paklaidomis. Taip pat palyginimui šiose lentelėse pateikiami parametru įverčiai, gauti remiantis (2.20), (2.22), (2.24) bei (2.25) formulėmis, t. y. visų keturių 2 skyriuje aprašytų interpoliavimo metodų rezultatai Palangos bei Trakų Vokės stotims atitinkamai (2.7) bei (2.12) modelių parametrams.

Visos įvertintos reikšmės yra nutolusios ne daugiau kaip per dvi standartines paklaidas nuo sugludintų reikšmių. Tai rodo, kad siūlomi erdviniai modeliai pakankamai tiksliai aprašo analizuojamą situaciją. Tiek (2.7), tiek (2.12) modelių atveju, kaip ir tikėtasi, arčiausiai tikrųjų reikšmių yra erdvinių svorių bei ordinaraus kriginio pagalba apskaičiuoti parametru įverčiai, t.y. įverčiai, gauti atsižvelgiant į erdvines koreliacijas.

	Palanga					Trakų Vokė				
	Įverčiai (SE)	ES	OK	TPM	IDW	Įverčiai (SE)	ES	OK	TPM	IDW
b_0	1,414 (0,006)	1,266	1,316	1,257	1,136	1,014 (0,005)	1,118	0,997	0,958	1,112
b_1	0,141 (0,009)	0,165	0,151	0,148	0,161	0,175 (0,007)	0,168	0,183	0,167	0,167
b_2	0,008 (0,009)	0,005	0,011	-0,029	0,024	0,310 (0,007)	0,360	0,395	0,483	0,224
ϕ_1	0,478 (0,014)	0,498	0,464	0,494	0,475	0,433 (0,010)	0,471	0,459	0,466	0,468
ϕ_2	-0,046 (0,015)	-0,033	-0,045	-0,031	-0,029	-0,020 (0,010)	-0,036	-0,015	-0,028	-0,029
ϕ_3	0,067 (0,015)	0,062	0,073	0,057	0,059	0,037 (0,010)	0,058	0,061	0,057	0,055
ϕ_4	0,055 (0,014)	0,026	0,035	0,023	0,023					
σ^2	0,173 (0,005)	0,198	0,193	0,186	0,181	0,219 (0,005)	0,232	0,301	0,198	0,193

11 lentelė. (2.7) lygtimi aprašyto modelio įverčiai su standartinėmis paklaidomis (SE) bei įverčiai, gauti erdviųjų svorių atveju (ES), formalaus ordinarus krigingo atveju (OK), trendo paviršiaus modelio atveju (TPM), bei atvirkštinio svorinio atstumo atveju (IWD).

	Palanga					Trakų Vokė				
	Įverčiai (SE)	ES	OK	TPM	IDW	Įverčiai (SE)	ES	OK	TPM	IDW
ϕ_1^*	0,429 (0,014)	0,450	0,437	0,418	0,461	0,415 (0,010)	0,446	0,438	0,419	0,487
ϕ_2^*	-0,062 (0,015)	-0,047	-0,051	-0,059	-0,048	-0,027 (0,011)	-0,049	-0,041	-0,022	-0,052
ϕ_3^*	0,065 (0,014)	0,042	0,071	0,019	0,034	0,018 (0,010)	0,035	0,027	0,034	0,055
θ_1^*	-0,977 (0,004)	-0,972	-0,981	-0,965	-0,971	-0,975 (0,003)	-0,970	-0,961	-0,976	-0,974
σ^{*2}	0,187 (0,013)	0,191	0,201	0,151	0,154	0,244 (0,009)	0,198	0,203	0,215	0,190

12 lentelė. (2.12) lygtimi aprašyto modelio įverčiai su standartinėmis paklaidomis (SE) bei įverčiai, gauti erdviųjų svorių atveju (ES), formalaus ordinarus krigingo atveju (OK), trendo paviršiaus modelio atveju (TPM), bei atvirkštinio svorinio atstumo atveju (IWD).

