

KAUNO TECHNOLOGIJOS UNIVERSITETAS

KRISTINA LUKOŠEVIČIŪTĖ

**CHAOTINIŲ PROCESŲ REKONSTRAVIMO  
BEI ALGEBRINIŲ SEKŲ MODELIAI  
LAIKO EILUČIŲ PROGNOZAVIME**

Daktaro disertacija  
Fiziniai mokslai, Informatika (09P)

2012, Kaunas

Disertacija rengta 2007 – 2011 metais Kauno technologijos universitete, Fundamentaliųjų mokslų fakultete, Matematinės sistemotyros katedroje, remiant Lietuvos valstybiniam mokslo ir studijų fondui.

**Mokslinis vadovas:**

Prof. habil. dr. **Minvydas Kazys Ragulskis** (Kauno technologijos universitetas, fiziniai mokslai, informatika – 09P).

# TURINYS

Ženklių, simbolių ir santrumpų sąrašas.....	5
<b>IVADAS.....</b>	<b>9</b>
<b>1. LITERATŪROS APŽVALGA.....</b>	<b>12</b>
1.1. Laiko eilučių modeliai .....	12
1.2. Laiko eilučių rekonstravimas .....	26
1.3. Hankelio matricos rango apžvalga .....	29
1.3.1. Hankelio matricos taikymų apžvalga .....	29
1.3.2. Hankelio matricos konstravimas algebrinėms sekoms .....	30
1.4. Programinis skaičiavimas kombinatorinio optimizavimo uždaviniuose.....	33
1.4.1. Evoliuciniai algoritmai .....	33
1.4.2. Dalelių spiečiaus algoritmas .....	34
1.4.3. Modeliuojamojo atkaitinimo algoritmas .....	35
<b>2. LAIKO EILUČIŲ PROGNOZAVIMAS REMIANTIS CHAOTINIŲ PROCESŲ SAVYBĖMIS .....</b>	<b>39</b>
2.1. Kriterijaus, nusakančio nereguliaraus vėlinimo optimalumą, konstravimas..39	
2.1.1. Rekonstrukcija į dvimatę vėlinimo koordinačių erdvę .....	40
2.1.2. Rekonstravimo į laiko vėlinimų erdvę kokybės parametro sudarymas .	41
2.1.3. Daugiamatė vėlinimo koordinačių erdvė, reguliaraus rekonstravimo į laiko vėlinimų erdvę atvejis .....	42
2.1.4. Daugiamatė vėlinimo koordinačių erdvė, nereguliaraus rekonstravimo į laiko vėlinimų erdvę atvejis. Tikslų funkcijos sudarymas. ....	45
2.1.5. Tikslų funkcijos savybės.....	50
2.2. Prognozavimo strategija, eksperimentai .....	52
2.2.1. Prognozės modelio taikymai Mackey-Glass laiko eilutės prognozei ....	54
2.2.1.1. Mackey-Glass laiko eilutės prognozė pagal MATLAB neraiškių logikų įrankių pavyzdį.....	54
2.2.1.2. Mackey-Glass laiko eilutės prognozavimas, pagrįstas laiko eilutės reguliariu rekonstravimu .....	55
2.2.1.3. Mackey-Glass laiko eilutės prognozavimas, pagrįstas laiko eilutės reguliariu rekonstravimu, esant optimaliam laiko vėlinimų erdvės matavimui57	
2.2.1.4. Mackey-Glass laiko eilutės prognozavimas, pagrįstas laiko eilutės nereguliariu rekonstravimu .....	59
2.2.2. Prognozės modelio taikymai Rössler atraktoriaus prognozei .....	60
2.2.2.1. Rössler laiko eilutės prognozavimas, pagrįstas laiko eilutės reguliariu rekonstravimu .....	61
2.2.2.2. Rössler laiko eilutės prognozavimas, pagrįstas laiko eilutės nereguliariu rekonstravimu .....	62
2.2.3 Prognozės modelio taikymai realaus pasaulio laiko eilutės prognozei... 63	
2.3. Evoliucinio algoritmo panaudojimas laiko eilutės rekonstravimo uždavinyje. ....	65
2.3.1. Evoliucinių algoritmų struktūra .....	65
2.3.2. Evoliucinių algoritmų parametrų parinkimas .....	68

2.3.3. Evoliucinių algoritmų rezultatų palyginimas su dalelių spiečiaus ir modeliuojamojo atkaitinimo modelio paieškos algoritmų rezultatais .....	71
2.3.4. Evoliucinių algoritmų pritaikymo laiko eilučių prognozei pavyzdžiai ..	73
2.3.4.1. <i>Mackey-Glass laiko eilutė</i> .....	73
2.3.4.2. <i>Stebimos kintamos žvaigždės kasdienio ryškumo laiko eilutė</i> .....	74
2.3.4.3. <i>Elektrokardiogramos (EKG) signalo laiko eilutė</i> .....	76
2.3.4.4. <i>S&amp;P (Standard &amp; Poor's) 500 akcijų svertinės vertės mėnesinės grąžos laiko eilutė</i> .....	78
2.3.4.5. <i>Rössler laiko eilutė</i> .....	80
2.4 Skyriaus išvados.....	82
<b>3. TRUMPŲ LAIKO EILUČIŲ PROGNOZAVIMAS.....</b>	<b>83</b>
3.1. Prognozės strategija .....	83
3.2. Skaitiniai eksperimentai su testine laiko eilute .....	85
3.3. Skaitiniai eksperimentai su PSO .....	86
3.4. Prognozavimo gerinimas ištrinant labiausiai nukrypusias reikšmes .....	90
3.5. Evoliucinių algoritmų taikymo palyginimas su PSO .....	92
3.6. Evoliucinių algoritmų parametrų derinimas.....	93
3.7. Skaitiniai eksperimentai su trumpomis laiko eilutėmis.....	95
3.7.1. Testinė laiko eilutė su pridėtu tolygiuoju triukšmu .....	95
3.7.2. Testinė laiko eilutė su pridėtu Gauso triukšmu.....	98
3.7.3. Andersono sausumos termometro temperatūros laiko eilutė .....	100
3.7.4. Australijos bazinės mėnesinės geležies gamybos laiko eilutė nuo 1956 m. sausio iki 1995 m. rugpjūčio.....	101
3.8. Adaptyvi rango nustatymo strategija .....	104
3.9. Australijos bazinės mėnesinės geležies gamybos laiko eilutės prognozė adaptyvaus rango parinkimo modeliu .....	107
3.10. Skyriaus išvados.....	108
<b>IŠVADOS.....</b>	<b>110</b>
<b>MOKSLINIŲ PUBLIKACIJŲ DISERTACIJOS TEMA SĄRAŠAS .....</b>	<b>119</b>
<b>PRIEDAS .....</b>	<b>121</b>

## Ženklų, simbolių ir santrumpų sąrašas

$\alpha$  – eksponentinio glodinimo faktorius

$\alpha_r$  – reikšmingumo lygmuo

$\beta$  – evoliucinių algoritmų panašumo parametras

$\beta_t$  – būsenų vektorius

$\gamma$  – trendo glodinimo faktorius

$\delta$  – intervalas tarp gretimų skaliarinės laiko eilutės taškų

$\Delta E$  – perėjimo iš energijos  $E_1$  į energiją  $E_2$  įvertis

$\overline{\Delta E}$  – teigiamų perėjimų energijos įvertis

$\Delta^m S_j(\rho)$  – algebrinės sekos charakteringasis determinantas

$\Delta x_t$  – laiko eilutės pokyčio greitis

$\varepsilon_i$  – lokalinės algebrinės sekos korekcijos,  $i = 1, 2, \dots, 2n$

$\theta_1, \theta_2, \dots, \theta_q$  – slenkančio vidurkio parametrai ARMA modelyje

$\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$  – įvertinti slenkančio vidurkio modelio parametrai.

$\theta(B)$  – slenkančio vidurkio  $q$ -tojo laipsnio daugianaris

$\Theta(x)$  – Heavisaido funkcija

$\Theta(B)$  – MA  $l \times l$  formato parametrų matricių  $\Theta_i$  daugianaris

$\kappa$  – evoliucinių algoritmų kryžminimo koeficientas

$\kappa_\alpha, \kappa_\beta, \lambda_T$  – temperatūros atšaldymo formulio koeficientai

$\lambda_i$  – lokaliųjų korekcijų tolerancijos koridorius, kai  $i = 1, 2, \dots, 2n$

$\mu$  – evoliucinių algoritmų mutacijos koeficientas

$\mu_{kl}$  – algebrinės progresijos tiesinių lygčių sistemos koeficientai

$\xi_k$  – atsitiktiniai skaičiai,  $\xi_k \in Z_0$

$\rho_k$  – algebrinės sekos charakteringosios lygties šaknys,  $k = 1, 2, \dots, r$

$\sigma_e$  – prognozės paklaidų standartinis nuokrypis

$\sigma_x$  – eilutės reikšmių standartinis nuokrypis

$\tau$  – laiko vėlinimas,  $\tau \in \mathbf{N}$

$\tau_i$  – nereguliaraus rekonstravimo laiko vėlinimas,  $\tau_i \in \mathbf{N}$ ,  $1 < i < d-1$

$\tau_{j,L}^{(k)}$  –  $k$ -tosios evoliucinių algoritmų generacijos  $j$ -sis poros kairiosios chromosomos tėvo genas

$\tau_{j,R}^{(k)}$  –  $k$ -tosios evoliucinių algoritmų generacijos  $j$ -sis poros dešinėsios chromosomos tėvo genas

$\varphi$  – harmoninio signalo fazė,

$\varphi_1, \varphi_2, \dots, \varphi_p$  – autoregresijos modelio parametrai

$\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_p$  – įvertinti autoregresijos modelio parametrai

$\varphi(B)$  – autoregresijos  $p$ -tojo laipsnio daugianaris

$\Phi(B)$  – AR  $l \times l$  formato parametrų matricių  $\Phi_i$  daugianaris

$\chi_0$  – atsitiktinis skaičius iš intervalo  $[0, 1]$   
 $\omega$  – harmoninio signalo ciklinis dažnis  
 $\nabla$  – integruotoji ARIMA dalis  
 $A$  – harmoninio signalo amplitudė  
 $a_1, a_2$  – tiesinio trendo modelio nežinomi parametrai  
 $a_e$  – proporcijos tarp Hankelio matricos determinanto ir lokaliųjų korekcijų koeficientas  
 $A(\omega)$  – laiko eilutės Furjė amplitudinis spektras  
ACF – autokoreliacijos funkcija  
 $AR(p)$  – autoregresijos modelis  
 $ARIMA(p, h, q)$  – autoregresinis integruotas slenkančio vidurkio modelis su laisvės laipsniais  $p, h$  ir  $q$   
 $ARMA(p, q)$  – autoregresinis slenkančio vidurkio modelis su laisvės laipsniais  $p$  ir  $q$   
 $B$  – atgalinio poslinkio operatorius  
 $b_\lambda$  – tolerancijos koridoriaus įvertinimo koeficientas  
 $b_{i,j}$  – neuroninio tinklo ryšio funkcijos nuokrypiai  
 $b_t$  – trendo įvertis laiko momentu  $t$   
 $c_1, c_2$  – dalelių pagreičio konstantos  
 $C_B$  – Bolcmano konstanta  
 $C^{(d)}(R)$  – koreliacinė suma  
 $d$  – rekonstruotos vėlinimų erdvės matavimas  
 $D_c(d)$  – sistemos koreliacijos matavimas  
 $d_s$  – sistemos būsenų erdvės matavimas  
 $d_j^{(m)}$  – Hankelio matricos minoro determinantas  
 $D_t$  – krypties keitimo nustatymo funkcija  
EA – evoliuciniai algoritmai  
 $e_t$  – laiko eilutės triukšmo narys, pasiskirstęs pagal Gauso skirstinį su vidurkiu, lygiu nuliui, ir standartiniu nuokrypiu, lygiu  $\sigma$   
 $\hat{e}_t$  – nustatytas laiko eilutės triukšmo narys laiko momentu  $t$   
 $\mathbf{e}_t, \tilde{\mathbf{e}}_t$  – atitinkamai  $l \times 1$  ir  $k \times 1$  formato baltojo triukšmo vektoriai  
 $f_k$  – neraiškos logikos tinklo  $k$ -osios taisyklės išėjimo parametras  
 $F(\tau_1, \tau_2, \dots, \tau_{d-1})$  – nereguliais laiko vėlinimais rekonstruoto į fazinę erdvę atraktoriaus tikslo funkcija  
 $F_1(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n})$  – lokaliųjų korekcijų minimizavimo tikslo funkcija  
 $F_{\max}$  – rekonstruoto atraktoriaus tikslo funkcijos geriausia reikšmė  
 $g$  – dalelių spiečiaus geriausias indeksas  
 $h$  – integruotosios ARIMA dalies laipsnis  
 $H$  – sekos Hankelio matrica  
 $H_j^{(m)}$  – Hankelio matricos minoras, kurio matavimas  $m, j = 0, 1, 2, \dots$

$\mathbf{H}_t$  – būsenų erdvės žinomų koeficientų  $l \times k$  matrica

$\tilde{H}_j^{(m)}$  – pakoreguotos algebrinės sekos Hankelio matricos minoras, kurio matavimas  $m$ ,  $j = 0, 1, 2, \dots$

$HrS_j$  – posekio  $S_j$  rangas

$y_p(d)$  – rekonstruotos laiko eilutės būsena laiko vėlinimų erdvėje diskrečiais laiko

momentais  $p \in \mathbf{N}$ ,  $1 \leq p \leq N - (d - 1) \tau$

$y'_t$  – neraiškos logikos neuroninio tinklo įėjimų vektorius

$\mathbf{Y}_t$  – išmatuojamų kintamųjų  $l \times 1$  formato vektorius

$k$  – konstanta apsprendžianti slenkančio vidurkio glodinimo gylį

$K$  – suvaržymo faktorius

$k_d$  – proporcijos koeficientas siejantis sistemos koreliacijos išmatavimą su koreliacine suma

$k_g$  – sėkmingų optimizavimo algoritmo bandymų skaičius

$l$  – vėlinimų periodų sklaičius

$L$  – neraiškos logikos sistemos taisyklių skaičius

$m$  – laiko eilutės vidurkis

$\tilde{m}$  – populiacijos dalelių skaičiaus dydis

MA – slenkantis vidurkis

MAPE – vidutinė absoliučioji procentinė paklaida

ME – prognozavimo paklaidų vidurkis

MSE – vidutinė kvadratinė paklaida

$N$  – laiko eilutės ilgis

$n_k$  – šaknų  $\rho_k$  pasikartojimų skaičius

$n_g$  – evoliucinių algoritmų generacijų skaičius

$N_p$  – chromosomų skaičius populiacijoje

$OX_i$  – vėlinimų koordinačių erdvės  $i$ -oji ašis

$\mathbf{P}_i$  – dalelių spiečiaus geriausia pozicija

$p_{i,j}$  – dalelių populiacijos elementai

$P_E$  – modeliuojamojo atkaitinimo sprendinio perėjimo tikimybė

PACF – dalinė autokoreliacijos funkcija

POCID – prognozavimo krypties keitimo metrika

PSO – dalelių spiečiaus optimizavimas

$Q_i$  – reguliariai rekonstruoto atraktoriaus laiko vėlinimų plokštumoje kokybės parametras,  $i = 1, 2, \dots, d$

$Q_{ij}$  – rekonstruoto atraktoriaus kokybės parametro projekcija  $X_i OX_j$  plokštumoje,  $i = 1, 2, \dots, d$ ,  $j = 1, 2, \dots, d$

$Q(\tau_1, \tau_2, \dots, \tau_{d-1}, \omega)$  – nereguliais laiko vėlinimais rekonstruoto atraktoriaus kokybės parametras vėlinimų erdvėje

$\hat{Q}$  – Ljung-Box statistika

$R$  – taškų porų  $y_i(d)$  ir  $y_j(d)$  didžiausias atstumas, kai  $i \neq j$   
 $r_1, r_2$  – tolygiai pasiskirstę atsitiktiniai skaičiai iš intervalo  $[0, 1]$   
 $\hat{r}_i$  –  $i$ -ojo vėlinimo autokoreliacijos įvertis  
 $R_1, R_2$  – elipsės pusašės  
 RMSE – vidutinės kvadratinės paklaidos kvadratinė šaknis  
 $s$  – modeliuojamojo atkaitinimo modelio sprendinys  
 $S$  – algebrinė seka  
 $S_j$  – algebrinės sekos posekis, kai  $j = 0, 1, 2, \dots$   
 $S_E$  – elipsės plotas vėlinimų koordinačių plokštumoje  
 $S_t$  – sugludinta laiko eilutės reikšmė laiko momentu  $t$   
 SA – modeliuojamojo atkaitinimo algoritmas  
 SES – viengubas eksponentinis glodinimo  
 SS – būvio būsenų modelis  
 $t$  – laiko momentas  
 $T$  – laiko vėlinimų didžiausia reikšmė  
 $T_e$  – atkaitinimo temperatūra  
 $T_0, T_f$  – atitinkamai pradinė ir galutinė atkaitinimo temperatūra  
 THEIL – Theil U statistika  
 $\mathbf{V}_i$  –  $i$ -tosios dalelės greitis  
 $v_{i,j}$  –  $i$ -tosios dalelės greičio elementas  
 VAR – vektorinis AR modelis  
 VARMA – vektorinis ARMA modelis  
 $w$  – inercijos svoris  
 $w_{i,j}$  – neuroninio tinklo ryšio funkcijos mazgams suteikiami svoriai  
 $w_k(\cdot)$  – neraiškios logikos sistemos  $k$ -osios taisyklės svoris  
 $w_{ik}(\cdot)$  – neraiškios logikos sistemos  $k$ -osios taisyklės  $i$ -ojo elemento sąryšio funkcija  
 $x_t$  – laiko eilutės reikšmė laiko momentu  $t$   
 $\hat{x}_t$  – laiko eilutės prognozė laiko momentu  $t$   
 $x_1, x_2, \dots, x_N$  – diskreti laiko eilutė  
 $\mathbf{X}$  – rekonstruotos dinaminės sistemos trajektorija  
 $\tilde{x}_i$  – algebrinės sekos nariai,  $i = 1, 2, \dots, 2n$   
 $\tilde{\mathbf{X}}_i$  – dalelių spiečiaus  $i$ -oji dalelė



### Temos aktualumas

Laiko eilučių prognozavimas, ypač ilgalaikis prognozavimas, yra nepaprastai svarbus uždavinys daugelyje mokslo ir inžinerijos sričių. Apibendrinant galima teigti, kad visų prognozės modelių pagrindinis tikslas yra tiksliau ar paprasčiau rekonstruoti laiko eilutę generuojančio proceso modelį iš praeities duomenų, ir tuomet jį panaudoti paskutinių laiko eilutės reikšmių ekstrapoliavimui į ateitį. Nors geriausio laiko eilučių prognozės modelio paieška niekada nesustoja, yra visuotinai sutinkama, kad niekada nebus sukurtas toks modelis, kuris bus geriausias lyginant su visais kitais modeliais visose situacijose.

Laiko eilučių prognozės uždavinius sąlyginai galima skirstyti į ilgų ir trumpų laiko eilučių prognozę. Ilgų laiko eilučių atveju modelio rekonstrukcijai yra prieinama ilga praeities duomenų istorija, tačiau paprastai tikimasi ir ilgų ateities horizontų, t.y. tikimasi atlikti prognozę gana toli į ateitį. Trumpų laiko eilučių atveju prieinama tik trumpa praeities duomenų istorija; tuo pačiu paprastai apsiribojama trumpu ateities horizontu (kai kurių uždavinių atveju pakanka prognozuoti tik vieną žingsnį į priekį). Natūralu, kad metodai skirti ilgų ir trumpų laiko eilučių prognozei, skiriasi ir naudojamais sistemos modelio rekonstrukcijos principais ir prognozavimui naudojamais mechanizmais.

Neplečiant diskusijos apie ilgų laiko eilučių prognozavimo modelių įvairovę, galima paminėti, kad pastaraisiais metais gana gerus prognozės rezultatus demonstruoja modeliai pagrįsti neraiškos logikos elementais bei dirbtiniais neuroniniais tinklais. Prognozės modelių apmokymui naudojami informacijos šrantai paprastai konstruojami iš skaliarinės laiko eilutės ją vektorizuojant laiko vėlinimo transformacijos operacijos pagalba. Galima pastebėti, kad šioje vietoje (atraktoriaus rekonstravimui iš skaliarinės laiko eilutės į daugiamatę fazinę vėlinimų erdvę) būtų galima panaudoti netiesinių sistemų atraktorių savybes aprašančius modelius. Dar daugiau, darbe pasiūlius naują optimizavimo kriterijų rekonstruoto atraktoriaus savybėms įvertinti, pasirodė, kad įmanoma žymiai pagerinti šių laiko eilučių prognozavimo modelių kokybę. Taigi, pasirodo, kad netiesinių dinaminių sistemų teorijos vystymas gali būti labai naudingas laiko eilučių prognozavimo uždaviniuose.

Jei ilgų laiko eilučių prognozės srityje šioje disertacijoje yra naudojami standartiniai prognozavimo įrankiai (keisčiant tik įeinančios informacijos duomenis bei tų duomenų kanalų skaičių), tai trumpų laiko eilučių srityje darbe siūlomas naujas prognozės modelis, paremtas algebrinės ekstrapoliacijos principais. Čia rėmtasi algebrine Hankelio rango sąvoka bei programinio skaičiavimo elementais algebrinių sekų identifikavimui iš praeities duomenų. Toks priėjimas prie trumpų laiko eilučių prognozės uždavinio leido pažvelgti į sistemos modelio rekonstravimo problemą visai kitu aspektu. Neabejotina tai, kad toks pasiūlytas prognozės modelis yra labai naudingas kai nagrinėjamos laiko eilutės prieinamas duomenų kiekis yra ribotas, ir kai svarbu tirti ne tik pačios laiko eilutės evoliuciją, bet ir tikėtinų lokalių minimumų bei maksimumų reikšmes.

Pasiūlyti ilgų ir trumpų laiko eilučių prognozės modeliai demonstruoja geras savybes savo taikymų srityse, o daugeliu atveju gaunamos prognozės paklaidos yra mažesnės lyginant su geriausiais modeliais šiuo metu naudojamais pasaulyje.

Galima paminėti, kad standartinei Mackey-Glass eilutei disertacijoje gautos prognozės paklaidos buvo daugiau nei du kartus mažesnės lyginant su geriausiu rezultatu pasaulyje 2010 metais. Visa tai ir apsprendžia šio disertacinio darbo aktualumą tiek teorine, tiek ir taikomąja prasme.

**Tyrimų objektas** – ilgų ir trumpų laiko eilučių prognozės modeliai, pagrįsti chaotinių atraktorių nereguliariu rekonstravimu į laiko vėlinimų fazinę erdvę ir skeletinės algebrinės sekos bazinio fragmento identifikavimu.

**Darbo tikslas** – sudaryti ir ištirti matematinius modelius, identifikuojančius laiko eilučių dinamiką aprašančius dėsnius, ir pritaikyti šiuos dėsnius eilučių reikšmių ekstrapoliavimui į ateitį.

**Suformuluotiems tikslams pasiekti darbe yra sprendžiami tokie uždaviniai:**

- 1) atlikti laiko eilučių prognozės modelių, su kai kuriais jų bus atliekamas darbe sukonstruotų modelių palyginimas, apžvalgą.
- 2) ilgų laiko eilučių prognozei *sukonstruoti naują laiko eilučių rekonstravimo į nereguliarių laiko vėlinimų erdvę matematinį modelį*, kuris turi atspindėti rekonstruotos laiko eilutės topologines ypatybes vėlinimų erdvėje;
- 3) laiko eilutės nereguliaraus rekonstravimo į laiko vėlinimų erdvę rezultatus *pritaikyti laiko eilučių prognozavimui neraiškių logikų sistemų pagrindu*;
- 4) laiko eilutės nereguliaraus rekonstravimo atveju optimalaus laiko vėlinimų rinkiniui *rasti panaudoti metaeuristinius kombinatorinius paieškos metodus*;
- 5) trumpų laiko eilučių prognozei *pritaikyti Hankelio matricos  $H$ -rango bei skeletinės algebrinės sekos fragmento koncepcijas*;
- 6) prognozuojant gana aukšto triukšmo lygio trumpas laiko eilutes *pritaikyti adaptyvaus kintamo rango prognozavimo strategiją*.

**Tyrimų metodai ir programinės priemonės:**

- 1) tyrimuose naudoti matematinės, geometrinės ir statistinės analizės metodai, dinaminių sistemų teorija, algebrinės analizės teorija. Toliau plėtotą dinaminių sistemų teoriją ir analizuotas šių sistemų bei algebrinės analizės praktinis pritaikymas;
- 2) ilgų laiko eilučių prognozės atveju naudoti ir kurti netiesinių dinaminių sistemų rekonstravimo metodai, kuriems sukonstruoti pasitelkta Furjė analizė, geometrinės ir operatorinės analizės elementai. Naudotos skaičiavimo terpės *Matlab v.7.0*. įrankiuose pateikiamos neraiškių logikų tinklo konstravimo programinės priemonės (angl. *Fuzzy Logic Toolbox*);
- 3) pačios autorės sukurtos programinės priemonės. Kuriant metaeuristines programines priemones atsižvelgta į klasikinių algoritmų rekomendacijas.

**Darbo mokslinis naujumas ir praktinė svarba:**

- sudarytas *naujas* dinaminių sistemų rekonstravimo į laiko vėlinimų erdvę *modelis, pagrįstas atraktoriaus užimamo tūrio* daugiamatėje vėlinimų erdvėje *įvertiu*. Toks nereguliarus atraktoriaus rekonstravimas atspindi atraktoriaus topologines ypatybes, kurias siūlomu metodu įvertinti nebėra taip sudėtinga, kaip skaičiuojant Liapunovo konstantas ar koreliacines integralines sumas. Pasiūlytas nereguliaraus

rekonstravimo metodas susietas su neraiškos logikos programinio skaičiavimo algoritmais ir labai pagerina laiko eilučių prognozes rezultatus;

- sveikaskaitinio perrinkimo uždaviniui spręsti naudojamiems evoliuciniams algoritmams sukonstruotas *naujas kryžminimo būdas*, įvestas papildomas parametras – panašumo parametras, aprašantis tam tikrų genų dominavimą chromosomų kryžminimo metu. Darbe pademonstruota, kad net turint nedidelę pradinę populiaciją šis panašumo parametras padeda išvengti kraujomaišos tarp chromosomų ir leidžia gana greitai surasti sprendinį artimą optimaliam sprendiniui;
- trumpų laiko eilučių prognozėje *pateiktas matematinis modelis, nustatantis skeletinės sekos pseudorangą*. Sukonstruota *modelio tinkamumo funkcija* įvertina laiko eilutės triukšmą. Sukonstruotas trumpų laiko eilučių prognozės modelis, paremtas identifikuotos skeletinės algebrinės sekos evoliucija. Praktinė tokio modelio svarba – gebėjimas prognozuoti gana triukšmingas laiko eilutes.

#### **Gynimui pateikiama:**

- 1) nauja dinaminų sistemų rekonstravimo metodika, kuri gali įvertinti nereguliariai į laiko vėlinimų erdvę rekonstruoto atraktoriaus užimamą tūrį. Sukurtas modelis pritaikomas prognozuojant chaotines laiko eilutes;
- 2) naujos evoliucinių algoritmų modifikacijos, kurios užtikrina nereguliaraus laiko vėlinimo rinkinio efektyvų kombinatorinį optimizavimą;
- 3) naujas trumpų laiko eilučių prognozavimo modelis, paremtas Hankelio matricos H-rango koncepcija ir skeletinių sekų identifikacija.

#### **Darbo rezultatų aprobavimas:**

Darbo tema pateikta 7 moksliniai straipsniai, iš jų 3 mokslinės informacijos instituto (ISI) pagrindinio sąrašo leidiniuose su citavimo indeksais, kitos publikacijos pristatytos keturių tarptautinių konferencijų medžiagoje. Disertacijos rezultatai eksponuoti parodoje-konkurse „KTU Technorama 2011“, kur pristatytas darbas „Chaotiniai procesai ir laiko eilučių prognozė“. Straipsnis „Non-Uniform Attractor Embedding for Time Series Forecasting by Fuzzy Inference System“ buvo apžvelgtas JAV gynybos departamento leidžiamame elektroniniame savaitraštyje „VerticalNews“ sekcijoje „Robotic&MachineLearning“ (2010-07-26).

#### **Darbo apimtis ir struktūra:**

Daktaro disertaciją sudaro įvadas, 3 pagrindiniai skyriai, išvados, literatūros sąrašas ir publikacijų sąrašas. Disertacijos apimtis 126 puslapių. Disertacijos pagrindinėje dalyje, apimančioje 99 puslapius, yra 52 paveikslai, 157 šaltinių cituojamos literatūros aprašas.

# 1. LITERATŪROS APŽVALGA

Prieš apžvelgdami su disertacijos tema susijusią literatūrą apibrėžkime keletą pagrindinių sąvokų.

Laiko eilute šiame darbe vadinama skaitinių duomenų seka, aprašanti tam tikro proceso evoliuciją. Laiko eilučių prognozavimui modeliuoti bus naudojamos dinaminės sistemos, atraktoriaus, algebrinės sekos bei skeletinės sekos sąvokos:

Dinaminėmis sistemomis vadinamos sistemos, kurioms galima sukonstruoti modelį, leidžiantį apskaičiuoti būdingą sistemos elgseną pagal tos sistemos matematinį modelį, šio modelio parametrų reikšmės bei pradines sąlygas.

Atraktoriumi vadinama aibė taškų, į kurią artėja (asimptotiškai arba neasimptotiškai) dinaminės sistemos sprendinys baigiantis pereinamiesiems procesams.

Algebrinė seka vadinama realiųjų ar kompleksinių skaičių seka, kurios evoliuciją apibrėžia algebrinis iteracinis matematinis modelis, ir kuriai egzistuoja  $H$ -rangas.

Bazinis algebrinis sekos fragmentas – tai toks minimalus baigtinis paeiliui einančių algebrinės sekos elementų rinkinys, iš kurio galima vienareikšmiškai rekonstruoti sekos evoliuciją apibrėžiantį algebrinį iteracinį matematinį modelį.

Skeletinė seka – tai tokia algebrinė skaičių seka, kuriai gauti iš tiriamosios nealgebrinės skaičių sekos (tokios sekos  $H$ -rangas neegzistuoja) būtina pirmiausiai panaikinti pridėtinį triukšmą, apie kurį paprastai nieko nežinoma.

Reikia pažymėti, kad laiko eilutės, dinaminės sistemos bei atraktoriaus sąvokos yra klasikiniai terminai plačiai naudojami tiek netiesinės dinamikos, tiek laiko eilučių teorijos srityse. Tuo tarpu algebrinės sekos bei bazinio algebrinio fragmento sąvokos yra gana nauji terminai, kurių kilmę būtų galima sieti su Kauno technologijos universiteto mokslininkų darbais pastarųjų metų bėgyje. Visi šie terminai ir atitinkami darbai apžvelgiami šiame skyriuje.

## 1.1. Laiko eilučių modeliai

Laiko eilučių prognozavimas, ypač ilgalaikis prognozavimas, yra nepaprastai svarbus uždavinys daugelyje mokslo ir inžinerijos sričių. Egzistuoja daugybė įvairiausių metodų laiko eilučių prognozavimui. Apibendrinant galima teigti, kad visų šių metodų pagrindinis tikslas yra sukonstruoti laiko eilutę generuojančio proceso modelį ir tuomet panaudoti tą modelį paskutinių laiko eilutės reikšmių ekstrapoliavimui į ateitį. Prognozavimo procedūras grubiai galima klasifikuoti pagal skirtingus metodus ir modelius: tai ir slenkančio vidurkio algoritmai [1, 2, 3], atsitiktinio klaidžiojimo ir trendo modeliai [4, 5, 6, 7, 8, 9, 10], eksponentinis glodinimo [1, 8, 10, 11, 12], būvio būsenos modeliavimas [13, 14], vektorių autoregresiniai modeliai [13, 15], kointegruoti ir priežastiniai modeliai [13, 16, 17, 18], metodai, paremti neuroniniais [19, 20], neraiškos logikos tinklais [21, 22], duomenų gavybos ir taisyklėmis paremti metodai [23, 24]. Galima paminėti, kad aukščiau pateiktos nuorodos apžvelgia tik kelis tipinius metodus iš daugybės efektyvių laiko eilučių prognozės metodų. Nors geriausio laiko eilučių prognozės metodo paieška niekada nesustoja, yra visuotinai sutinkama, kad nebus tokio metodo, kuris bus geriausias iš visų metodų visose situacijose.

Vienas iš paprasčiausių laiko eilučių prognozavimo būdų yra pagrįstas dabartinio laiko eilutės stebėjimo priskyrimu ateities reikšmei, priimant tai, kaip geriausią prognozę ( $x_{t+1} = x_t$ ). Tokios rūšies modelis istoriškai žinomas kaip *atsitiktinio klaidžiojimo* (Random Walk) modelis [4, 5, 9], kurio schema apibrėžiama taip:

$$x_t = x_{t-1} + e_t; \quad (1.1)$$

arba

$$\Delta x_t = x_t - x_{t-1} = e_t; \quad (1.2)$$

čia  $x_t$  yra dabartinė laiko eilutės reikšmė,  $x_{t-1}$  laiko eilutės reikšmė, esanti prieš  $x_t$ ,  $e_t$  yra triukšmo narys, pasiskirstęs pagal Gauso skirstinį su vidurkiu, lygiu nuliui, ir standartiniu nuokrypiu, lygiu  $\sigma$ . Kitais žodžiais tariant, laiko eilutės pokyčio greitis ( $\Delta x_t$ ) yra baltasis triukšmas [6].

Šis modelis gali būti panaudotas bent jau kaip pirmoji aproksimacija daugeliui ekonominių ir finansinių laiko eilučių. Pavyzdžiu galima laikyti, kad tam tikrų akcijų kaina dabartiniu momentu yra lygi šių akcijų kainai praeitą prekybos dieną prie tos kainos pridėdam/atimam tam tikrą akcijų kainos pokytį [7].

Ką tik aprašytas modelis aiškiai leidžia suprasti, kad jei informacijos aibė susideda iš laiko eilutės praeities duomenų, ateities duomenys yra sunkiai prognozuojami. Bendru atveju,  $x_{t-1}$  reikšmė yra geriausia  $x_t$  reikšmės prognozė [6]. Kitas paprastas laiko eilučių prognozės modelis yra *tiesinio trendo modelis* [18]. Šį modelį aprašanti lygtis yra:

$$x_t = a_1 + a_2 t + e_t; \quad (1.3)$$

čia  $a_1$  ir  $a_2$  yra nežinomi parametrai, kurie turi būti įvertinti iš turimų laiko eilutės duomenų.

Jeigu parametrai  $a_1$  ir  $a_2$  bėgant laikui išlieka pastovūs, tokiu atveju turime globalaus trendo modelį [25]. Tuomet laiko eilutės reikšmėms prognozuoti gali būti naudojamas tiesinės regresijos modelis [7]. Tuo tarpu kai šie parametrai laikui bėgant keičiasi lėtai, modelis vadinamas lokaliajo trendo modeliu. Lokalūs trendas gali keistis laiko eigoje. Prognozuojama naujausia kryptis, kurią norima ekstrapoliuoti į ateitį [25]. Lokalaus trendo modeliui sukonstruoti dažniausiai naudojamas Holt-Winters dvigubas eksponentinis glodinimo [10]. Prognozuojant tiek globalaus trendo modeliu, tiek lokalaus trendo modeliu gaunami rezultatai dažniausiai būna ganėtinai prasti [26].

Klasikinis plačiai naudojamas laiko eilučių prognozės modelis yra paremtas paprastu *slenkančio vidurkio* (Moving Average) algoritmu [1]. Šiuo plačiai paplitusiu modeliu iš laiko eilutės duomenų  $x_i$ ,  $i = 0, 1, 2, \dots$  šalinami charakteringi atsitiktiniai pokyčiai [2]:

$$S_t = \frac{1}{k} \sum_{i=0}^{k-1} x_{t-i} = S_{t-1} + \frac{x_t - x_{t-k}}{k}; \quad (1.4)$$

čia  $S_t$  suglontinta reikšmė laiko momentu  $t$ , o  $k$  yra iš anksto nustatyta konstanta apsprendžianti glodinimo gylį. Šis algoritmas dažniausiai vadinamas glodinimo algoritmu. Aišku, algoritmo, naudojamo laiko eilučių prognozei efektyvumas priklauso nuo glodinimo gylio parametro  $k$ . Kuo daugiau laiko eilutės reikšmių yra naudojama aritmetinio vidurkio skaičiavimui, tuo labiau prarandama informacija apie lokalias prognozuojamo proceso variacijas ateityje. Tuo atveju, kai glodinimo gylio parametras  $k$  yra priimamas lygus vienetui, gaunamas paprastas vieno žingsnio kartojimo algoritmas [3]. Prognozavimo rezultatai, gaunami naudojant slenkančio vidurkio algoritmą, daugeliu atveju yra gana prasti, jei prognozavimo tikslas yra ne tiksliai nustatyti globalias prognozuojamo proceso tendencijas, bet ir įvertinti tikslesnį proceso vystymąsi ateityje.

Kitas gana paprastas ir pragmatiškas algoritmas skirtas gana nesudėtingų laiko eilučių prognozei yra vadinamasis *eksponentinis glodinimo* [10]. Šiuo atveju laiko eilutės prognozavimas konstruojamas iš eksponentiniu svoriu paveikto praeities duomenų vidurkio [11]. Viengubo eksponentinio glodinimo (Single Exponential Smoothing) metodas priskiria eksponentiškai didėjančius svorius  $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \alpha(1-\alpha)^3, \dots$  vis senesniems duomenims:

$$S_t = \alpha x_{t-1} + (1-\alpha)S_{t-1}; \quad (1.5)$$

čia  $t=1,2,3,\dots$ ,  $S_1 = x_0$ , o  $\alpha$  yra glodinimo faktorius,  $0 \leq \alpha \leq 1$  [1, 12]. Glodinimo faktoriaus  $\alpha$  parinkimas yra gana specifinė procedūra, bendru atveju šio faktoriaus skaitinė reikšmė priderinama prie konkrečios laiko eilutės skaitinių eksperimentų pasekoje [11]. Iš praktinių sumetimų laikoma, kad faktoriaus  $\alpha$  reikšmė yra pasirinkta tinkamai, jei prognozuojamo proceso vidutinė kvadratinė paklaida yra mažiausia lyginant su prognozės rezultatais, taikant kitas faktoriaus reikšmes [26]. Procesui evoliucionuojant optimali faktoriaus  $\alpha$  reikšmė gali keistis.

Eksponentinio glodinimo algoritmas plačiai taikomas įvairių laiko eilučių prognozavimui pirmiausiai dėl savo paprastumo ir pritaikomumo įvairiuose automatizuotose prognozavimo sistemose [27, 28]. Šis eksponentinio glodinimo algoritmas buvo adaptuotas laiko eilučių su sezoniškumu ir trendu prognozavimui XX a. šeštajame dešimtmetyje (Braun (1959, 1963), Holt (1957), Winter (1960) darbai [12]). Kaip tipinį pavyzdį galima pateikti *dvigubo eksponentinio glodinimo algoritmą*, plačiai taikomą laiko eilučių prognozei, kai duomenys su trendu yra suglodinami ribojant vidutinį augimą kiekvieno periodo pabaigoje [29]:

$$S_t = \alpha x_{t-1} + (1-\alpha)(S_{t-1} + b_{t-1}), \quad 0 \leq \alpha \leq 1, \quad (1.6)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, \quad 0 \leq \gamma \leq 1; \quad (1.7)$$

Čia  $\gamma$  yra trendo glodinimo faktorius, kuris parenkamas kartu su glodinimo faktoriumi  $\alpha$ . Šioms konstantoms ieškoti dažniausiai naudojami netiesinio optimizavimo metodai [11].  $b_t$  yra trendo įvertis laiko momentu  $t$ . Modelio pradinės sąlygos gali būti priimanamos taip:

$$S_1 = x_0, \quad b_1 = x_1 - x_0. \quad (1.8)$$

Esant stipriai išreikštam sezoniškumo faktoriui, dažnai naudojamas trigubo eksponentinio glodinimo modelis [28].

Laiko eilučių prognozavimo metodai, naudojantys sudėtingus, netiesinius ryšius, paprastai remiasi statistiniais regresiniais modeliais, ARIMA modeliais, Kalmano filtravimo modeliais [30, 31].

*Autoregresijos*  $AR(p)$  modelyje, prognozuojama laiko eilutės reikšmė išreiškiama kaip tiesinis darinys iš  $p$  ankstesnių reikšmių ir atsitiktinių triukšmų [32]:

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + e_t; \quad (1.9)$$

čia  $p$  yra  $AR(p)$  modelio laipsnis,  $x_t$  - laiko eilutės reikšmė laiko momentu  $t$ ,  $e_t$  - atsitiktinis triukšmas, pasiskirstęs pagal Gauso skirstinį su vidurkiu lygiu nuliui ir standartiniu nuokrypiu lygiu  $\sigma$ , o  $\varphi_1, \varphi_2, \dots, \varphi_p$  yra autoregresijos modelio parametrai [33].

Kai paprasčiausiame pirmos eilės  $AR(1)$  modelyje parametras  $\varphi$  lygus vienetui, gaunamas atsitiktinio klaidžiojimo modelis, o autoregresijos procesas laikomas nestacionariu. Procesas laikomas nestacionariu taipogi, jeigu parametras  $\varphi$  įgyja reikšmes  $|\varphi| > 1$ . Tuo tarpu, kai  $|\varphi| < 1$ , autoregresijos procesas laikomas stacionariu [7].

Jungiant autoregresijos modelį su slenkančio vidurkio modeliu gaunamas  $ARMA(p,q)$  modelis. Slenkančio vidurkio modelyje ieškoma laiko eilutės reikšmė išreiškiama tiesiniu dariniu iš ankstesnių  $q$  momentų bei balto triukšmo reikšmių:

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}; \quad (1.10)$$

čia  $q$  –  $MA(q)$  modelio laipsnis, o  $\theta_1, \theta_2, \dots, \theta_q$  yra slenkančio vidurkio modelio parametrai [32].

Įvedus atgalinio poslinkio (vėlinimo) operatorių  $B$  ir apjungiant (1.9) bei (1.10) lygtis,  $ARMA(p,q)$  modelį galima užrašyti tokiu pavidalu:

$$\varphi(B)x_t = \theta(B)e_t; \quad (1.11)$$

čia autoregresijos  $p$ -tojo laipsnio daugianaris  $\varphi(B) = 1 - \sum_{i=1}^p \varphi_i B^i$ , slenkančio vidurkio

$q$ -tojo laipsnio daugianaris  $\theta(B) = 1 - \sum_{j=1}^q \theta_j B^j$  [32, 33].

Laiko eilučių prognozavimui dažnai taikomas *autoregresinis integruotas slenkančio vidurkio (ARIMA) modelis* [34,35], kurį pirmą kartą pristatė Box ir Jenkins (1976). ARIMA modelis kilo iš autoregresijos (AR) modelio, slenkančio vidurkio (MA) modelio ir AR bei MA kombinuoto (ARMA) modelio [32].

$ARIMA(p,h,q)$  modelyje kintamojo ateities reikšmė gaunama kaip tiesinis darinys iš tam tikro skaičiaus praeities duomenų ir atsitiktinių triukšmų [33]. Tai yra procesas, kuris generuoja laiko eilutę su vidurkiu  $m$  ir turi pavidalą:

$$\varphi(B)\nabla^h(x_t - m) = \theta(B)e_t; \quad (1.12)$$

čia integruotoji dalis  $\nabla = (1-B)$ . Sveikieji skaičiai  $p$ ,  $q$  ir  $h$  yra modelio parametrai.

ARIMA modeliavimo procedūra, kurią suformulavo Box ir Jenkins (1976), susideda iš tokių iteracinių žingsnių [36]:

1 žingsnis. *Duomenų paruošimas*. Reikia atsižvelgti į tai, ar laiko eilutė yra stacionari [32], o jei ne, ar yra įmanoma atlikti atitinkamas duomenų transformacijas stabilizuojant laiko eilutės sklaidą.

2 žingsnis. *Modelio parinkimas*. Sistemos identifikavimo sąvoka siejama su principine galimybe nustatyti modelio optimalią struktūrą bei įverčius iš apriorinės ir eksperimentinės informacijos [38]. Tam tikro ARIMA modelio parinkimas paprastai yra nelengva užduotis [37]. Dažniausiai modelio parinkimui naudojamos atitinkamos diagramos, grafikai, statistikos, laiko eilutės autokoreliacijos funkcijos (ACF) arba dalinės autokoreliacijos funkcijos (PACF) [32].

3 žingsnis. *Parametrų įvertinimas*, t.y. tokių modelio koeficientų reikšmių radimas, su kuriomis yra įmanomas geriausias modelio apsimokymas. Nuo 1968 metų pasirodė labai daug publikacijų ARIMA modelio parametrų įvertinimo klausimu. Tai rodo problemos aktualumą ir tai, kad ji dar neišspręsta [36]. Vienas iš parametrų vertinimo metodų – didžiausio tikėtimumo metodas [38].

4 žingsnis. *Modelio tikrinimas*. Kai parametrai yra tinkamai parinkti, tikrinamas modelio adekvatumas, tiriant ACF ir PACF paklaidas [32]. Paklaidos turėtų imituoti baltąjį triukšmą. Jeigu nustatoma, kad modelis yra neadekvatus, grįžtame prie 2 žingsnio ir ieškome geresnio modelio [36]. Sudaryto modelio adekvatumo vertinimui naudojamos įvairūs statistiniai kriterijai. Literatūroje dažniausiai pasitaikantys vertinimo kriterijai yra: modelio paklaidų autokoreliacijos AC grafiko vertinimas, Ljung-Box testas (Q statistika), AIC ir Schwarz kriterijus ir kiti [36, 3940]. Šiame tyrime nustatytiems ARIMA modelių adekvatumo tyrimui naudojama Ljung-Box statistika:

$$\tilde{Q} = N(N+2) \sum_{i=1}^l \frac{\hat{r}_i^2}{N-i} \quad (1.13)$$

čia  $N$  – laiko eilutės stebėjimų skaičius,  $l$  – vėlavimoperiodų skaičius,  $\hat{r}_i$  –  $i$ -ojo vėlavimo autokoreliacijos įvertis, kuris randamas pagal formulę:

$$\hat{r}_k = \frac{\sum_{t=k+1}^N \hat{e}_t \hat{e}_{t-k}}{\sum_{t=1}^N \hat{e}_t^2} \quad (1.14)$$

čia  $\hat{e}_t$  nustatyta prognozės paklaida laiko momentu  $t$ :  $\hat{e}_t = x_t - \hat{x}_t$  ( $\hat{x}_t$  laiko eilutės prognozė laiko momentu  $t$ ).

$\tilde{Q}$  statistika naudojama testuoti reikšmingas paklaidų koreliacijas didėjant vėlinimui  $l$ . Jeigu apskaičiuota  $\tilde{Q}$  statistikos reikšmė yra mažesnė už teorinio  $\chi^2_{\alpha, l-p-q}$  skirstinio kritinę reikšmę (arba pagal  $\tilde{Q}$  statistiką nustatyta reikšmingumo tikimybė yra didesnė už pasirinktą reikšmingumo lygmenį), daroma



$\alpha_r$  reikšmingumo išvada, kad paklaidos neautokoreliuoja ir ARMA modelis sudarytas adekvačiai [40].

5 žingsnis. *Prognozavimas*. Kai modelis jau yra parinktas, įvertintas ir patikrintas, tuomet galima tiesiogiai skaičiuoti laiko eilutės prognozuojamas reikšmes.

Prognozės modelis identifikuotam ARMA modeliu yra išvedamas iš (1.11) išraiškos tokiu būdu [38]:

$$x_t = \frac{\theta(B)}{\varphi(B)} e_t.$$

Laikant, kad sudarytas modelis prognozuoja geriausiai, kai įvertinamas prognozės triukšmas lygus  $\hat{e}_{t+1} = x_{t+1} - \hat{x}_{t+1}$ . Tokiu atveju prognozė:

$$\hat{x}_{t+1} = \frac{\theta(B)}{\varphi(B)} \hat{e}_{t+1} - x_{t+1}.$$

Iš čia

$$\varphi(B)\hat{x}_{t+1} = \theta(B)\hat{e}_{t+1} - \varphi(B)x_{t+1}.$$

Pertvarkius šį reiškinį, gaunamas prognozės modelis:

$$\hat{x}_{t+1} = (1 - \varphi(B))x_{t+1} - (1 - \theta(B))\hat{e}_{t+1};$$

arba

$$\hat{x}_{t+1} = \sum_{i=1}^p \hat{\varphi}_i B^i x_{t+1-i} - \sum_{j=1}^q \hat{\theta}_j B^j \hat{e}_{t+1-j};$$

čia  $\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_p$  ir  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$  atitinkamai įvertinti autoregresijos ir slenkančio vidurkio modelio parametrai. Į modelio prognozės išraišką įstatomos žinomos  $(x_t, \dots, x_{t-p+1})$  reikšmės ir pagal modelio išraišką įvertinamos  $\hat{e}_t, \dots, \hat{e}_{t-q+1}$  reikšmės.

Ekonomikos, medicinos, mokslo, inžineriniuose ir kai kuriuose kituose uždaviniuose galima susidurti su *daugiamačiais* laiko eilučių prognozavimo modeliais [13, 41, 42, 43]. Kartais kintamųjų grupė, kurią tyrėjas prognozuoja, gali priklausyti tokioms grupėms, kurių stebėjimai matuojami tuo pačiu laiku. Tarkime, kad valstybinė turizmo organizacija nagrinėja tarptautinio turizmo paklausos laiko eilutę iš keleto generuojančiųjų šalių. Tokia laiko eilutė gali būti paveikta tam tikrų globalių jėgų (šalies ekonominės ir politinės padėties, socialinių reiškinų ir pan.). Tokiu atveju tam tikros situacijos prognozavimo tikslumas gali būti pagerintas tiriant „lygiagrečių“ laiko eilučių kovariaciją [44]. Dažniausiai daugialypiai modeliai konstruojami tam tikro kintamojo priklausomybe nuo kito atskiro kintamojo. Prognozės uždaviniuose daugiamačiai modeliai ne visada būna geresni nei vienmačiai. Tačiau daugiamačiai patogūs modeliuojant abipuses priklausomybes ir geriau tinka duotai imčiai, bet dažnai pasitaiko, kad vienmačiai modeliai geriau prognozuoja nei daugialypiai metodai už imties ribų. Tai gali būti dėl kelių priežasčių. Viena jų – daugiamačiai modeliai naudoja daugiau parametrų nei vienmačiai. Todėl jų radimas padidina prognozavimo paklaidas. Daugiamačiai modeliai yra sudėtingesni ir labai jautrūs paklaidoms [13].

Modeliuojant daugiamatį laiko eilutės prognozavimo modelį svarbus aiškinamasis įrankis yra kros-koreliacija, kuri yra apibendrintas autokoreliacijos funkcijos daugiamatis atvejis. Pagrindinis įrankio tikslas rasti tiesinius dinaminis sąryšius laiko eilutės duomenyse, kurie buvo gauti iš stacionaraus proceso [45].

Daugiamačio prognozavimo atvejui gali būti priskiriami dviejų tipų vadinamieji vienos lygties modeliai: regresijos modeliai ir perdavimo funkcijos modeliai. Abu tipai yra modeliai be grįžtamojo ryšio, t. y. kai modelio prognozavimas konstruojamas taip, kad  $X$  tipo (aiškinamųjų) vieno ar daugiau kintamųjų prognozės generuojamos, kai  $Y$  tipo (priklausomi) kintamieji modeliuojami priklausomai nuo  $X$  kintamųjų. Šie modeliai yra naudingi prognozavime tik tuo atveju, jei aiškinamųjų kintamųjų ateities reikšmės yra žinomos, jeigu aiškinamųjų kintamųjų prognozavimas yra dalinai lengvas, arba jeigu nėra grįžtamojo ryšio iš priklausomų kintamųjų aiškinamiesiems kintamiesiems [45].

Dažnai taikymų uždaviniuose galima rasti daugiamačius prognozavimo modelius, tokius kaip *vektorinį ARMA modelį* (VARMA) arba vektorinį autoregresijos modelį (VAR) [46, 47, 48]. Tai apibendrinti vienmačių ARMA bei AR modelių daugiamačiai atvejai. VARMA proceso populiacijos charakteristikos pirmą kartą buvo išvestos Quenouille (1957), tačiau tokio modelio programinė realizacija tapo įmanoma tik nuo 1980 ir nuo 1990 metų [15]. VARMA modelis vietoje įprastų ARMA parametrų naudoja parametrų matricas bei daugiamatį baltojo triukšmo vektorius:

$$\Phi(B)Y_t = \Theta(B)e_t; \quad (1.15)$$

čia  $Y_t$  yra  $l \times 1$  kintamųjų vektorius;  $B$  žymimas vėlinimų operatorius;  $e_t$  yra  $l \times 1$  mato baltojo triukšmo vektorius, kurio  $E(e_t e'_{t-l}) = 0$  kai  $l \neq 0$ .  $\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p$  ir  $\Theta(B) = I + \Theta_1 B + \dots + \Theta_q B^q$  yra AR ir MA parametrų matricų  $\Phi_i$  ir  $\Theta_j$  (kurių išmatavimai  $l \times l$ ) daugianariai;  $i = 1, 2, \dots, p$ ,  $j = 1, 2, \dots, q$ . Sakoma, kad modelis yra stabilus, jei lygties  $\det(\Phi(z)) = 0$  visų šaknų absoliučiosios reikšmės yra didesnės už vienetą. Modelis yra atpažįstamas ir apverčiamas (invertible), jei lygties  $\det(\Theta(z)) = 0$  visų šaknų moduliai yra lygūs arba didesni už vienetą. Deja, šios sąlygos yra būtinos, bet ne pakankamos, todėl vektorinis ARMA modelis retai naudojamas praktiniuose taikymuose [45, 15].

Dėl didelio VARMA modelio parametrų skaičiaus iškyla keletas keblumų [46]:

- a) mažas įverčių statistinis tikslumas, taigi sumažėja galimybė teisingai interpretuoti prognozuojamą modelį;
- b) skaitinio optimizavimo procedūra turi polinkį į klaidas;
- c) ilgas skaičiavimo laikas.

Vektorinis autoregresijos modelis yra atskiras VARMA atvejis, kai  $\Theta(z) \equiv 1$ :

$$\Phi(B)x_t = e_t. \quad (1.16)$$

Tokiame modelyje parametrų skaičiaus nustatymui geru kriterijumi laikomas kros-koreliacijos funkcija [45].

Daugiamačiuose laiko eilučių prognozavimo modeliuose reikalaujama, kad jos būtų stacionarios [13]. Laiko eilutės komponentai gali būti integruoti  $I(h)$  (nestacionarūs), t.y. kai laiko eilutėje yra  $h > 0$  laipsnio integracija  $(1-B)^h x_t$ . Tada rekomenduojama naudoti kointegruotą modelį, t.y. VAR modelyje naudojamas

paklaidos tikslinimas  $\Delta x_t = x_t - x_{t-1}$ . Tuomet VAR modelis tampa vektoriniu paklaidos tikslinimo modeliu (Vector Error Correction Model) [16]. Jei modelio kintamieji nėra kointegruoti, jie pirmiausia turi būti  $h$  kartų diferencijuojami ir tada gali būti taikomas VAR modelis [45].

*Būvio būsenų (State Space – SS) modeliavimas* suteikia vieningą metodiką sprendžiančią daugybę problemų laiko eilučių analizėje, ir yra lankstus tikslinant parametrų struktūrą laiko eilučių procesuose [49]. Tokių algoritmų pagrindinis tikslas rasti neišmatuojamų kintamųjų įverčius, pagrįstus išmatuojamais kintamaisiais, per aibę lygčių, kurios vadinamos būsenų erdvės modeliu [14]. Toks modelis apibūdinamas lygtimis:

$$\mathbf{Y}_t = \mathbf{H}_t \boldsymbol{\beta}_t + \mathbf{e}_t, \quad (1.17)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \tilde{\mathbf{e}}_t. \quad (1.18)$$

Lygtis (1.17) vadinama stebėjimų lygtimi ir sieja  $l \times 1$  mato išmatuojamų kintamųjų vektorių  $\mathbf{Y}_t$  ir  $k \times 1$  formato neišmatuojamų kintamųjų vektorių  $\boldsymbol{\beta}_t$ , kurie vadinami būsenomis. Matrica  $\mathbf{H}_t$  ( $l \times k$  matavimo) yra žinomų koeficientų matrica, o  $\mathbf{e}_t$  ir  $\tilde{\mathbf{e}}_t$  yra atitinkamai balto triukšmo  $l \times 1$  ir  $k \times 1$  formato vektoriai, kurie vadinami matavimų paklaidomis, su kovariacijos matrica  $E(\mathbf{e}_t \mathbf{e}_t') = \Sigma_e$ . Be to, būsenų vektorius  $\boldsymbol{\beta}_t$  kinta laike pagal (1.18) lygtį, perėjimo arba būsenos lygtį. Čia trikdžiai  $\mathbf{e}_t$  ir  $\tilde{\mathbf{e}}_t$  laikomi nekoreliuotais, t.y.  $E(\mathbf{e}_t \tilde{\mathbf{e}}_s') = \mathbf{0}$  su visomis  $t$  ir  $s$  reikšmėmis [14, 50]. Yra laikoma, kad pradinis būsenų vektorius nekoreliuoja nė su viena triukšmo išraiška, kas reiškia, kad būsenų vektorius turi Markovo savybių. Bendru atveju būsenų erdvės pateikime nei stebėjimų vektorius, nei būsenų vektorius nelaikomas stacionariu. Daugybė laiko eilučių modelių naudoja būsenų erdvės modelį. Būsenų vektoriaus įverčiams rasti būvio būsenų modelis naudoja Kalmano filtrą (1960) [49]. Tai rekurentinis algoritmas, kuris skaičiuoja eilutės reikšmę laiko momentu  $t+1$  iš praeities duomenų laiko momentais  $t, t-1, \dots, 1$ . Kalmano filtras suteikia vienodą požiūrį į prognozavimą ir įvertinimą visų procesų, kurie gali būti gauti iš būsenų erdvės reprezentacijos.

Kita klasė modelių, skirta laiko eilučių prognozavimui, yra aiškinamieji modeliai. Juose mėginama surasti ryšį tarp ieškomojo kintamojo ir kai kurių aiškinamųjų kintamųjų, tokių kaip kaina, eksportas, laikas, kurie valdo ieškomą kintamąjį. Tokie modeliai dažnai vadinami *priežastiniais modeliais* (Causal Model) [18].

Šių modelių yra keletas. Paprastas tiesinės regresijos metodas: dominančio dydžio prognozė gaunama aiškinant jo kitimo priežastį kito nepriklausomo kintamojo kitimu. Netiesinės kintamųjų priklausomybės gali būti išreiškiamos taikant daugianarių, logistines, eksponentines funkcijas. Tačiau prognozuojant realaus pasaulio eilutes, tenka taikyti sudėtingesnius daugiamačius regresijos modelius (kai dominantis dydis prognozuojamas aiškinant jo kitimo priežastį daugiau nei vieno kintamojo dydžiais), tokius kaip daugiamatė logistinė regresija [17], tiesinis ne Gauso ne ciklinis (acyclic) priežastinis modelis [16] ir pan.

*Duomenų gavyba* (Data mining), dar žinoma kaip „žinių suradimas duomenų bazėse“, yra procesas, ieškantis reikšmingų pavyzdžių iš duomenų bazių. Be to,

duomenų gavyba taip pat yra taikymas, kuris leidžia priimti reikšmingus, teisingus sprendimus [51]. Tai tiriamasis ir sudėtingas procesas, kuriame atliekami daugiamačiai iteraciniai žingsniai [23]:

1 žingsnis. *Taikymo srities nustatymas*: ištirti ir suprasti taikymų sritį ir atitinkamas išankstines žinias.

2 žingsnis. *Tikslinio rinkinio parinkimas*: parinkti tinkamą rinkinį, arba sutelkti dėmesį į kintamųjų poaibį arba duomenų imtis, kur duomenys tiesiogiai susiję su analizuojama sritimi yra gauti iš duomenų bazės.

3 žingsnis. *Pirminis duomenų apdorojimas*: pašalinti triukšmą arba reaguoti į trūkstamas sritis, sumažinti nereikalingą dimensiją arba panaudoti naudingus transformacijos metodus. Šio žingsnio tikslas padidinti efektyvių kintamųjų skaičių palyginimui.

4 žingsnis. *Duomenų gavybos pagrindinis etapas*. Šio proceso metu taikomi dirbtinio intelekto metodai siekiant rasti reikšmingus arba trokštamus duomenų rinkinius patogioje pateikimo formoje. Dirbtinio intelekto modeliai gali būti tokie kaip asocijuotų taisyklių parinkimas, klasifikacijų medžiai, klasterių metodai.

5 žingsnis. *Žinių gavyba*: prieš tai vykusių žingsnių gautos informacijos pagrindu vizualizuoja gautus duomenų pavyzdžius, arba vaizduoja duomenis, priklausomus nuo gautų duomenų pavyzdžių. Be to, šis procesas taip pat patikrina arba panaikina bet kokią potencialų konfliktą, vykusį su buvusiomis pažintomis žiniomis.

6 žingsnis. *Žinių taikymas*: šiuo atveju rastos žinios panaudojamos tiesiogiai į taikymo sritį ar į kitas sritis ateities veiksams.

7 žingsnis. *Žinių vertinimas*: randame įdomiausius pavyzdžius, kurie reprezentuoja žiniomis paremtus duomenis dominančiu mato tikslumu.

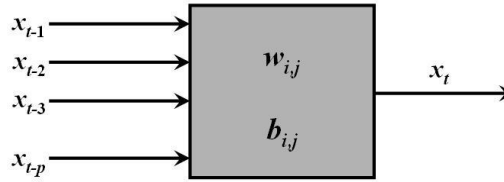
Duomenų gavybos modeliu gali būti vienas iš modelių: duomenų klasifikacija, regresija, klasterizavimas, sąryšio taisyklės, priklausomumo modeliavimas, sekų analizė ir t. t. [23].

Duomenų gavybos modeliai buvo sėkmingai pritaikyti finansinių laiko eilučių analizei, prognozavimui [52, 53], medicininiams duomenims [51].

*Dirbtiniai neuroniniai tinklai* (ANN) yra sėkmingai taikomi daugelyje sričių. Viena iš sėkmingų ANN panaudojimo sričių yra laiko eilučių prognozė ir modeliavimas, įskaitant finansinių [19], biomedicinių [20], chaotinių [54] bei kito tipo laiko eilučių prognozavimą. ANN modeliai dažniausiai geriau veikia nei standartiniai tiesiniai metodai (tokie kaip Box-Jenkins) [55], ypač kai eilutė yra triukšminga arba nagrinėjama dinaminė sistema yra netiesinė. Tokiais atvejais geresnis prognozavimo vykdymas gali būti paaiškinamas ANN modeliui būdingu netiesiškumu ir jo adaptyviu prisitaikymu prie sudėtingų užtriukšmintų duomenų. ANN modeliai imituoja biologinę nervų sistemą, kuri pasižymi saviorganizacijos funkcijomis, adaptyvaus duomenų valdymo prisitaikomumu ir asociatyvine atmintimi [56]. ANN gali apmokyti save iš duotos duomenų sekos ir automatiškai nustatyti paslėptas funkcines priklausomybes analizuojamuose duomenyse, net jei tos funkcinės priklausomybės nėra iš anksto žinomos ar jas sunku nustatyti [56].

Laiko eilučių modeliavimui ir prognozavimui plačiausiai naudojamas vieno paslėpto sluoksnio tiesioginio sklidimo (feed forward) neuroninis tinklas [56]. Dirbtinio neuroninio tinklo modelio schema laiko eilučių prognozavimui vaizduojama 1.1 paveiksle. Modeliui paduodamas laiko eilutės praeities duomenų

vektorius  $\{x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-p}\}$ , tuomet juodoji dėžė grąžina prognozuojamą reikšmę  $x_t$ . Juodojoje dėžėje naudojami ryšio tarp mazgų svoriai  $w_{i,j}$  ir nuokrypiai  $b_{i,j}$ , kurie tikslinami tinklo mokymo metu [57].



**1.1 pav.** Dirbtinio neuroninio tinklo pritaikymas laiko eilučių prognozėje.

ANN algoritmo optimaliai struktūrai parinkti yra žinoma daugybė metodų, tokių kaip pašalinimo algoritmas, laiko daugianarių algoritmas, kanoninės dekompozicijos metodas, tinklų informacijos kriterijus [57]. Po modelio struktūros parinkimo vykdomas neuroninio tinklo apmokymo procesas. Tinklo parametrai adaptiviai parenkami tokie, kad neuroninio tinklo paklaidų funkcija būtų minimizuota. ANN apmokymui paprastai naudojamas atgalinio sklidimo (backpropagation) apsimokymo algoritmas [57, 58].

Chaotinių laiko eilučių prognozavimui gali būti naudojami įvairūs neuroninių tinklų modeliai, tokie kaip daugiasluoksniai perceptroniniai neuroniniai tinklai (MLP) [59, 60], rekurentiniai neuroniniai tinklai (RNN) [61], netiesiniai autoregresiniai neuroninio tinklo modeliai su išoriniais įėjimais (NARX) [55].

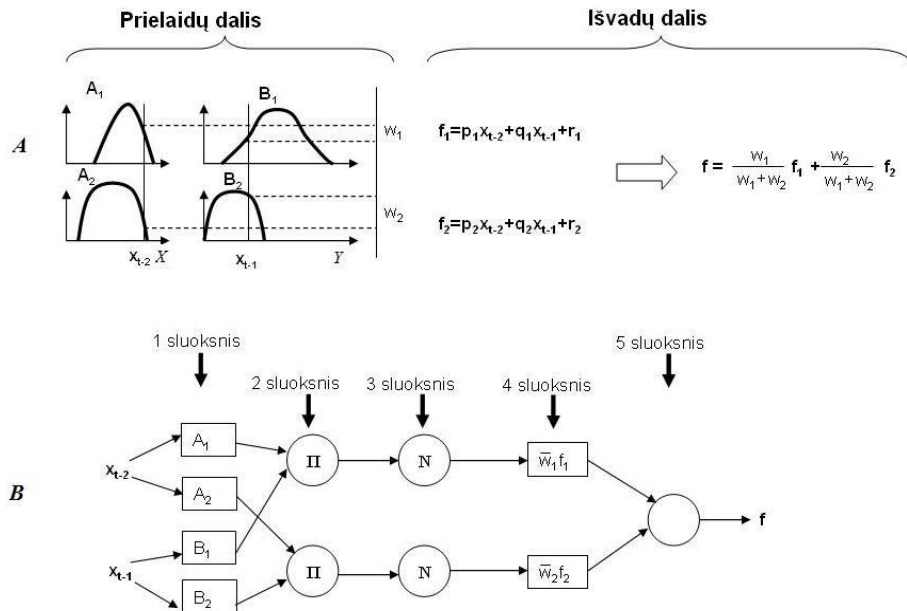
Laiko eilučių prognozavimui taip pat dažnai naudojamos *neraiškos logikos* ir *neuroninės neraiškos logikos sistemos* (fuzzy logic, neuro-fuzzy logic system). Daugybė neraiškos logikos modelių yra taikoma spręsti įvairių sričių uždavinius, tokius kaip finansinių laiko eilučių prognozavimas [21, 62, 63, 64, 65], temperatūros prognozė [22], transporto analizė ir prognozavimas [66] ir kitus. Siekiant tiksliau prognozuoti laiko eilutes, ar dėl kitų svarių priežasčių kai kurie statistiniai įrankiai, tokie kaip regresinė analizė, slenkančio vidurkio, integruoto slenkančio vidurkio, ARIMA ir daug kitų modelių, tokių kaip neuroniniai tinklai [67], buvo jungiami su neraiškių logikų sistemomis [68]. Kaip tipinį pavyzdį galima paminėti ANFIS architektūrą, kurioje panaudoti ANN ir neraiškių sprendimų sistemos (FIS) modeliai [69]. FIS optimalaus sprendinio paieškai naudoja tam tikrą pradinę patirtį ir sugeneruotas žinias. ANFIS naudoja daugiareikšmę loginę sistemą, vadinamą neraiškia logika, su kuria siekiama atsižvelgti į paslėptą netikslumą duomenyse ir atitinkamai tuos duomenis tiksliai atvaizduoti [70]. Tikslas pasiekiamas atliekant įėjimo duomenų fuzifikaciją per sąryšio funkcijas [24]. ANFIS gali gana greitai pritaikyti sąryšio funkcijų parametrus, ir priklausomai nuo įėjimo duomenų juos adaptiviai optimizuoti. Kad sistemą būtų galima apmokyti, joje nustatomos neraiškos kalbos taisyklės arba sąlyginiai (jei-tai) sakiniai [24].

Tarp visų FIS modelių, dažniausiai yra naudojamas neraiškių logikų Takagi-Sugeno (1985) neraiškos logikos modelis [71]. Neuroninio neraiškių sistemų modelio scheminė architektūra yra pavaizduota 1.2 paveiksle. Šiame pavyzdyje laikoma, kad FIS turi du įėjimo kintamuosius  $x_{t-2}$  ir  $x_{t-1}$  bei vieną išėjimo kintamąjį  $f$ . Neraiškių išvadų tipinė taisyklių aibė su dviem neraiškiomis jei-tai taisyklėmis išreiškiama tokiu būdu:

1 taisyklė: Jei  $x_{t-2}$  yra  $A_1$ , o  $x_{t-1}$  yra  $B_1$ , tai  $f_1 = p_1 x_{t-2} + q_1 x_{t-1} + r_1$ ;

2 taisyklė: Jei  $x_{t-2}$  yra  $A_2$ , o  $x_{t-1}$  yra  $B_2$ , tai  $f_2 = p_2 x_{t-2} + q_2 x_{t-1} + r_2$ ;

čia  $A_1$ ,  $A_2$  ir  $B_1$ ,  $B_2$  yra įėjimo duomenų  $x_{t-2}$  ir  $x_{t-1}$  sąryšio funkcijos. Atitinkamai  $p_1$ ,  $q_1$ ,  $r_1$  ir  $p_2$ ,  $q_2$ ,  $r_2$  yra išėjimo funkcijos parametrai. Takagi-Sugeno modelio neraiškaus argumentavimo mechanizmas iliustruojamas 1.2 paveikslu A dalyje. Atitinkamai ekvivalenti ANFIS architektūra vaizduojama 1.2 paveikslu B dalyje.



**1.2 pav.** A – Takagi-Sugeno jei-tai taisyklės ir neraiškių išvadų priėmimo mechanizmas; B – atitinkama ANFIS architektūra [68].

Neuroninis neraiškių sistemų modelis yra penkių sluoksnių neuroninis tinklas, kurio pirmasis yra įėjimų sluoksnis, t.y. praeities duomenų rinkinys. Pirmojo sluoksnio išėjimams yra priskiriamos sąryšio funkcijos antrajame sluoksnyje (vyksta duomenų fuzifikacija). Trečiasis sluoksnis yra taisyklių sluoksnis, t.y. aktyvuojamos tam tikras neraiškos logikos taisyklių kompleksas ir su tam tikru algebriniu operatoriumi priimamos išvados. Ketvirtasis sluoksnis yra normalizacijos sluoksnis, o paskutinis sluoksnis yra defuzifikacijos sluoksnis. Paskutiniojo sluoksnio išėjimas ir yra prognozuojama laiko eilutės reikšmė. Tokio modelio architektūros ir apmokymo bei naudojimo detales aprašymus galima rasti [69, 71, 72, 73, 74].

Galima tik paminėti, kad modelio konstravime optimalus tinklas gaunamas tuomet, kai apmokymo metu prognozės paklaidos yra mažiausios. Tad nustatant, kada sudarytas modelis veikia geriausiai, yra matuojamos tokios paklaidos metrikos kaip MSE, RMSE, MAPE ir kt. Skirtingai nei ARIMA modeliuose, čia dažnai nėra tikrinamas modelio adekvatumas, nes pats tinklas savaiminio apsimokymo procese naudoja struktūrizuotus duomenis ir ryšio funkcijos parametrai nustatomi minimizuojant modelio išėjimo paklaidos metrikas [33, 75, 77]. Kai kuriais atvejais modelio adekvatumo problema tampa sunkiai išsprendžiama – pvz. trumpose laiko

eilutėse apmokymo duomenų kiekis gali būti per mažas statistiškai pagrįstai hipotezei apie paklaidų pasiskirstymą pagal balto triukšmo dėsnį įrodymui. Tačiau dažnai tenka sutikti netgi ilgų laiko eilučių inovatyvius prognozės įrankius, kuriuose akcentas teikiamas ne modelio adekvatumui, o daromų paklaidų statistikiniam įverčiams. Pavyzdžiui, darbuose [33,37,7578] pateikiamos prognozės schemos demonstruoja geriausius standartinių sekų prognozės rezultatus pasaulyje MSE, RMSE, MAPE prasme, tačiau šių schemų generuojamos paklaidos nėra pasiskirsčiusios pagal balto triukšmo dėsnį. Tas ypač pasakytina apie laiko eilučių prognozės įrankius pagrįstus neuroninių ar fuzzy tinklų pagrindu. Kaip minėta prieš tai, šių tinklų savaiminis apsimokymas pagrįstas prognozės paklaidų minimizavimu, vertinant tinklo funkcionalumą praeities duomenų atžvilgiu. Tokio tinklo adaptacija savaiminiam apsimokymui ne daromų paklaidų minimizavimo prasme, o daromų paklaidų statistikinio skirtinio prasme būtų nekorektiška [75].

Laiko eilučių prognozei pirmasis ANFIS modelį pritaikė Jang [79]. Jis pristatė hibridinį atgalinio sklido ir mažiausių kvadratų apmokymo algoritmą, susidedantį iš dviejų mokymo etapų, vadinamų tiesioginiu sklidimu ir atgaliniu sklidimu. Tiesioginio sklido apmokyme yra fiksuojami tiesiniai parametrai, o tuo tarpu atgalinio sklido vykdymo etape netiesiniai parametrai yra atnaujinami atgalinio apmokymo algoritmu. ANFIS modelio efektyvumas buvo tikrinamas imituojant Mackey-Glass laiko eilutę. Lyginant su kitais neraiškos logikos architektūros modeliais, ANFIS gautos prognozavimo paklaidos buvo labai mažos. Prognozuojant chaotines laiko eilutes, ANFIS vidutinės kvadratinės paklaidos kvadratinė šaknis (Root Mean Square Error) buvo mažiausia [80]. Kim ir Kasabov savo darbe [81] pristatė hibridinį HyFIS modelį, kuriame ANFIS struktūra suderinta su dviejų fazių hibridine mokymo schema. Modelio pirmoje fazėje žinių kasybos modeliu ieškomos neraiškos logikos lingvistinės taisyklės ir pradinės neuroninio tinklo struktūros. Antroje fazėje neraiškos taisyklės derinamos gradientinio nusileidimo algoritmu. Šis dviejų fazių mokymo metodas apjungia ir skaitinę, ir lingvistinę informaciją. Tokiu būdu pagerinamas neuroninio tinklo skaidrumas ir neraiškos sistemos savarankiškas prisitaikymas ANFIS modelyje.

Tačiau kaip ir daugelis neraiškaus modeliavimo metodų, ANFIS adresuoja imties duomenis neatsižvelgdamas į koreliaciją tarp duomenų komponentų. Dėl to gaunamas neefektyvus įėjimo erdvės duomenų skaidymas ir neveiksmingas modeliavimas [82]. Šiai problemai spręsti autoriai pasiūlė TDANFIS modelį, kuris panaudoja Kharhunen Leave transformaciją, kad ANFIS įėjimo erdvės duomenys koreliuotų. Įėjimo duomenų pirminis apdorojimas parodė geresnį laiko eilutės prognozavimą lyginant su tradiciniais ANFIS etaloniniais uždaviniais. Kitas modelis, kuris rodė geresnius RMSE rezultatus ir sutaupė daugiau laiko nei tradicinis ANFIS modelis, buvo MANFIS (modified ANFIS) [83]. MANFIS architektūra pagrįsta FIS taisyklių skaičiaus derinimu su neraiškių aibių įėjimų skaičiumi [83, 84].

Laikinas ANFIS modelis, kuriame atsitiktinai sukonstruojamas neuroninis tinklas apmokomas laikinu atgalinio sklido algoritmu, buvo pristatytas [85]. Laikinas ANFIS modelis buvo sukurtas dubliuojant ANFIS struktūrą tam tikram

laiko intervalų skaičiui. Nors šis modelis, testuojant su daugeliu etaloninių laiko eilučių prognozavimu, rodė truputį didesnes paklaidas nei tradicinis ANFIS modelis, bet atpažinimo paklaidos šiuo atveju buvo mažesnės. Vienas iš šio modelio rūpesčių, tai rasti geriausią laiko eilutės laiko intervalų skaičių. Kitas mėginimas pagerinti ANFIS tikslumą ir tuo pačiu metu palengvinti skaičiavimo sudėtingumą, buvo Chen ir Zhang sukonstruotas En-ANFIS (an ensemble ANFIS) [86] modelis, kuriame integruojamas aibių mokymas (atsitiktinis ir plėtros atrinkimo metodas) daugialypėje individualioje ANFIS penkių sluoksnių struktūroje. Laivo šoninio supimo judesio laiko eilutės prognozavimui buvo pristatytas hibridinis modelis [87], kuriame apjungiami bangelių transformacijos (Wavelet Transform) ir ANFIS modeliai. Šiame modelyje prieš prognozuojant ANFIS tinklu laiko eilutės duomenys skaidomi į reguliaraus periodo signalus, pagrįstus bangų transformacija. Dar vienas modelio ANFIS pagerinimas laiko eilučių prognozei gautas įtraukiant į modelį savarankišką atnaujinimą ir mokymą [88].

Daugelis pristatytų ANFIS modelio pagerinimų laiko eilučių prognozei yra atlikti per mokymosi algoritmą, išankstinį įėjimo duomenų apdorojimą ir „jei – tai“ taisyklių pakeitimą [88]. Šiame tyrime ANFIS modelis bus naudojamas, kaip juoda dėžė, kurios įėjimai bus laiko eilutės vektoriai ne iš eilės einančių duomenų, o duomenų, tarp kurių bus tam tikras laiko vėlinimas. Tokiu būdu, prieš prognozuojant laiko eilutę, reikės atlikti skaliarinės duomenų eilutės rekonstrukciją į atitinkamo mato vektorių eilutę. Laiko eilutės rekonstravimas grindžiamas originalia metodika, kuri bus detalai aprašyta 1.2 skyrelyje.

Taigi, ką tik buvo apžvelgta keletas laiko eilučių modelių, kurie dažnai yra naudojami prognozės uždaviniuose. Šiame darbe bus pasiūlytas dar vienas laiko eilučių prognozės modelis, kuris remsis Hankelio matricos sąvoka. Plačiau apie šią sąvoką bus aptarta atskirame 1.3. skyrelyje, o kaip ji panaudojama laiko eilučių prognozavime, bus aprašyta visame 3 skyriuje.

Laiko eilučių trumpalaikis prognozavimas ir ilgalaikis modeliavimas nebūtinai yra identiški tikslai: pagrindinių lygčių radimas su timkamomis ilgalaikėmis savybėmis nebūtinai yra patikimiausias kelias aprašyti parametrus gero trumpalaikio prognozavimo, o modelis tinkantis trumpalaikiam prognozavimui gali turėti neteisingas ilgalaikio prognozavimo savybes [89]. Todėl šiame darbe yra išskiriami ilgalaikis ir trumpalaikis laiko eilučių prognozavimas.

Prognozavimo kokybei tikrinti yra įvedamos įvairios *paklaidų matavimo metrikos*. Kai kuriuose šaltiniuose, norint įsitikinti, kokio dydžio daromos paklaidos, yra naudojamas paprasčiausias paklaidų vidurkis [18]:

$$ME = \frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t); \quad (1.19)$$

čia  $x_t$  ir  $\hat{x}_t$  atitinkamai yra laiko eilutės tikroji reikšmė ir prognozė laiko momentu  $t$ , kai laiko eilutėje yra  $N$  duomenų. Tačiau ši metrika yra tik įvertio metrika, nes ji teparodo, ar prognozės procesas yra vidutiniškai teigiamas ar neigiamas. Ji dažnai naudojama kartu su kitomis metrikomis, nes  $ME$  parametras daug informacijos apie modelio tikslumą nesuteikia.