3.5. Erdvės-laiko modelio parinkimas kryžminio patikrinimo metodu

Įvertinus (2.7) bei (2.12) lygtimis aprašytus laiko eilučių modelių parametrus ir sudarius erdvinius modelius buvo gauti aštuoni modeliai. Modelio parinkimui buvo pritaikytas „palikti vieną“ (leave-one-out) kryžminio

patikrinimo metodas. Metodą pritaikius kiekvienai iš 20 stočių, buvo apskaičiuotos vieno žingsnio į priekį 2007 metų prognozės. Kadangi modeliuojant laiko eilutes pradiniai duomenys buvo logaritmuoti, kryžminio patikrinimo metodui taip pat buvo naudojami logaritmuoti duomenys. Galima pastebėti, kad juos būtų nesudėtinga atstatyti, pritaikius eksponentinę funkciją.

13 lentelėje pateikti kryžminio patikrinimo metodo rezultatai rodo, kad geriausiai duomenis aprašo $\hat{Z}_{ES}(s_0; t)$ modelis, tai yra modelis, kuriame buvo pašalinta sezoninė komponentė ir gautos stacionarios liekanos, kurioms įvertinti autoregresiniai modeliai $AR(p)$, o erdviniam modeliavimui pritaikytas erdviųjų svorių metodas, tai yra metodas, atsižvelgiantis į erdvinę koreliaciją (semivariogramas) tarp stebėjimų.

Modelis	RMSE
$\hat{Z}_{TPM}(s_0; t)$	0,57 m/s
$\hat{Z}_{ES}(s_0; t)$	0,14 m/s
$\hat{Z}_{IDW}(s_0; t)$	0,21 m/s
$\hat{Z}_{OK}(s_0; t)$	0,16 m/s
$\hat{Z}_{TPM}^*(s_0; t)$	0,48 m/s
$\hat{Z}_{ES}^*(s_0; t)$	0,17 m/s
$\hat{Z}_{IDW}^*(s_0; t)$	0,28 m/s
$\hat{Z}_{OK}^*(s_0; t)$	0,23 m/s

13 lentelė. Kryžminio patikrinimo metodo rezultatai

Visų siūlomų modelių paklaidos gana nedidelės, tačiau matome, kad tikslesnius rezultatus duoda metodai, atsižvelgiantys į erdvinę koreliaciją.

Išvados

Pirmoje darbo dalyje aptarti erdvės-laiko duomenys, aprašytos pagrindinės jų charakteristikos, duomenų tipai bei modeliai. Antroje darbo dalyje aprašyti siūlomi erdvinių laiko eilučių modeliavimo metodai, kurių realizacija aptarta trečioje darbo dalyje.

Pasiūlyti gana paprasti metodai, interpoliuojantys vienmačių laiko eilučių parametrus erdvėje. Šie metodai pakankamai tiksliai aprašo vėjo greičio dinamiką erdvėje ir laike.

Pirmiausia, kiekviename erdvės taške, nagrinėjamos vienmatės laiko eilutės. Siūlomi du laiko eilučių modeliai. Pirmasis siūlomas modelis sudarytas iš sezoninės komponentės, $AR(p)$ proceso bei sezoninės liekanų dispersijos. Laiko eilučių dekompozicijos taikymas leidžia gana lengvai sumodeliuoti vėjo greičio dinamiką laike. Pastebėsime, kad pašalinus visas determinuotas komponentes, liko tik liekanos, kurios buvo asimptotiškai nepriklausomos laike ir vienodai pasiskirsčiusios pagal normalųjį dėsnį.

Antrasis siūlomas vienmačių laiko eilučių modelis – klasikinis integruotas autoregresijos slenkamojo vidurkio modelis $ARIMA(p, d, q)$. Šio modelio liekanos po dispersijos sezoniškumo įvertinimo taip buvo asimptotiškai nepriklausomos laike ir vienodai pasiskirsčiusios.

Erdvinės struktūros nustatymui, vertinama empirinė smeivariograma. Empirinė semivariograma vertinama nepriklausomiems ir vienodais pasiskirsčiusiems stebėjimams (replikacijoms).