Prognozavimo tikslumui labiau tinkamos metrikos, kuriose imamos prognozės absoliučiosios paklaidos arba paklaidų kvadratai, kaip antai vidutinis absoliutus nuokrypis (Mean Absolute Deviation), vidutinė kvadratinė paklaida (Mean Squared Error), kvadratinė šaknis iš vidutinės kvadratinės paklaidos (Root Mean Square Error), paklaidų dispersijos ar vidutiniai kvadratiniai nuokrypiai [18, 57, 68, 88, 90]. Laiko eilučių prognozėje viena iš dažniausiai naudojamų metrikų yra vidutinė kvadratinė paklaida:

$$MSE = \frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2 ; \quad (1.20)$$

Kvadratinė šaknis iš šios metrikos yra kita labai dažnai eksperimentuose naudojama metrika:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2} . \quad (1.21)$$

Ši metrika yra daugiau tiesiogiai susijusi su paklaidų variacija ir standartiniu nuokrypiu. Todėl neretai vertinant eilutės prognozės paklaidas yra vertinamas būtent paklaidų išsibarstymas apie jų vidurkį. Šios dvi metrikos yra tiesiogiai pririštos prie laiko eilutės matavimo skalės, todėl kai kurie autoriai laiko, kad jos netinkamos palyginti laiko eilučių modelių funkcionalumui [6]. Todėl dažnai lyginant vieno ir kito modelio prognozės tikslumą yra pasirenkamos metrikos, kuriose vertinamas paklaidos santykis su tikrosios laiko eilutės reikšmėmis, t.y. vidutinė procentinė paklaida (Mean Percentage Error), vidutinė absoliučioji procentinė paklaida (Mean Absolute Percentage Error) [18, 32, 88, 90, 91]:

$$MAPE(\%) = \frac{1}{N} \sum_{t=1}^N \left| \frac{x_t - \hat{x}_t}{x_t} \right| \cdot 100 . \quad (1.22)$$

Jeigu tyrimo metu norima gauti prognozės perspektyvos tikslumą, kuris susieją modelio vystymą su atsitiktinio klaidžiojimo modeliu, naudojama Theils U statistika:

$$THEIL = \frac{\sum_{t=1}^N (x_t - \hat{x}_t)^2}{\sum_{t=1}^N (x_t - x_{t-1})^2} . \quad (1.23)$$

Priimta laikyti, kad prognozės modelis veikia taip pat kaip ir atsitiktinio klaidžiojimo modelis, jei statistika *THEIL* lygi vienetui. Jei statistika didesnė už vienetą, tuomet modelis veikia prasčiau už atsitiktinio klaidžiojimo modelį. Tik tuo atveju, kai statistikos reikšmė artėja į nulį, laikoma, kad tiriamas prognozės modelis veikia puikiai [32].

Paklaidų metriką taip pat galima susieti su laiko eilutės vidurkiu. Ši metrika yra įdomi tuo, kad ji prognozuoja laiko eilutės krypties pasikeitimą (Prediction Of Change In Direction):

$$POCID = \frac{100}{N} \sum_{t=1}^N D_t , \quad (1.24)$$

kai

$$D_t = \begin{cases} 1, & \text{jei } (x_t - x_{t-1})(x_t - \hat{x}_t) \geq 0, \\ 0, & \text{kitu atveju.} \end{cases} \quad (1.25)$$

Šios aukščiau išvardintos metrikos yra dažniausiai pasitaikančias laiko eilučių prognozės kokybės vertinimo metrikos. Kiekvienas autorius, tirdamas savo modelio funkcionalumą, priima naudoti vis kitokią paklaidų vertinimo metriką atsižvelgdamas į sprendžiamo uždavinio specifiką [18, 32, 57]. Šiame darbe paklaidos vertinamos skaičiuojant paklaidų standartinę nuokrypį arba dažniausiai naudojamą RMSE metriką.

## 1.2. Laiko eilučių rekonstravimas

Laiko eilučių prognozės modeliai skirstomi į ilgų laiko eilučių prognozę ir trumpų laiko eilučių prognozę. Ilgų laiko eilučių prognozeje yra kaupiami dideli duomenų kiekiai ir prognozavimo horizontas yra ilgas. Prognozuojant laiko eilutes visada tikimasi, kad dinaminės sistemos ateities elgsena bus panaši į sistemos elgseną praeityje.

Norint ištirti dinaminės sistemos elgseną ir įsitikinti, ar jos permainingumas iš tikrųjų yra chaotinis, reikėtų kiekybiškai ištirti tiriamą procesą [92]. Dėl tam tikrų priežasčių vienų dinaminių sistemų, tokių kaip švytuoklė, būsenas galima stebėti tiesiogiai. Tokioms sistemoms nėra sunku nustatyti būsenų laisvės laipsnių skaičių. Pavyzdžiui, trijų dimensijų sistemai, tokiai kaip Lorencio modelis, reikia registruoti tris kintamuosius. Kai sistemos būsenų erdvės matas išauga, išauga ir kintamųjų skaičius, kurie turi būti aprašyti [92]. Tačiau tiriant atvejus su kontroliuojamais eksperimentais, kuriuose ne visus kintamuosius galima registruoti, ir/arba stebint nekontroliuojamas sistemas (tokias kaip atmosferinės ir hidrologinės sistemos) kurių matematinis formalizavimas ir bendras kintamųjų skaičius nėra tiksliai žinomas, tuomet tokiose dinaminėse sistemose sudėtinga nustatyti chaoso egzistavimą [93]. Daugeliui sistemų, pavyzdžiui degalų sistemoms, arba daugelio režimų lazerio sistemoms yra nežinoma kiek laisvės laipsnių yra aktyvių (t.y. nežinomas kintamųjų skaičius) [92]. Tokiu atveju geriausia analizuoti duomenis iš tiksliai išmatuotų kintamųjų. Tokiose sistemose chaoso reiškinių nustatymui buvo pasiūlyta daugybė metodų, kurių populiariausi yra koreliacijos dimensijos metodas, didžiausių Liapunovo eksponenčių metodas, Kolmogorovo entropijos metodas, pakaitinių duomenų metodas, lokalinės srities metodas, netiesinis prognozavimas ir pagrindinių komponentų metodas [93, 94]. Reikia paminėti, kad nepaisant jų privalumų, nei vienas šių metodų negali teikti patikimo skirtumo tarp chaotinės ir atsitiktinės sistemos. Minėti metodai yra detalai aprašyti [95-97], todėl šiame darbe jie plačiai neapšarinėjami.

Dinaminių sistemų teorijoje gana žinomas metodas, rekonstruojantis stebimos netiesinės sistemos vieno kintamojo laiko eilutę į laiko vėlinimų erdvę [98]. Šis metodas paremtas principu, kad net daugiamatei dinaminei sistemai vieno kintamojo laiko eilutės duomenys yra pakankami nustatyti pilnos sistemos dinaminėms charakteristikoms [92].

Dar 1980 metais, straipsnio [95] autoriai pademonstravo, kad rekonstruojant sistemos vieno kintamojo laiko eilutę į laiko vėlinimų erdvę, naudojant laiko

vėlinimo kintamąjį, galima pastebėti tas pačias atraktoriaus topologines charakteristikas ir geometrines formas, kokios būtų pastebimos tiriant visus sistemos nepriklausomus kintamuosius būsenų erdvėje. Trajektorių evoliucija rekonstruotoje vėlinimų erdvėje atkartojama tokia, kokia yra tikroji trajektorių evoliucija pilnoje būsenų erdvėje [92].

Rekonstruota trajektorija,  $\mathbf{X}$ , gali būti išreikšta kaip matrica, kurios kiekviena eilutė yra lako vėlinimų erdvės vektorius:

$$\mathbf{X} = [\mathbf{y}_1(d), \mathbf{y}_2(d), \dots, \mathbf{y}_n(d)]^T; \quad (1.26)$$

čia  $\mathbf{y}_p(d)$  yra sistemos būsena diskrečiais laiko momentais  $p$ . Laiko eilutei  $x_i$  ( $i = 1, 2, \dots, N$ ) kiekvienas  $\mathbf{y}_p(d)$  konstruojamas tokiu būdu:

$$\mathbf{y}_p(d) = (x_p, x_{p+\tau}, \dots, x_{p+(d-1)\tau}) \quad p = 1, 2, \dots, N - (d-1)\tau. \quad (1.27)$$

Šis vektorius formuoja  $d$ -matę rekonstruotą laiko vėlinimų erdvę, kurioje  $\tau$  vadinamas rekonstravimo į vėlinimų erdvę laiko vėlinimu,  $d$  – laiko vėlinimų erdvės matavimas. Matrica  $\mathbf{X}$  yra  $n \times d$  formato matrica, o konstantos  $d$ ,  $N$ ,  $n$  ir  $\tau$  siejamos lygybe  $N = n + (d-1)\tau$ .

Taigi galima paklausti, kiek dimensijų reikia panaudoti rekonstruojant dinaminės sistemos kintamąjį į laiko vėlinimų erdvę, kad rekonstruota erdvė kuo geriau atspindėtų visos dinaminės sistemos elgseną.

Takens [99] ir Mane [100] parodė, kad jei sistemos pagrindinių būsenų (turima omenyje sistemos visų kintamųjų, kuriuos reikia įvertinti) erdvė yra  $d_s$  matavimo, o rekonstruota fazinė erdvė yra  $2d_s + 1$  matavimo, tuomet galima pilnai aprašyti sistemos dinamiką naudojant rekonstruotos vėlinimų erdvės vektorius. Priešingai, atraktoriaus projekcija į mažesnės dimensijos erdvę gali iškreipti sistemos atraktoriaus įvėčius (pvz. Liapunovo eksponentės įvertį). Šį rezultatą galima suvokti paprasčiausiu būdu: jei būsenų erdvė yra  $d_s$  matavimo, tada galima sistemos dinamiką aprašyti  $d_s$  eilės diferencialinių lygčių sistema [92]. Reikia surasti dvigubai didesnę rekonstruotos erdvės dimensiją už tikrąją atraktoriaus būsenų erdvės dimensiją, kad galėtume atkartoti jo dinamiką.

Rekonstravimo metodui realizuoti galima naudoti koreliacinę sumą [92]:

$$C^{(d)}(R) = \frac{1}{N(N-1)} \sum_{i,j=1, i \neq j}^N \Theta[R - \|\mathbf{y}_i(d) - \mathbf{y}_j(d)\|]. \quad (1.28)$$

Jeigu vektorius  $\mathbf{y}_i(d)$  duoda tašką rekonstruotoje laiko vėlinimų erdvėje, tai koreliacijos suma pasako, kiek taškų porų vienas nuo kito yra nutolę ne daugiau kaip atstumu  $R$ .

Koreliacijos sumai apskaičiuoti naudojama Heavisaido funkcija  $\Theta$  :

$$\Theta(x) = \begin{cases} 0, & \text{jei } x < 0; \\ 1, & \text{jei } x \geq 0. \end{cases} \quad (1.29)$$

Vektorių skirtumo modulis šiuo atveju reiškia didžiausią koordinacių atstumą, kas leidžia ženkliai pagreitinti skaičiavimus, daug nenusižengiant skaičiavimų tikslumui [101]:

$$\|y_i(d) - y_j(d)\| = \max_k |x_{i+k\tau} - x_{j+k\tau}|. \quad (1.30)$$

Kitais žodžiais tariant, randamas didžiausias skirtumas tarp atitinkamų komponentių, ir tas dydis naudojamas kaip vektorių skirtumo ilgis.

Sistemos koreliacinė dimensija  $D_c(d)$  randama iš sąryšio:

$$C^{(d)}(R) = kR^{D_c(d)}; \quad (1.31)$$

čia  $k$  pasirenkamas proporcijos koeficientas.

Ieškant rekonstravimo dimensijos praktiškai, tai galima atlikti dviem būdais:

1. Iš (1.30) formulės išsireikšti:

$$D_c(d) = \frac{\lg(C^{(d)}(R)) - \lg k}{\lg R}. \quad (1.32)$$

Skaičiuoti  $D_c(d)$  reikšmes kintant rekonstravimo dimensijai  $d$ . Skaičiavimuose  $\lg R$  fiksuoti ir pagal (1.28) formulę surandame koreliacijos sumą.

2. Brėžti  $\lg(C^{(d)}(R))$  ir  $\lg R$  priklausomybės grafikus. Taškų aibę  $(\lg R, \lg(C^{(d)}(R)))$  aproksimuoti tiesę, panaudodami mažiausių kvadratų metodą. Gautas tiesės krypties koeficientas ir yra laiko vėlinimų erdvės dimensiškumo įvertis.

Laiko vėlinimų erdvės konstravimui būtina parinkti laiko vėlinimą  $\tau$ . Tik optimali  $\tau$  reikšmė leidžia optimaliai atskirti trajektoriją neperžiangiant rekonstruotos vėlinimų erdvės minimumo [93]. Jeigu laiko vėlinimo parametras  $\tau$  yra per mažas, visos vėlinimų erdvės koordinatės yra beveik vienodos, ir rekonstruotos trajektorijos suspaudžiamos laiko vėlinimų erdvėje. Jei laiko vėlinimo parametras  $\tau$  yra per didelis, vieno laiko momento dinamika bus labai stipriai atskirta nuo kito laiko momento. Taigi rekonstruotas geometrinis objektas atrodys ypač sudėtingas ir atpažinti sistemos elgseną bus labai sunku [94].

Šioje tyrimų kryptyje buvo pasiūlyta daugybė metodų ir rekomendacijų. Populiariausios yra autokoreliacijos metodas ir bendros informacijos metodas [103, 104].

Yra pastebėta, kad laiko eilutės rekonstravimas nereguliais laiko vėlinimais gali turtingiau perteikti rekonstruoto atraktoriaus topologines savybes. Tačiau dažniausiai šis klausimas giliau nebūdavo nagrinėjamas dėl sudėtingų geometrinio bei topologinio šio uždavinio formuluočių [102]. Šiame darbe bus pristatyta metodika, kuri suteiks galimybę efektyviai (skaičiavimų prasme) nagrinėti atraktoriaus optimalaus neregularaus rekonstravimo į vėlinimų erdvę uždavinį (optimalumo kriterijai bus įvesti vėliau). Deja, neregularaus rekonstravimo laiko vėlinimų  $\tau_1, \tau_2, \dots, \tau_{d-1}$  parinkimui tenka spręsti pilno perrinkimo uždavinį. O jeigu vėlinimų erdvės dimensija yra gana didelė, toks uždavinys tampa sunkiai

išsprendžiamu skaičiavimų apimties prasme. Šią problemą išspręsti padeda metaeuristiniai metodai, kurie trumpai aprašyti 1.4 skyrelyje.

### **1.3. Hankelio<sup>1</sup> matricos rango apžvalga**

#### **1.3.1. Hankelio matricos taikymų apžvalga**

Hankelio matricos (Hermann Hankel (1839-1873)) plačiai naudojamos sistemos atpažinimui, kai turint tik sistemos išėjimo duomenis norima sukonstruoti sistemos būsenų erdvės modelį. Pirmasis šios sudėtingos problemos sprendimas, įgavęs būsenų erdvės realizacijos uždavinio pavadinimą, buvo įgyvendintas 1965 metais [105]. Šio uždavinio sprendimas pagrįstas Hankelio matrica, kurios skaidymas į stebėjimų matricą ir kontrolės matricą yra žinomas kaip Ho-Kolman realizacijos metodas [105]. Hankelio matrica pagrįsti modeliai yra tinkami aiškinti tiesinius įėjties/išeities atvaizdavimus su begaline aibe parametų, nes matrica gaunama tiesiogiai iš sistemos įėjties/išeities duomenų. Praėjo nemažai laiko, kol buvo surastas būdas kaip iš teorinių rezultatų, pateiktų [105] šaltinyje, pereiti prie patikimo programinio algoritmo realizavimo [106]. Deterministinės realizacijos teorijos kombinacija, pagrįsta Hankelio matricos faktorizacija kartu su Markovo sistemų teorija, leido sukurti minimalių realizacijų stochastinę teoriją. Stochastinės teorijos uždavinys kartu su inovacijų teorija ir spektrinės faktorizacijos teorija intensyviai tirtas septintajame devyniolikto amžiaus dešimtmetyje [107, 108].

Buvo sukurta daugybė Hankelio matrica pagrįstų naujų taikymų skirtingose mokslo ir inžinerijos srityse. Impulso/atsako modeliavimo išeities rezultatų gavimas apibendrintai Hankelio matricai konstruoti ir šios matricos singuliariųjų reikšmių dekompozicija padėjo atrasti redukuotos eilės modelius aukštesnio matavimo tiesinėse dinaminėse sistemose [109]. Hankelio matrica buvo panaudota išplėsti laiko eilutes į sistemos trajektorijų matricas [110]. Singuliariųjų reikšmių dekompozicija padeda prognozuoti atsitiktinius nekontroliuojamus šuolius. Sveikų skaičių sekos Hankelio transformacijos, apibrėžtos [111], panaudojamos klasifikuoti tam tikras sveikaskaitines sekas. Hankelio matricos metodas taip pat buvo panaudotas dinaminės sistemos atpažinimui iš laiko eilutės duomenų [112]. Visi šie modeliai pagrįsti prielaida, kad vykdomą sistemos būsenų erdvės realizaciją galima aprašyti tam tikros rūšies dinaminių sistemų analiziniais modeliais.

Naujas skaitinių sekų atpažinimo modelis ir Hankelio matricos rango sąvoka sukonstruota [113]. Svarbu pabrėžti, kad sekos Hankelio rangas yra koncepcija, nepriklausanti nuo sistemos būsenų erdvės realizacijos. Hankelio rangas tik paaiškina algebrinius sąryšius tarp sekos elementų nemėginant rekonstruoti tiriamos dinaminės sistemos analizinių modelių. Be to, šie išvedami algebriniai sąryšiai yra tikslūs.

Hankelio rangas buvo panaudotas išreikšti netiesinių diferencialinių lygčių sprendinius formomis, susidedančiomis iš baigtinių standartinių funkcijų sumų santykių [114, 115, 116]. Šioje disertacijoje Hankelio rangas bus naudojamas skeletinės algebrinės progresijos tiriamoje laiko eilutėje atpažinimui ir ši informacija bus panaudota šios laiko eilutės ateities reikšmių prognozei.

---

<sup>1</sup> Hermann Hankel (1838-1873)

### 1.3.2. Hankelio matricos konstravimas algebrinėms sekoms

Trumpų laiko eilučių prognozės modeliui konstruoti šioje disertacijoje naudojama algebrinių sekų Hankelio matricos konstravimo teorija, kuri yra pristatyta [113] darbe. Tarkime, kad duota seka  $S$  sudaryta iš realiųjų, arba kompleksinių skaičių:

$$S := (x_0, x_1, x_2, \dots) = (x_k; k \in N_0). \quad (1.33)$$

Ši  $S$  seka yra vadinama *algebrine seka*. Jos posekis žymimas  $S_j$ ;  $j = 0, 1, 2, \dots$ :

$$S_j := (x_j, x_{j+1}, x_{j+2}, \dots). \quad (1.34)$$

Galima pastebėti, kad  $S = S_0$ . Iš sekos  $S$  elementų galima sukonstruoti Hankelio matricą  $H$ :

$$H = \begin{bmatrix} x_0 & x_1 & x_2 & \dots \\ x_1 & x_2 & x_3 & \dots \\ x_2 & x_3 & x_4 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}. \quad (1.35)$$

Matricos  $H$  minorai  $H_j^{(m)}$  apibrėžiami tokiu būdu:

$$H_j^{(m)} := [x_{r+s-2+j}]_{1 \leq r, s \leq m} = \begin{bmatrix} x_j & x_{j+1} & \dots & x_{j+m-1} \\ x_{j+1} & x_{j+2} & \dots & x_{j+m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{j+m-1} & x_{j+m} & \dots & x_{j+2m-2} \end{bmatrix}. \quad (1.36)$$

Šių minorų determinantai žymimi  $d_j^{(m)}: \det H_j^{(m)} = d_j^{(m)}$ .

Sekos  $S_j$  rangas apibrėžiamas tokiu būdu: rangą yra laikomas toks natūralusis skaičius  $m_j$ , kuris tenkina sąlygą (jei tik tas rangas egzistuoja): su visais  $k \in N$  yra teisinga

$$d_j^{(m_j+k)} = 0; \quad (1.37)$$

kai tuo tarpu  $d_j^{(m_j)} \neq 0$ .

Toliau pažymima:

$$m_j = Hr(x_j, x_{j+1}, \dots) = HrS_j. \quad (1.38)$$

Jeigu neįmanoma rasti tokio skaičiaus  $m_j$ , tuomet laikoma, kad posekis  $S_j$  neturi rango:  $HrS_j := +\infty$ .

Tarkime, kad posekio  $S_j$  rangas yra  $HrS_j = m$ ;  $m < +\infty$ . Tada sudaromas determinantas:

$$\Delta^{(m)}S_j(\rho) := \begin{vmatrix} x_j & x_{j+1} & \dots & x_{j+m} \\ x_{j+1} & x_{j+2} & \dots & x_{j+m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{j+m-1} & x_{j+m} & \dots & x_{j+2m-1} \\ 1 & \rho & \dots & \rho^m \end{vmatrix}. \quad (1.39)$$

Lygtis

$$\Delta^{(m)}S_j(\rho) = 0; \quad (1.40)$$

vadinama algebrinės sekos  $S_j$  charakteristine lygtimi. Jos šaknų  $\rho_1, \rho_2, \dots, \rho_r$  ( $\rho_k \in C$  su visais  $k=1, 2, \dots, r$ ) kartotinumų eilė atitinkamai yra  $n_1, n_2, \dots, n_r$  ( $n_k \in N$ );  $n_1 + n_2 + \dots + n_r = m$ .

Toliau iš sistemos

$$x_n = \sum_{k=1}^r \sum_{l=0}^{n_k-1} \mu_{kl} \binom{n}{l} \rho_k^{n-l}; \quad n = j, j+1, j+2, \dots; \quad (1.41)$$

apskaičiuojami koeficientai  $\mu_{kl} \in C$ ;  $k=1, 2, \dots, r$ ;  $l=0, 1, \dots, n_k-1$ . [104] įrodoma, kad (1.41) sistemos sprendinys yra vienintelis.

Jei (1.38) lygties visos šaknys yra skirtingos, (1.41) sistema įgyja formą:

$$x_n = \sum_{k=1}^m \mu_{k0} \rho_k^n. \quad (1.42)$$

**Apibrėžimas.** Posekis  $S_j$  vadinamas *algebrine progresija*, jei jo visi elementai išreikšti (1.41) lygybe.

Elementų  $x_j, x_{j+1}, x_{j+2}, \dots, x_{j+m}$  aibė, kuri tenkina (1.41) lygtį, vadinama *algebrinės progresijos fragmentu* [113].

Remiantis šiuo apibrėžimu yra formuluojama išvada: Tegu sekos  $(x_k; k \in N_0)$  rangas yra  $m$ , o seka  $(\xi_k; k \in N_0)$  yra atsitiktinių skaičių seka. Tuomet, rangas  $Hr(x_k + \xi_k; k \in N_0) = +\infty$ .

**Pavyzdys.** Tarkim, kad duota periodinė seka  $S := (3, 1, 0, 3, 1, 0, \dots)$ . Posekio

$$S_0 \text{ determinantai yra } d_0^{(1)} = |3| = 3; \quad d_0^{(2)} = \begin{vmatrix} 3 & 1 \\ 1 & 0 \end{vmatrix} = -1; \quad d_0^{(3)} = \begin{vmatrix} 3 & 1 & 0 \\ 1 & 0 & 3 \\ 0 & 3 & 1 \end{vmatrix} = -28 \text{ ir}$$

$d_0^{(m)} = 0$ , kai  $m = 4, 5, \dots$  Taigi gaunasi, kad sekos rangas yra  $HrS = 3$ .

Tuomet įmanoma sukonstruoti algebrinę lygtį:

$$\Delta^{(3)}S_0(\rho) = \begin{vmatrix} 3 & 1 & 0 & 3 \\ 1 & 0 & 3 & 1 \\ 0 & 3 & 1 & 0 \\ 1 & \rho & \rho^2 & \rho^3 \end{vmatrix} = 28(1 - \rho^3) = 0;$$

kurios šaknys  $\rho_1 = 1; \rho_2 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i; \rho_3 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$ . Visos šaknys skirtingos, todėl koeficientų  $\mu_{k0}$  ( $k=1, 2, 3$ ) radimui naudojama (1.42) lygtis. Sudaroma tiesinių algebrinių lygčių sistema:

$$\begin{cases} \mu_1 + \mu_2 + \mu_3 = 3; \\ \mu_1 + \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)\mu_2 + \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)\mu_3 = 1; \\ \mu_1 + \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)^2\mu_2 + \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)^2\mu_3 = 0. \end{cases}$$

Sistemos sprendinys:  $\mu_1 = \frac{4}{3}, \mu_2 = \frac{5}{6} - \frac{\sqrt{3}}{6}i, \mu_3 = \frac{5}{6} + \frac{\sqrt{3}}{6}i$ . Taigi duotosios sekos elementai gali būti išreikšti tokia algebrine lygtimi:

$$x_n = \frac{4}{3} + \left(\frac{5}{6} - \frac{\sqrt{3}}{6}i\right)\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)^n + \left(\frac{5}{6} + \frac{\sqrt{3}}{6}i\right)\left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)^n.$$

## 1.4. Programinis skaičiavimas kombinatorinio optimizavimo uždaviniuose

### 1.4.1. Evoliuciniai algoritmai

Kombinatorinio optimizavimo uždaviniai plačiai taikomi įvairiuose inžineriniuose, paieškos operacijų ir socialinių moksluose. Daugelio kombinatorinių uždavinių sprendimui buvo sėkmingai pritaikyti *evoliuciniai algoritmai* (*Evolutionary Algorithms*). Evoliuciniai algoritmai turi gebėjimą apimti visą paieškos aibę, gali tuo pačiu metu manipuluoti su grupe sprendinių ir adaptyviai gerinti dinaminių sistemų potencialą [117]. Evoliucinių algoritmų taikymas laiko eilutės rekonstravimo į vėlinimų erdvę laiko vėlinimų parinkimui nėra naujas uždavinys [118]. Laiko vėlinimų įtraukimas į evoliucinį prognozavimą demonstruojamas [119], kuriame evoliuciniais algoritmais ieškoma mažiausias laiko vėlinimų erdvės matavimas. Dalinis laiko vėlinimų paieškos metodas su modifikuotu genetiniu algoritmu pristatytas [120] straipsnyje. Šioje disertacijoje evoliucinis algoritmas pritaikomas jau nusistatyto matavimo vėlinimų erdvei rekonstruoti reikalingiems laiko vėlinimams rasti, t.y. spręsti sveikaskaitiniam kombinatoriniam optimizavimo uždaviniui.



Ankstyvasis genetinio paieškos algoritmo realizavimas buvo pasiūlytas Bremermann (1962) ir Recherberg (1965), nors Holland darbas 1975 metais suformavo bazę daugelio šiuolaikinių genetinių algoritmų vystymui [121]. Adaptyvieji genetinių algoritmų taikymai optimizacijai buvo išplėsti Goldberg [122]. Genetiniai algoritmai yra įkvėpti evoliucijos ir pritaikyti ieškant globalaus optimumo įvairiuose taikymo uždaviniuose. Šie algoritmai sprendžiamo uždavinio sprendimą užšifruoja į paprastą chromosomos pavidalo duomenų struktūrą ir tai struktūrai taiko rekombinacijos operaciją, taip išsaugant kritinę uždavinio informaciją [123]. Goldberg pasiūlytas algoritmas atliekamas tokia eiga:

1 žingsnis. *Iniciacija*. Generuojama pradinė populiacija susidedanti iš  $N_p$  chromosomų, kurios naudojamos uždavinio globalaus optimalaus sprendinio radimui. Šiame žingsnyje taip pat nusistatomi kiti evoliucinio algoritmo parametrai: kryžminimo koeficientas, mutacijos koeficientas, generacijų skaičius.

2 žingsnis. *Įvertinimas*. Po to, kai sugeneruojama pradinė populiacija, kiekviena chromosoma yra įvertinama vartotojo prisiskirta tinkamumo funkcija. Kiekvienos charakteristikos tinkamumo funkcijos reikšmė yra uždavinio rodiklis, kuris suformuotas chromosomos tinkamumui gerinti, didinti jos išlikimo tikimybę.

3 žingsnis. *Algoritmo nutraukimo kriterijus*. Genetinio algoritmo vykdymo procesas nuo 2 iki 7 žingsnio kartojamas tol, kol yra pasiekiamas užsibrėžtas nutraukimo kriterijus. Dažniausiai nutraukimo kriterijumi laikoma viena iš sąlygų: kol pasiekiamas nusistatytas generacijų skaičiaus arba kai sprendinys nebesikeičia nuo praeitos generacijos.

4 žingsnis. *Elitizmo mechanizmas*. Siekiant užtikrinti geriausių chromosomų propagavimą, genetinis algoritmas naudoja elitizmo mechanizmą [124]. Šis mechanizmas atranka tam tikrą procentą individų, kurie bus naudojami vykdant individų genetines operacijas.

5 žingsnis. *Atranka*. Atrankos procese yra atrankamos tinkamos chromosomos iš tėvų populiacijos. Tai dažniausiai būna turnyrinė atranka [124]. Chromosomos, kurių tinkamumo funkcijos reikšmės geresnės, turi didesnę šansą patekti į naują populiaciją. Atrankos metu parenkama tiek chromosomų, kokio dydžio populiacija dalyvaus kryžminimo procese.

6 žingsnis. *Kryžminimas*. Šis procesas vykdomas sukeičiant eilės tėvų chromosomų atitinkamus segmentus. Taip genetiniame algoritme pratęsiamas naujų sprendinių ieškojimas.

7 žingsnis. *Mutacija*. Tai išskirtinai genetinio algoritmo mechanizmas, kuriame atsitiktinai pasirenkamas genas, kuris pakeičiamas kitu genu.

#### **1.4.2. Dalelių spiečiaus algoritmas**

*Dalelių spiečiaus optimizavimas (Partial Swarm Optimization – PSO)* yra skaitinis evoliucinis metodas pagrįstas socialinės elgsenos imitavimu, pirmą kartą pristatytas Eberhart ir Kennedy 1995 m. [125]. Daugelis autorių PSO metodą pritaikė spręsti kombinatorinio optimizavimo uždaviniams [126, 127, 128]. Per bendradarbiavimą ir kompetenciją tarp populiacijos, PSO optimizavimo metodas (algoritmas) gali rasti gerus sprendimus produktyviai ir efektyviai. Kiekvienas PSO algoritme individas traktuojamas kaip taškinė (volume-less) dalelė (materialusis taškas)  $d$ -matėje erdvėje. PSO algoritme kiekviena dalelė juda paieškos erdvėje greičiu, kuris dinamiškai koreguojamas dalelės judėjimo patirtimi ir grupės judėjimo

patirtimi. Kiekviena  $i$ -oji dalelė aprašoma koordinatėmis  $\tilde{\mathbf{X}}_i = (x_{i1}, x_{i2}, \dots, x_{id})$ ;  $i=1, 2, \dots, \tilde{m}$ , čia  $\tilde{m}$  – populiacijos dydis. Ankstesnė pozicija, kuri duoda geriausią  $i$ -osios dalelės jos judėjimo erdvėje tinkamumo funkcijos reikšmę, yra registruojama ir vaizduojama kaip  $\mathbf{P}_i = (p_{i1}, p_{i2}, \dots, p_{id})$ . Dalelių populiacijos geriausios dalelės indeksas žymimas simboliu  $g$ . Kiekvienos  $i$ -osios dalelės greitis žymimas  $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$ . PSO populiacijos greičio ir pozicijos atnaujinimo lygtys:

$$\begin{aligned} v_{id} &= w \cdot v_{id} + c_1 \cdot r_1 \cdot (p_{id} - x_{id}) + c_2 \cdot r_2 \cdot (p_{gd} - x_{id}), \\ x_{id} &= x_{id} + v_{id}, \quad i=1, 2, \dots, d; \end{aligned} \quad (1.43)$$

čia  $r_1$  ir  $r_2$  yra du tolygiai pasiskirstę atsitiktiniai skaičiai iš intervalo  $[0, 1]$ ; konstantos  $c_1$  ir  $c_2$  yra teigiami sveikieji skaičiai, vadinamosios pagreičio konstantos, kurios reiškia statinio pagreičio sąlygos svorį, kuris traukia kiekvieną dalelę prie geriausių dalelių ir globalaus geriausio sprendinio;  $w$  vadinamas inercijos svoris balansuojantis globalų ir lokalų sprendinius.

Ankstesni PSO optimizavimo tyrimai (dažniausiai bandymų ir paklaidų būdu) parodė, kad daugelyje taikymų patogų laikyti pagreičio konstantas  $c_1 = c_2 = 2$ , o  $w=1$  [129]. Inercijos svoris įvestas kontroliuoti balansui tarp globalaus ir lokalaus tyrimo galimybių [130]; rekomenduojama naudoti  $w=0,9$ . Didelė inercijos svorio reikšmė lengvina globalaus sprendinio paiešką, o štai maža reikšmė – lokalaus sprendinio.

1999 metais Clerc nurodė, kad būtinybė naudoti suvaržymo faktorių  $K$  ((1.44) lygtis) atsirado norint užtikrinti PSO algoritmo konvergavimą [131]:

$$v_{id} = K(w \cdot v_{id} + c_1 \cdot r_1 \cdot (p_{id} - x_{id}) + c_2 \cdot r_2 \cdot (p_{gd} - x_{id})), \quad (1.44)$$

$$\text{čia } K = \frac{2}{\left| 2 - \varphi_K - \sqrt{\varphi_K^2 - 4\varphi_K} \right|} \text{ ir } \varphi_K = c_1 + c_2; \quad \varphi_K > 4.$$

Visais atvejais, kai buvo panaudotas Clerc suvaržymo metodas,  $\varphi_K = 4,1$ , tada konstanta  $K = 0,729$ . Shi ir Eberhart [131] nurodė, kad tą patį gali gauti naudojant (1.43) lygtyje  $w = 0,729$  ir  $c_1 = c_2 = 1,494$ .

PSO algoritmo konvergavimo ir stabilumo analizė buvo paskelbta Trelea [132], kur imamas inercijos svoris 0,6, ir  $c_1 = c_2 = 1,7$  pagerino parametrų parinkimą rekomenduojamą Shi ir Eberhart.

Bet kokių atveju, nepaisant daugybės ieškojimo pastangų, parametrų parinkimas dažniausiai laikomas empiriniu uždaviniu ir priklauso nuo objekto funkcijos topologijos ir/arba tinkamumo funkcijos struktūros. Atsižvelgiant į rekomendacijas [129, 131, 132], šioje disertacijoje PSO parametrai parinkti trimis siūlomais atvejais: pirma parametrų aibė  $w = 0,9$  ir  $c_1 = c_2 = 2$ , antra parametrų aibė  $w = 0,729$  ir  $c_1 = c_2 = 1,494$  pagal Clerc, trečia –  $w = 0,6$  ir  $c_1 = c_2 = 1,7$  pagal Trelea.

Shi ir Eberhart nurodė, kad PSO algoritmo atlikime populiacijos dydžio efektas yra mažai reikšmingas. Daugelis tyrėjų spiečiaus dydį pasirenka nuo 10 iki

60 [133]. Siekiami palyginti PSO ir evoliucinių algoritmų efektyvumą, laikome, kad dalelių spiečiaus dydis PSO algoritme yra toks pat kaip ir evoliucinių algoritmų populiacijos dydis.

### 1.4.3. Modeliuojamojo atkaitinimo algoritmas

*Modeliuojamojo atkaitinimo algoritmas (Simulated Annealing – SA)* formaliai kombinatorinio optimizavimo uždaviniams spręsti buvo pristatytas 1983 metais Kirkpatrick ir kitų autorių [134] darbuose. Šio modelio pradininkai buvo Metropolis, Rosenbluth ir kiti (1953) [135]. Šis algoritmas remiasi analogija su fizikiniu procesu – atkaitinimu (grūdiniu), iš statistinės mechanikos modeliuojant procesus sistemose, sudarytose iš didelio skaičiaus mažų diskrečių dalelių. Modeliuojant sistemos atkaitinimą, iš pradžių sistemai suteikiama aukšta temperatūra, kuri palaipsniui mažinama, kol sistema pereina į stacionarią būseną (plienas užsigrūdina) [136]. Optimizavimo uždaviniui spręsti, sistemos perėjimas iš vieno energijos lygio ( $E_1$ ) į kitą ( $E_2$ ) priimamas su tikimybe :

$$P_E = \begin{cases} 1, & \Delta E < 0, \\ e^{-\Delta E/C_B T_e}, & \Delta E \geq 0; \end{cases} \quad (1.45)$$

čia  $T_e$  yra kontrolės parametras atitinkantis temperatūrą,  $\Delta E = E_2 - E_1$ , o  $C_B$  – Bolcmano konstanta, kurią optimizavimo metu autoriai dažnai prilygina vienetui.

Atkaitinimo metodo (SA) principai paremti mechanizmu, kuris padeda su tam tikra tikimybe ištrūkti iš lokaliajo sprendinio, ir suteikia galimybę rasti kitus (gal net ir globalius) sprendinius [136, 137], t.y. Metropolio kriterijuje (1.45) su nedidele tikimybe gali būti priimtas blogesnis sprendinys. Šiame mechanizme svarbiausias parametras yra temperatūra, kuri algoritmo pradžioje prisiskiriama labai didelė ir su kiekvienu proceso vykdymo žingsniu yra mažinama. Tačiau pasirenkama temperatūros pradinė reikšmė turi būti nei per didelė, nei per maža (žiūrėti 1.3 paveikslą A dalį). Jei temperatūra parenkama per didelė, yra labai didelė tikimybė, kad bus rastas globalus sprendinys, tačiau jo ieškojimas gali užtrukti labai ilgai. Tuo atveju, kai temperatūra parenkama per maža, sumažėja tikimybė surasti globalų sprendinį, nors paieškos laikas ir yra sutaupomas (pavyzdys pateiktas 1.3 paveikslą B dalyje).

Pradinės temperatūros parinkimui yra siūloma įvairių metodų. Labiau klasikinius ir intuityvius temperatūros parinkimo metodus yra paaiškinęs Kirkpatrick su kitais autoriais [138]. Skaičiavimas susideda iš priimamų sprendinių santykio palyginimo su tam tikra duota reikšme  $\chi_0$  iš intervalo  $(0, 1)$ . Vienas iš būdų rasti pradinės temperatūros reikšmę, tai priskirti didelę reikšmę ir tuomet tikrinti ar priimamų sprendinių skaičiaus santykis su tikrinamų sprendinių skaičiumi yra nemažesnis už  $\chi_0$ . Jei šis santykis yra mažesnis, temperatūra mažinama pusiau. Taip procesas tęsiamas kol pasiekama tinkama temperatūros reikšmė. Kirkpatrick pasiūlė, kad pradinė temperatūra  $T_0$  bus reikšminga jei sprendinių, kurių energija auga  $\Delta E$ , vidurkis bus priimtas su tikimybe  $\chi_0$ . Tuomet

$$T_0 = -\frac{\overline{\Delta E}}{\ln(\chi_0)}; \quad (1.46)$$

čia  $\overline{\Delta E}$  – teigiamų perėjimų energijos įvertis, kuris randamas atsitiktinai generuojant keletą perėjimų [138].

Šis parametras kontroliuojamas įvedant specialią temperatūros „aušinimo“ funkciją. Tokių funkcijų gali būti įvairūs variantai. Dažniausiai naudojami yra [135, 136]:

1. mažėjimas geometrine progresija:

$$T_k = \kappa_\alpha T_{k-1}; \quad (1.47)$$

čia  $T_k$  esamoji temperatūros reikšmė, iteracijų skaičius  $k=1, 2, \dots, n$ ;  $0,8 \leq \kappa_\alpha \leq 0,99$ .

2. Lundy-Mees formulė:

$$T_k = \frac{T_k}{1 + \kappa_\beta \cdot T_{k-1}}; \quad (1.48)$$

čia  $\kappa_\beta \leq T_0$  ( $T_0$  – atkaitinimo proceso pradžioje prisiskiriama temperatūra).

3. logaritminė atšaldymo formulė:

$$T_k = \frac{T_{k-1}}{1 + \lambda_T T_{k-1}}; \quad (1.49)$$

kai

$$\lambda_T = \frac{T_0 - T_f}{c T_f T_0}; \quad (1.50)$$

čia konstantai  $c$  [136] šaltinyje priskiriama reikšmė 200.  $T_f$  apibrėžiama galutinė temperatūros reikšmė.

Klasikinis modeliujamo atkaitinimo metodo algoritmas turi tokią schemą[136, 137]:

1 žingsnis. Sugeneruojamas pradinis sprendinys  $s$ . Nustatoma pradinė temperatūros reikšmė, nustatomas minimalus leistinas atstumas, apskaičiuojama tikslo funkcijos reikšmė  $f(s)$ .

2 žingsnis. Generuojamas naujas kaimyninis atsitiktinis sprendinys  $s'$ . Atstumas nuo  $s$  iki  $s'$  turi būti ne mažesnis nei duotas minimalus atstumas (priklauso nuo naudojamos metrikos), tokiu atveju iš naujo generuojamas naujas kaimyninis sprendinys. Tada skaičiuojamas skirtumas  $\Delta f = f(s) - f(s')$ .

3 žingsnis. Jei tenkinamas Metropolio kriterijus, t. y.  $\min\{1, \exp(\Delta f / T_k)\} > \eta$ , tada priimamas sprendinys  $s'$ . Priešingu atveju paliekamas einamasis sprendinys  $s$  ( $\eta$  yra atsitiktinis skaičius, tolygiai pasiskirstęs intervale  $[0, 1]$ ).

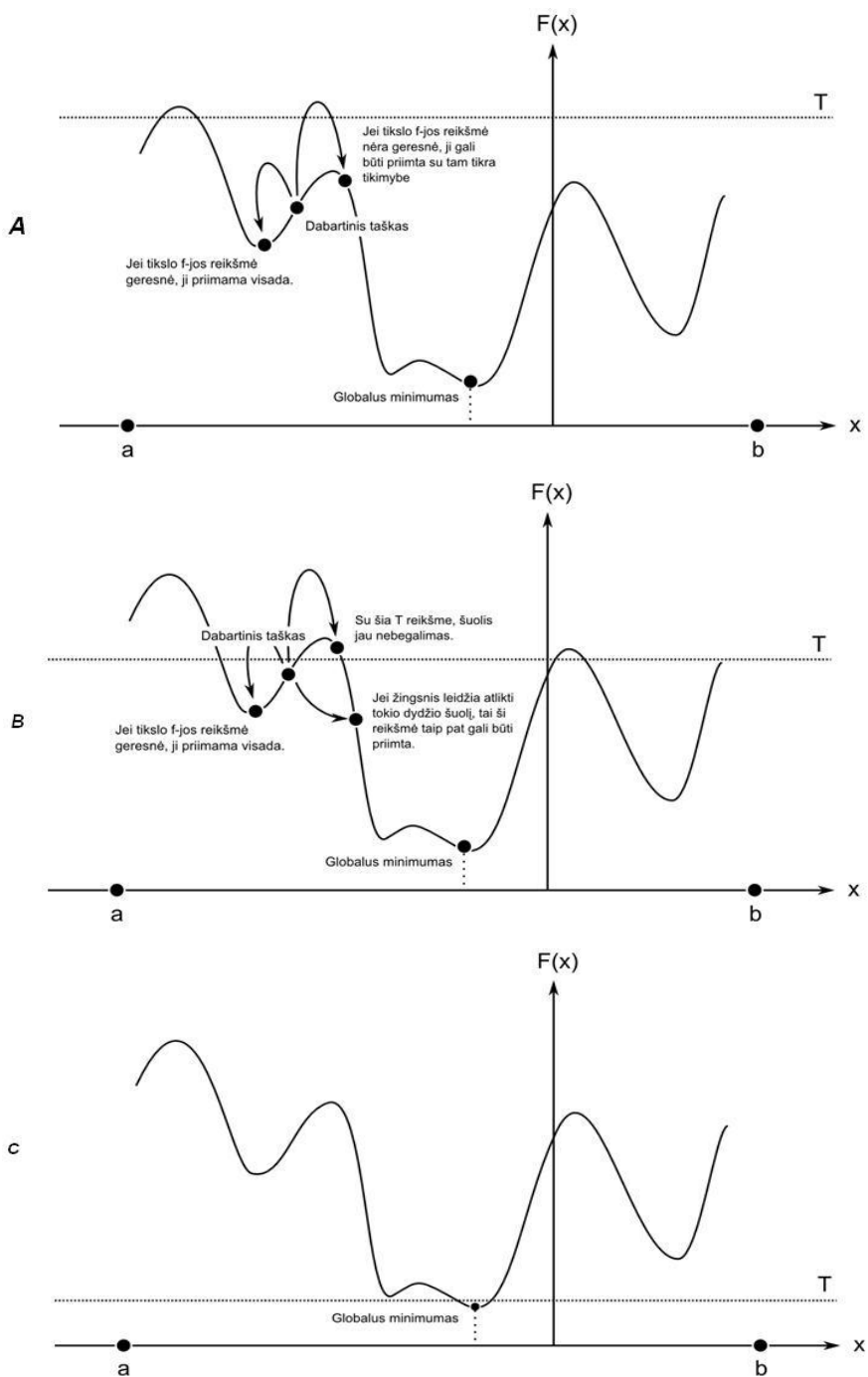
4 žingsnis. Kartais gali būti tikrinama „Metropolio pusiausvyra“ temperatūros mažinimo momentui nustatyti. Pusiausvyros kriterijumi gali būti, pavyzdžiui, tikslo funkcijos reikšmių svyravimų amplitudė.

5 žingsnis. Jeigu tenkinamas algoritmo stabdymo kriterijus, uždavinio optimaliu sprendiniu priimamas paskutinis gautas sprendinys  $s'$ . Kitu atveju mažinama temperatūra ir grįžtama į antrą žingsnį. Algoritmo baigimo kriterijumi gali būti užsibrėžtas iteracijų skaičius, pasiekiamą galutinę temperatūros mažinimo ribą (1.3. paveikslo C dalis) ar kai atkaitinimo procesas nusistovi.

Algoritme kaimyninio sprendinio parinkimas yra svarbus žingsnis. Nuo šio žingsnio priklauso atkaitinimo proceso vykdymo funkcionalumas. Jei pasirenkami turtingi kaimynai, kurie saugo didelį kiekį informacijos, padidėja tikimybė rasti globalų sprendinį, tačiau pailgėja sprendinio ieškojimo laikas. Kombinatorinio perstatymo uždaviniams siūloma artimiausius kaimynus generuoti pagal formulę:

$$s' = s \times [1 + \text{Rand}(-1,1)]. \quad (1.51)$$

Sprendžiant bet kurį uždavinį kiekvienam jų nėra vieno konkretaus geriausio algoritmo rasti optimalų sprendinį. Kiekvienai konkrečiai situacijai tenka parinkti geriausią tam uždaviniui spręsti variantą, kuris su tam tikru parametru rinkiniu rastų priimtinausią globalinį ekstremumą. Ši problema ypatingai aktuali, kai nagrinėjamas ne vieno kintamojo optimizavimo uždavinys. Laiko eilučių prognozėje ieškomas ne vienas laiko vėlinimas, o  $d-1$  optimalių laiko vėlinimų, kurių kiekvienas gali būti skirtingas dydis, apsprendžiantis to rekonstruojamo atraktoriaus topologiją laiko vėlinimų erdvėje. Dėl to sprendžiant kiekvieną konkretų uždavinį reikės atitinkamai parinkti geriausiomis savybėmis pasižymintį optimizavimo algoritmą, tinkantį konkrečiam uždaviniui, konkrečioje situacijoje, prie konkrečių parametru rinkinių.



**1.3 pav.** Atkaitinimo algoritme parenkamos pradinės temperatūros reikšmė: A – kai parenkama nei per didelė, nei per maža temperatūra; B – kai parenkama per maža temperatūra; C – kai pasiekiami temperatūra, prie kurios sistema „užšąla“.

## 2. LAIKO EILUČIŲ PROGNOZAVIMAS REMIANTIS CHAOTINIŲ PROCESŲ SAVYBĖMIS

### 2.1. Kriterijaus, nusakančio nereguliaraus vėlinimo optimalumą, konstravimas

Laiko eilučių prognozėje šiame darbe planuojama pritaikyti nereguliaraus rekonstravimo į laiko vėlinimų erdvę metodą, t.y. kai rekonstruojant laiko eilutę tarp gretimų koordinačių bus parinkti nebūtinai vienodi laiko vėlinimai. Tokio pasirinkimo priežastis yra įsitikinimas, jog nereguliarių laiko vėlinimų rekonstravimas gali leisti geriau reprezentuoti chaotinio atraktoriaus geometrinę realizaciją laiko vėlinimų erdvėje. Ne vienas autorius yra pastebėjęs, kad nereguliaraus rekonstravimas gali pagerinti stebimos sistemos funkcionalumą, tačiau dažniausiai šis uždavinys giliau nebūdavo nagrinėjamas [102]. Schreiber [139] parodė, kad nereguliaraus rekonstravimas yra geresnis rekonstruojant atraktorius daugiamatėje koordinačių erdvėje, kai laiko eilutės iššaukia keletą neproporcingų dažnių. Neseniai toks nereguliaraus rekonstravimas buvo panaudotas laiko eilučių prognozavimui [140]. Deja, nereguliaraus rekonstravimo laiko vėlinimų  $\tau_1, \tau_2, \dots, \tau_{d-1}$  parinkimui lengviau naudoti euristinius skaičiavimo metodus (pavyzdžiui, perspektyviniai neuroniniai tinklai su struktūriniu apmokymu [140]) nei autokoreliaciją ar bendra informaciją paremtus metodus, kurie tiko reguliaraus rekonstravimo atveju dinaminiams ryšiams tarp kintamųjų nustatyti. Autokoreliacijos ir kitų panašių metodų funkcijų skaičiavimas tampa nepraktiškas nereguliaraus rekonstravimo atveju, ypač kai rekonstruotos vėlinimų erdvės matavimas yra didelis. Vien tik iš geometrinių samprotavimų, pavyzdžiui, kokį turį užima rekonstruotas atraktorius nereguliariais laiko vėlinimais rekonstruotoje vėlinimų erdvėje, tokių kriterijų geometrine prasme taikymas yra nekorektiškas ir yra labai sunkiai įgyvendinamas. Taip yra dėl to, kad rekonstruojamo atraktoriaus laiko vėlinimų erdvėje trajektorijos konstruojamas objektas ganėtinai sudėtingas. Algoritmų, kurie nustatinėja tokio objekto turį, teorinis pagrindimas yra komplikuotas, o realizacija būtų labai sudėtinga.

Planuojant ieškoti optimalaus laiko vėlinimų vektoriaus rekonstruotai vėlinimų erdvei, atsiranda poreikis *kurti naujus modelius*, kurie leistų greitai, efektyviai ir pakankamai tiksliai nustatyti optimalius laiko vėlinimų parametrus.

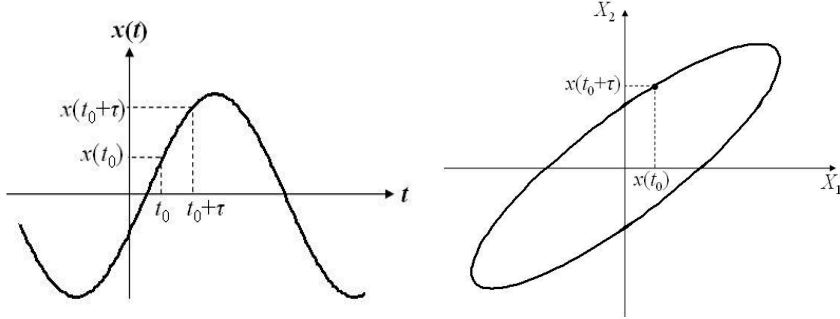
Tarkime, kad duota skaliarinė laiko eilutė  $x_1, x_2, \dots, x_N$ . Tuomet laiko vėlinimo vektoriai, gaunami nereguliariai rekonstruojant į vėlinimų erdvę, gali būti konstruojami taip ([141]):

$$\begin{aligned} \mathbf{y}_p(d) &= (x_p, x_{p+\tau_1}, \dots, x_{p+\tau_1+\dots+\tau_{d-1}}), \\ p &= 1, 2, \dots, N - \sum_{i=1}^{d-1} \tau_i. \end{aligned} \quad (2.1)$$

čia  $d$  yra rekonstravimo į vėlinimų erdvę matas;  $\mathbf{y}_p(d)$  yra  $p$ -tasis rekonstruotas vektorius, kurio matavimas  $d$ ;  $\tau_i \in \mathbb{N}$  vadinami laiko vėlinimais. Jei laiko intervalas tarp dviejų gretimų taškų skaliarinėje laiko eilutėje yra  $\delta$  (diskretinio matavimo dažnis  $1/\delta$ ), tai  $\delta \sum_{i=1}^{d-1} \tau_i$  yra bendras rekonstravimo į laiko vėlinimų erdvę langas.

### 2.1.1. Rekonstrukcija į dvimatę vėlinimo koordinačių erdvę

Tikslo funkcijos konstravimui šame darbe naudojami geometriniai atitikmenys. Tikslas yra kaip galima plačiau išskleisti atraktorių laiko vėlinimų erdvėje. Taigi reikia sukonstruoti analitinį kriterijų, kuris paaikšintų atraktoriaus išskleidimo vėlinimų erdvėje kokybę (reikšmingumą). Pradžioje nagrinėjame dvimatę vėlinimų koordinačių erdvę ir tirime harmoninės laiko eilutės sugeneruoto atraktoriaus geometrinę formą, kuri gaunama eilutę rekonstruojant į vėlinimų erdvę (2.1 pav.).



**2.1. pav.** Laiko eilutės rekonstravimo schema į dvimatę laiko vėlinimų erdvę

Diskrečioji harmoninė laiko eilutė gali būti išreiškiama tokiu pavidalu

$$x_s = A \sin(\omega \delta(s-1) + \varphi), \quad s = 1, 2, \dots, N \quad (2.2)$$

čia  $A$  yra amplitudė,  $\omega$  ciklinis dažnis,  $\varphi$  harmoninės funkcijos fazė, o  $\delta$  – laiko intervalas tarp dviejų gretimų taškų skaliarinėje laiko eilutėje. Funkcijos reikšmės  $x_i$  ir  $x_{i+\tau}$  ( $i = 1, 2, \dots, N - \tau$ ) vėlinimų plokštumoje  $X_1 O X_2$  pažymimos tašku  $(x_i, x_{i+\tau})$ , kur  $\tau$  yra laiko vėlinimas, o  $X_1$  ir  $X_2$  yra vėlinimų koordinačių plokštumos ašys. Elementarūs matematiniai patvarkymai parodo, kad harmoninė laiko eilutė rekonstruotoje laiko vėlinimų plokštumoje atvaizduojama į elipsę, kurios lygtis turi išraišką (2.2. pav.):

$$X_2 = X_1 \cos(\omega \tau \delta) + \sqrt{A^2 - X_1^2} \sin(\omega \tau \delta), \quad (2.3)$$

Galima pastebėti, kad pagrindinės elipsės simetrijos ašys yra  $X_2 = X_1$  ir  $X_2 = -X_1$ . Jeigu rekonstravimo į vėlinimų erdvę langas lygus

$$\tau \delta = \frac{2\pi}{\omega}(n-1), \quad n = 1, 2, \dots; \quad (2.4)$$

elipsė yra suspaudžiama į liniją ant diagonalės  $X_2 = X_1$ . Lygiai taip pat, kai

$$\tau \delta = \frac{\pi}{\omega} + \frac{2\pi}{\omega}(n-1), \quad n = 1, 2, \dots; \quad (2.5)$$

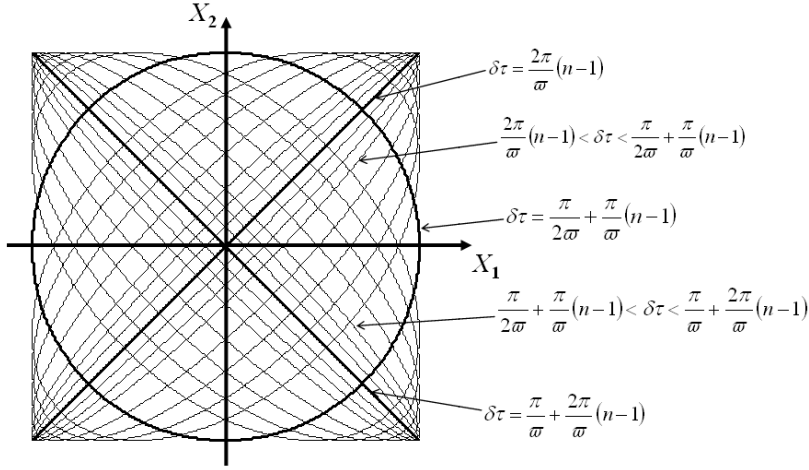
elipsė suspaudžiama į liniją ant diagonalės  $X_2 = -X_1$ . Galiausiai, jei

$$\tau \delta = \frac{\pi}{2\omega} + \frac{\pi}{\omega}(n-1), \quad n = 1, 2, \dots; \quad (2.6)$$



harmoninė laiko eilutė laiko vėlinimo koordinačių plokštumoje atvaizduojama į apskritimą.

Toliau atraktoriaus geometrinė forma bus panaudota sudarant rekonstravimo į vėlinimų erdvę kokybės parametą.



**2.2. pav.** Harmoninio signalo rekonstravimo į dvimatę fazinę erdvę galimos geometrinės realizacijos priklausomai nuo parenkamo laiko vėlinimo).

### 2.1.2. Rekonstravimo į laiko vėlinimų erdvę kokybės parametro sudarymas

Skyriaus pradžioje buvo minėta, kad rekonstruojant atraktorių į laiko vėlinimų erdvę siekiama, jog jo vaizdas erdvėje būtų kuo turtingesnis. Šiuo atveju rekonstruojant harmoninį signalą į dvimatę vėlinimų plokštumą, siekiama, kad gautos elipsės plotas būtų kuo didesnis. Taigi apibrėžiant parametą  $Q_1$ , charakterizuojantį atraktoriaus dinamiką rekonstruotoje erdvėje, naudojami geometriniai skaičiavimai:

$$Q_1 = \frac{S_E}{\pi A^2}; \quad (2.7)$$

čia  $S_E$  rekonstruotos dvimatės vėlinimų erdvės plokštumoje atvaizduotos elipsės plotas. Akivaizdu, kad  $0 \leq Q_1 \leq 1$ . Jei  $Q_1 = 1$ , tai gaunamas optimalus rekonstruoto objekto atvejis, kai harmoninė laiko eilutė atvaizduojama į apskritimą. Blogiausias atvejis gaunamas su  $Q_1 = 0$ , kai elipsė suspaudžiama į liniją ant įstrižainių. Atvaizduotos elipsės pusašės  $R_1$  ir  $R_2$  iš 2.3 lygties gali būti išreikštos tokiu būdu:

$$\begin{aligned} R_1 &= A \sin(\omega \tau \delta) / \sqrt{1 - \cos(\omega \tau \delta)}; \\ R_2 &= A \sin(\omega \tau \delta) / \sqrt{1 + \cos(\omega \tau \delta)}. \end{aligned} \quad (2.8)$$

Tada elipsės plotas:

$$E = \pi \cdot R_1 R_2 = \pi A^2 |\sin(\omega \tau \delta)|. \quad (2.9)$$

Tokiu būdu kokybės parametras  $Q_1$  įgyja pavidalą

$$Q_1 = |\sin(\omega\tau\delta)|. \quad (2.10)$$

### 2.1.3. Daugiamatė vėlinimo koordinačių erdvė, reguliaraus rekonstravimo į laiko vėlinimų erdvę atvejis

Jeigu vėlinimo koordinačių erdvės matavimas yra  $d$ , tuomet egzistuoja  $\frac{d(d-1)}{2}$  skirtingų rekonstruoto atraktoriaus plokštuminių projekcijų. Jeigu vėlinimo koordinačių erdvės ašys numeruojamos tokiu būdu  $X_i$ ,  $i=1,\dots,d$ , tuomet harmoninės laiko eilutės projekcija bus elipsė  $X_iOX_j$  plokštumoje, kur  $1 \leq i \leq d$ ;  $1 \leq j \leq d$ ;  $i \neq j$ . Reguliarus laiko eilutės rekonstravimas į laiko vėlinimų erdvę gaunamas tuomet, kai

$$\tau = \tau_1 = \dots = \tau_{d-1}. \quad (2.11)$$

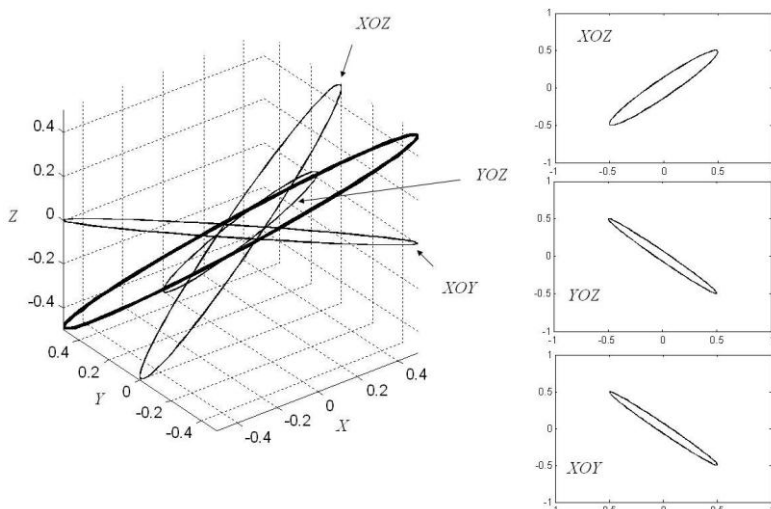
Apibrėžkime rekonstravimo į vėlinimų erdvę kokybės parametro reikšmes kiekvienai galimai  $i$  ir  $j$  kombinacijai:

$$Q_{ij} = |\sin(\omega(j-i)\tau\delta)|. \quad (2.12)$$

Rekonstruojant harmoninę laiko eilutę į trimatę laiko vėlinimų erdvę 2.3 paveiksle iliustruojama kaip atrodo tos plokštuminės projekcijos erdvėje ir kiekvienoje projekcijoje atskirai. Galima pastebėti, kad kai tarp vėlinimų koordinačių yra vienodas laiko atsilikimas, projekcijos yra visiškai vienodos ( $XOY$  ir  $YOZ$  plokštumose laiko vėlinimas yra  $\tau$ ). Tuo atveju, kai laiko vėlinimas yra kitoks, vėlinimų erdvės projekcija atrodo kitaip ( $XOZ$  plokštumoje laiko vėlinimas yra lygus  $2\tau$ ).

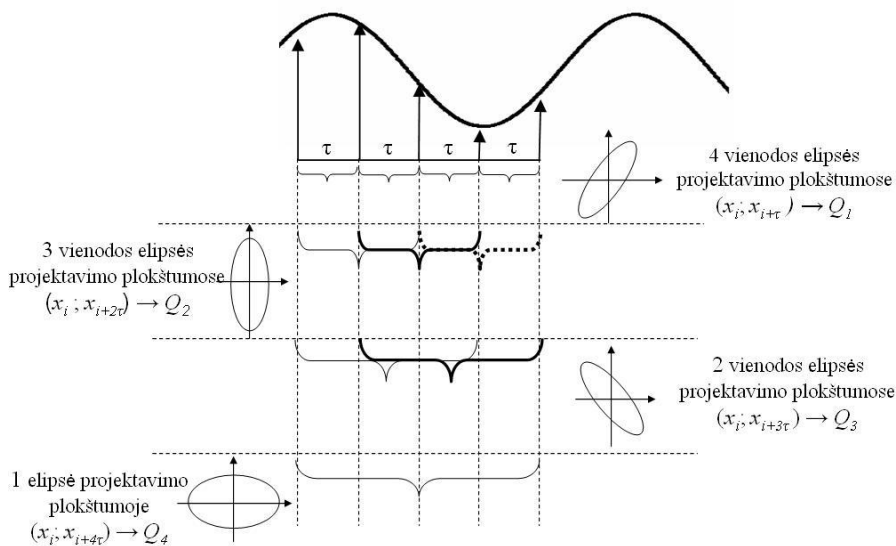
Kadangi  $Q_{ij}$  reikšmė priklauso tik nuo skirtumo tarp  $j$  ir  $i$ ; tai žymime:

$$\begin{aligned} Q_1 &= Q_{ij} \Big|_{|j-i|=1}, \\ Q_2 &= Q_{ij} \Big|_{|j-i|=2}, \\ &\dots \\ Q_{d-1} &= Q_{ij} \Big|_{|j-i|=d-1}. \end{aligned} \quad (2.13)$$



**2.3. pav.** Harmoninės laiko eilutės rekonstruotos trimatės laiko vėlinimų erdvės objekto visos galimos projekcijos į plokštumas

Tarkime, kad harmoninę laiko eilutę reikia rekonstruoti į penkiamatę laiko vėlinimų erdvę, be to ši laiko eilutė rekonstruojama reguliariais laiko vėlinimais. 2.4 paveiksle demonstruojama schema, kad prie vienodų laiko vėlinimo reikšmių laiko eilutės vėlinimų erdvės plokštuminės projekcijos yra vienodos. Į šį faktą atsižvelgiama konstruojant kokybės parametro įvertinimo funkciją.



**2.4. pav.** Kokybės parametrų reikšmių priklausomybės nuo laiko vėlinimo schema

Toliau darbe rekonstravimo į vėlinimų erdvę kokybės funkcija apibrėžiama kaip visų galimų plokštuminių projekcijų rekonstravimo kokybės parametrų vidurkis:

$$Q(\tau, \omega) = \frac{2}{d(d-1)} \sum_{k=1}^{d-1} (d-k) Q_k = \frac{2}{d(d-1)} \sum_{k=1}^{d-1} (d-k) |\sin(k\omega\tau\delta)|. \quad (2.14)$$

Akivaizdu, kad  $0 \leq Q(\tau, \omega) \leq 1$ . Lygybė  $Q=0$  reiškia, kad harmoninė eilutė suspaudžiama į atkarpą visose galimose projekcijose. Harmoninei laiko eilutei  $Q=0$  reikšmė įgaunama, kai

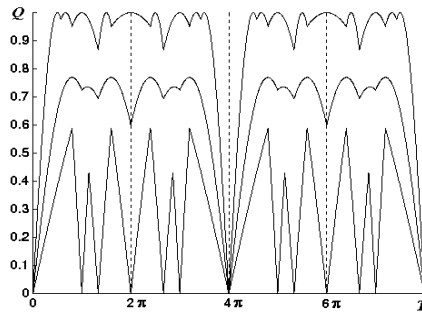
$$\omega = \frac{\pi}{\tau\delta} \cdot m, \quad m = 1, 2, \dots \quad (2.15)$$

Galima pastebėti, kad šie dažniai nepriklauso nuo vėlinimų erdvės matavimo  $d$ . Dar daugiau, integralas

$$\int_{\frac{\pi}{\tau\delta} m}^{\frac{\pi}{\tau\delta} (m+1)} Q(\tau, \omega) d\omega = \frac{2}{d(d-1)} \sum_{k=1}^{d-1} \left( (d-k) \int_{\frac{\pi}{\tau\delta} m}^{\frac{\pi}{\tau\delta} (m+1)} |\sin(k\omega\tau\delta)| d\omega \right) = \frac{2}{\tau\delta}, \quad m \in \mathbf{N}, \quad (2.16)$$

yra invariantas  $d$  atžvilgiu.

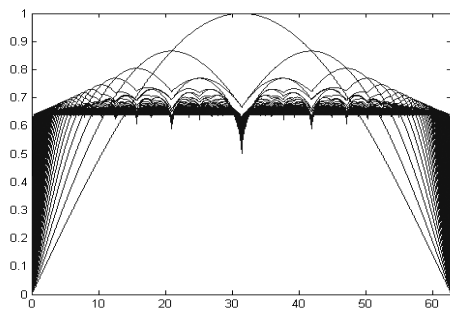
Žemiau pateiktame 2.5. paveikslo grafike yra vaizduojami harmoninio signalo kokybės funkcijos keli galimi pavyzdiniai atvejai, kai rekonstruotos laiko vėlinimų erdvės matavimas  $d=5$ . Rekonstravimo į vėlinimų erdvę kokybės funkcija suformuota kaip visų dvimačių kokybės funkcijų aritmetinis vidurkis visose galimose vėlinimų erdvės plokštumų projekcijose. Šiame pavyzdyje buvo atlikta  $\frac{d(d-1)}{2} = 10$  kokybės funkcijos realizacijų. Kokybės funkcijos didžiausių reikšmių kitimas kintant laiko dažniui vaizduojamas viršutine kreive (periodas  $T = (d-1)\tau = 4\tau$ ). Mažiausios kokybės funkcijos reikšmės pavaizduotos grafiko apatinė kreivė. Viduryje vaizduojamos kokybės funkcijos vidutinės reikšmės.



**2.5. pav.** Rekonstravimo į penkiamatę vėlinimų erdvę kokybės funkcijos grafinis vaizdavimas.

Rekonstravimo į laiko vėlinimų erdvę kokybės funkcijos priklausomybė nuo ciklinio dažnio prie skirtingų rekonstruotos erdvės dimensijų vaizduojama 2.6 paveiksle. Čia matomas kokybės funkcijos vaizdas viename periode esant skirtingam vėlinimų erdvės matavimui  $d$ . Kai vėlinimų erdvės matas 2,  $Q$  funkcija yra sinusoidės vieno lanko funkcija. Kai kokybės funkcijos reikšmė pasiekia vienetą, rekonstruojamas harmoninis signalas bus išplėstas maksimaliai

– iki apskritimo. Didinant matavimo skaičių, kokybės funkcijos forma darosi vis sudėtingesnė. Kai  $d$  artėja į begalybę, kokybės funkcija įgauna į fraktalinių savybių.



**2.6. pav.** Rekonstravimo į vėlinimų erdvę kokybės funkcijos priklausomybė nuo laiko vėlinimų ir vėlinimų erdvės mato.

#### **2.1.4. Daugiamatė vėlinimo koordinačių erdvė, nereguliaraus rekonstravimo į laiko vėlinimų erdvę atvejis. Tikslų funkcijos sudarymas.**

Šiame skyriuje nagrinėjamas nereguliaraus rekonstravimo į laiko vėlinimų erdvę atvejis. Skirtingų projekcijų skaičius  $d$ -matėje vėlinimo koordinačių erdvėje vis dar  $\frac{d(d-1)}{2}$ , bet apibendrinimas, panaudotas (2.13) sistemoje, kai projekcijų kokybės funkcijos  $Q_{ij}$  reikšmės priklausė tik nuo skirtumo tarp  $j$  ir  $i$ , šiuo atveju nebetinkamas. Todėl rekonstravimo į vėlinimų erdvę kokybės funkcija yra priklausoma nuo parametrų  $\tau_1, \dots, \tau_{d-1}, \omega$ . Tuomet

$$Q(\tau_1, \dots, \tau_{d-1}, \omega) = \frac{2}{d(d-1)} \left( \sum_{i=1}^{d-1} |\sin(\omega \delta \tau_i)| + \sum_{i=1}^{d-2} |\sin(\omega \delta (\tau_i + \tau_{i+1}))| + \dots + \left| \sin \left( \omega \delta \sum_{j=1}^{d-1} \tau_j \right) \right| \right). \quad (2.17)$$

Reikėtų atkreipti dėmesį, kad iki šiol nagrinėtos tik harmoninės funkcijos. Akivaizdu, kad darbo tikslas yra prognozuoti bet kokius signalus. Šiuo atveju, jeigu būtų nagrinėjamas kurios nors laiko eilutės Furjė amplitudinis spektras, būtų galima pastebėti, kad tas signalas susideda iš daugelio harmoninių komponentų. Įvertinant harmoninio signalo transformaciją rekonstruojant harmoninį procesą į elipsę dvimatėje vėlinimų erdvėje galima pastebėti, kad visiškai nesvarbu, nuo kurios pradinės fazės pradedamas šis rekonstrukcijos procesas. Taigi, visą informaciją apie atraktoriaus formą teikia rekonstruojamos harmoninės eilutės amplitudė, o ne fazė. Lygiai ta pati idėja perkeliama sudėtingam signalui, kuris susideda iš visos eilės harmoninių komponentų. Galima pastebėti, kad kiekvieną iš atskirai paimtų harmoninių komponentų prie atitinkamai parinktų laiko vėlinimų  $\tau_1, \tau_2, \dots, \tau_{d-1}$  specifiskai paveikia rekonstravimo kokybės funkcija  $Q(\tau_1, \dots, \tau_{d-1}, \omega)$ . Konkretaus dažnio harmoninis signalas atitinkamai atvaizduojamas į tam tikrą elipsės visose skirtingose projektavimo plokštumose. Jeigu nagrinėjamą signalą galima vertinti kaip tiesinį darinį iš įvairių harmoninių komponentų, tai galima pritaikyti šį

optimalaus harmoninės eilutės rekonstravimo įvertį būtent tam sudėtingo signalo rekonstrukcijos įverčiui kurti, t.y. konstruojama tokia tikslo funkcija, kuri įvertina pirmiausia, kaip atskiros diskrečiosios amplitudinio spektro komponentės yra paveikiamos rekonstrukcijos kokybės funkcijos prasme. Tai reprezentuoja netiesioginis integralas šio įverčio skaitiklyje, kai perbėgami visi dažniai ir suskaičiuojama kokias Furjė amplitudinio spektro komponentes labiausiai paveikė kokybės funkcija  $Q$ . Laikantis šių prielaidų apibrėžiama tikslo funkcija  $F$ , kuri charakterizuoja atraktoriaus vidutinį tūrį rekonstruotoje vėlinimų erdvėje:

$$F(\tau_1, \dots, \tau_{d-1}) = \frac{\pi}{2} \frac{\int_0^\infty A(\omega) Q(\tau_1, \dots, \tau_{d-1}, \omega) d\omega}{\int_0^\infty A(\omega) d\omega}. \quad (2.18)$$

čia  $A(\omega)$  yra laiko eilutės Furjė amplitudinis spektras. Netiesioginiai integralai skaičiuojami visoje dažnių juostoje (praktiškai tegalima skaičiuoti tik apibrėžtinius integralus laikant, kad rėžiai turi siekti viršutinę dažnių ribą). Vardiklyje  $A(\omega)$  integralas naudojamas tik tam, kad būtų normuojama tikslo funkcijos reikšmė (šis integralas konkrečiai laiko eilutei skaičiuojamas vieną kartą prieš optimizacijos procedūrą). Tikslo funkcijos skaitinę reikšmę, gautą pagal (2.18) formulę, darbe naudojama kaip rodiklis rodantis prie kokio nereguliaraus laiko vėlinimų rinkinio

laiko eilutės atraktorius užima didžiausią tūrį vėlinimų erdvėje. Daugiklis  $\frac{\pi}{2}$  yra naudojamas tam, kad tikslo funkcijos reikšmė rekonstruojant baltąją triukšmą į  $d$ -matę vėlinimų erdvę taptų lygi vienetui.

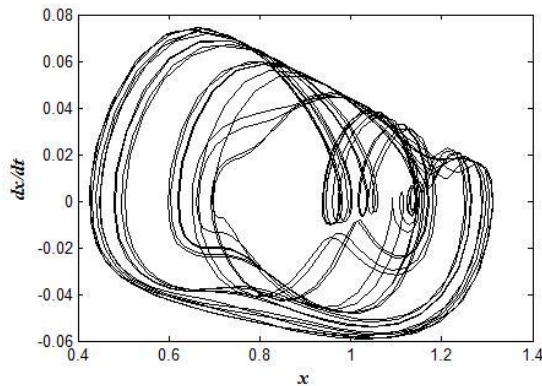
Taigi tikslo funkcijos uždavinys yra parinkti tokį laiko vėlinimų rinkinį, su kuria tikslo funkcijos reikšmė bus maksimizuojama, t.y. rekonstruotas atraktorius duos turtingiausią vaizdą laiko vėlinimų erdvėje, t.y. šioje erdvėje atraktoriaus tūris bus didžiausias. Tokiu būdu yra gaunama daugiausia informacijos apie tiriamos sistemos kompleksiskumą. Minimizavimo uždavinys čia nesprenžiamas dėl to, kad minimalaus tūrio suspausta sistemos evoliucija bus beveik identiška visose koordinatėse, todėl informacija apie sistemos kompleksiskumą bus prarasta.

Aiškumo dėlei, toliau demonstruojamas paprastas pavyzdys, kuriame skaičiuojama rekonstravimo į laiko vėlinimų erdvę kokybės funkcija. Taip pat parodoma tikslo funkcijos reikšmės priklausomybė nuo rekonstruotos tam tikro matavimo erdvės laiko vėlinimo. Tarkime, kad duota netiesinė diferencialinė lygtis su vėluojančiu argumentu

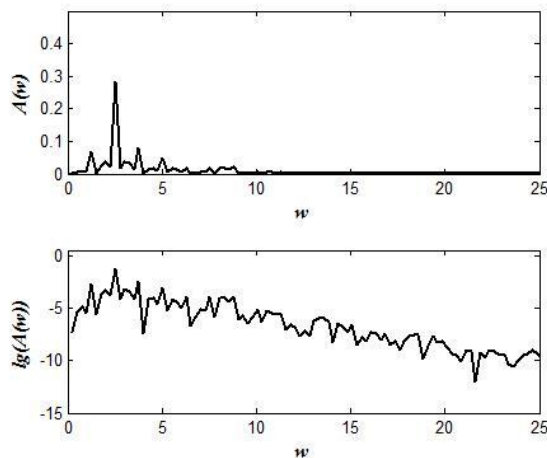
$$\frac{dx(t)}{dt} = \frac{0,2x(t-17)}{1+x^{10}(t-17)} - 0,1x(t). \quad (2.19)$$

Tai etaloninės Mackey-Glass chaotinės laiko eilutės konkretus atvejis, kuris bus minimas vėliau. Tarkime, kad duotas ir diferencialinės lygties sprendinys fazinėje plokštumoje  $\left\{x; \frac{dx}{dt}\right\}$  (žiūrėti 2.7 paveikslą). Sprendiniui  $x(t)$  suformuojamas Furjė amplitudinis spektras skaičiuojamas panaudojant Matlab v.7.0.

skaičiavimo terpės įrankiuose pateikiamas programines priemonės. Šio amplitudinio spektro grafikas pavaizduotas 2.8 paveiksle.



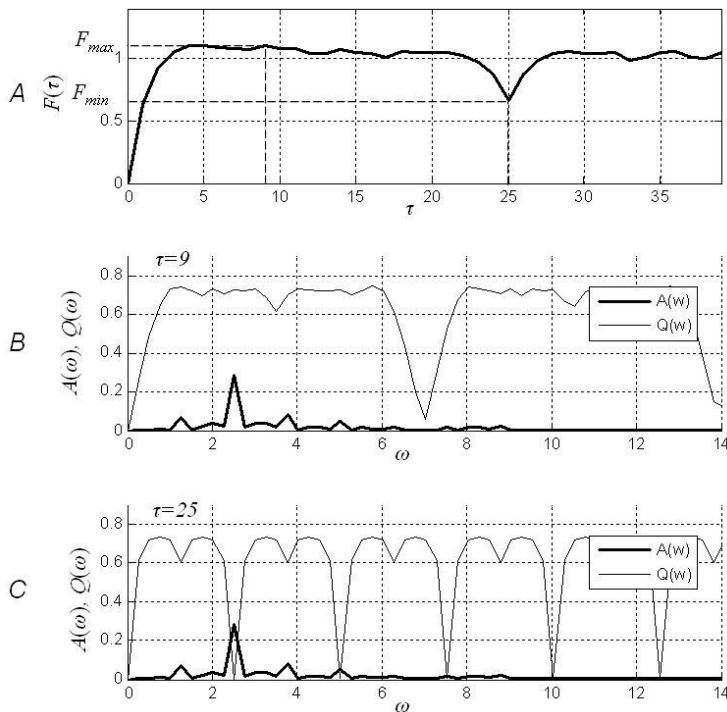
**2.7. pav.** Chaotinės Mackey-Glass diferencialinės lygties dalinis sprendinys fazinėje plokštumoje  $x$ ;  $dx/dt$ .



**2.8 pav.** Dalinio sprendinio  $x(t)$  Furjė amplitudinis spektras tiesinėje ir logaritminėje skalėse

Tarkime, kad chaotinis atraktorius reguliariais laiko vėlinimais rekonstruojamas į vėlinimų erdvę (kai visi laiko vėlinimai tarp gretimų rekonstruojamų komponentų yra vienodi). Toliau skyreliuose bus parodyta, kad rekonstruotos vėlinimų erdvės matavimas  $d = 6$ . Kokybės funkcija įgyja pavidalą, demonstruojama 2.9 paveikslų B ir C dalyse. Jeigu skaičiuojant tikslo funkciją pasirenkamas laiko vėlinimas, kuriame ši funkcija duoda globalų maksimumą tiriamoje srityje, tuomet kokybės funkcija išskleidžiama plačiausiai. Furjė amplitudinis spektras tai pačiai laiko eilutei visada vienodas. Taigi šiuo atveju gaunama, jog jei laiko vėlinimas parenkamas optimalus, Furjė amplitudinio spektro pikas sutampa su kokybės funkcijos pirmuoju pikų (2.9 pav. B atvejis). Jei kokybės funkcijos braižymui parenkamas laiko vėlinimas, prie kurio tikslo funkcija įgijo lokalų minimumą, tada gaunamas suspaustas grafikas, o Furjė amplitudinio spektro pikas sutampa su kokybės funkcijos mažiausia reikšme (2.9 paveikslų B atvejis).

Tikslo funkcijos apibrėžime Furjė amplitudinis spektras  $A(\omega)$  nepriklauso nuo kintamojo  $\tau$ . Tačiau kokybės funkcija  $Q(\tau, \dots, \tau, \omega)$  priklauso, ir nors  $\tau$  kitimas nekeičia funkcijos  $Q(\tau, \dots, \tau, \omega)$  formos, bet skalė kinta ( $Q(\tau, \dots, \tau, \omega)$  grafikas yra suspaudžiamas arba išplečiamas  $\omega$  ašies kryptimi). Tikslas yra parinkti tokį  $\tau$ , kad  $A(\omega)Q(\tau, \dots, \tau, \omega)$  rezultatas būtų maksimalus.