Erdviniam modeliavimui pasiūlyti keturi metodai: du iš jų atsižvelgia į erdvinę koreliaciją, t. y. naudojama erdvinių svorių formulė bei formalaus ordinaraus kriginio metodas, o kiti du atsižvelgia tik į taškų tarpusavio

išsidėstymą, t. y. naudojami trendo paviršiaus bei svertinio atvirkštinio atstumo metodai.

Vėjo greitis yra gana triukšmingas meteorologinis rodiklis, todėl jį gana sudėtinga modeliuoti. Dėl šios priežasties siūlomi modeliai turi gana daug parametrų, tačiau iš kitos pusės modeliai yra gana aiškūs ir juos lengva taikyti erdvės-laiko prognozei, kas yra labai svarbu įvairiuose taikymuose.

Nagrinėjamuose modeliuose galima atskirti erdvę ir laiką, taikant gerai žinomą *ARIMA* procesą laike ir Gauso atsitiktinį lauką erdvėje. *ARIMA* parametrai yra gana stabilūs erdvėje, o visa erdvinė priklausomybė išreikšta liekanose. Tokie modeliai gali būti taikomi tuo atveju, kai analizuojamas dydis visoje teritorijoje yra homogeniškas. Didesnėje ar labiau komplikuojuose geografinėje vietoje, modeliai būtų daug sudėtingesni ir turėtų dar daugiau komponentų.

Pritaikius kryžminio patikrinimo metodą bei apskaičiavus šaknį iš vidutinės kvadratinės paklaidos (RMSE) buvo nustatyta, kad analizuojamą vėjo greitį Lietuvos respublikoje geriausiai aprašo sezoninis $AR(p)$ modelis, kurio parametrų interpoliavimui buvo pritaikytas erdvinių svorių metodas.

Literatūra

1. Abt M. (1999). Estimation the Prediction Mean Squared Error in Gaussian Stochastic Processes with Exponential Correlation Structure. *Scandinavian Journal of Statistics*, 26, pp. 563–578.
2. Abt M., Welch W. J. (1998). Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes. *The Canadian Journal of Statistics*, 26(1.1), pp. 127–137.
3. Ailliot P., Monbet V. and Prevosto M. (2006). An autoregressive model with time-varying coefficients for wind fields. *Environmetrics*, 17, pp. 107–117.
4. Anderson T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley & Sons, New York.
5. Di Battista T., Di Filippo R., Visini G. (2004). The environmental Long – Memory Space–Time Series prediction. *Environmental Informatics Archives*, 2, pp. 289–296.
6. Benth F.E. and Šaltytė Benth. J. (2007). The volatility of temperature and pricing of weather derivatives. *Quantitative Finance*, 7(5), pp. 553–561.
7. Boshnakov G. N. (2005). On the asymptotic properties of multivariate sample autocovariances. *Journal of Multivariate Analysis*, 92, pp. 45–52.
8. Bouette J.-Ch., Chassagneux J.-F., Sibai D., Terron R. and Charpentier A. (2006). Wind in Ireland: long memory or seasonal effect? *Stochastic Environmental Research and Risk Assessment*, 20, pp. 141–151.
9. Box G. E. P., Jenkins G. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
10. Brett A.C. and Tuller S.E. (1991). The autocorrelation of hourly wind speed observations. *Journal of Applied Meteorology*, 30(6), pp. 823–833.
11. Brown B.G., Katz R.W. and Murphy A.H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, 23(8), pp. 1184–1195.