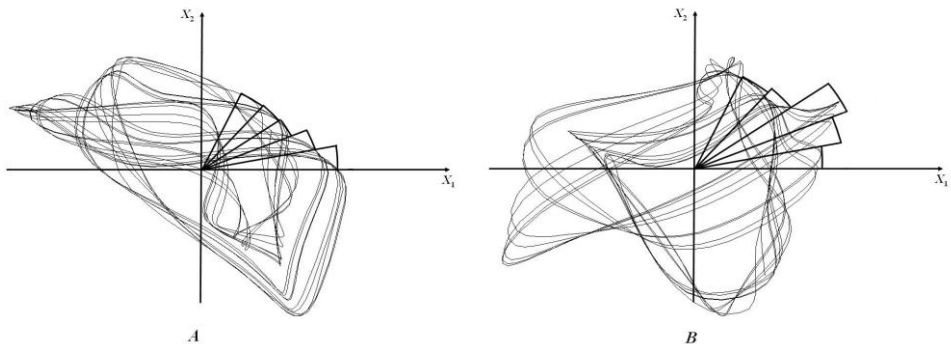


**2.9 pav.** Atraktoriaus skleidimas laiko vėlinimų erdvėje: A - tikslo funkcijos kitimas keičiantis laiko vėlinimo reikšmei; B - kokybės funkcijos kreivė, kai parenkamas laiko vėlinimas esant pirmai didžiausiai tikslo funkcijos reikšmei; C - kokybės funkcijos kreivė, kai parenkamas laiko vėlinimas esant mažiausiai tikslo funkcijos reikšmei

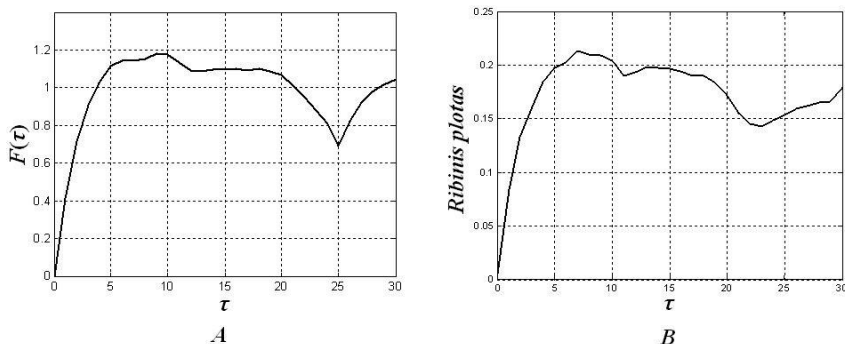
Sukonstravus tokią metodiką optimaliam vėlinimų rinkiniui parinkti, kurią galima pritaikyti konkrečiai laiko eilutei, galima pabandyti tą metodiką patikrinti, konkrečiam signalui. Pavyzdžiui, tiriama Mackey-Glass laiko eilutė. Tikrinimo procesą atliekamas rekonstruojant laiko eilutę į trimatę laiko vėlinimų erdvę. Gaunamos dvi skirtingos projekcijos, kurios vaizduojamos 2.10 paveiksle. Dabar signalo amplitudinis spektras neskaičiuojamas, o tiesiog tiesmukiškai matuojamas to atraktoriaus užimamas plotas plokščioje projekcijoje. Tam visa polinė koordinatų sistema suskirstoma į atitinkamas skritulio išpjovas – „skilteles“, ir kiekvienoje „skilteleje“ randamas atraktoriaus taškas, kuris toje srityje yra labiausiai nutolęs nuo koordinatų pradžios taško. Gautas atstumas laikomas tos „skiltelės“ spinduliu. Tokiu būdu sukonstruojama iš skiltelių 2.10 paveiksle vaizduojama struktūra, kuriai galima tiesiogiai išmatuoti gaunamos projekcijos plotą. Toks skaičiavimas leidžia palyginti siūlomo kriterijaus tikslo funkcijos rezultatus su ploto matavimo rezultatais. Šių kriterijų palyginimas pateikiamas 2.11 paveiksle. Galima pastebėti, kad tikslo funkcijos kriterijaus grafikas šiek tiek skiriasi nuo plotų skaičiavimo



kriterijaus grafiko, tačiau reikia nepamiršti, kad atlikant Furjė analizę buvo priimta prielaida, kad signalas yra periodinis ir imtas tik tam tikro ilgio laiko eilutės duomenų vektorius, t.y. skaičiavimuose naudoti tik tam tikro periodo duomenys. Tačiau žinoma, kad reali laiko eilutė yra ne periodinė, o chaotinė. Tai vien dėl to 2.11 paveikslą grafikai skiriasi. Be to, čia buvo atlikta šio signalo dekompozicija į harmonines komponentes nevertinant fazių skirtumo tarp tų signalų. Iš tikrųjų tie fazių skirtumai yra, ir jie apibrėžti atitinkamame faziniame spektre. Tie lokaliniai skirtumai irgi gali turėti tam tikros įtakos, todėl šie grafikai ir skiriasi. Tačiau galima pastebėti, kad pagrindiniai ekstremumai lieka tokie patys. Turint omenyje, kad ši metodika bus taikoma laiko eilučių prognozei, kur ne visada galima nagrinėti labai ilgas eilutes ir negalima konstruoti ilgų atraktoriaus geometrinių realizacijų. Tai viena iš priežasčių kodėl darbe taikoma Furjė analizė. Antra priežastis, svarbus skaičiavimo greitis. Negalima sau leisti kiekvieną kartą prognozuojant sekantį eilutės elementą ilgai ieškoti kriterijaus reikšmių. Šiuo atžvilgiu sukurtas kriterijus greičio prasme turi didelį pranašumą lyginant su tiesmukišku plotų skaičiavimu plokščiose projekcijose. Nagrinėjant tą patį signalą, Furjė transformaciją tereikia paskaičiuoti tik vieną kartą ir galima toliau dirbti su skirtingais laiko vėlinimo parametrais. Jeigu būtų dirbama su realiais plotais, kiekvieną kartą tektų perskaičiuoti jų sumas keičiant laiko vėlinimų rinkinį. Tas yra visiškai nepriimtina kombinatorinio optimizavimo atžvilgiu.



**2.10 pav.** Scheminė diagrama, iliustruojanti Mackey-Glass laiko eilutės plokštuminės projekcijos ploto skaičiavimą: A - laiko vėlinimas tarp vėlinimų koordinatų yra  $\tau$  ; B - laiko vėlinimas tarp vėlinimų koordinatų yra  $2\tau$  .



**2.11 pav.** Mackey-Glass atraktoriaus rekonstravimo į trimatę vėlinimų erdvę optimalaus laiko optimalaus vėlinimo radimas su A – tikslo funkcijos pagalba, B – ribinės sumos pagalba.

Reikia pažymėti, kad tiek kokybės funkcijos, tiek kriterijaus, nustatančio atraktoriaus atvaizdavimo į daugiamatę erdvę (tiek reguliaraus, tiek ir nereguliaraus atvaizdavimo atveju) optimalius vėlinimus, sąvokos yra šios disertacijos siūlomos originalios sąvokos.

### 2.1.5. Tikslų funkcijos savybės

a)  $F(0, \dots, 0) = 0$ . Ši savybė išplaukia iš rekonstravimo į laiko vėlinimų erdvę kokybės funkcijos (2.17) apibrėžimo.

b) Jei  $A(\omega) = L$  („baltas triukšmas“), kur  $L$  yra teigiama konstanta, o rekonstruojamos vėlinimų erdvės matavimas  $d = 2$ , tuomet visiems  $\tau \geq 1$  teisinga  $F(\tau) = 1$ .

Išrodykime šį teiginį. Tarkime, kad laiko vėlinimų erdvės matavimas  $d = 2$ , tuomet rekonstravimo į vėlinimų erdvę kokybės funkcija yra  $Q(\tau, \omega) = |\sin(\omega\tau\delta)|$ .

Funkcijos  $A(\omega)Q(\tau, \omega)$  grafiko ir  $O\omega$  užimamas plotas  $S_1$ , kurio forma gali būti iliustruojama 2.12 paveikslu, randamas skaičiuojant tokį apibrėžtinį integralą:

$$S_1 = \int_0^{\frac{\pi}{\tau\delta}} A(\omega)Q(\tau, \omega) d\omega = \int_0^{\frac{\pi}{\tau\delta}} L \sin(\omega\tau\delta) d\omega = \frac{2L}{\tau\delta}. \quad (2.20)$$

Tarkime, kad harmoninio signalo vienos srities  $S_1$ , kurios periodas  $\left[0, \frac{\pi}{\tau}\right]$ ,

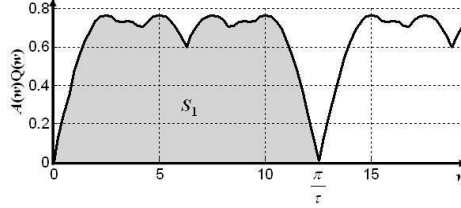
plotas lygus:  $S_1 = \int_0^{\frac{\pi}{\tau\delta}} \sin(\omega\tau\delta) d\omega = \frac{2}{\tau\delta}$ . Tuomet tikslo funkcija skaičiuojama tokiu būdu:

$$\begin{aligned} F(\tau) &= \frac{\int_0^{\frac{\pi}{\tau\delta}} A(\omega)Q(\tau, \omega) d\omega}{\int_0^{\frac{\pi}{\tau\delta}} A(\omega) d\omega} = \lim_{T \rightarrow \infty} \frac{\int_0^T |\sin(\omega\tau\delta)| d\omega}{\int_0^T d\omega} = \\ &= \lim_{n \rightarrow \infty} \frac{\int_0^{\frac{\pi}{\tau\delta} n} |\sin(\omega\tau\delta)| d\omega}{\int_0^{\frac{\pi}{\tau\delta} n} d\omega} + \frac{\int_{\frac{\pi}{\tau\delta} n}^T |\sin(\omega\tau\delta)| d\omega}{\int_{\frac{\pi}{\tau\delta} n}^T d\omega} = \lim_{n \rightarrow \infty} \frac{S_1 \cdot n}{\frac{\pi}{\tau\delta} n} = \frac{2}{\pi}. \end{aligned}$$

Sumoje, prieš suskaičiuojant integralą, vertinami tik sveikų sričių  $S_1$  plotai ir pridedama liekana, kuri konverguoja į nulį.

Tada tikslo funkcijos reikšmė lygi:

$$F(\tau) = \frac{\pi}{2} \lim_{n \rightarrow \infty} \frac{\int_0^{\frac{n\pi}{\tau\delta}} L |\sin(\omega\tau\delta)| d\omega}{\int_0^{\frac{n\pi}{\tau\delta}} L d\omega} = \frac{\pi}{2} \lim_{n \rightarrow \infty} \frac{n S_1 \tau \delta}{\pi n L} = 1 \quad (2.21)$$



**2.12 pav.**  $A(\omega)Q(\tau, \omega)$  grafiko ribojamas plotas  $S_1$ .

Taigi, jei į gautos tikslo funkcijos reikšmę įvedamas daugiklis  $\frac{\pi}{2}$ , įrodomas teiginio teisingumas, kai visi  $\tau \geq 1$ . Tada tikslo funkcija nepriklauso nuo laiko vėlinimų, jei tik skaliarinė laiko eilutė neša baltos triukšmo duomenis – rekonstruojamo atraktoriaus forma bet kuriam laiko vėlinimui yra ta pati.

Teiginys įrodytas, kai  $d = 2$ .

c) Analogiškai, jei  $A(\omega) = L$  („baltas triukšmas“), kur  $L$  yra teigiama konstanta, o rekonstruojamos laiko vėlinimų erdvės matavimas  $d \geq 2$ , ir  $\tau_1 = \dots = \tau_{d-1} = \tau$ ,  $\tau \geq 1$  (reguliaraus rekonstravimo atvejis), tuomet  $F(\tau, \dots, \tau) = 1$ .

Teiginys įrodomas analogiškai, kaip ir  $d = 2$  atveju:

$$F(\tau, \dots, \tau) = \frac{\pi}{2} \sum_{k=1}^{d-1} \left( \frac{2(d-k)}{d(d-1)} \frac{\int_0^{\infty} A(\omega) |\sin(k\omega\tau\delta)| d\omega}{\int_0^{\infty} A(\omega) d\omega} \right) = \quad (2.22)$$

$$= \frac{\pi}{2} \sum_{k=1}^{d-1} \frac{2(d-k)}{d(d-1)} \frac{2}{\pi} = 1.$$

d) Analogiškai rezultatai gaunami ir nereguliaraus rekonstravimo atveju: jei  $A(\omega) = L$  („baltas triukšmas“), kur  $L$  yra teigiama konstanta, o rekonstruojamos laiko vėlinimų erdvės matavimas  $d \geq 2$ , tuomet  $F(\tau_1, \dots, \tau_{d-1}) = 1$ .

Kitaip tariant remiantis b) atvejo įrodymu galima parodyti, kad:

$$F(\tau_1, \dots, \tau_{d-1}) = \frac{2}{d(d-1)} \cdot \frac{\pi}{2} \cdot \frac{1}{\int_0^\infty L d\omega} \left( \sum_{i=1}^{d-1} \int_0^\infty L |\sin(\omega \delta \tau_i)| d\omega + \dots \right. \\ \left. \dots + \int_0^\infty L \left| \sin \left( \omega \delta \sum_{j=1}^{d-1} \tau_j \right) \right| d\omega \right) = \frac{2}{d(d-1)} ((d-1) + \dots + 1) = 1. \quad (2.23)$$

Kiekvienas neapibrėžtinis integralas šioje lygtyje buvo skaičiuojamas lygiai taip pat kaip ir (2.21).

Galima paminėti, kad jei skaliarinė laiko eilutė pasiskirsčiusi pagal tolygųjį skirstinį intervale  $[a, b]$ , ji gali būti rekonstruota į hyper-kubą. Rekonstravimo į vėlinimų erdvę taškų pasiskirstymas bet kurioje hyper-kubo vietoje yra pastovus ir nepriklauso nuo laiko vėlinimų.

e)  $A(\omega)$  yra delta funkcija;  $d = 2$ . Suprantama, kad  $A(\omega) = \delta(\omega_0) = \begin{cases} \infty, & \text{kai } \omega = \omega_0, \\ 0, & \text{kitu atveju,} \end{cases}$  ir  $\int_0^\infty \delta(\omega_0) d\omega = 1$ . Tada,

$$F(\tau) = \frac{\pi}{2} Q(\tau, \omega_0). \quad (2.24)$$

## 2.2. Prognozavimo strategija, eksperimentai

Sukurtas modelis, ieškant optimalios laiko vėlinimų aibės bus taikomas laiko eilutės prognozavimui. Sukonstruoto laiko eilutės prognozavimo modelio, paremto laiko eilutės rekonstravimu nereguliais laiko vėlinimais į vėlinimų erdvę, taikymas bus demonstruojamas dviem chaotinėms laiko eilutėms Mackey-Glass ir Rössler bei realaus pasaulio stebimos žvaigždės kasdienio kintamo ryškumo laiko eilutei [142]. Pirmiausia skaičiuojant 1.2 skyrelyje (1.26) lygtyje apibrėžtą koreliacinę sumą bus surasta, į kokio matavimo vėlinimų erdvę geriausia rekonstruoti laiko eilutę. Tada darbe pasiūlyta metodologija bus nustatomas kiekvienos laiko eilutės rekonstravimo į vėlinimų erdvę optimalus laiko vėlinimų rinkinys. Galiausiai turimi duomenys bus panaudoti laiko eilutės prognozei su MATLAB įrankiuose pateikiama adaptivia neuroniniais tinklais paremta neraiškių išvadų sistemos funkcija ANFIS. ANFIS architektūros trumpą aprašymą galima rasti 1.1 skyrelyje, o detalesnį [24, 74].

Bendras laiko eilutės prognozės modelis demonstruojamas 2.13 paveiksle. Neraiškių logikų sistema prognozės uždavinyje naudojama kaip juodoji dėžė, kurios įėjimų vektorius yra rekonstruotos laiko eilutės reikšmės, o išėjimo reikšmė  $\hat{x}(t + \tau_{d-1})$  yra prognozuojama laiko eilutės reikšmė:

$$\hat{x}(t + \tau_{d-1}) = f(y'_t) \quad (2.25)$$

čia

$$y'_t = \left( x \left( t - \sum_{i=1}^{d-2} \tau_i \right), x \left( t - \sum_{i=2}^{d-2} \tau_i \right), \dots, x(t) \right). \quad (2.26)$$

Bendru atveju neraiškos logikos išėjimas yra funkcija [69, 71, 74]:

$$f(y'_t) = \frac{\sum_{k=1}^L w_k(y'_t) f_k}{\sum_{k=1}^L w_k}; \quad (2.27)$$

kur  $L$  – neraiškos logikos sistemos taisyklių skaičius, kuris priklauso nuo įėjimų vektoriaus ilgio ir užsibrėžtų sąryšio funkcijų skaičiaus (šiam tyrimui naudotos po dvi sąryšio funkcijos kiekvienam  $d-1$  tinklo įėjimui ir taisyklių skaičius  $L = 2^{d-1}$ );  $f_k$  –  $k$ -osios taisyklės išeities parametras;  $k$ -osios taisyklės svoris

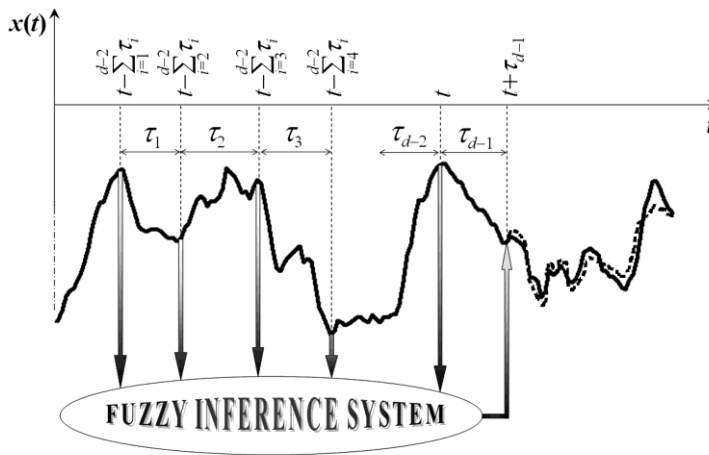
$$w_k(y'_t) = \prod_{i=1}^{d-1} w_{ik}(y'_{ti}), \quad k = 1, 2, \dots, L; \quad w_{ik}(\cdot) \text{ } i\text{-ojo elemento ir } k\text{-osios taisyklės}$$

sąryšio funkcija. Šiame darbe naudojama vadinamoji varpo formos sąryšio funkcija:

$$w_{ik} = \frac{1}{1 + \left[ \left( \frac{y'_{ti} - c_{ik}}{a_{ik}} \right)^2 \right]^{b_{ik}}}, \quad (2.28)$$

čia  $\{a_{ik}, b_{ik}, c_{ik}\}$  yra nežinomų parametrų aibė. Paprasčiausia tokio tinklo veikimo schema buvo pateikta 1.2 paveiksle.

Kai laiko eilutė rekonstruojama į vėlinimų erdvę reguliariais laiko vėlinimais, tuomet tiesiog prognozuojama  $x(t + \tau)$  reikšmė iš praeities duomenų  $\{x(t - (d - 2)\tau), x(t - (d - 3)\tau), \dots, x(t)\}$ . Bet kuriuo atveju (reguliaraus ar nereguliaraus rekonstravimo atveju) „juodoje dėžėje“ vyksta įeinančių duomenų fuzifikacija, informacijos apdorojimas penkių sluoksnių neuroniniame tinkle, defuzifikacija, ir informacijos išvedimas į vienintelį išėjimo kanalą.



**2.13 pav.** Scheminė diagrama, iliustruojanti laiko eilutės prognozavimo modelio, pagrįsto neraiškių išvadų sistemomis ir laiko eilutės nereguliariu rekonstravimu į vėlinimų erdvę, funkcionalumą.

### 2.2.1. Prognozės modelio taikymai Mackey-Glass laiko eilutės prognozei

Chaotinė Mackey-Glass vėlinimo diferencialinė lygtis yra etaloninis uždavinys, naudojamas daugelyje tyrimų diferencialinių modelių palyginimui su mokymo ir apibendrinimo galimybėmis [143-145]. Mackey-Glass vėlinimo diferencialinės lygties išraiška yra:

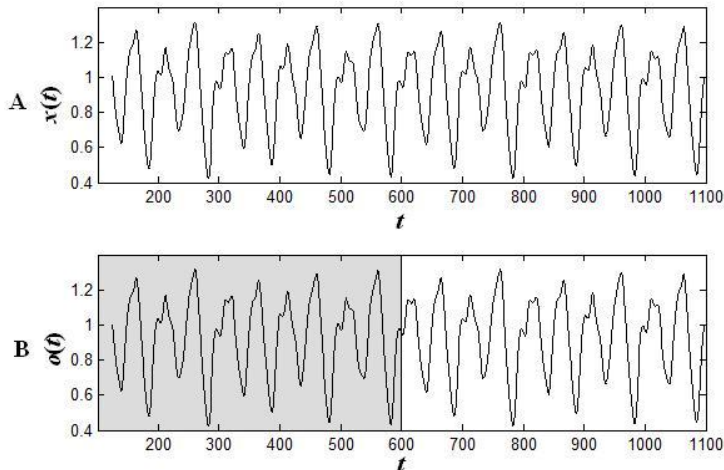
$$\frac{dx(t)}{dt} = \frac{ax(t-c)}{1+x^{10}(t-c)} - bx(t), \quad (2.29)$$

čia  $a$ ,  $b$  ir  $c$  yra konstantos;  $t$  yra laikas. Kai  $a=0,2$ ;  $b=0,1$  ir  $c=17$  Mackey-Glass lygtis generuoja chaotinę laiko eilutę.

#### 2.2.1.1. Mackey-Glass laiko eilutės prognozė pagal MATLAB neraiškių logiką įrankių pavyzdį

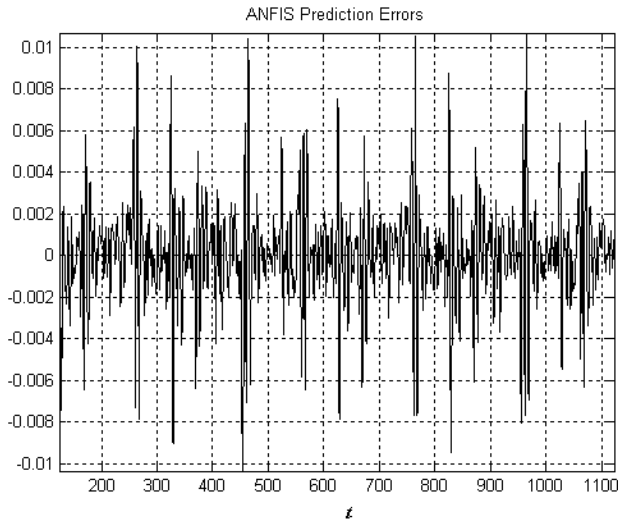
Norint palyginti sukurto modelio funkcionalumą pirmiausia laiko eilutė rekonstruojama į vėlinimų erdvę reguliariais laiko vėlinimais. Šiame poskyryje naudojamas standartinis Matlab neraiškių logiką (Fuzzy Logic) įrankių funkcionalumo demonstravimo pavyzdys, kuriame naudojami tokie sistemos parametrai: rekonstruojamos erdvės matavimas  $d=5$ ; reguliarūs laiko vėlinimai  $\tau=6$  [146]. Taigi yra prognozuojama  $x(t+6)$  reikšmė iš laiko eilutės praeities  $x(t-18)$ ,  $x(t-12)$ ,  $x(t-6)$ , ir  $x(t)$ .

Mackey-Glass diferencialinės lygties sprendinio  $x(t)$  dinamika laike vaizduojama 2.14 paveikslo A dalyje. Pagal minėtą pavyzdį konstruojamas laiko eilutės 1000 duomenų vektorius nuo  $t=119$  iki 1118; pirmieji 500 naudojami Matlab'o funkcijos ANFIS (Adaptive Network-Based Fuzzy Inference System) apmokymui, o kiti duomenys naudojami prognozės tikslumo tikrinimui. ANFIS prognozės rezultatai pavaizduoti 2.14 paveikslo B dalyje. Pilkojoje srityje vaizduojami prognozės rezultatai apmokymo metu.



2.14 pav. Mackey-Glass chaotinio atraktoriaus laiko eilutė: A – tikrasis signalas; B – ANFIS prognozė

Tikrinant ar susikurtas modelis gerai prognozuoja, stebimos prognozės skirtuminės paklaidos, kurios demonstruojamos 2.15 paveiksle. Tolesnių skyrelių skaičiavimo struktūra susideda iš tokių žingsnių: pirmiausia nustatomas rekonstruojamos vėlinimų erdvės matavimas, tada randamas laiko vėlinimų rinkinys; trečiame žingsnyje ANFIS adaptuojama uždavinio sprendimui. Laiko vėlinimai tarp kanalų (įėjimo ir išėjimo) turi priklausyti laiko vėlinimų aibe, rastai rekonstruojant laiko eilutę į vėlinimų erdvę pagal 2.12 paveikslo schemą. Be to, jei keičiamas rekonstravimo vėlinimų erdvės matavimas, tai turi keistis ir neraiškių išvadų sistemos kanalų skaičius. Kadangi darbe naudojamosi Matlab standartiniu „Fuzzy Logic“ įrankiu, šis pakeitimas vykdomas standartinių programinių priemonių pagalba. Galiausiai, kai jau į ANFIS sistemą adaptuojama laiko vėlinimų aibė, galima prognozuoti laiko eilutės ateities reikšmes.

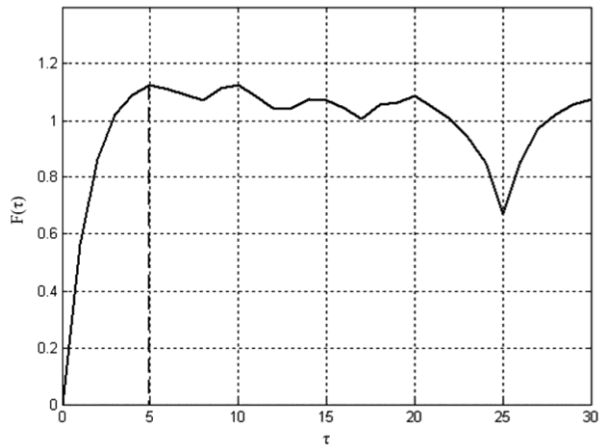


**2.15 pav.** Mackey – Glass laiko eilutės skirtuminės paklaidos tarp originalios ir prognozuotos pagal ANFIS.

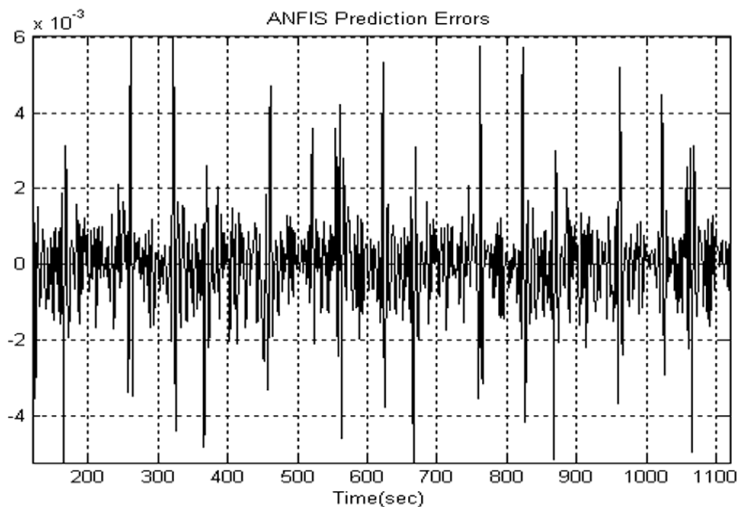
### **2.2.1.2. Mackey-Glass laiko eilutės prognozavimas, pagrįstas laiko eilutės reguliariu rekonstravimu**

Darbas tęsiamas toliau su Mackey-Glass laiko eilute. Atliekamas reguliarius rekonstravimas į  $d = 5$  matavimo vėlinimų erdvę. Tačiau dabar optimalų laiko vėlinimą surandamas naudojantis (2.18) lygtyje sukonstruota tikslo funkcija. Su šia funkcija buvo gauta, kad optimalus laiko vėlinimas  $\tau = 5$  (2.16 pav.);  $F(5,5,5,5) = 1,1249$  (kai standartiniame ANFIS pavyzdyje naudojamos vėlinimui  $\tau = 6$  tikslo funkcijos reikšmė  $F(6,6,6,6)$  yra lygi 1,1098). Pasirenkant optimalų laiko vėlinimą  $\tau = 5$  iš kelių tikslo funkcijos maksimalių reikšmių (maksimali reikšmė gaunama ir kai  $\tau = 10$ ) pasirenkama ta, kurios laiko vėlinimas mažiausiais, kad laiko eilutės optimaliam rekonstravimui būtų galima panaudoti kuo mažiau praeities duomenų. Prognozuojama  $x(t+5)$  iš praeities reikšmių  $x(t-15)$ ,  $x(t-10)$ ,  $x(t-5)$ , ir  $x(t)$ . Vėl konstruojamas 1000 duomenų vektorius nuo  $t = 116$  iki 1115, pirmi 500 naudojami sistemos apmokymui, kiti – prognozės tikslumo nustatymui. Absoliučiosios prognozavimo paklaidos parodytos 2.17 paveiksle.

Akivaizdu, kad laiko eilutės prognozavimo kokybė pagerėjo – paklaidos sumažėjo daugiau nei du kartus lyginant su pavyzdinio uždavinio rezultatais 2.15 paveiksle.



**2.16 pav.** Mackey – Glass laiko eilutės tikslo funkcijos grafikas, kai reguliariais laiko vėlinimais rekonstruojame laiko eilutę į  $d = 5$  matavimo erdvę. Gaunamas optimalus laiko vėlinimas  $\tau = 5$ , su kuriuo tikslo funkcijos reikšmė yra  $F(5,5,5,5) = 1.1249$ .



**2.17 pav.** Mackey – Glass laiko eilutės prognozavimo su ANFIS paklaidos, kai reguliaraus rekonstravimo laiko vėlinimai  $\tau = 5$ , laiko vėlinimų erdvės matavimas  $d = 5$ .

Kad aiškiau būtų galima pastebėti prognozavimo kokybę vertinamas standartinio paklaidų nuokrypio santykis su standartiniu tikrosios reikšmės nuokrypiu procentais. 2.1. lentelėje galima pastebėti, kad šis santykis lyginant su praeitos dalies santykiu sumažėjo per pus. Paklaidoms vertinti taip pat pasirenkama MAPE metrika, kuri yra pateikiama 2.1 lentelėje. Ši metrika taip pat rodo prognozės tikslumo pagerėjimą.



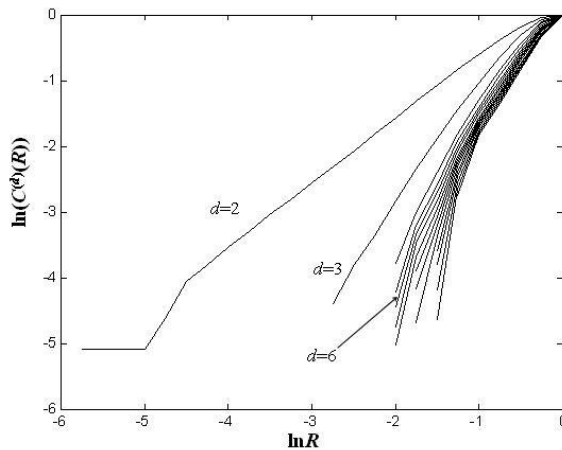
## 2.1 lentelė. Mackey-Glass chaotinės laiko eilutės prognozės paklaidų palyginimas

Rekonstravimo būdas	Vėlinimų erdvės matavimas $d$	Laiko vėlinimų rinkinys $(\tau_1, \dots, \tau_{d-1})$	$F(\tau_1, \dots, \tau_{d-1})$	Prognozavimo paklaidų ir tikrų reikšmių standartinių nuokrypių santykis $\frac{\sigma_e}{\sigma_x}$ procentais.	Prognozavimo paklaidų MAPE(%)
Reguliarus	5	(6, 6, 6, 6)	1,1098	1,1169%	0,1930 %
Reguliarus	5	(5, 5, 5, 5)	1,1249	0,5775%	0,0983%
Reguliarus	6	(9, 9, 9, 9, 9)	1,0990	0,3220%	0,0590 %
Nereguliarus	6	(6, 5, 5, 5, 5)	1,1026	0,1356%	0,0295 %

### 2.2.1.3. Mackey-Glass laiko eilutės prognozavimas, pagrįstas laiko eilutės reguliarium rekonstravimu, esant optimaliam laiko vėlinimų erdvės matavimui

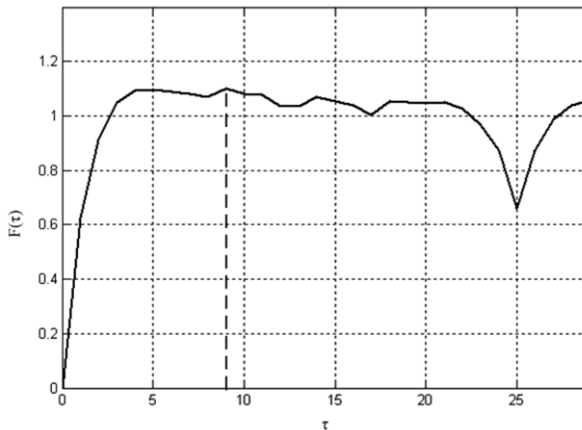
Šioje dalyje tęsiamas Mackey-Glass lygties sprendinio reguliarus rekonstravimas į laiko vėlinimų erdvę. Tačiau, dabar reikia nustatyti į kokio matavimo erdvę reikia rekonstruoti laiko eilutę, kad ji būtų optimaliai išskleista. Koreliacinės sumos pagalba, pagal 1.2 skyriuje (1.28) ir (1.31) sąryšius, brėžiami  $\ln R$  ir  $\ln(C^{(d)}(R))$  priklausomybės grafikai kintant rekonstravimo vėlinimų erdvės matavimui  $d$ . Taškų aibę  $(\ln R, \ln C^{(d)}(R))$  aproksimuojant tiesę, jos krypties koeficientas ir yra vėlinimų erdvės dimensiško įvertis. Šios priklausomybės nuo rekonstravimo erdvės matavimo pateikiamos 2.18 paveiksle. Grafike brėžtos linijos, kai matavimas kito nuo 2 iki 15. Pagal jį galima apytiksliai spręsti, kad kai rekonstruojamos erdvės matavimas yra  $d = 6$  tiesių krypties koeficientai jau mažai keičiasi.

Dabar galime nustatyti optimalų reguliaraus rekonstravimo į vėlinimų erdvę laiko vėlinimą. Funkcijos  $F(\tau, \tau, \tau, \tau, \tau)$  optimizavimas duoda rezultatą  $\tau = 9$  (Nors 2.19 paveiksle galima susidaryti vaizdą, kad tikslo funkcijos maksimali reikšmė gaunama ir kai  $\tau = 4$ , tačiau skaičiavimai parodė, kad prie  $\tau = 9$  tikslo funkcijos reikšmė truputį didesnė);  $F(9,9,9,9,9)=1,0990$ . Reikia pastebėti, kad šios reikšmės nederėtų lyginti su kito matavimo tikslo funkcijos optimalia reikšme. Ši reikšmė šiuo atveju yra mažesnė, nei  $F(5,5,5,5,5)$  (tuo labiau nei su  $F(6,6,6,6,6)$ ) iš prieš tai buvusio taikymo atvejo. Tačiau nustatyta laiko vėlinimo reikšmė yra geriausia kai vėlinimų erdvės matavimas  $d = 6$ .

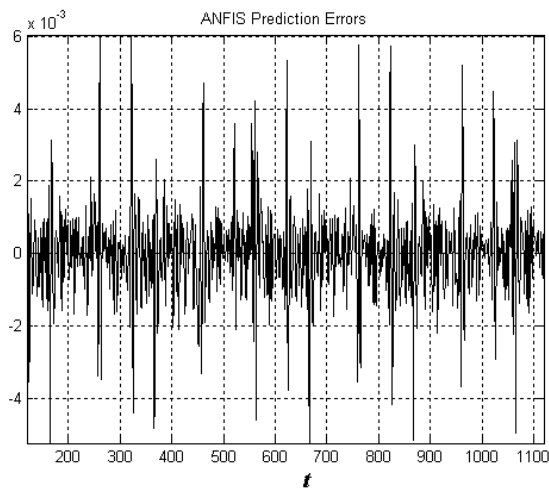


**2.18 pav.** Koreliacinės sumos natūrinio logaritmo funkcijos grafiko priklausomybė nuo  $\ln R$  kintant rekonstruojamos vėlinimų erdvės matavimui.

Prieš Mackey-Glass laiko eilutės prognozavimą reikia pakoreguoti ANFIS neraiškių logikų tinklo struktūrą. Dabar įėjimų į tinklą skaičius yra didesnis nei prieš tai nagrinėtu atveju. Prognozuojama  $x(t+9)$  iš praeities reikšmių  $x(t-36)$ ,  $x(t-27)$ ,  $x(t-18)$ ,  $x(t-9)$ , ir  $x(t)$ . Analogiškai, kaip prieš tai buvusiame skyrelyje, laiko eilutės prognozėje pakartojamos apmokymo ir tikrinimo procedūros; ANFIS absoliučiosios prognozavimo paklaidos pavaizduotos 2.20 paveiksle. Stebint prognozavimo paklaidų skalę, galima pastebėti, jog jos sumažėjo vidutiniškai 3 kartus, lyginant su 2.15 paveikslu duomenimis. Standartinių paklaidų nuokrypių santykis 2.1 lentelėje yra maždaug dvigubai mažesnis, nei kai laiko eilutė buvo rekonstruojama į penkiamatę vėlinimų erdvę.



**2.19 pav.** Mackey – Glass laiko eilutės tikslo funkcijos grafikas, kai laiko eilutė rekonstruojama į  $d = 6$  matavimo vėlinimų erdvę. Gaunamas optimalus laiko vėlinimas  $\tau = 9$ , su kuriuo tikslo funkcijos reikšmė yra  $F(9,9,9,9,9) = 1,0990$ .

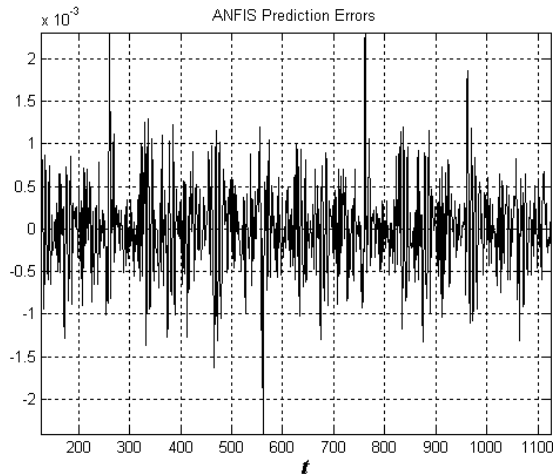


**2.20 pav.** Mackey – Glass laiko eilutės prognozavimo su ANFIS paklaidos, kai reguliaraus rekonstravimo optimalūs laiko vėlinimai  $\tau = 9$  esant vėlinimų erdvės matavimui  $d = 6$ .

#### **2.2.1.4. Mackey-Glass laiko eilutės prognozavimas, pagrįstas laiko eilutės nereguliariu rekonstravimu**

Darbas tęsiamas su Mackey-Glass lygties sprendiniu, tik dabar laiko eilutė rekonstruojama į fazinę erdvę nereguliais laiko vėlinimais. Ankstesniame poskyryje suradome, kad šios eilutės rekonstravimo fazinės erdvės optimalus matas yra  $d = 6$ . Dabar reikia surasti optimalią laiko vėlinimų aibę  $\{\tau_i\}$ ,  $i = 1, \dots, 5$ . Pilnai perrinkus visus įmanomus laiko vėlinimų derinius intervale  $[1, 30]$  gauta, kad geriausia aibė  $\{6, 5, 5, 5, 5\}$ . Su tokia laiko vėlinimų aibe gauta didžiausia tikslo funkcijos reikšmė  $d = 6$  matavimo fazinėje erdvėje:  $F(6, 5, 5, 5, 5) = 1,1026$ . Deja, tokios tikslo funkcijos kitimą grafiškai pavaizduoti jau sunku, nes reiktų vaizduoti penkiamatę laiko vėlinimų erdvę.

Taigi prognozuojama  $x(t+5)$  reikšmė iš praeities reikšmių  $x(t-21)$ ,  $x(t-15)$ ,  $x(t-10)$ ,  $x(t-5)$ , ir  $x(t)$ . ANFIS skirtuminės prognozavimo paklaidos parodytos 2.21 paveiksle. Galima pastebėti, kad maksimalus skirtumas tarp pradinės ir prognozuotos laiko eilutės 2.20 ir 2.21 paveiksluose beveik vienodas, bet standartinių nuokrypių santykis 2.1 lentelėje skiriasi: 2.20 paveikslo pakaldų ir tikrųjų reikšmių standartinių nuokrypių santykis procentais lygus 0,3220%, o 2.21 paveikslo pakaldų ir tikrųjų reikšmių standartinių nuokrypių santykis procentais lygus 0,1356%.



**2.21 pav.** Mackey – Glass laiko eilutės prognozavimo su ANFIS paklaidos, kai laiko eilutė nereguliariais laiko vėlinimais  $\tau = \{6, 5, 5, 5, 5\}$  rekonstruojama į šešiamatę vėlinimų erdvę.

Kaip jau minėta anksčiau, prieš prognozuojant laiko eilutę, reikėjo atlikti pilną laiko vėlinimų rinkinio perrinkimą. Tai pristatytos koncepcijos būdinga savybė. Tikslo funkcija (2.18) lygtyje reprezentuoja atraktoriaus išskleidimą daugiamatėje fazinėje erdvėje, bet ši funkcija nediskriminuoja tarp atraktoriaus konkrečių projekcijų. Tikslo funkcija duoda atraktoriaus tūrio įvertį; tam reikalingi skaičiavimai yra gana paprasti ir efektyvūs. Jei laiko vėlinimų reikšmės apribotos intervalu  $1 \leq \tau_i \leq T$ , kai  $i = 1, \dots, d - 1$ , tai optimalios laiko vėlinimų aibės radimui reikia patikrinti  $T^{d-1}$  rinkinių.

### 2.2.2. Prognozės modelio taikymai Rössler<sup>2</sup> atraktoriaus prognozei

Rössler chaotinis žemėlapis yra vienas iš pavyzdinių netiesinės dinamikos uždavinių [147]. Rössler sistema susideda iš trijų diferencialinių lygčių:

$$\begin{aligned}\frac{dx}{dt} &= -y - z; \\ \frac{dy}{dt} &= x + ay; \\ \frac{dz}{dt} &= b + z(x - c),\end{aligned}\tag{2.30}$$

čia  $x$ ,  $y$  ir  $z$  yra būsena kintamieji;  $t$  yra laikas;  $a$ ,  $b$  ir  $c$  yra kontrolės parametrai. Tradicinė Rössler bifurkacijos diagrama kuriama varijuojant  $c$  reikšmę, kai parametrai  $a = b = 0,1$  fiksuoti [148]. Ši bifurkacijos diagrama rodo, kad atraktorius yra periodinis mažiems  $c$ , bet greitai tampa chaotiniu, didėjant parametro  $c$  reikšmėms. Paskutiniaisiais metais Rössler atraktorius plačiai buvo panaudotas biomediciniuose tyrimuose [149, 150]. Rössler žemėlapis tinkamas charakterizuoti skirtingus biologinius ritmus, ypač neuronų virpesius. Šis atraktorius bus

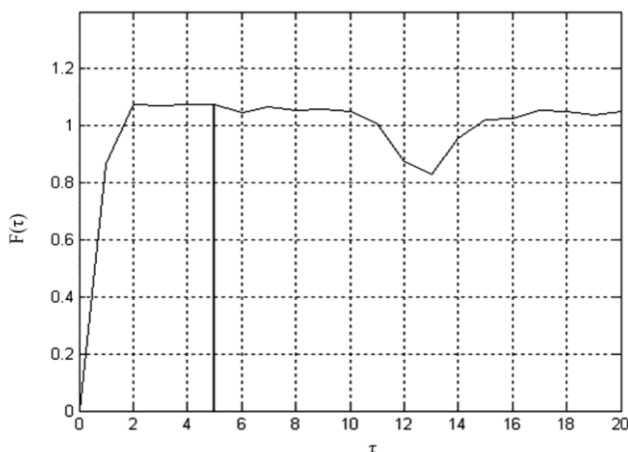
<sup>2</sup> Otto E. Rössler – vokiečių biochemikas, gimęs 1940 metais Berlyne.

testuojamas nereguliaraus rekonstravimo į laiko vėlinimų erdvę modelio pagrindu ir jo reikšmės bus prognozuojamos ANFIS sistemos pagalba.

### 2.2.2.1. Rössler laiko eilutės prognozavimas, pagrįstas laiko eilutės reguliariu rekonstravimu

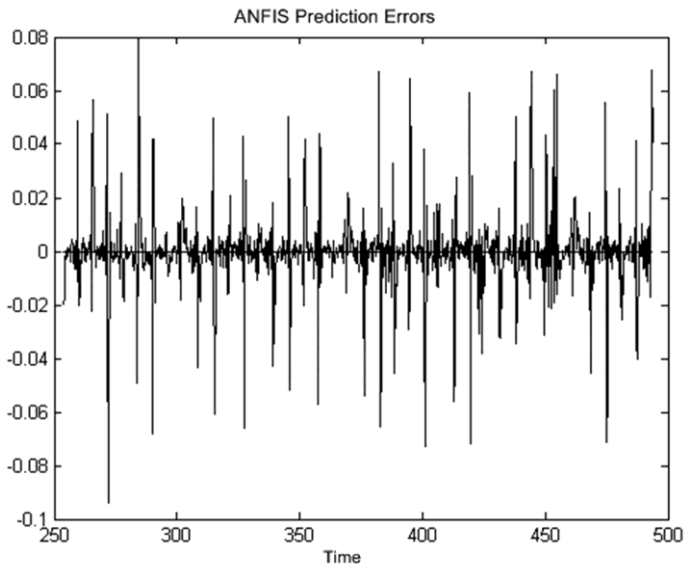
Konstruojamas chaotinis Rössler atraktorius su tokiais parametru reikšmėmis:  $a = b = 0,1$ ;  $c = 20$ . Konstruojant skaitinį (2.30) lygties sprendinį, naudojamas integravimo laike metodas su pastoviu žingsniu 0,02. Tada atrenkamas kas 12-tas duomenų taškas iš skaitinio sprendinio ir sukonstruojamas nauja duomenų eilutė, kuri bus naudojama prognozavimui (taigi turime laiko žingsnius  $\delta = 0,24$ ). Toks dirbtinis laiko eilutės duomenų atrinkimas pakeičia laiko skalę išsaugant pradinį integravimo tikslumą.

Pradžioje, sudaryta laiko eilutė reguliariais laiko vėlinimais rekonstruojama į vėlinimų erdvę. Analogiškai, kaip ir Mackey-Glass laiko eilutės atveju nustatoma, į kokio matavimo erdvę reikia rekonstruoti tą laiko eilutę. Vėl pasinaudojant koreliacinės sumos natūrinio logaritmo funkcijos priklausomybe nuo  $\ln R$  yra nustatoma, kad optimalus vėlinimų erdvės matavimas yra  $d = 6$ . Toliau ieškant optimalaus laiko vėlinimo yra gaunama, kad tikslo funkcijos reikšmė didžiausia, kai  $\tau = 5$ , t. y.  $F(5,5,5,5) = 1,0760$  (žiūrint į 2.22 paveikslą susidaro įspūdis, kad maksimali tikslo funkcijos reikšmė gaunama jau kai  $\tau = 2$ , tačiau skaičiavimai parodė, kad kai  $\tau = 5$  tikslo funkcijos reikšmė yra truputį didesnė). Prognozuojama reikšmė  $x(t+5)$  iš praeities duomenų  $x(t-20)$ ,  $x(t-15)$ ,  $x(t-10)$ ,  $x(t-5)$ , ir  $x(t)$ . Konstruojamas 1000 duomenų vektorius nuo  $t = 252,80$  iki 492,56. Pirmi 500 duomenų naudojami ANFIS tinklo apmokymui, o kiti – nustatyti prognozavimo tikslumui. Gaunamos santykinės prognozavimo paklaidos pavaizduotos 2.23 paveiksle. Šių paklaidų standartinis nuokrypio bei pradinių duomenų standartinio nuokrypio santykis procentais lygus 3,3180%.



**2.22 pav.** Rössler laiko eilutės reguliaraus rekonstravimo į šešiamatę vėlinimų erdvę tikslo funkcija. Optimalus laiko vėlinimas yra  $\tau = 5$ , optimali tikslo funkcijos reikšmė

$$F(5,5,5,5) = 1,0760.$$

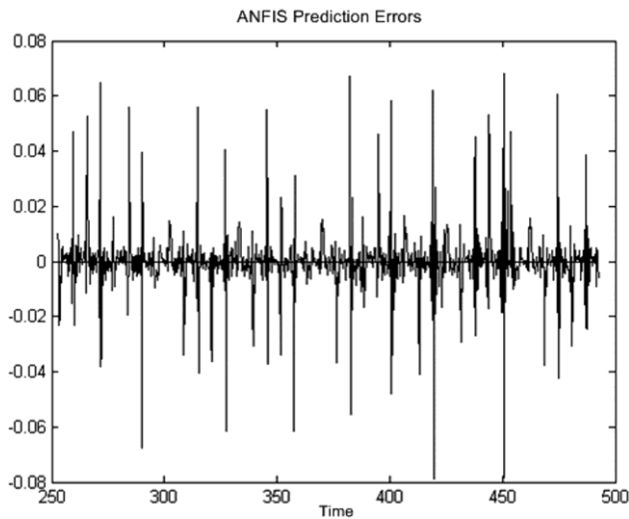


**2.23 pav.** Rössler laiko eilutės prognozavimo su ANFIS paklaidos, kai reguliaraus rekonstravimo laiko vėlinimai  $\tau = 5$ , o laiko vėlinimų erdvės matavimas  $d = 6$ .

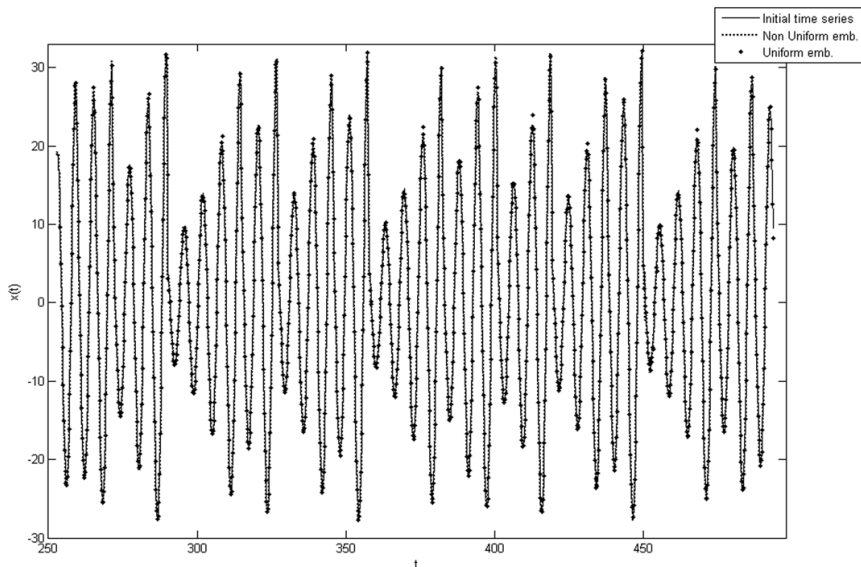
#### ***2.2.2.2. Rössler laiko eilutės prognozavimas, pagrįstas laiko eilutės nereguliariu rekonstravimu***

Kadangi prieš tai pavyko nustatyti, kad Rössler laiko eilutę optimaliai galima rekonstruoti į  $d = 6$  vėlinimų erdvę, toliau nagrinėjamas nereguliaraus rekonstravimo atvejis. Pilnojo perrinkimo būdu surandama, kad optimalus laiko vėlinimų rinkinys  $\{4,6,4,3,3\}$  geriausiai reprezentuoja rekonstruojamą atraktorių šešiamatėje vėlinimų erdvėje. Tikslų funkcijos reikšmę gaunama  $F(4,6,4,3,3) = 1,0823$ . Taigi prognozuojama  $x(t+3)$  reikšmė iš praeities duomenų  $x(t-17)$ ,  $x(t-13)$ ,  $x(t-7)$ ,  $x(t-3)$ , ir  $x(t)$ . Konstruojamas 1000 duomenų vektorius, kai laikas kinta nuo  $t = 252,08$  iki  $491,84$ . Pirmi 500 duomenų panaudojami ANFIS tinklo apmokymui, likę prognozės tikslumo nustatymui. Santykinės prognozavimo paklaidos parodytos 2.24 paveiksle. Paklaidų bei pradinių duomenų standartinių nuokrypių santykis procentais šiuo atveju lygus 2,6266%.

Vizualizuojant skirtumą tarp reguliaraus ir nereguliaraus rekonstravimo Rössler sistemos prognozavimo atvejų viename grafike yra palyginami skaitiniai sprendiniai su pradine laiko eilute 2.25 paveiksle. Šiame grafike ištisine linija vaizduojami tikrosios laiko eilutės duomenys, punktyrine linija žymima nereguliaraus rekonstravimo atveju gaunama prognozė ir taškeliais – nereguliaraus rekonstravimo atveju gaunama prognozė. Iš esmės grafike visiškai nesimato skirtumo tarp tikrosios laiko eilutės ir jos prognozės reikšmių nereguliaraus rekonstravimo atveju. Reguliaraus rekonstravimo atveju galima pastebėti tik kelis taškus, kurie nesutampa su tiriamos laiko eilutės duomenimis. Didesnį skirtumą prognozės duomenų nuo pradinės laiko eilutės parodo tik paklaidų grafikai 2.23 ir 2.24 paveiksluose. Tai tik dar kartą patvirtina, kad nereguliaraus rekonstravimo atveju galima pagerinti laiko eilutės prognozės rezultatus (nors ir nežymiai).



**2.24 pav.** Rössler laiko eilutės prognozavimo su ANFIS paklaidos, kai netolygaus rekonstravimo į vėlinimų erdvę laiko vėlinimai  $\tau = \{4, 6, 4, 3, 3\}$ , o dimensija  $d = 6$ .



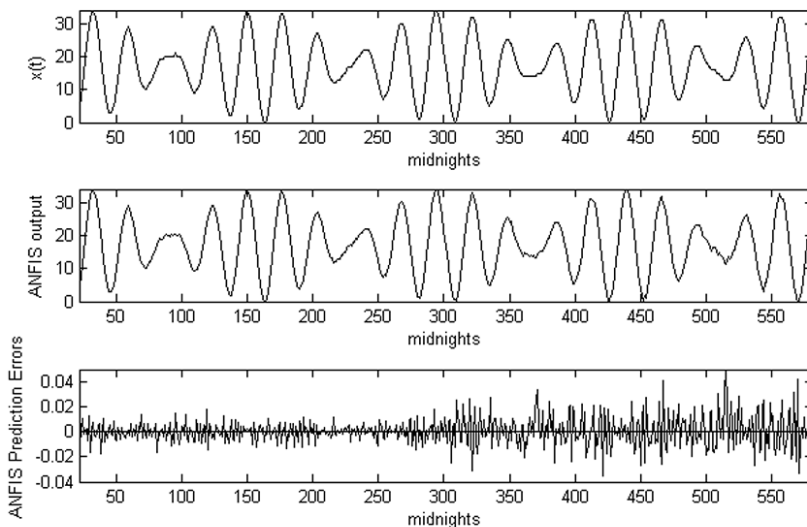
**2.25 pav.** Rössler laiko eilutės (ištininė linija) palyginimas su jos prognozuojamomis reikšmėmis reguliaraus rekonstravimo (punktų linija) ir nereguliaraus rekonstravimo (taškeliai) atvejais.

### 2.2.3 Prognozės modelio taikymai realaus pasaulio laiko eilutės prognozei

Iki šiol darbe buvo nagrinėjami diferencialinių lygčių chaotiniai sprendiniai ir jų prognozė. Gana didelis iššūkis yra realaus pasaulio laiko eilučių prognozė. Šiose eilutėse įprastai būna gana didelis triukšmo lygis, o pati sistemos dinamika yra aprašoma dažniausiai labai sudėtingais modeliais. Dažnai nutinka ir taip, kad net neaiški būna to modelio kilmė. Tada sunku nustatyti, kokios diferencialinių lygčių

sistemos galėtų aprašyti tokių sistemų modelius. Taigi toliau darbe bus mėginama pritaikyti sukurtą laiko eilučių nereguliariais laiko vėlinimais rekonstravimo į vėlinimų erdvę prognozės modelį, pagrįstą ANFIS tinklais, realioms laiko eilutėms prognozuoti. Šioame darbe pasirinkta laiko eilutė, kurioje yra 600 stebimos žvaigždės kasdieninis kintamo ryškumo matavimo duomenys. Kaip ir anksčiau pagal 1.3 temoje apibrėžtą koreliacinės sumos formulę (1.28) skaičiuojant jos natūrinio logaritmo priklausomybes nuo  $\ln R$ , nustatoma, kad šią laiko eilutę geriausiai galima reprezentuoti, jei ji rekonstruojama į  $d=6$  matavimo erdvę. Reguliaraus rekonstravimo atveju, optimalūs laiko vėlinimai pasirodo yra  $\{4,4,4,4,4\}$ ;  $F(4,4,4,4,4)=1,0719$ ; prognozavimo ANFIS algoritmu prognozės paklaidų bei pradinių duomenų standartinių nuokrypių santykis procentais lygus 0,1434%.

Darbas tęsimas toliau rekonstruojant laiko eilutę nereguliariais laiko vėlinimais. Pilno perrinkimo būdu randama, kad su laiko vėlinimų rinkiniu  $\{4,4,5,4,4\}$ , tikslo funkcijos reikšmė  $F(4,4,5,4,4)=1,0747$ . Prognozuojama  $x(t+4)$  reikšmė, kai žinomos praeities reikšmės  $x(t-17)$ ,  $x(t-13)$ ,  $x(t-9)$ ,  $x(t-4)$  ir dabarties reikšmė  $x(t)$ . Konstruojamas 560 duomenų ilgio vektorius  $y_p^{(5)}$ , kurio laikas kinta nuo  $t=18$  iki 579. Pirmi 280 duomenų panaudojami ANFIS tinklo apmokymui, o likę – prognozavimo tikslumo tikrinimui. 2.26 paveiksle yra palyginti tikrieji duomenys ir su duomenimis, kuriuos išveda ANFIS prognozavimo algoritmas. Čia taip pat parodytas eilutės santykinų paklaidų grafikas, kurių standartinio nuokrypio santykis su pradinių duomenų standartiniu nuokrypiu lygus 0,1273%. Tai dar kartą parodo, kad nereguliaraus rekonstravimo atveju galima pasiekti geresnių rezultatų nei reguliaraus atraktoriaus rekonstravimo atveju.



**2.26 pav.** Realios laiko eilutės palyginimas su ANFIS prognozės duomenimis nereguliaraus rekonstravimo į vėlinimų erdvę atveju, kai  $d=6$ , o laiko vėlinimai  $\{4,4,5,4,4\}$ .



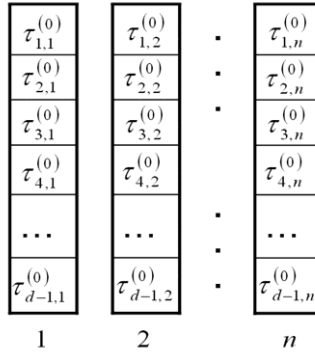
## 2.3. Evoliucinio algoritmo panaudojimas laiko eilutės rekonstravimo uždavinyje

Apibendrinami prieš tai nagrinėtų laiko eilučių rekonstravimo uždavinio ir prognozės rezultatus, galima pasakyti, kad laiko eilutės prognozės rezultatai yra geresni eilutę rekonstruojant į vėlinimų erdvę nereguliariais laiko vėlinimais. Kiekvieną kartą nustatant optimalų laiko vėlinimų rinkinį reikėjo atlikti pilną visų įmanomų vėlinimų derinių perrinkimą. Kadangi buvo nuspresta, jog kiekvienas laiko vėlinimas turi tenkinti sąlygą  $1 \leq \tau \leq T$ , kai  $T = 30$ , reikia patikrinti  $T^{d-1}$  rinkinį, t. y. reikia apskaičiuoti  $T^{d-1}$  tikslo funkcijos reikšmę pagal (2.18) formulę. Kuo didesnio matavimo vėlinimų erdvė reprezentuoja tiriamą laiko eilutę, tuo ilgiau užtrunka toks pilno perrinkimo uždavinys. Taigi šiuo atveju tenka spręsti kombinatorinio optimizavimo uždavinį. Daugelio kombinatorinio optimizavimo uždavinių sprendimui yra taikomi sveikaskaitinio optimizavimo metodai, tokie kaip modeliujamojo atkaitinimo algoritmas (SA), dalelių spiečiaus algoritmas (PSO), evoliuciniai algoritmai (EA) ir kiti. Šiame skyrelyje bus tikrinama ar galima supaprastinti laiko eilučių rekonstravimo uždavinį taikant vieną iš šių algoritmų.

### 2.3.1. Evoliucinių algoritmų struktūra

Evoliucinių algoritmų uždavinys yra rasti tokį laiko vėlinimų rinkinį, su kuria tikslo funkcijos reikšmė būtų didesnė nei reguliaraus rekonstravimo atveju. Darbe evoliuciniai algoritmai sudaromi taip, kad kiekvienoje chromosomoje būtų saugoma laiko vėlinimų seka. Kiekvienas genas yra natūralusis skaičius, patenkantis į intervalą  $1 \leq \tau \leq T$ ; įvertinant atraktorių kompleksiskumą priimama, kad  $T = 30$ . Tuomet kiekvienos chromosomos ilgis yra  $(d-1)$ . Parametras  $T$  nėra evoliucinio algoritmo parametras (nors jis turi tiesioginės įtakos šių algoritmų funkcionalumui). Kadangi, dažniausiai laiko vėlinimų erdvės optimalus matavimas gaunamas 6 arba didesnis skaičius (tuo galima įsitikinti vėliau, tiriant realaus pasaulio laiko eilutes), tai rekonstravimo į vėlinimų erdvę langas (taškų skaičius tarp ANFIS tinklo pirmo įėjimo ir išėjimo) bus lygus arba didesnis už  $(d-1)T = 150$ . Dalis tam tikro ilgio pradinių duomenų vektoriaus turi būti panaudota ANFIS tinklo apmokymui. Dėl šios priežasties nustatomas tam tikras laiko eilutės mažiausio ilgio apribojimas. Kita vertus, skaičiavimo eksperimentai rodo, kad optimalūs laiko vėlinimai retai būna didesni už 15. Tai pateisina parinktą reikšmę  $T = 30$  visoms analizuojamoms laiko eilutėms.

Evoliucinio algoritmo populiacija susideda iš  $n$  chromosomų, kuriose genų reikšmės sugeneruojamos atsitiktinai (2.27 paveikslas). Anksčiau buvo minėta, kad tikslo funkcija  $F$  charakterizuoja atraktoriaus skleidimo rekonstruotoje vėlinimų erdvėje turį (pagal (2.18) lygtį), todėl ji naudojama kaip evoliucinio algoritmo tinkamumo funkcija (kurios reikšmės yra realūs skaičiai), skaičiuojama kiekvienai populiacijos chromosomai. Sudarant evoliucinio algoritmo vykdymo etapus, darbe buvo atsižvelgta į 1.4.1 skyrelyje aprašytą klasikinio algoritmo schemą. Vykdamas evoliucinį algoritmą daugiausia dėmesio buvo skirta šiems etapams: atranka, kryžminimas ir mutacija.



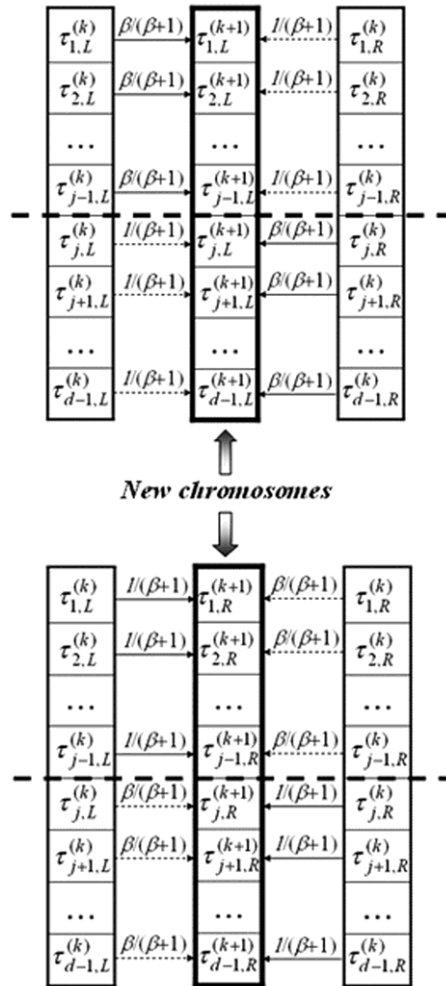
**2. 27 pav.** Scheminė diagrama, iliustruojanti pradinę populiaciją iš  $n$  chromosomų;  $\tau_{i,j}^{(0)}$  reiškia  $i$ -tojo elemento  $j$ -ąją chromosomą,  $\tau_{i,j}^{(0)} \in N$ ,  $1 \leq \tau_{i,j}^{(0)} \leq T$ ,  $1 \leq i \leq d-1$ ,  $j = 1, 2, \dots, n$ .

Atrankos procese iš pradinės populiacijos imamas lyginis chromosomų skaičius, kad sudarant naują populiaciją būtų galima kiekvienai chromosomai rasti porą. Chromosomų atrinkimui į naują populiaciją naudojamas turnyrinės atrankos metodas, kuriame chromosoma turi tuo daugiau šansų patekti į porų populiaciją, kuo didesnė yra jos tinkamumo funkcijos reikšmė. Vis dėl to, chromosomai, kurios tinkamumo funkcijos reikšmė yra maža, tikimybė būti atrinktai nėra nulinė. Be to, ta pati chromosoma į naują populiaciją gali būti atrinkta ne vieną kartą. Atrankos proceso pabaigoje naujoje populiacijoje visos chromosomos yra grupuojamos poromis.

Naujoje porų populiacijoje visos chromosomų poros yra kryžminamos. Darbe naudojamas vieno taško kryžminimo algoritmas [151] (kryžminimo taško padėtis kiekvienos poros chromosomoms randamas atsitiktiniu būdu). Dar daugiau, šiame darbe klasikinis kryžminimo algoritmas modifikuojamas įvedant panašumo parametą  $\beta$  ( $\beta \geq 0$ ). Panašumo parametro vaidmuo kryžminimo procese pavaizduota 2.28 paveiksle. Chromosomose sukeičiami ne tik genuai, bet taip pat leidžiame dviem naujoms chromosomoms daugiau ar mažiau būti panašioms vienai į kitą ir į jų tėvus. Toks kryžminimo algoritmas panašumo parametro atžvilgiu konverguoja į klasikinį kryžminimo procesą, kai parametras  $\beta$  artėja į 0 arba į begalybę (abi palikuonių chromosomos yra lygios, kai  $\beta = 1$ ). Bendru atveju, šis algoritmas aprašomas lygybėmis:

$$\begin{aligned}
 \tau_{j,L}^{(k+1)}|_- &:= \text{round} \left( \frac{\beta}{\beta+1} \tau_{j,L}^{(k)} + \frac{1}{\beta+1} \tau_{j,R}^{(k)} \right), \\
 \tau_{j,R}^{(k+1)}|_- &:= \text{round} \left( \frac{1}{\beta+1} \tau_{j,L}^{(k)} + \frac{\beta}{\beta+1} \tau_{j,R}^{(k)} \right), \\
 \tau_{j,L}^{(k+1)}|_+ &:= \text{round} \left( \frac{1}{\beta+1} \tau_{j,L}^{(k)} + \frac{\beta}{\beta+1} \tau_{j,R}^{(k)} \right), \\
 \tau_{j,R}^{(k+1)}|_- &:= \text{round} \left( \frac{\beta}{\beta+1} \tau_{j,L}^{(k)} + \frac{1}{\beta+1} \tau_{j,R}^{(k)} \right);
 \end{aligned} \tag{2.31}$$

čia  $k$  yra vykdomos generacijos skaičius,  $\tau_{j,L}^{(k)}$  yra  $j$ -sis poros kairiosios chromosomos tėvo genas,  $\tau_{j,R}^{(k)}$  yra  $j$ -sis poros dešinėsios chromosomos tėvo genas,  $\tau_{j,L}^{(k+1)}$   $j$ -tasis kairiosios vaiko chromosomos genas,  $\tau_{j,R}^{(k+1)}$  yra  $j$ -tasis dešinėsios vaiko chromosomos genas, indeksuose žymimas minuso ženklas reiškia, kad  $j$ -tasis genas randasi virš kryžminimo taško, o pliuso ženklas – genas yra žemiau kryžminimo taško. Bet kuriuo atveju, vaiko chromosomos  $j$ -tojo geno reikšmė tilps į intervalą  $[\tau_{j,L}^{(k)}; \tau_{j,R}^{(k)}]$ .



**2.28 pav.** Kryžminimo algoritmo iliustracijos schematinė diagrama;  $\beta$  yra panašumo parametras; punktyrinės linijos rodo kryžminimo tašką.

Evoliuciniuose algoritmuose naudojamas kryžminimo koeficientas charakterizuoja tikimybę, kad atrinktoje populiacijoje chromosomų porai bus atlikta kryžminimo procedūra. Reikėtų atkreipti dėmesį, kad ši kryžminimo procedūra yra vienas faktorių, padedančių išvengti kraujomaišos (nusistovėjusios sprendinių aibės), kuri ypač pavojinga sveikaskaitiniam laiko vėlinimui rinkiniui. Pradinės populiacijos generuojami chromosomų genai yra atsitiktiniai natūralieji skaičiai.

Kryžminimas be panašumo parametro įvedimo neleistų atsirasti kitokiems genams (naujiems laiko vėlinimams), nei jau yra populiacijoje. Net pilnas sugeneruotų genų perrinkimas tuo atveju galėtų duoti tik tokius laiko vėlinimų rinkinius, kurie gali būti labai nutolę nuo globalaus optimalaus sprendinio.

Kryžminimo procedūra su panašumo parametru gali generuoti naujus genus. Galima įsitikinti, kad *bet kokių atveju, ši procedūra nepadengs visos argumentų erdvės, jei pradinės genų aibės minimali reikšmė bus didesnė už 1, o maksimali – mažesnė už T* ( griežtai kalbant, tai yra tik būtina, bet nepakankama sąlyga).

Teiginys įrodomas prieštaros būdu. Tarkime, kad visų genų pradinėje populiacijoje mažiausia reikšmė yra  $a$ , o didžiausia –  $b$ , ir tarkime, kad po eilės generacijų genų reikšmės *padengs visą intervalą*  $[1, T]$ . Tačiau, nepriklausomai nuo to, kokia yra parametro  $\beta$  reikšmė, bet kurios vaiko chromosomos naujų genų reikšmės priklausys intervalui  $[a, b]$ . Be to, po kryžminimo su panašumo parametru niekada negausime chromosomos su genu  $a$  (minimaliu laiko vėlinimu) kitoje vietoje (vaiko chromosomos viduje) negu jo padėtis buvo pradinės populiacijos chromosomoje. Kitais žodžiais tariant, didžiausias įmanomas laiko vėlinimų vektorių padengiamas poerdvis po kryžminimo proceso su panašumo parametru yra

$$\tau_{i \min} = \min_{1 \leq j \leq n} (\tau_{i,j}^{(0)}), \quad \tau_{i \max} = \max_{1 \leq j \leq n} (\tau_{i,j}^{(0)}), \quad i = 1, 2, \dots, d - 1. \quad (2.32)$$

Taigi teiginys įrodytas.

Mutacijos procedūra sukliudo algoritmui konverguoti į lokalųjį sprendinį ir padeda pasiekti globalųjį sprendinį. Šiame darbe naudojama klasikinė mutacijos procedūra. Mutacijos parametras  $\mu$  ( $0 \leq \mu < 1$ ) apibrėžia mutacijos proceso intensyvumą. Algoritmo vykdymo metu yra perbėgama per visus chromosomos genus  $k$ -tojoje generacijoje ir kiekvienam genui generuojamas atsitiktinis skaičius iš intervalo  $[0, 1]$ . Jei tas sugeneruotas skaičius yra mažesnis už  $\mu$ , genas yra pakeičiamas atsitiktiniu natūraliuoju skaičiumi iš intervalo  $[1; T]$ . Dažniausiai, mutacijos parametro  $\mu$  reikšmė imama labai maža ( $\mu < 0,01$ ). Didesnis mutacijos intensyvumas gali kliudyti evoliuciniam algoritmui evoliucionuoti [123, 124].

### 2.3.2. Evoliucinių algoritmų parametrų parinkimas

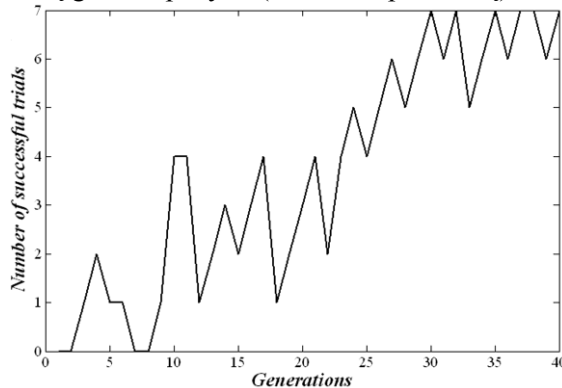
Bendruoju atveju evoliucinių algoritmų parametrų parinkimas yra empirinis procesas, nors kai kurie bendri principai yra paaiškinti [151, 152] šaltiniuose. Ankstesniuose skyreliuose buvo pademonstruota, kad reguliariais laiko vėlinimais rekonstruojant Mackey-Glass chaotinę laiko eilutę į šešiamatę vėlinimų erdvę, didžiausia tikslo funkcijos reikšmė  $F(9, 9, 9, 9, 9) = 1,0990$  gaunama su laiko vėlinimų rinkiniu  $\{9, 9, 9, 9, 9\}$ . Taigi, evoliucinio algoritmo tikslas yra surasti tokį laiko vėlinimų rinkinį, su kuriais Mackey-Glass eilutei tikslo funkcijos reikšmė būtų didesnė nei 1,0990. Šis rezultatas bus naudojamas kaip kriterijus, pagal kurį bus atrinkti optimalius evoliucinio algoritmo parametrai. Kadangi Mackey-Glass eilutės prognozavimas yra nagrinėtas straipsniuose [88, 143-145, 153], tai prognozuojant šią eilutę pagal šiame darbe sukurtą modelį bus su kuo palyginti gautus rezultatus.

Reikia nuspręsti, kokios turi būti šių parametrų reikšmės: kryžminimo koeficientas  $\kappa$ , mutacijos koeficientas  $\mu$ , panašumo parametras  $\beta$ , populiacijos

dydis  $n$  ir generacijų skaičius. Parenkant parametrų reikšmes yra vadovaujamosi klasikinio evoliucinio algoritmo sudarymo rekomendacijomis [151]. Tikrinamos kryžminimo koeficiento reikšmės  $\kappa \in [0,6; 0,8]$ , ir mutacijos koeficiento  $0 \leq \mu \leq 0,1$ . Kaip jau minėta anksčiau, jei panašumo parametro skaitinė reikšmė  $\beta = \infty$  (arba  $\beta = 0$ ), tai gaunamas klasikinis evoliucinio algoritmo modelis. Tyrimo eigoje bus pademonstruota, kad  $\beta$  reikšmė turi įtakos algoritmui, kai  $1 \leq \beta \leq 10$ .

Šio disertacinio darbo eksperimentuose populiacijos dydis pasirenkamas  $n_g = 24$  (ir pradinei, ir darbinei populiacijoms). Kaip minėta anksčiau, tikslas yra sugeneruoti nereguliaraus rekonstravimo į vėlinimų erdvę optimalų laiko vėlinimų rinkinį. Atlikus skaitinius eksperimentus su Mackey-Glass'o chaotinė laiko eilute ir pasirinkę  $n_g = 16, 20, 24, 28$  ir  $32$ , gauta, kad daugiausia sėkmingų rinkinių pavyksta susigeneruoti, kai  $n_g = 24$ .

Nėra sukurtų metodų, nustatančių kiek evoliuciniame algoritme turi būti atliekama generacijų. Sprendžiant paprastus uždavinius, algoritmas paprastai konverguoja į gerą sprendinį jau po 20 ar 30 generacijų. Sudėtingesniuose uždaviniuose reikia atlikti daugiau generacijų. Sprendžiant sudėtingus uždavinius (pvz. Tvarkaraščių sudarymo uždavinius) nėra neįprasta evoliuciniuose algoritmuose naudoti daugiau kaip 400 generacijų [154]. Patikimiausias metodas surasti tinkamą generacijų skaičių yra bandymų ir klaidų metodas, nors daugybė autorių siūlė metodų, padedančių nustatyti, kiek ilgai sprendinio ieškojimas turi tęstis [151, 154]. Savo modelyje naudojamos 40 generacijų, nes ilgiau vykdant populiacijos evoliuciją daugiau geresnių rezultatų gauti nepavyko (žiūr. 2.29 paveikslą).



**2.29 pav.** Ryšys tarp Evoliucinio algoritmo sėkmingų bandymų (iš 100 galimų) ir generacijų skaičiaus. Sėkmingu bandymu laikėme tuos atvejus, kai gauta nereguliarių laiko vėlinimų aibė, su kuria tikslo funkcijos reikšmė didesnė nei reguliaraus rekonstravimo atveju. Skaitiniai eksperimentai atlikti su Mackey-Glass laiko eilute;  $\mu = 0,002$ ,  $\kappa = 0,8$ ,

$$\beta = 2, n_g = 24.$$

Vieną kartą įvykdžius evoliucinį algoritmą, toje populiacijoje yra randama viena geriausia laiko vėlinimų aibė. Akivaizdu, kad rezultatas priklauso nuo pradinės populiacijos chromosomų (tarp visų kitų atsitiktinių faktorių). Todėl, evoliucinis algoritmas vykdomas 100 kartų ir yra stebima, kiek kartų gaunamas toks laiko vėlinimų rinkinys, su kuriuo tikslo funkcijos reikšmė būtų didesnė nei 1,0990 (tą skaičių pažymime  $k$ ). Kitaip tariant yra sukurama 100 populiacijų ir stebimos,

kelios iš jų yra sėkmingos. Sprendimo rezultatai yra pateikti 2.2 lentelėje. Reikia pažymėti, kad klasikinis vieno taško kryžminimas (kai  $\beta = 0$ ) neduoda nei vieno tinkamo rezultato.

Evoliuciniu algoritmu, su parametrais  $\mu = 0,001$ ,  $\kappa = 0,8$  ir  $\beta = 3$ , iš 100 bandymų buvo gauta laiko vėlinimų aibė, kuri buvo penkis kartus geresnė už  $\{9, 9, 9, 9, 9\}$ . Tačiau dažniausiai sėkmingi rinkiniai pasirodė, kai  $\beta = 2$  (lentelėje 2.2 pateikiama sėkmingų rinkinių skaičiaus suma esant skirtingoms  $\mu$  ir  $\kappa$  reikšmėms). Be to, geriausias laiko vėlinimų rinkinys  $\{7, 4, 5, 5, 6\}$  sugeneruotas, kai  $\mu = 0,002$ ,  $\kappa = 0,8$ ,  $\beta = 2$ . Su šiuo rinkiniu tikslo funkcijos reikšmė lygi  $F(7, 4, 5, 5, 6) = 1,1049$  (rezultatas pateiktas 2.3 lentelėje).

**2.2 lentelė.** Ryšys tarp sėkmingų bandymų ir panašumo koeficiento  $\beta$ .

Panašumo koeficientas $\beta$	0	1	2	3	5	10	15	20	30
Sėkmingų bandymų suma	0	3	27	12	6	4	0	0	0

**2.3 lentelė.** Evoliucinių algoritmų imitavimo rezultatai.

$F_{\max}$  – geriausia tikslo funkcijos reikšmė (didžiausia tikslo funkcijos reikšmė), gauta po 100 bandymų su atitinkamais algoritmo parametrais;  $k_g$  sėkmingų bandymų skaičius, kai tikslo funkcija bandyme viršija 1,0990.