12. Brown P. E., Diggle P. J., Lord M. E., Young P. C. (2001). Space–time calibration of radar rainfall data. *Applied Statistics*, 50(2), pp. 221–241.
13. Bruno F., Guttorp P., Sampson P. D., Cocchi D. (2003). Non-separability of space–time covariance models in environmental studies. *The ISI International Conference on Environmental Statistics and Health*, pp. 153–161.
14. Castino F., Festa R. and Ratto C.F. (1998). Stochastic modelling of wind velocities time series. *Journal of Wind Engineering and Industrial Aerodynamics*, 74-76, pp. 141–151.
15. Celik A.N. (2003). A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renewable Energy*, 29, pp. 593–604.
16. Chandler R.E., Bate S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1), pp. 167–183.
17. Chatfield, C. (2000). *Time-Series Forecasting*. Chapman & Hall /CRC Press, Boca Raton.
18. Chiles J. P. Delfiner P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.
19. Christensen R. (2001). *Advanced Linear Modeling*. Springer – Verlag, New York.
20. Cliff A. D., Ord J. K. (1975). Space–Time Modeling with an Application to Regional Forecasting. *Transactions of the Institute of British Geographers*, 64, p.p. 119–128.
21. Coenen G. (2005). Asymptotic confidence bands for the estimated autocovariance and autocorrelation functions of vector autoregressive models. *Empirical Economics*, 30, pp. 65–75.
22. Cressie N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17 (5), pp. 563–586.
23. Cressie N., Huang H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of American Statisticians Asociacion*, 94 (448), pp. 1330–1340.
24. Cressie N., Wikle Ch. K. (2011). *Statistics for Spatio-Temporal Data*. Willey, New York.

25. Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.
26. Cripps E., Nott D., Dunsmuir W.T.M. and Wikle Ch. (2005). Space–time modelling of Sydney Harbour winds. *Australian and New Zealand Journal of Statistics*, 47(1), pp. 3–17.
27. De Cesare S., Myers D.E., Posa D. (2001). Product–sum covariance for space–time modeling: an environmental application. *Environmetrics* 12 (1), p.p. 11–23.
28. De Iaco S., Myers D.E., Posa D. (2002). Nonseparable space–time covariance models: some parametric families. *Mathematical Geology*, 34(1), pp. 23–43.
29. De Iaco S., Myers D.E., Posa D. (2002). Space–time variograms and a functional form for total air pollution measurements. *Computational Statistics & Data Analysis*, 41 p.p. 311–328.
30. De Iaco S., Palma M., Posa D. (2005). Modeling and prediction of multivariate space–time random fields. *Computational Statistics & Data Analysis*, 48, pp. 525–547.
31. De Iaco S., Posa D. (2012). Predicting spatio-temporal random fields: Some computational aspects. *Computers and Geosciences*, 41, pp. 12–24.
32. Dučinskis K., Šaltytė Benth J. (2003). *Erdvinė statistika*. Klaipėdos universiteto leidykla, Klaipėda.
33. Dykes J.A., Mountain D.M. (2003). Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics & Data Analysis*, 43, 581–603.
34. Fonseca T. C. O., Steel M. F. J. (2011). A general class of nonseparable space–time covariance models. *Environmetrics*, 22(2), pp. 224–242.
35. Fuentes M. (2006). Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference*, 136, pp. 447–466.
36. Garcia A., Torres J.L., Prieto E. and De Francisco A. (1998). Fitting wind speed distributions: a case study. *Solar Energy*, 62(2), pp. 139–144.
37. Genton M. G. (2007). Separable approximations of space–time covariance matrices. *Environmetrics*, 18, pp. 681–695.