$\mu$	$\kappa$	$\beta=0$		$\beta=1$		$\beta=2$		$\beta=3$		$\beta=5$		$\beta=10$	
		$F_{\max}$	$k_g$	$F_{\max}$	$k_g$	$F_{\max}$	$k_g$	$F_{\max}$	$k_g$	$F_{\max}$	$k_g$	$F_{\max}$	$k_g$
0,001	0,6	1,0853	0	1,0959	0	1,0993	2	1,0983	0	1,0963	0	1,0992	1
0,002	0,6	1,0866	0	1,101	1	1,1044	3	1,1014	1	1,0963	0	1,0956	0
0,005	0,6	1,0875	0	1,1007	1	1,1044	2	1,1004	1	1,0997	1	1,0997	1
0,01	0,6	1,0898	0	1,0957	0	1,1023	3	1,0988	0	1,0955	0	1,0914	0
0,001	0,7	1,0941	0	1,0974	0	1,101	2	1,0986	0	1,0988	0	1,0973	0
0,002	0,7	1,0965	0	1,0999	1	1,0969	0	1,0993	2	1,1014	2	1,097	0
0,005	0,7	1,0885	0	1,0981	0	1,1005	1	1,1004	1	1,099	1	1,0936	0
0,01	0,7	1,0963	0	1,0926	0	1,1026	2	1,1018	1	1,101	1	1,097	0
0,001	0,8	1,0936	0	1,0964	0	1,1017	3	1,103	5	1,0968	0	1,1014	2
0,002	0,8	1,0956	0	1,0962	0	<b>1,1049</b>	4	1,097	0	1,1002	1	1,0956	0
0,005	0,8	1,0842	0	1,0988	0	1,1029	3	1,0992	1	1,0963	0	1,0925	0
0,01	0,8	1,0820	0	1,0981	0	1,1012	2	1,0968	0	1,093	0	1,0973	0

Taigi, minėtos evoliucinio algoritmo parametrų reikšmės fiksuojamos ( $\mu = 0,002$ ,  $\kappa = 0,8$ ,  $\beta = 2$ ). Jos liks tos pačios prognozuojant realias laiko eilutes. Tačiau, pirma reikia patikrinti prognozavimo su neraiškių logikų sistemomis paklaidas, kurios gaunamos Mackey-Glass laiko eilutę rekonstruojant į vėlinimų erdvę su laiko vėlinimų rinkiniu  $\{7, 4, 5, 5, 6\}$ . Iš praeities duomenų  $x(t-21)$ ,  $x(t-14)$ ,  $x(t-10)$ ,  $x(t-5)$  ir  $x(t)$  prognozuojama viena reikšmė  $x(t+6)$ . Konstruojamas 1000 duomenų vektorius laiko intervale nuo  $t=122$  iki  $t=1121$ , iš kurių pirmieji 500 duomenų naudojami ANFIS tinklo apmokymui, o likę –

prognozavimo tikslumui nustatyti. Prognozės rezultatų absoliučiosios paklaidos pavaizduotos 2.30 paveiksle, o šių paklaidų standartinis nuokrypis yra lygus  $3,4989 \cdot 10^{-4}$ . Reguliaraus rekonstravimo į vėlinimų erdvę atveju laiko eilutės prognozavimo paklaidų standartinis nuokrypis buvo lygus  $7,2862 \cdot 10^{-4}$ .

### 2.3.3. Evoliucinių algoritmų rezultatų palyginimas su dalelių spiečiaus ir modeliavimo atkaitinimo modelio paieškos algoritmų rezultatais

Toliau darbe palyginimas evoliucinių algoritmų funkcionalumas su dalelių spiečiaus (PSO) ir modeliavimo atkaitinimo (SA) paieškos algoritmais. Ieškomas optimalus laiko vėlinimų rinkinys su PSO algoritmu. Tikrinama, kaip veikia algoritmas, kai naudojamas vienas parametrų rinkinys iš trijų tipinių rinkinių, aprašytų 1.4.2 skyrelyje. Papildomai nagrinėjama, kaip sprendinio gerumas priklauso nuo dalelių spiečiaus dydžio. Dalelių spiečiaus skaičius vieną sykį bus imamas toks, koks yra evoliucinių algoritmų populiacijos dydis, kitą sykį 40 dalelių (reikšmė iš siūlomo intervalo  $[10, 60]$ ), na ir pagaliau imamas didžiausias rekomenduojamas spiečiaus dalelių skaičius 60. Sprendinio paieškos žingsnių skaičius taip pat bus bandoma susieti su evoliuciniuose algoritmuose nusistatytu generacijų skaičiumi 40. Šioje dalyje tikrinama ar didesnis žingsnių skaičius turi įtakos algoritmo funkcionalumui.

Toliau nagrinėjama ta pati Mackey-Glass laiko eilutė. Ieškant šios eilutės laiko vėlinimų rinkinio, kuris būtų artimas optimalam, vykdomas PSO algoritmas su kiekvienu parametrų rinkiniu po 100 kartų. Rasti laiko vėlinimų rinkiniai bus laikomi sėkmingais, kai jų tikslo funkcijos reikšmė viršys 1,0990. Visi PSO sprendinio paieškos rezultatai yra pateikti 2.4 lentelėje.

### 2.4 lentelė. PSO algoritmo imitavimo rezultatai.

$F_{\max}$  – geriausia tikslo funkcijos reikšmė, gauta po 100 bandymų su atitinkamais algoritmo parametrais;  $k$  – sėkmingų bandymų skaičius, kai tikslo funkcija bandyme viršija 1,0990 Mackey-Glass laiko eilutei.

Žingsnių skaičius	Spiečiaus dydis	PSO 1 rinkinys		PSO 2 rinkinys		PSO 3 rinkinys	
		$k$	$F_{\max}$	$k$	$F_{\max}$	$k$	$F_{\max}$
40	24	0	1,0826	0	1,0896	0	1,0794
40	40	0	1,0754	2	1,1019	2	1,0997
40	60	0	1,0971	4	1,1044	0	1,0986
100	24	0	1,0926	0	1,0952	1	1,1007
100	40	0	1,0903	3	1,1001	0	1,0866
100	60	0	1,0919	4	1,1019	1	1,1033

Pagal lentelės 2.4 rezultatus galima pastebėti, kad PSO algoritmas sėkmingiau veikia, kai turimas parametrų rinkinys  $w = 0,729$  ir  $c_1 = c_2 = 1,494$ . Tačiau ypatingai geresnių rezultatų už evoliuciniais algoritmais sugeneruotus rezultatus čia nėra gaunama. Nors didinant spiečiaus dydį galima tikėtis daugiau sėkmingų rinkinių, bet tada nukenčia skaičiavimo laikas, nes jau su 60 dalelių spiečiaus kompiuterinio laiko sąnaudos yra nepalyginamai didesnės nei sąnaudos, reikalingos

vykdyti evoliucinius algoritmus. Lygiai taip pat nukenčia skaičiavimo laikas, kai didinamas spiečiaus vystymosi žingsnių skaičius.

Kitas atsitiktinio perrinkimo algoritmas, kuris pritaikomas laiko eilutės nereguliaraus rekonstravimo į vėlinimų erdvę laiko vėlinimų rinkinio radimui, yra modeliuojamojo atkaitinimo modelis (SA). Šiuo atveju reikia išspręsti dvi pagrindines problemas: (a) kokia turi būti pasirenkama pradinė temperatūra; (b) kokia bus jos mažinimo formulė.

Pradinės temperatūros radimui pasirenkamas klasikinis metodas, ieškant atsitiktinai sugeneruotų sprendinių priėmimo santykio su visais sugeneruotais sprendiniais ir reikalaujama, kad tas santykis būtų ne mažesnis už 0,5. Temperatūra bus mažinama periodiškai po kiekvieno bandymo. Temperatūros mažinimo formule imama 1.4.3 skyrelyje minima (1.47) geometrinės progresijos bendrojo nario formulė, kurioje temperatūra po kiekvieno žingsnio dauginama iš parametro  $\kappa_\alpha$ . Stebimos tikslo funkcijos reikšmės temperatūrą mažinant trimis atvejais: kiekvieną kartą dauginant ją iš 0,95, 0,97 ir 0,99. Temperatūra bus mažinama užsibrėžtą iteracijų skaičių. SA optimizavimo rezultatai pateikiami 2.5 lentelėje. Modeliuojamojo atkaitinimo algoritmas yra pradedamas prisiskiriant pradinę laiko vėlinimų aibę, kurios elementai yra generuojami atsitiktiniai natūralieji skaičiai, priklausantys intervalui  $[1, 30]$ .

## 2.5 lentelė. SA algoritmo imitavimo rezultatai.

$F_{\max}$  – geriausia tikslo funkcijos reikšmė, gauta po 100 bandymų su atitinkamais algoritmo parametrais;  $k_g$  sėkmingų bandymų skaičius, kai tikslo funkcija bandyme viršija 1,0990 Mackey-Glass laiko eilutei.

Iteracijų skaičius	$\kappa_\alpha$	$k_g$	$F_{\max}$
100	0.95	0	1,071
500	0.95	3	1,1004
1000	0.95	0	1,0811
100	0.97	0	1,0613
500	0.97	0	1,0932
1000	0.97	7	1,101
100	0.99	0	1,0236
500	0.99	0	1,071
1000	0.99	3	1,1009

Tiriamo modeliuojamojo atkaitinimo algoritmo pritaikymas keliamam rekonstravimo uždavinio sprendimui yra pakankamai sudėtingas, nes tikslo funkcija yra labai jautri laiko vėlinimų rinkinio pasikeitimais. Modeliuojamojo atkaitinimo algoritmas gerai tinkamas sudėtingų tolydžių funkcijų globalių ekstremumų paieškai. Dėl savo gebėjimo kartais priimti blogesnę sprendinį šis algoritmas nelabai tinkamas pasiūlyto nereguliaraus rekonstravimo uždaviniui spręsti vien dėl to, kad vieno laiko vėlinimo pakeitimas per vieną poziciją gali iššaukti gana didelį tikslo funkcijos pokytį. Taigi dažnai tokia savybė tiesiog pablogina padėtį ir geresnio sprendinio radimas tampa visiškai atsitiktiniu procesu. Negelbsti šiuo atveju nei didelis iteracijų skaičius, nei pakankamai lėtas temperatūros mažinimas. Dėl šios



priežasties sukonstruoto uždavinio sprendimui šiame disertaciniame darbe pasirenkami evoliuciniai algoritmai.

### 2.3.4. Evoliucinių algoritmų pritaikymo laiko eilučių prognozei pavyzdžiai

Kaip jau buvo minėta anksčiau, pristatytas laiko eilutės prognozavimo modelis, pagrįstas nereguliariu atraktoriaus rekonstravimu į vėlinimų erdvę, gerai veikia, kai chaotinė laiko eilutė yra stacionari. Toliau darbe šis modelis bus taikomas skirtingoms laiko eilutėms prognozuoti, panaudojant evoliucinį algoritmą kaip kombinatorinį paieškos algoritmą laiko vėlinimų rinkinio radimui. Gauti rezultatai palyginami su kitų prognozės modelių rezultatais.

#### 2.3.4.1. Mackey-Glass laiko eilutė

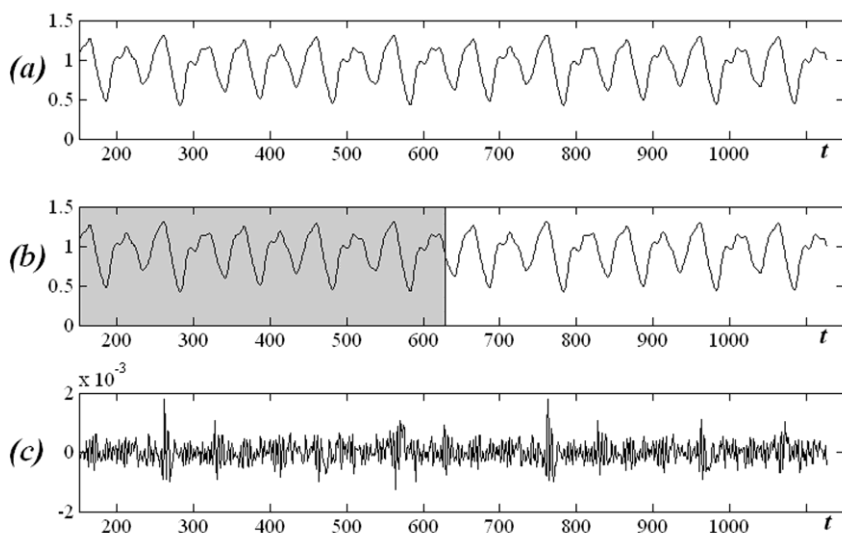
Mackey-Glass laiko eilutė naudojama siekiant sureguliuoti evoliucinio algoritmo parametrus. Vis dėl to, sukonstruoto modelio rezultatai palyginami su prognozavimo rezultatais [36, 153], kur taip pat buvo naudojama 50%/50% mokymo/testavimo laiko eilutės duomenų padalijimas. Atsižvelgiant į kitų autorių prognozės rezultatų kokybei vertinti naudojamas metrikas, šiame darbe gaunamų rezultatų tikslumui patikrinti naudojama RMSE metrika. Skirtingų modelių prognozavimo paklaidų RMSE palyginimas yra pateiktas 2.6 lentelėje. Galima pastebėti, kad mažiausios paklaidos gautos prognozuojant ANFIS tinkle, kurio įėjimo ir išėjimo duomenys tarpusavyje atsilieka nereguliais laiko vėlinimais, gautais nereguliariai rekonstruojant laiko eilutę į vėlinimų erdvę.

Darbe sukonstruotu modeliu gautos paklaidos yra 7 kartus mažesnės už prognozavimo ANN ar ARMA modeliais gautas paklaidas pademonstruojamas [36, 88, 153] šaltiniuose (ir žymiai geresnės paklaidos už kitų 2.6 lentelėje pateiktų metodų paklaidas).

**2.6 lentelė.** Mackey-Glass laiko eilutės prognozės paklaidų įverčio palyginimas

Prognozės modelis	AR	Atgalinio sklaidinio NN	Klasikinis RBF (su 23 neuronais)	Hibridinis ANN ir ARMA	Evoliucionuojantis RBF su įėjimų atranka	ANFIS su reguliariu rekonstravimu	ANFIS su savaiminiu apsimo-kymu	ANFIS su nereguliariu rekonstravimu
RMSE	0,19	0,02	0,0114	0,0025	0,00081	0,0007284	0,00055	0,0003497

Sukonstruoto modelio prognozavimo kokybė gali būti paašikinta tuo, kad tiriamą eilutę yra stacionarus chaotinis procesas. Atrinkinėjant artimą optimaliam laiko vėlinimų rinkinį, sukonstruojamas proceso modelis ir tuomet jis panaudojamas ekstrapoliuojant praeities elgseną į ateitį. Tikslų funkcijos išraiškoje, kuri naudojama kaip rekonstravimo į laiko vėlinimų erdvę kriterijus, nėra narių, kurie būtų kaip nors susiję su paklaidų vertinimo metrikomis. Taigi čia dirbama tik su rekonstruoto atraktoriaus į daugiamatę vėlinimų erdvę metrikomis. Toks skaitinis kriterijus yra sudaromas tam, kad būtų galima prognozuoti ilgalaikę stacionarios chaotinės sistemos elgseną. Eksperimentai su Mackey-Glass laiko eilute demonstruoja tokio metodo efektyvumą.



**2.30 pav.** Mackey-Glass laiko eilutės prognozė: (a) - pradinis signalas; (b) - ANFIS prognozavimas su nereguliais laiko vėlinimais; (c) - prognozavimo paklaidos.

Šiame darbe eksperimentiniais skaičiavimai buvo atlikti su MATLAB programiniu įrankiu, naudojantis AMD Sempron™ Procesoriumi 3400+, 1,81 GHz, 512 MB RAM. Sukonstruoto modelio skaičiavimo kaštai susideda iš 4 dalių: (i) optimalaus rekonstravimo į vėlinimų erdvę matavimo nustatymas; (ii) optimalaus reguliaraus rekonstravimo į vėlinimų erdvę laiko vėlinimų rinkinio nustatymas; (iii) artimo optimaliam laiko vėlinimų rinkinio parinkimas evoliuciniais algoritmais; (iv) laiko eilutės prognozavimas ANFIS tinklu.

Visų šių 4 punktų skaičiavimo sąnaudų sujungimas neleistų mums teisingai įžvelgti uždavinio skaičiavimo sudėtingumo. Šiame darbe vertinamos skaičiavimo sąnaudos visoms 4 dalims atskirai.

Rekonstravimo į vėlinimų erdvę matavimo  $d$  nustatymas iš mūsų pareikalauja maždaug 7 sekundžių, kai skaliarinės laiko eilutės ilgis yra 1000 taškų. Skaičiavimo sąnaudos parinkti optimalių laiko vėlinimų rinkinį tolygiai rekonstruojant laiko eilutę priklauso nuo rekonstruojamos vėlinimų erdvės matavimo  $d$  (kai  $d = 6$ , tai užima 7 sekundes). Evoliuciniu algoritmu generuojant 100 populiacijų atrinkti geriausius laiko vėlinimų rinkinius Mackey-Glass eilutei užėmė 97 sekundes. Daugiausia laiko, 8 min, trunka laiko eilutės prognozavimui naudojamos ANFIS sistemos mokymas ir testavimas.

Pasiūlytu modeliu sudėtinga prognozuoti atsitiktinius impulsus, šuolius duomenyse, sezoniskumą, jei tokie reiškiniai nepasirodė ANFIS sistemos apmokymo metu. Kaip minėta anksčiau, šis metodas gerai veikia tiriant ergodinius stacionarius procesus. Sekančiuose skyreliuose palyginamas sukonstruoto modelio ir kitų prognozavimo modelių funkcionalumas.

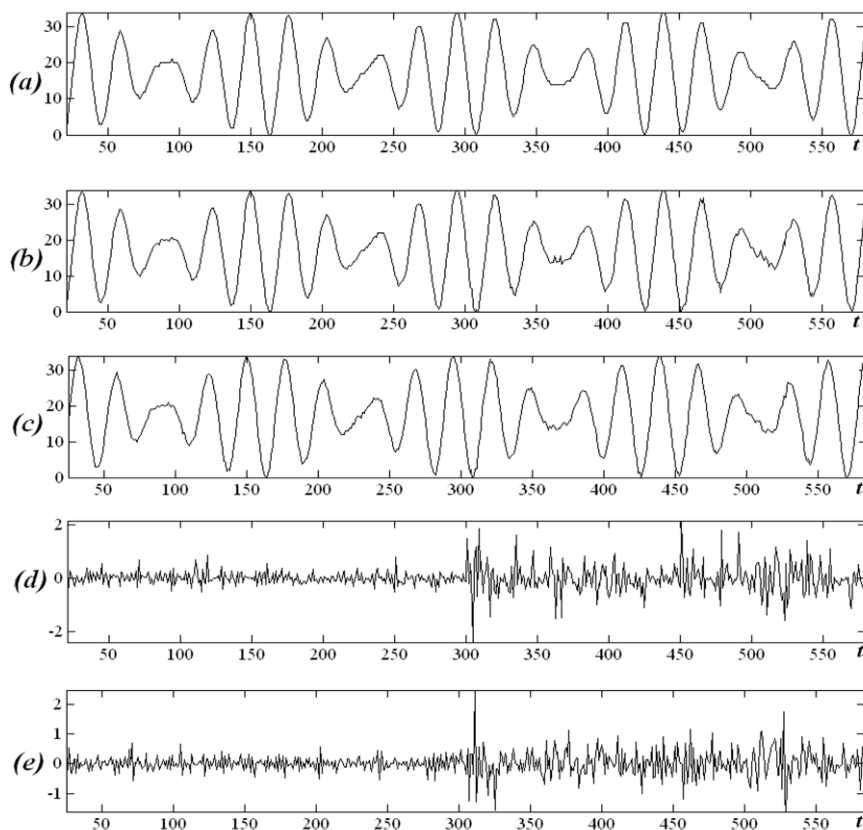
#### **2.3.4.2. Stebimos kintamos žvaigždės kasdienio ryškumo laiko eilutė**

Testuojamas evoliucinio algoritmo funkcionalumas su fiksuotais parametrais (optimaliu laikomas Mackey-Glass eilutei naudojamų evoliucinių algoritmų parametrų rinkinys) prognozuojant realią laiko eilutę, kuri apibūdina kintamos žvaigždės kasdienį ryškumą, matuotą 600 vidurnakčių iš eilės. 2.2.2 skyrelyje buvo

naustatyta, kad šią laiką eilutę geriausia rekonstruoti į šešiamatę laiko vėlinimų erdvę. Gautas reguliaraus rekonstravimo optimalus laiko vėlinimų rinkinys yra  $\{4, 4, 4, 4, 4, 4\}$ , o tikslo funkcijos reikšmė  $F(4, 4, 4, 4, 4, 4) = 1,0719$ . Su tokiais parametrais prognozuojant laiko eilutės ateities reikšmes gauta, kad paklaidų RMSE lygi 0,4259 (pateikta 2.7 lentelėje). Galima pastebėti, kad su reguliaraus rekonstravimo gautais laiko vėlinimais ANFIS tinklo prognozavimas duoda 2 kartus geresnius rezultatus, nei prognozavimas su hibridiniu ANN ir ARMA modeliu, demonstruojamu [36] šaltinyje.

**2.7 lentelė.** Kintamos žvaigždės kasdienio ryškumo laiko eilutė

Prognozės modelis	ANN	Hibridinis ANN ir ARMA modelis	ANFIS su reguliariu rekonstravimu	ANFIS su nereguliaru rekonstravimu
RMSE	1,438	0,868	0,4259	0,3815



**2.31 pav.** Kintamos žvaigždės ryškumo laiko eilutės prognozė: (a) - pradinis signalas; (b) - ANFIS prognozavimas su reguliariais laiko vėlinimais; (c) - ANFIS prognozavimas su nereguliais laiko vėlinimais; (d) - ANFIS prognozavimo paklaidos su reguliariais laiko vėlinimais; (e) - ANFIS prognozavimo paklaidos su nereguliais laiko vėlinimais.

Toliau nagrinėjama laiko eilutės nereguliarių rekonstravimą į vėlinimų erdvę. Iš pradžių evoliucinio algoritmo pagalba randamas laiko vėlinimų rinkinys artimas optimaliam rinkiniui. Tokiu būdu randama, kad geriau prognozuoti laiko eilutę

galima turint laiko vėlinimų rinkinį  $\{5, 4, 5, 4, 6\}$  (tikslo funkcijos reikšmė  $F(5, 4, 5, 4, 6)=1,0729$ ). Šiuo atveju prognozavimo paklaidų RMSE lygi 0,3815 (prognozavimo rezultatai ir paklaidos pavaizduotos 2.31 paveiksle). Skaičiavimo sąnaudos prognozuojant šią laiko eilutę su ANFIS sistema papildomai užtruko dar beveik vieną minutę, lyginant su Mackey-Glass eilute (tam tikriausiai turėjo įtakos realaus pasaulio laiko eilutės sudėtingesnė struktūra).

#### 2.3.4.3. Elektrokardiogramos (EKG) signalo laiko eilutė

Toliau darbe pristatytas modelis taikomas labai nestacionariam ir triukšmingam elektrokardiogramos (EKG) signalui (iš [155] šaltinio) prognozuoti. Stebėtos laiko eilutės bandymų dažnis yra  $1/\delta = 250\text{Hz}$ . Prieš rekonstruojant laiko eilutę reikia nustatyti, į kokio matavimo vėlinimų erdvę geriausia rekonstruoti eilutės atraktorių optimizuojant jo tūrį. Pagal 1.2 skyrelyje aprašytą koreliacinės sumos (1.28) formulę vėl tikrinama, kada nusistovi šios sumos natūraliojo logaritmo funkcija, priklausanti nuo  $\ln R$ . Gaunama, kad EKG signalo laiko eilutę parankiausia rekonstruoti į  $d = 8$  laiko vėlinimų erdvę. Laiko eilutė rekonstruojama į vėlinimų erdvę nereguliais laiko vėlinimais. Tačiau, kaip minėta anksčiau, norint rasti optimalų laiko vėlinimų rinkinį šiuo atveju sudėtinga atlikti laiko vėlinimų pilną perrinkimą. Todėl čia panaudojami evoliuciniai algoritmai ir ieškomas laiko vėlinimų rinkinys, kuris būtų artimas optimaliam rinkiniui.

Evoliucinio algoritmo pagalba randamas artimas optimaliam laiko vėlinimų rinkinys  $\{14, 12, 15, 13, 12, 14, 13\}$ . Su šiuo rinkiniu gauta tikslo funkcijos reikšmė  $F(14, 12, 15, 13, 12, 14, 13)=1,0530$ . Prognozuojama  $x(t+13\delta)$  reikšmė, kai turimi praeities duomenys  $x(t-80\delta)$ ,  $x(t-66\delta)$ ,  $x(t-54\delta)$ ,  $x(t-39\delta)$ ,  $x(t-26\delta)$ ,  $x(t-14\delta)$  ir  $x(t)$ . Konstruojamas 1000 duomenų vektorius  $y_p^{(8)}$ , kai laikas  $t$  kinta nuo  $275+180\delta$  iki  $275+1179\delta$ . Pirmi 500 duomenų naudojami ANFIS tinklo apmokymui, o kiti 500 – prognozavimo paklaidų tikrinimui. Duoto signalo pradiniai duomenys, jo prognozavimo rezultatai ir paklaidų grafikai pateikti 2.32 paveiksle. Prognozės paklaidų RMSE reikšmė siekia 0,1131. Nors prognozavimo paklaidos bendroju atveju yra pakankamai didelės, tačiau ANFIS metodas sugeba pagauti EKG signalo pagrindinius pikus (nors ateities reikšmės prognozuojamos iš duomenų, kurie jau prieš tai buvo suprognozuoti tinklo mokymosi metu). Šiuo signalu gauti rezultatai patvirtino, kad siūlomas modelis geriausiai tinka stacionariems chaotiniams signalams prognozuoti.

Laiko vėlinimų erdvės, į kurią rekonstravome signalą, matavimo  $d$  nustatymas užtruko 7 s. Reguliaraus rekonstravimo laiko vėlinimų rinkinys buvo ieškomas 9 s. Šimto populiacijų generavimas ir evoliucionavimas užtruko 177 s. Daugiausia laiko atėmė laiko eilutės mokymas ir paklaidų tikrinimas ANFIS sistema – 14 min.

Sukonstruoto modelio rezultatai palyginami su kitų prognozavimo modelių rezultatais. Laiko eilutė prognozuojama su Boks-Dženkins (Box-Jenkins) procedūra ARIMA(4, 1, 3). Eksperimentais pavyko nustatyti, kad šiam signalui geriausiai tinka modelio architektūra 4-1-3 ( autoregresijos parametras 4, integruotosios dalies parametras 1 ir slenkančio vidurkio parametras yra 3). Parinkto modelio adekvatumui tikrinti naudota Ljung-Box testas (Q statistika), pateiktas priede

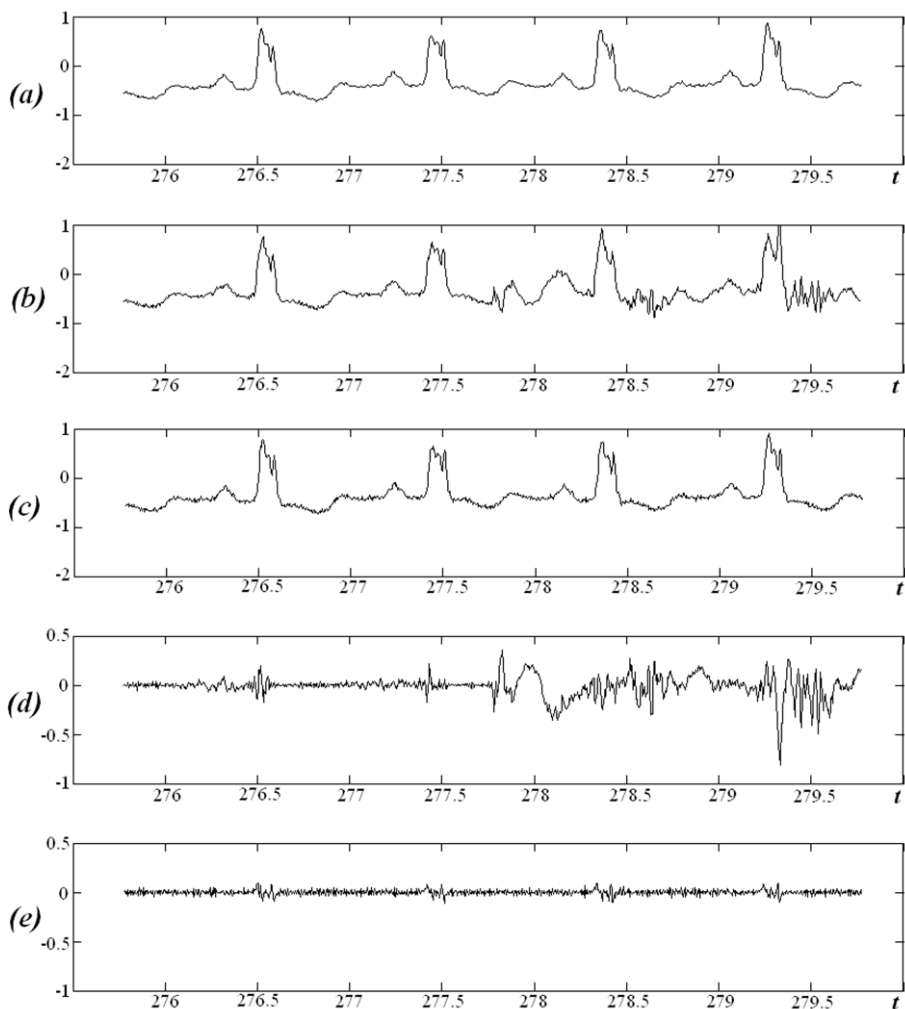
pirmoje lentelėje. Pati statistika detaliai aprašyta [403940], kur teigiama, kad modelis yra sudarytas adekvačiai, jei kriterijaus reikšmingumo tikimybė yra didesnė už pasirinktą reikšmingumo lygmenį. Reikšmingumo lygmuo šiuo atveju yra parinkta 0,05, kaip rekomenduojama [40].

Prognozavimo paklaidų RMSE reikšmės pateiktos 2.8 lentelėje. Iš lentelės galima pastebėti, kad ARIMA modelio prognozavimo rezultatai 5 kartus geresni, nei darbe sukonstruoto modelio. Nagrinėto modelio prognozavimo paklaidų grafikas pateiktas 2.32 paveikslo *e* punkte.

**2.8 lentelė.** EKG laiko eilutė

Prognozės modelis	ARIMA(4,1,3)	ANFIS su nereguliariu rekonstravimu
RMSE	0,0257	0,1131

Tiesiogiai palyginti ARIMA modelio skaičiavimo sąnaudas su mūsų siūlomo modelio sudėtinga. ARIMA skaičiavimo sąnaudas tektų skaidyti į dvi dalis: (i) optimalios modelio architektūros nustatymas, t.y. reikia nustatyti kokio sudėtingumo bus autoregresijos, integruotoji ir slenkančio vidurkio dalys. (ii) laiko eilutės prognozavimas. Sudedant šių dviejų atskirų dalių skaičiavimo sąnaudas, ARIMA prognozavimo efektyvumas nediskriminuojamas, iš tikrųjų efektas gali būti priešingas. Bendruoju atveju ARIMA modelio prognozavimas yra žymiai greitesnis nei mūsų modelio. Tačiau ARIMA modelio architektūros nustatymas gali užimti daug daugiau laiko nei mūsų modelyje, ieškant optimalaus laiko vėlinimų rinkinio, perduodamo ANFIS sistemos prognozavimui.



**2.32 pav.** EKG signalo laiko eilutės prognozė: (a) - pradinis signalas; (b) - ANFIS prognozavimas su nereguliais laiko vėlinimais; (c) – ARIMA prognozavimas; (d) - ANFIS prognozavimo paklaidos; (e) - ARIMA prognozavimo paklaidos.

#### 2.3.4.4. S&P (Standard & Poor's) 500 akcijų svertinės vertės mėnesinės grąžos laiko eilutė

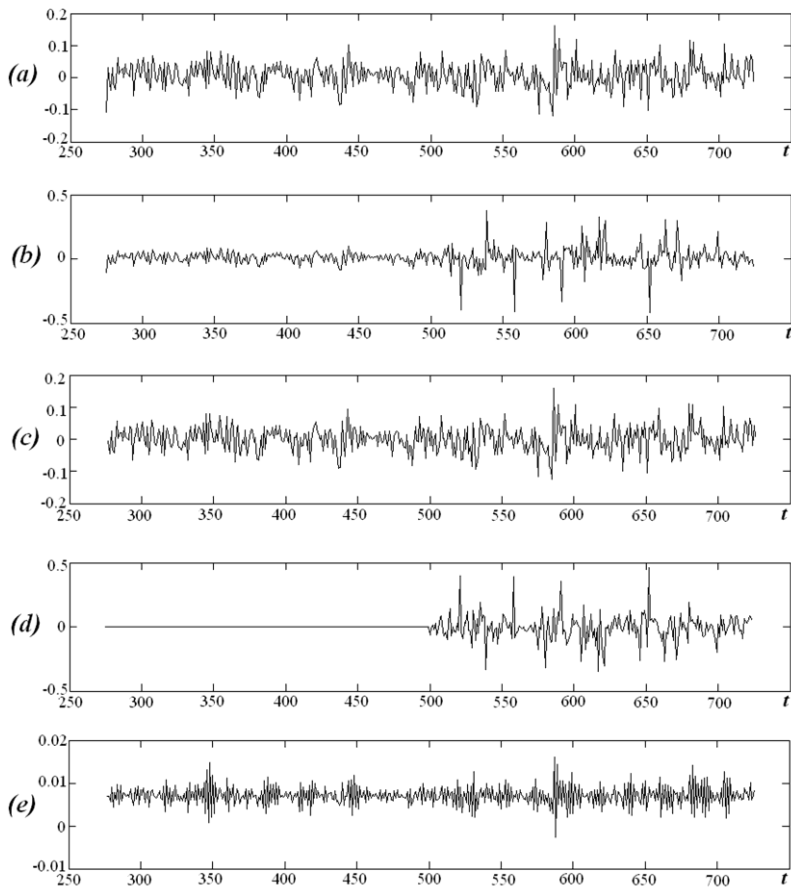
Toliau prognozuojama laiko eilutė, sudaryta iš S&P 500 akcijų svertinės vertės mėnesinių grąžų nuo 1926 iki 1991 metų [142]. Anksčiau aprašyta koreliacinės sumos pagalba yra randama, kad šią laiko eilutę geriausia rekonstruoti į  $d=8$  matavimo vėlinimų erdvę. Optimali reguliaraus rekonstravimo į vėlinimų erdvę laiko vėlinimų aibė yra  $\{5, 5, 5, 5, 5, 5, 5\}$ . O evoliucinio algoritmo pagalba randama artima optimaliai neregularaus rekonstravimo laiko vėlinimų aibė  $\{9, 16, 10, 11, 9, 11, 10\}$ . Su šia aibe apskaičiuojama tikslo funkcijos reikšmė  $F(9, 16, 10, 11, 9, 11, 10)=1,0198$ . Taigi prognozuojama reikšmė  $x(t+10\delta)$  (kai  $\delta=1$ ) turint praeities duomenis  $x(t-66\delta)$ ,  $x(t-57\delta)$ ,  $x(t-41\delta)$ ,  $x(t-31\delta)$ ,  $x(t-20\delta)$ ,  $x(t-11\delta)$  ir  $x(t)$ . Konstruojami vektoriai  $y_p^{(8)}$  iš 450 eilutės duomenų,

nuo  $t = 267$  iki  $t = 716$ . Pirmieji 225 duomenys panaudojami ANFIS tinklo apmokymui, o likę prognozavimo paklaidoms tikrinti. Duotosios laiko eilutės duomenys, jos prognozavimo rezultatai ir paklaidų grafikas pateikti 2.33 paveiksle. Paklaidų RMSE šiuo atveju lygi 0,0660. Skaičiavimo sąnaudos šiai laiko eilutei yra tokios pačios kaip ir EKG signalo, išskyrus ANFIS prognozavimo (truko 1 min. trumpiau).

Prognozavimo tikslumui nustatyti, palyginkime sukonstruoto prognozės modelio rezultatus su ARIMA modelio prognozės rezultatais. Eksperimentiniu būdu buvo nustatyta, kad Box-Jenkins procedūros ARIMA(1, 0,1) architektūra 1-0-1 yra geriausias modelis šiems duomenims prognozuoti (modelio adekvatumo tyrimas pateiktas priede lentelėje Nr. 2). Prognozavimo paklaidų RMSE pateikta 2.9 lentelėje. Galima pastebėti, kad ARIMA modelio RMSE maždaug 1,5 karto geresnės, nei darbe sukonstruoto modelio. Be to, šiuo atveju nepasiteisino nereguliaraus rekonstravimo laiko eilutės prognozė, nes RMSE paklaidos gaunamos prastesnės. Iš to peršasi išvada, kad jei laiko eilutė yra labai triukšminga bei neergodinė, tai joje yra sudėtinga atrasti tam tikrus dinامينius sąryšius. Todėl sukonstruotas modelis ne visada duoda puikius rezultatus, tiriant tokias eilutes. Šioje vietoje galima prisiminti įvade paminėtą teiginį, kad neįmanoma sukurti vieno metodo, kuris būtų geresnis už visus kitus metodus visose galimose situacijose. Prognozavimo paklaidos demonstruojamos 2.33 paveiksle.

**2.9 lentelė.** S&P 500 akcijų laiko eilutė

Prognozės modelis	ARIMA(1,0,1)	ANFIS su reguliarium rekonstravimu	ANFIS su nereguliarium rekonstravimu
RMSE	0,0387	0,0635	0,066



**2.33 pav.** S&P 500 signalo prognozė: (a) - pradinis signalas; (b) - ANFIS prognozavimas su nereguliais laiko vėlinimais; (c) – ARIMA prognozavimas; (d) - ANFIS prognozavimo paklaidos; (e) - ARIMA prognozavimo paklaidos.

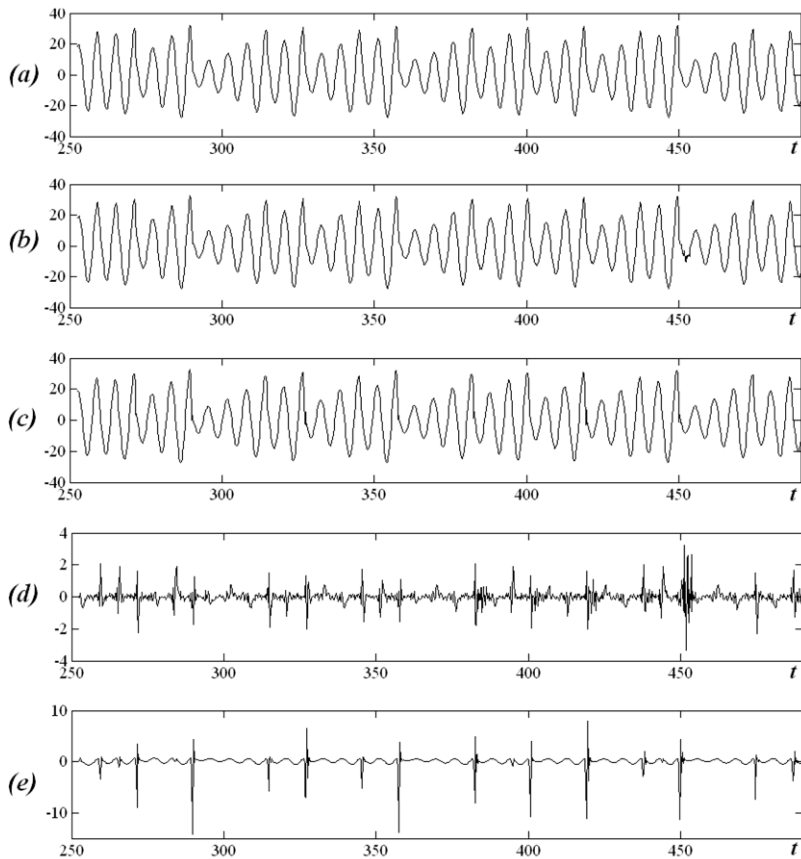
#### 2.3.4.5. Rössler laiko eilutė

Rössler chaotinis atraktorius buvo nagrinėtas 2.2.2 skyrelyje. Nustatyta, kad šį atraktorių geriausia rekonstruoti į šešiamatę erdvę. Evoliucinio algoritmo pagalba randamas artimas optimaliam laiko vėlinimų rinkinys  $\{4, 4, 3, 4, 4\}$ , su kuriuo tikslo funkcijos reikšmė lygi  $F(4, 4, 3, 4, 4) = 1,0805$ , ir yra geresnė, nei reguliaraus rekonstravimo atveju  $F(5, 5, 5, 5, 5) = 1,0760$ . Tačiau ši reikšmė nesiekia optimalios pilno perrinkimo reikšmės  $F(4, 6, 4, 3, 3) = 1,0823$ . Taigi, su evoliuciniu algoritmu bet kuriuo atveju galima tikėtis gauti geresnes prognozes nei reguliaraus rekonstravimo atveju. Šiuo atveju su nustatytu laiko vėlinimų rinkiniu iš praeities reikšmių  $x(t-15\delta)$ ,  $x(t-11\delta)$ ,  $x(t-7\delta)$ ,  $x(t-4\delta)$  ir  $x(t)$  yra prognozuojama reikšmė  $x(t+4\delta)$ . Konstruojamas 1000 duomenų vektorius  $y_p^{(6)}$ , kai laikas kinta nuo  $t = 252,80$  iki  $t = 252,80 + 1000\delta$ . Pirmi 500 šių duomenų naudojami ANFIS tinklo apmokymui, o likę 500 prognozavimo paklaidų tikslumui tikrinti. Sukonstruota pradinių duomenų laiko eilutė, jos prognozavimas ir prognozavimo paklaidos



demonstruojami 2.34 paveiksle. Prognozavimo RMSE lygi 0,4839. Šios eilutės prognozavimui skaičiavimo sąnaudos tokios pat, kaip ir Mackey-Glass eilutės prognozavimo atveju.