38. Giacomini R., Granger C. W. J. (2004). Aggregation of space–time processes. *Journal of Econometrics*, 118, pp. 7–26.
39. Hagnell M. (1991). A Multivariate Time Series Analysis of Fertility, Adult Morality, Nuptiality, and Real Wages in Sweden 1751 – 1850: a Comparison of two different approaches. *Journal of Official Statistics*, 7(4), pp. 437–455.
40. Haslett J., Raftery A.E. (1989). Space–time modelling with long-memory dependence: assessing Ireland’s wind power resource. *Applied Statistics*, 38(1), pp. 1–50.
41. Huang H.-C., Martinez F., Mateu J., Montes F. (2007). Model comparison and selection for stationary space–time models. *Computational Statistics & Data Analysis*, 51, p.p. 4577–4596.
42. Im H. K, Rathouz P. J., Frederick J. E. (2009). Space–time modeling of 20 years of daily air temperature in the Chicago metropolitan region. *Environmetrics*, 20(5). pp. 494–511.
43. Journel A. G., Huijbregts C. J. (1978). *Mining Geostatistics*. Academic Press, London.
44. Jun M, Stein M.L. (2007). An approach to producing space–time covariance functions on spheres. *Technometrics*, 49(4). pp. 468–479.
45. Kamarianakis Y. (2003). Spatial–Time Series Modeling: a Review of the Proposed Methodologies. *Regional Economics Applications Lab Technical Series, University of Illinois at Champaign-Urbana*, REAL 03-T-19.
46. Kooperberg Ch., O’Sullivan F. (1996). Predictive Oscillation Patterns: A Synthesis of Methods for Spatial-Temporal Decomposition of Random Fields. *Journal of the American Statistical Association*, 91 (436), pp. 1485–1495.
47. Lahiri S. N. et. all. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference*, 103, pp. 65–85.
48. Landagan O. Z., Barrios, E. B. (2007). An Estimation procedure for a spatial-temporal model. *Statistics & Probability Letters*, 77, pp. 401–406.

49. Lee E., Moon M.-S., Gunst R.F. (2006). A Spatiotemporal Model for Pollutant Concentrations. *Communications in Statistics – Theory and Methods*, 35, pp. 489–497.
50. Le N. D., Zidek J. V. (2006). *Statistical Analysis of Environmental Space–Time Processes*. Springer, New York.
51. Lertxundi-Manterola A., Saez M. (2009). Modelling of nitrogen dioxide (NO₂) and fine particulate matter (PM₁₀) air pollution in the metropolitan areas of Barcelona and Bilbao, Spain. *Environmetrics*, 20, pp. 477–493.
52. Li B., Genton M. G., Sherman M. (2008). On the asymptotic joint distribution of sample space–time covariance estimators. *Bernoulli*, 14, pp. 228–248.
53. Lun I.Y.F. and Lam J.C. (2000). A study of Weibull parameters using long-term wind observations. *Renewable Energy*, 20, pp. 145–153.
54. Ma Ch. (2003). Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis*, 86, pp. 97–107.
55. Martin M., Cremades L.V. and Santabarbara J.M. (1999). Analysis and modelling of time series of surface wind speed and direction. *International Journal of Climatology*, 19, pp. 197–209.
56. Milligan M., Schwartz M. and Wan Y. (2003). Statistical wind power forecasting models: results for U.S. wind farms. Conference paper presented May 18-21 at Windpower 2003, Austin, Texas.
57. Nagarsenker B., Nagarsenker P. B. (1982). On the distribution of independence test for the residuals of Space–Time ARMA models. *Communications in Statistics – Theory and Methods*, 11 (6), pp. 653–661.
58. Nott D. J., Dunsmuir W. T. M. (2002). Estimation of nonstationary spatial covariance structure. *Biometrika*, 89 (4), pp. 819–829.
59. Orasi A., Lasinio G. J., Ferrari C. (2009). Comparison of calibration methods for the reconstruction of space–time rainfall fields during a rain enhancement experiment in Southern Italy. *Environmetrics*, 20(7), pp. 812–834.
60. Pfeifer P. E., Deutsch S. J. (1980). A Three–Stage Iterative Procedure for Space–Time Modeling. *Technometrics*, 22(1), pp. 35–47.

61. Pfeifer P. E., Deutsch S. J. (1981). Seasonal Space–Time ARIMA Modeling. *Geographical Analysis*, 13(2), pp. 117–132.
62. Rathouz P. J. Longitudinal and spatial analysis of short-term respiratory health effects of air pollution. (2004). Presented at the Eastern North America Region (ENAR) of the International Biometric Society Meeting, Pittsburgh, PA, March 28–31.
63. Rehman S. and Halawani T.O. (1994). Statistical characteristics of wind in Saudi Arabia. *Renewable Energy*, 4(8), pp. 949–956.
64. Ripley B. D. (1981). *Spatial Statistics*. Wiley, New York.
65. Romanowicz R, Young P., Brown P., Diggle P. A recursive estimation approach to the spatio-temporal analysis and modelling of air quality data. (2006). *Environmental Modelling & Software*, 21, pp. 759–769.
66. Sahin M. A. D. and Sen Z. (2004). A new spatial prediction model and its application to wind records. *Theoretical and Applied Climatology*, 79, pp. 45–54.
67. Šaltytė Benth J. and Benth F.E. (2010). Analysis and modelling of wind speed in New York. *Journal of Applied Statistics*, 6 (37), pp. 893–909.
68. Šaltytė Benth J., Benth F.E. and Jalinskas P. (2007). A spatial-temporal model for temperature with seasonal variance. *Journal of Applied Statistics*, 34(7), pp. 823–841.
69. Šaltytė Benth J., Dučinskas K. (2005). Linear Discriminant Analysis of Multivariate Spatial-Temporal Regressions. *Scandinavian Journal of Statistics*, 32, pp. 281–294.
70. Šaltytė Benth J., Šaltytė L. (2011). Spatial-temporal model for wind speed in Lithuania. *Journal of Applied Statistics*, 41(7), pp. 531–563.
71. Shumway R. H., Stoffer D. S. (2006). *Time Series Analysis and Its Applications With R Examples*. Wiley, New York.
72. Torres J.L., Garcia A., De Blas M. and De Francisco A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, 79, pp. 65–77.
73. Wikle CH. K., Milliff R. F., Nychka D., Berliner L. M. (2001). Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds. *Journal of the American Statistical Association*, 96 (454), pp. 382–397.

74. Yan Z., Bate S., Chandler R.E., Isham V. and Wheeler H. (2002). An analysis of daily maximum wind speed in Northwestern Europe using generalized linear models. *Journal of Climate*, 15, pp. 2073–2088.
75. Yao Q., Brockwell P. J. (2006). Gaussian Maximum Likelihood Estimation for ARMA Models II: Spatial Processes. *Bernoulli*, 12(3), pp. 403–429.
76. Zhang H. (2002). On Estimation and Prediction for Spatial Generalized Linear Mixed Models. *Biometrics*, 58, pp. 129–136.
77. Zhang H. (2004). Inconsistent Estimation and Asymptotically Equal Interpolation in Model-Based Geostatistics. *Journal of the American Statistical Association*, 99 (465), pp. 250–261.
78. Zhang Z, Switzer P. (2007). Stochastic space–time regional rainfall modeling adapted to historical rain gauge data. *Water Resources Research*; 43:W03441.
79. Zhou M., Boungiorno J. (2006). Space–time Modeling of Timber Prices. *Journal of Agricultural and Resource Economics*, 31(1), pp. 40–56.

Publikacijų sąrašas

Šaltytė L, Dučinskas K. Spatial time series modeling with system R. *Lietuvos Matematikos Rinkinys*, T. 44, Spec. nr., 770-773, 2004.

L. Šaltytė. Modified STARIMA model for space–time data, *Lietuvos Matematikos Rinkinys*, T. 45, Spec. nr., 375-379, 2005.

Šaltytė L., Dučinskas K., Modified STARIMA model. Abstracts of Communications of 9th international Vilnius conference, June 25-30, 2006, p. 283.

Šaltytė L.. Adequacy testing of two methods for modeling spatial time series. The International Conference “Applied Stochastic Models and Data Analysis (ASMDA)” , Vilnius, Lietuva, 2009. (ISI proceedings)

Šaltytė-Benth J., Šaltytė L. Spatial-temporal model for wind speed in Lithuania. *Journal of Applied Statistics*. Vol 38. No 6, 2011.

Paulionienė L. Spatial ARMA model for wind speed data. *Lietuvos Matematikos Rinkinys*, Vol. 53, Spec. nr., 102–107, 2012.

Priedas: Erdvinių laiko eilučių modeliavimo algoritmas