Sukonstruoto modelio prognozavimo rezultatai palyginami su kito modelio rezultatais. Konstruojama Box-Jenkins procedūrą ARIMA(2,0,3). Bandymų keliu nustatoma, kad ARIMA geriausiai prognozuoja, kai modelio architektūra 2-0-3 (modelio adekvatumo tyrimas pateiktas priede lentelėje Nr. 3). ARIMA prognozavimo paklaidų (kurių grafikas demonstruojamas 2.34 paveiksle) RMSE galime rasti 2.10 lentelėje. Taigi lyginant su darbe sukonstruoto modelio prognozės paklaidų RMSE, ARIMA modelio yra 3 kartus didesnės.



**2.34 pav.** Rossler chaotinės laiko eilutės prognozė: (a) - pradinis signalas; (b) - ANFIS prognozavimas su nereguliariais laiko vėlinimais; (c) – ARIMA prognozavimas; (d) - ANFIS prognozavimo paklaidos; (e) - ARIMA prognozavimo paklaidos.

**2.10 lentelė.** Rossler laiko eilutė

Prognozės modelis	ARIMA(2,0,3)	ANFIS su reguliarium rekonstravimu	ANFIS su nereguliariu rekonstravimu
RMSE	1,2723	0,4926	0,4839

## 2.4 Skyriaus išvados

Šiame skyriuje sukonstruotas prognozavimo modelis labai gerai tinka dirbant su stacionariomis laiko eilutėmis, kurios yra stipriai paveiktos triukšmo. Sukurtas modelis mėgina konstruoti netiesinį sistemos modelį. Bendru atveju, sukonstruotoje tikslo funkcijoje (ir evoliucinių algoritmų optimizavimo procedūroje) laiko eilutės dinamiką reprezentuojama per Furjė amplitudinį spektrą (2.18 lygtis). Iš tiesų beveik nėra skirtumo (išskyrus dažnių langą) ar Furjė amplitudinis spektras yra skaičiuojamas tik su duomenimis, naudojamais ANFIS procedūros apmokymui ar su visais duomenimis, skirtais ir prognozavimo testavimui, jei tik procesas yra stacionarus. Nors ANFIS prognozavimas yra atskira procedūra, vykdoma po to kai randamas artimas optimaliam laiko vėlinimų rinkinys, puikūs Mackei-Glass ir Rosslerio laiko eilučių prognozavimo rezultatai gali būti aiškinami tuo, kad abi laiko eilutės yra stacionarios. Kitais žodžiais tariant, optimali laiko vėlinimų aibė tinka tiek apmokymo aibei, tiek testavimo aibei (kas, pavyzdžiui, netinka EKG eilutei, nes tai nėra stacionarus procesas).

Šiame darbe laiko eilučių rekonstravimas vyksta skirtingo dimensiškumo vėlinimų erdvėse. Mato nustatymui, naudojamas žinomas koreliacinių sumų skaičiavimo metodas. Šiame darbe laiko vėlinimų erdvės matavimas yra tiesiogiai susijęs su ANFIS tinklo įėjimų skaičiumi (naudojami  $d-1$  įėjimai ir vienas išėjimas, pagal kurį prognozuojama sistemos būsena ateityje). Dinaminių sistemų teorijoje rekonstruoto atraktoriaus erdvės matavimas naudojamas visai kitiems tikslams (apibūdinant nagrinėjamos sistemos kompleksiskumo matą).

Ieškant artimo optimaliam laiko vėlinimų rinkinio evoliucinio algoritmo parametrai derinami konkrečiai Mackey-Glass eilutei. Taip daroma dėl to, kad jau yra žinomas tikslus optimalus sprendinys (atlikus pilną laiko vėlinimų rinkinio perrinkimą). Tai nereiškia, kad suderinti evoliucinio algoritmo parametrai tinkami tik šiai laiko eilutei. Evoliucinių algoritmų pagalba yra randamas ne globalus Mackey-Glass laiko eilutės optimalus sprendinys. Panašiai, evoliucinis algoritmas neranda globalaus tikslo funkcijos maksimumo Rosslerio laiko eilutei; vis dėlto ANFIS prognozavimo rezultatai yra geresni nei ARIMA rezultatai. Jeigu būtų galima nepaisyti skaičiavimo sąnaudų būtų galima atmesti evoliucinių algoritmų naudojimą, o vietoje to atlikti pilną visų galimų laiko vėlinimų rinkinių perrinkimą. Tada ANFIS prognozavimas būtų dar tikslesnis.

Natūralu, kad darbe sukonstruotas modelis nėra geriausias pasirinkimas, jei norima prognozuoti stipriai laike kintančią laiko eilutę. Kaip minėta anksčiau, vienas modelis negali dirbti geriau už kitus modelius visose situacijose. Tačiau jei būtų mėginama jungti sukurtą modelį su kitų prognozavimo modelių funkcionalumu (kai, pavyzdžiui, tikslo funkcija įtrauktų ir paklaidos, ir atraktoriaus metrikas), tektų atlikti dar vieną įdomų laiko eilučių prognozės tyrimą.

### 3. TRUPŲ LAIKO EILUČIŲ PROGNOZAVIMAS

Trumpų laiko eilučių prognozei šiame disertaciniame darbe yra naudojama Hankelio matricos rango sąvoka. Tad pirmame skyrelyje apžvelgiama tyrimui aktuali jau žinoma Hankelio matricos sekų konstravimo teorija, kuri yra tyrinėjama [113, 156] šaltiniuose.

#### 3.1. Prognozės strategija

Tarkim, kad duotos  $2n+1$  laiko eilutės reikšmės, kurios bus naudojamos sukonstruoti proceso modelį. Naudojant šį modelį bus galima ekstrapoliuoti proceso praeities elgseną ateityje:

$$x_0, x_1, x_2, \dots, x_{2n-1}, x_{2n}; \quad (3.1)$$

čia  $x_{2n}$  yra dabartinio momento reikšmė. Turint nelyginį stebėjimų skaičių sukonstruojamas Hankelio matricos minoras  $H_0^{(n+1)}$ :

$$H_0^{(n+1)} = \begin{bmatrix} x_0 & x_1 & \dots & x_n \\ x_1 & x_2 & \dots & x_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n+1} & \dots & x_{2n} \end{bmatrix}. \quad (3.2)$$

Jei  $\det H_0^{(n+1)} = 0$ , tai seka yra algebrinė progresija ir  $x_{2n+1}$  galima rasti tiesiogiai pagal 1.3.2. skyrelio (1.41) formulę. Tačiau, jei seka  $x_0, x_1, x_2, \dots, x_{2n}$  nėra algebrinė progresija, tai įmanoma prognozuoti  $x_{2n+1}$  pagal (1.41) formulę?

Priimama kita prielaida. Tarkime, kad (3.1) išraiškos seka gauta prie algebrinės progresijos narių pridėdant atsitiktinį triukšmą. Kitais žodžiais tariant, formuluojama hipotezė, kad

$$x_k := \tilde{x}_k + \varepsilon_k; \quad k = 0, 1, 2, \dots, 2n; \quad (3.3)$$

čia  $\varepsilon_k$  yra pridėtas triukšmas. Tuomet determinantas:

$$\det \tilde{H}_0^{(n+1)} = \det \begin{bmatrix} x_0 - \varepsilon_0 & x_1 - \varepsilon_1 & \dots & x_n - \varepsilon_n \\ x_1 - \varepsilon_1 & x_2 - \varepsilon_2 & \dots & x_{n+1} - \varepsilon_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_n - \varepsilon_n & x_{n+1} - \varepsilon_{n+1} & \dots & x_{2n} - \varepsilon_{2n} \end{bmatrix} = 0; \quad (3.4)$$

čia

$$\tilde{H}_j^{(m)} := [x_{r+s-2+j} - \varepsilon_{r+s-2+j}]_{1 \leq r, s \leq m}. \quad (3.5)$$

Dar daugiau, tegu begalinė seka  $\tilde{x}_k$ ;  $k = 0, 1, 2, \dots$  yra algebrinė progresija, ir ši seka yra tam tikro tipo skeletinė seka, apibrėžianti globalinę laiko eilutės dinamiką.

Darbo tikslas išskirti skeletinę algebrinę seką  $\tilde{x}_k$ ;  $k=0,1,2,\dots$  iš duotosios algebrinės sekos  $x_k$ ;  $k=0,1,2,\dots$  atimant atsitiktinį triukšmą  $\varepsilon_k$ ,  $k=0,1,2,\dots, 2n$ .

Tokiu atveju kyla klausimas, kaip būtų galima identifikuoti šį pridėtinį triukšmą  $\varepsilon_k$ ,  $k=0,1,2,\dots,2n$ ? Akivaizdu, kad šis uždavinys turi be galo daug sprendinių. Tikslas yra minimizuoti bet kokius nuokrypius nuo tikrosios laiko eilutės, t. y. nuo tariamos algebrinės progresijos. Todėl, korekcijų sekai  $\{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}\}$  darbe siūloma tokia tinkamumo funkcija, kurią reikia maksimizuoti:

$$F_1(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}) = \frac{1}{a_e \left| \det \left( \tilde{H}_0^{(n+1)} \right) \right| + \sum_{k=0}^{2n} \lambda_k |\varepsilon_k|}; \quad a_e > 0; \quad (3.6)$$

čia

$$\lambda_k = \frac{\exp(b_\lambda(k+1))}{\sum_{j=0}^{2n} \exp(b_\lambda(j+1))}; \quad k=0,1,\dots,2n; \quad b_\lambda > 0. \quad (3.7)$$

Akivaizdu, kad  $\sum_{k=0}^{2n} \lambda_k = 1$ ,  $0 < \lambda_0 < \lambda_1 < \dots < \lambda_{2n-1} < \lambda_{2n}$ .

Jeigu  $\det \tilde{H}_0^{(n+1)} = 0$ , tuomet tinkamumo funkcijos reikšmė bus didžiausia, kai  $\varepsilon_0 = \varepsilon_1 = \dots = \varepsilon_{2n} = 0$  ir tuomet  $F(0,0,\dots,0) = +\infty$ . Parametras  $a_e$  apibrėžia proporcijos tarp determinanto reikšmės ir korekcijų su svoriais sumos baudą (abi baudos turi tą pačią reikšmę, kai  $a_e = 1$ ). Koeficientai  $\lambda_0, \lambda_1, \dots, \lambda_{2n}$  nustato tolerancijos koridorių lokalinėms korekcijoms  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$ . Visų lokaliųjų korekcijų svoriai bus vienodi, jei  $b_\lambda = 0$ . Kuo didesnis  $b_\lambda$  imamas, tuo didesnis svoris suteikiamas lokalinėms korekcijoms dabarties momentui, lyginant su korekcijų svoriais praeities momentais. Kitais žodžiais tariant, kuo koreguojamasis taškas yra toliau praeityje, tuo labiau toleruojama tame taške esančios reikšmės korekcija. Taip daroma prielaida, kad kuo duomenys yra arčiau dabarties, tuo jie svarbesni.

Akivaizdu, kad lokaliųjų korekcijų aibės  $\{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}\}$  radimas nėra paprastas skaičiavimo uždavinys. Pirmiausia reikia pabrėžti, kad suformuluotas uždavinys neturi vienintelio sprendinio. Žinoma, korekcijos turi būti kuo mažesnės, bet pagal (3.6) lygtį taip pat yra leidžiama, kad pakoreguotų duomenų Hankelio matricos determinantas nebūtinai būtų tiksliai lygus nuliui. Kitais žodžiais tariant, skaičiuojamas pakoreguotos sekos pseudo rangas ir mėginama balansuoti tarp determinanto reikšmės ir lokaliųjų korekcijų. Ieškant optimalios lokaliųjų korekcijų aibės toliau yra naudojamosi programiniu skaičiavimu remiantis tinkamumo funkcija, kuri išreiškta (3.6) formule.

Taigi uždavinys – maksimizuojant tinkamumo funkciją (3.6) rasti tokią korekcijų  $\varepsilon_k$  seką, kuria paveikus originalią seką, gaunamas duomenų rinkinys, iš kurio suformuotos Hankelio matricos determinantas yra beveik lygus nuliui. Tuo pat metu reikalaujama, kad ir korekcijų svorinė suma būtų

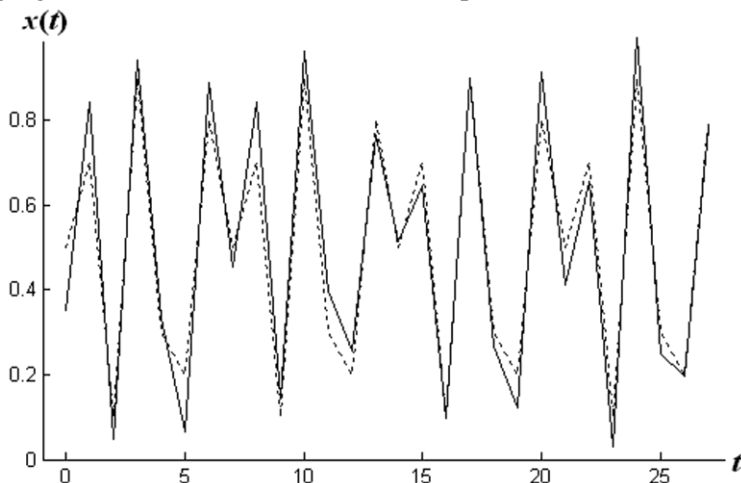
kuo mažesnė. Tai leidžia išskirti apytikslį algebrinį dėsnį, aprašantį analizuojamos sekos modelį.

Pakoregavus seką lokalinėmis korekcijomis, toliau ieškomos sekos ateities reikšmės pagal (1.41) formulę (net jei pakoreguotos sekos determinantas nėra visiškai lygus nuliui).

Galima pastebėti, kad pristatytas prognozavimo modelis (nors ir grindžiamas Hankelio matricomis) nenaudoja būsenų erdvės modelio koncepcijos, arba dinaminių sistemų koncepcijos. Tai netiesinis algebrinis modelis, kadangi Hankelio matricos rango koncepcija apima netiesinius algebrinius sąryšius. Kitais žodžiais tariant, šis modelis naudoja netiesinį algebrinį skeletinės sekos atpažinimo metodą. Kita svarbi pristatyto prognozavimo modelio savybė yra būtinybė atpažinti sekos bazinį fragmentą, t.y. prieš pradedant koreguoti sekos narius būtina fiksuoti parametą  $n$ . Akivaizdu, kad koreguoti galima skirtingo ilgio sekas. Tikslas yra parinkti tokio ilgio seką (bazinį fragmentą), kuriai būtų reikalingos minimalios korekcijos. Šie klausimai, kartu su detalizuotomis diskusijomis apie programinio skaičiavimo strategijas yra nagrinėjami tolesniuose skyreliuose.

### 3.2. Skaitiniai eksperimentai su testine laiko eilute

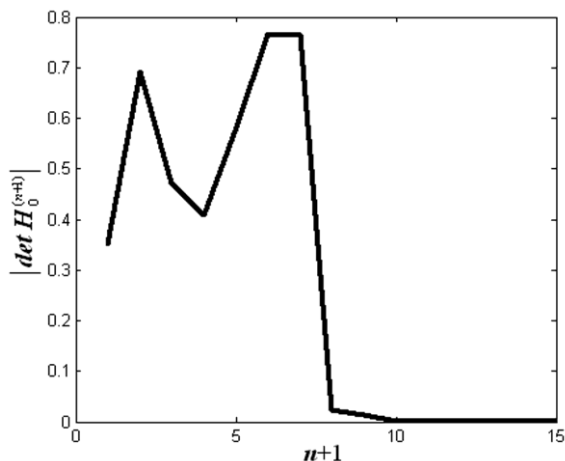
Optimalios lokaliųjų korekcijų aibės  $\{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}\}$  paieškai šiame darbe naudojami evoliuciniai algoritmai arba dalelių spiečiaus algoritmas. Tačiau prieš tai konstruokime dirbtinę testinę laiko eilutę, kuri naudojama derinant šių algoritmų parametrus. Pirmiausia suformuojama periodinė seka (periode yra septynių elementų skaitinės reikšmės 0,5; 0,7; 0,1; 0,9; 0,3; 0,2; 0,8). Ši seka reprezentuoja skeletinę algebrinę seką. Toliau kiekvienam sekos elementui pridedami atsitiktiniai skaičiai, pasiskirstę pagal tolygųjį skirstinį intervale  $[-0,15; 0,15]$  (testinė eilutė demonstruojama 3.1 paveiksle punktyrine linija). Testinė laiko eilutė naudojama pristatyto prognozavimo modelio funkcionalumo patikrinimui.



**3.1 pav.** Testinės laiko eilutės (ištiesinė linija) ir periodinės laiko eilutės (punktyrinė linija) grafinis vaizdas

Pirmoji užduotis – rasti laiko eilutės bazinį fragmentą, kuris bus naudojamas skeletinės algebrinės sekos rekonstravimui. Sukonstruotai skeletinei sekai šis

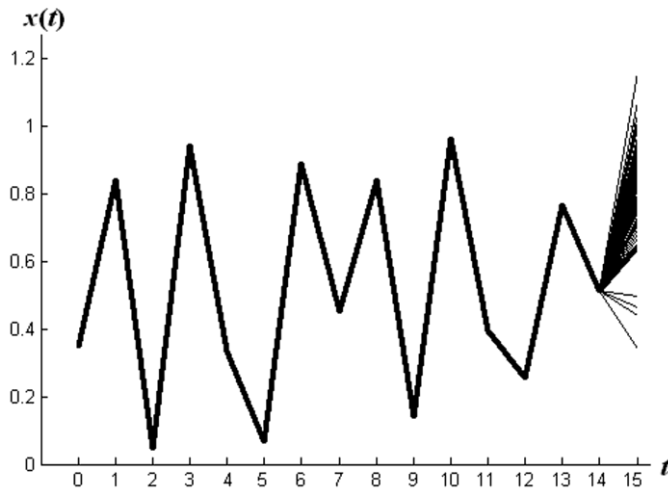
uždavinys yra trivialus ( $H$ -rangas šiuo atveju lygus 8), tačiau laiko eilutei su triukšmu  $H$ -rangą reikia nustatyti (nėra toks akivaizdus). Ryšys tarp Hankelio matricos determinanto absoliučiosios reikšmės ir Hankelio matricos mato yra pavaizduotas 3.2 paveiksle. Galima pastebėti, kad determinantas nelygus nuliui, kai matricos matas yra 8 (taip yra dėl pridėto triukšmo). Vis dėlto šiame darbe pasirenkamas Hankelio matricos matas 8 ( $n=7$ ; laiko eilutės bazinio fragmento ilgis būtų  $2n+1=15$ ). Laiko eilutės prognozavimo modelio funkcionalumas, kai Hankelio matricos matas parenkamas kitoks (tai pačiai laiko eilutei), bus aptartas kitame skyrelyje.



**3.2 pav.** Hankelio matricos determinanto absoliučiosios reikšmės priklausomybė nuo matricos rango.

### 3.3. Skaitiniai eksperimentai su PSO

Kaip jau minėta ankstesniame skyrelyje, optimizavimo algoritmo tikslas maksimizuoti (3.6) lygtimi aprašomą tinkamumo funkciją. Šioje lygtyje fiksuojamas parametras  $a_e = 1$ , taip suvienodinant balansą tarp Hankelio matricos determinanto dydžio ir lokalinių korekcijų dydžio. Akivaizdu, kad nauja korekcijų  $\{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}\}$  aibė yra generuojama kiekvieną kartą iš naujo, kai yra vykdomas PSO algoritmas. Taigi, darbe PSO algoritmas vykdomas 100 kartų ir kiekvieną kartą skaičiuojama prognozuojama reikšmė  $x_{15}$  (tiksliau proceso metu gaunamos 100 skirtingų  $x_{15}$  reikšmių). Tada skaičiuojama vidutinės kvadratinės paklaidos kvadratinė šaknis(RMSE) tarp tikrosios  $x_{15}$  reikšmės ir 100 prognozuotų  $x_{15}$  reikšmių (prognozuojamos reikšmės 3.3 paveiksle yra vaizduojamos kai  $t=15$ ). Dalelių spiečiaus skaitinių eksperimentų, kai imami skirtingi parametrų rinkiniai (jie paminėti 1.4.2 skyrelyje), rezultatai pateikiami 3.1 lentelės pirmose trijose eilutėse. Geriausias rezultatas gaunamas trečiam parametrų rinkiniui, kai  $b_\lambda = 0,5$ .



**3.3 pav.** Laiko eilutės bazinis fragmentas (stora ištisinė linija) ir 100 prognozuojamų  $x_{15}$  stebėjimo reikšmių.

**3.1 lentelė.** Skirtingais algoritmais gaunamų laiko eilutės prognozės paklaidų RMSE;  $b_\lambda$  yra parametras nustatantis korekcijų  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$  tolerancijos koridorių tinkamumo funkcijoje. Paryškinti skaičiai rodo kiekvienam algoritmui labiausiai tinkamas  $b_\lambda$  reikšmes.

$b \cdot$ Metodas	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	3	4
PSO pirmas rinkinys	0,1952	<b>0,1950</b>	0,1997	0,2071	0,2030	0,2029	0,1997	0,2022	0,2010	0,2038	0,2017
PSO antras rinkinys	0,1892	0,1912	0,1903	0,1931	0,1954	0,1999	<b>0,1890</b>	0,1957	0,1979	0,2029	0,1953
PSO, trečias rinkinys	0,1889	0,1911	<b>0,1883</b>	0,1929	0,1926	0,1906	0,1953	0,1970	0,1994	0,1948	0,1955
GA, $\beta=0$	0,1899	0,1916	0,1919	0,1939	<b>0,1878</b>	0,1883	0,1886	0,1921	0,1948	0,1904	0,2061
GA, $\beta=2$	0,1870	0,1885	0,1886	0,1883	<b>0,1861</b>	0,1886	0,1893	0,1882	0,1896	0,1888	0,1893

**3.2 lentelė.** Korekcijų svoriai  $\lambda_k$ ; periodinės laiko sekos elementai  $s_k$ ; testinės laiko eilutės elementai  $x_k$ ; lokalinės korekcijos  $\varepsilon_k$ ; nustatytos skeletinės sekos elementai  $x_k - \varepsilon_k$ ; bazinio fragmento ilgis  $2n+1=15$ .

$k$	$\lambda_k$	$s_k$	$x_k$	$\varepsilon_k$	$x_k - \varepsilon_k$
0	0,0003590	0,5	0,3515	0,1133	0,2381
1	0,0005919	0,7	0,8419	-0,1460	0,9879
2	0,0009759	0,1	0,0464	-0,0958	0,1422
3	0,0016089	0,9	0,9416	0,0188	0,9229
4	0,0026526	0,3	0,3353	-0,0976	0,4330
5	0,0043735	0,2	0,0663	-0,0860	0,1523
6	0,0072106	0,8	0,8884	0,0538	0,8345
7	0,0118883	0,5	0,4526	0,0841	0,3684
8	0,0196005	0,7	0,8421	0,0316	0,8105
9	0,0323158	0,1	0,1423	-0,0053	0,1476
10	0,0532798	0,9	0,9644	-0,0521	1,0165
11	0,0878435	0,3	0,3967	-0,0007	0,3975
12	0,1448294	0,2	0,2573	-0,0127	0,2700
13	0,2387833	0,8	0,7659	-0,0057	0,7716
14	0,3936871	0,5	0,5141	0,0017	0,5124

Pradžioje, atsižvelgiant į (3.6) lygties reikalavimus, skaičiuojamos  $2n+1=15$   $\lambda_k$  reikšmių (3.2 lentelės antras stulpelis). Sekančiame lentelės stulpelyje pateikiamos skeletinės laiko eilutės reikšmės, pati testinė (pridėjus triukšmą) laiko eilutė – ketvirtame 3.2 lentelės stulpelyje. Vieno PSO algoritmo įvykdymo sugeneruotos lokalinės korekcijos pateiktos šios lentelės penktame stulpelyje. Iš šių pataisytų duomenų  $x_k - \varepsilon_k = \tilde{x}_k$  sudaroma charakteristinė lygtis:

$$\det \tilde{H}_0^{(8)} = \det \begin{bmatrix} 0,2381 & 0,9879 & 0,1422 & 0,9229 & 0,4330 & 0,1523 & 0,8345 & 0,3684 \\ 0,9879 & 0,1422 & 0,9229 & 0,4330 & 0,1523 & 0,8345 & 0,3684 & 0,8105 \\ 0,1422 & 0,9229 & 0,4330 & 0,1523 & 0,8345 & 0,3684 & 0,8105 & 0,1476 \\ 0,9229 & 0,4330 & 0,1523 & 0,8345 & 0,3684 & 0,8105 & 0,1476 & 1,0165 \\ 0,4330 & 0,1523 & 0,8345 & 0,3684 & 0,8105 & 0,1476 & 1,0165 & 0,3975 \\ 0,1523 & 0,8345 & 0,3684 & 0,8105 & 0,1476 & 1,0165 & 0,3975 & 0,2700 \\ 0,8345 & 0,3684 & 0,8105 & 0,1476 & 1,0165 & 0,3975 & 0,2700 & 0,7716 \\ 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 \end{bmatrix} = 0. \quad (3.8)$$

**3.3 lentelė.** Charakteristinės lygties šaknys  $\rho_i$  ir koeficientai  $\mu_{i0}$ .

$i$	$\rho_i$	$\mu_{i0}$
1	-0,8856 + 0,3868i	0,2106 + 0,1458i
2	-0,8856 - 0,3868i	0,2106 - 0,1458i
3	-0,2407 + 1,0041i	0,0236 - 0,0738i
4	-0,2407 - 1,0041i	0,0236 + 0,0738i
5	1,0037	0,5216
6	0,5843 + 0,6880i	-0,0227 - 0,0656i
7	0,5843 - 0,6880i	-0,0227 + 0,0656i



Charakteristinės lygties (3.8) visos septynios šaknys yra pateikiamos 3.3 lentelės antrame stulpelyje. Galima pastebėti, kad visos šaknys yra skirtingos (visi rekurentiniai indeksai  $n_i = 1; i = 1, 2, \dots, 7$ ) ir tuomet (1.41) lygtis įgauna pavidalą:

$$x_n = \sum_{k=1}^7 \mu_{k0} \rho_k^n; n = 0, 1, 2, \dots \quad (3.9)$$

Algebrinių lygčių sistema parametrų  $\mu_{k0}$  radimui,  $k = 1, 2, \dots, 7$  sudaroma tokiu būdu:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 & \rho_6 & \rho_7 \\ \rho_1^2 & \rho_2^2 & \rho_3^2 & \rho_4^2 & \rho_5^2 & \rho_6^2 & \rho_7^2 \\ \rho_1^3 & \rho_2^3 & \rho_3^3 & \rho_4^3 & \rho_5^3 & \rho_6^3 & \rho_7^3 \\ \rho_1^4 & \rho_2^4 & \rho_3^4 & \rho_4^4 & \rho_5^4 & \rho_6^4 & \rho_7^4 \\ \rho_1^5 & \rho_2^5 & \rho_3^5 & \rho_4^5 & \rho_5^5 & \rho_6^5 & \rho_7^5 \\ \rho_1^6 & \rho_2^6 & \rho_3^6 & \rho_4^6 & \rho_5^6 & \rho_6^6 & \rho_7^6 \end{bmatrix} \cdot \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \mu_{30} \\ \mu_{40} \\ \mu_{50} \\ \mu_{60} \\ \mu_{70} \end{bmatrix} = \begin{bmatrix} 0,2381 \\ 0,9879 \\ 0,1422 \\ 0,9229 \\ 0,4330 \\ 0,1523 \\ 0,8345 \end{bmatrix}. \quad (3.10)$$

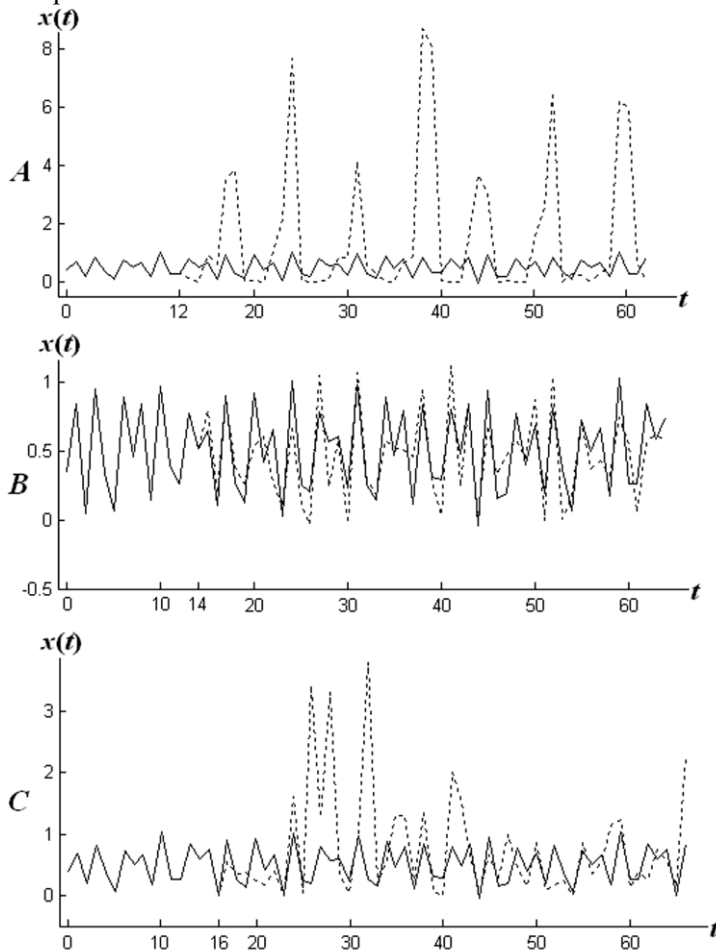
Nagrinėjamo pavyzdžio parametrai  $\mu_{k0}$ , kai  $k = 1, 2, \dots, 7$  yra demonstruojami 3.3 lentelės trečiame stulpelyje. Galiausiai, prognozuojama reikšmė

$$x_{15} = \sum_{k=1}^7 \mu_{k0} \rho_k^{15} = 0,6510 \text{ (kai tikroji testinės laiko eilutės reikšmė } x_{15} = 0,6437).$$

Prognozavimo horizontas yra viena reikšmė į ateitį, kai imamas stebėjimų langas yra laiko eilutės bazinio fragmento aibė. Žingsnis po žingsnio keliaujama su fiksuoto ilgio langu į ateitį. Fiksuotam stebėjimų langui prognozuojama kiekviena sekanti viena reikšmė 100 kartų. Kai jau yra gauta 100 to paties taško suprognozuotų reikšmių, skaičiuojamas jų algebrinis vidurkis ir gautoji reikšmė yra laikoma aktualia prognozuota reikšme. Testinės laiko eilutės prognozės rezultatai yra pateikti 3.4 paveikslo B dalyje.

Iki šiol skaitiniai eksperimentai buvo atlikti, kai  $n = 7$  (laiko eilutės bazinio fragmento ilgis yra 15). Bazinio fragmento ilgio keitimas yra svarbus klausimas. Toliau bus atlikti skaitiniai eksperimentai, kai  $n = 6$  ir  $n = 8$  (demonstracijos pateiktos 3.4 paveikslo A ir C dalyse). Galima pastebėti, kad sudarytas prognozės modelis prastai dirba kai fragmento ilgis trumpesnis nei skeletinės laiko eilutės mažiausias ilgis, kuris reikalingas, kad Hankelio matricos determinantą būtų lygus 0 (3.4 paveikslo A dalis). Iš tikrųjų, neįmanoma ką nors prognozuoti, jei proceso modelis nėra adekvatus. Iš kitos pusės, jei laiko eilutės bazinio fragmento ilgis yra didesnis nei skeletinės eilutės periodas, prognozavimo rezultatai nėra tokie prasti (pateikta 3.4 paveikslo C dalyje), bet vis tiek žymiai prastesni, nei parinkus optimalaus ilgio bazinį fragmentą. Galima pažymėti, kad algebrinė seka turi būti tik periodinė seka, bei aritmetinė arba geometrinė progresija. Algebrinės progresijos padengia daug platesnę nei minėtos laiko eilučių sritį (kitas klausimas, koks yra tokios laiko eilutės rangas?). Bendru atveju, laiko eilutė yra algebrinė progresija, jei jos vystymasis laike reglamentuojamas tam tikros rūšies deterministiniu modeliu.

Taigi iš esmės, tokia prognozavimo koncepcija, pagrįsta algebrinių sekų atpažinimu, turėtų būti plačiai taikoma.



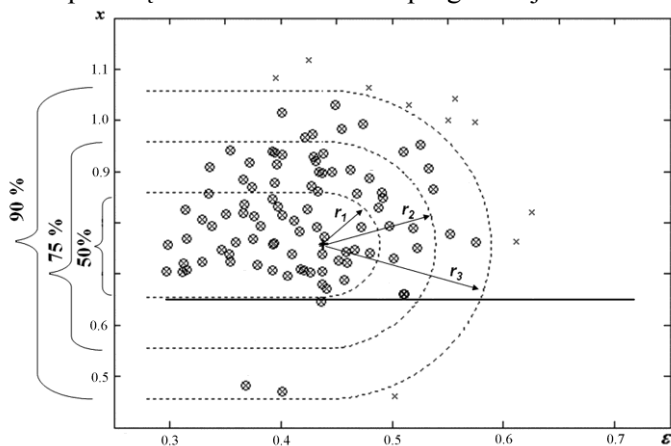
**3.4 pav.** Testinės laiko eilutės prognozė (punktyrinė linija), kai algebrinės sekos bazinio fragmento ilgis: A – per trumpas; B – nustatytas teisingai; C – per ilgas.

### 3.4. Prognozavimo gerinimas ištrinant labiausiai nukrypusias reikšmes

Anksčiau buvo minėta, kad prognozuojama reikšmė imamas aritmetinis vidurkis iš 100 prognozių bandymų. Natūralu, kad kai kurios šių bandymų prognozės gali būti labai netikslios, pavyzdžiui, dėl nesėkmingo pradinės populiacijos pasiskirstymo. Vienas iš būdų išmesti tas prognozuojamas išskirtis, tai interpoliuoti tas 100 prognozavimo bandymų reikšmes pagal Gauso skirstinio funkciją ir tuomet pasinaudoti  $3\sigma$  arba  $2\sigma$  taisykle. Deja, atlikti eksperimentai su testine eilute parodė, kad hipotezės apie prognozuojamų reikšmių pasiskirstymą pagal Gauso skirstinį negalima statistiškai įrodyti su pakankamu patikimumu.

Be to, reikėtų atsižvelgti į faktą, kad pristatytas metodas bando rasti skeletinę algebrinę seką koreguodamas laiko eilutės bazinio fragmento reikšmes. Jei tie pataisymai yra labai dideli, prognozavimas gali būti tiesiog blogas dėl to, kad nustatyta skeletinė seka gali būti labai nutolusi nuo pirminės laiko eilutės. Todėl, šiame darbe yra išskiriami tie labiausiai nutolę bandymai, kurių prognozuota

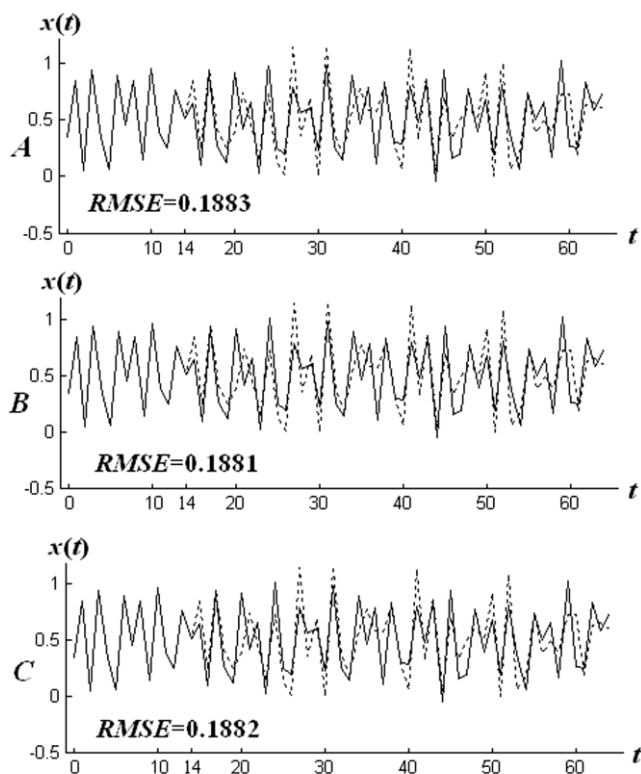
skaitinė reikšmė stipriai nutolusi nuo aritmetinio visų bandymų prognozuojamų reikšmių vidurkio, arba kurių bazinio fragmento korekcijų absoliučią reikšmių suma yra didelė palyginus su kitais bandymais. Akivaizdu, kad reikalingas specialus algoritmas, kuris atpažintų labiausiai nutolusias prognozuojamas reikšmes.



**3.5 pav.** Scheminė diagrama iliustruojanti labiausiai nutolusių prognozės bandymų išmetimą;  $\varepsilon$  atitinka korekcijų  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$  absoliučią reikšmių sumą; x – atitinka ieškomas  $x_{15}$  reikšmės prognozes; storas „+“ ženklas vaizduoja 100-to prognozės reikšmių debesio masės centrą; stora horizontali linija žymi tikrąją  $x_{15}$  reikšmę; spinduliai  $r_1, r_2$  ir  $r_3$  rodo atitinkamai 50%, 25% ir 10% reikšmių išmetimą; ploni kryžiai rodo tuos 10% išmetamų reikšmių; skaičiavimo rezultatai vaizduoja 3 paveikslą duomenis.

Galima pasiūlyti gana paprastą algoritmą, kuris yra iliustruojamas 3.5 paveiksle. Čia demonstruojami  $x_{15}$  reikšmės 100 prognozavimo bandymų. Horizontalioje  $\varepsilon$ -ašyje žymimos korekcijų  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$  absoliučią reikšmių suma, vertikalioje  $x$ -ašyje atidedamos faktinės reikšmės  $x_{15}$  prognozuojamos reikšmės. Pradžioje skaičiuojami lokalinių korekcijų modulių sumos bei prognozuojamų reikšmių vidurkis ir dispersija visiems 100 prognozės bandymų. Tuomet keičiama vertikaliosios ašies skalė, kad būtų galima suvienodinti dispersijas matematinėse koordinatėse 3.5 paveikslo abiejose ašyse (fizinės koordinatės nesikeičia). Kitaip tariant, 3.5 paveiksle reikšmių debesies forma išplečiama daugiau ar mažiau į skritulio tipo objektą. Toliau skaičiuojamas debesies masės centras (3.5 paveiksle vaizduojamas pliuso ženklu). Tada nuo masės centro brėžiamas toks skritulys, kad santykis reikšmių pakliūnančių į plotą, kurį riboja pusė skritulio ir horizontalios linijos, ir visų reikšmių skaičiaus yra lygus tam tikram pasirinktam trupmeniniam skaičiui (3.5 paveiksle spindulys  $r_1$  atitinka 0,5,  $r_2$  – 0,75,  $r_3$  – 0,9).

Skaitiniai eksperimentai su testine laiko eilute vykdant PSO algoritmą ( $n = 7$ , su trečia parametru aibe,  $b_\lambda = 0,5$ ) rodo, kad išmetant ribotą skaičių labiausiai nutolusių prognozavimo reikšmių galima pagerinti prognozavimo tikslumą (pavyzdys demonstruojamas 3.6 paveiksle).



**3.6 pav.** Testinės laiko eilutės prognozė, kai korekcijų  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$  radimui panaudotas PSO ( $n = 7$ ; 3 parametrų aibė,  $b_\lambda = 0,5$ ) algoritmas: A – neišmetant nei vieno prognozės bandymo; B – išmetant 10% prognozės bandymų; C – išmetant 25% prognozės bandymų.

### 3.5. Evoliucinių algoritmų taikymo palyginimas su PSO

Natūralu klausti, ar ieškant lokaliųjų korekcijų  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$  aibės galima pasinaudoti kitu optimizavimo metodu. Tam bus panaudoti evoliuciniai algoritmai (EA) ir palygintas jų funkcionalumas su PSO gautais rezultatais. Tačiau nors EA taikymas galbūt nėra geriausias būdas spręsti kiekvieną uždavinį, bet kai kurie EA privalumai patvirtina, kad algoritmas susidoroja su dideliu kiekiu kintamųjų, dirba su eksperimentiniais duomenimis ir analizinėmis funkcijomis, optimizuoja kintamuosius su ypač sudėtingais paviršiais [135].

Kiekviena chromosoma mūsų skaičiavimuose susideda iš aibės lokaliųjų korekcijų ( $2n+1$  genai pateikiami kaip realieji skaičiai). Pradinė populiacija susideda iš  $m$  chromosomų, kuriose genai generuojami atsitiktinai (mūsų eksperimentuose naudojamas  $m=50$ ). Kiekvieno geno reikšmė priklauso intervalui  $[-0,2; 0,2]$ . Tinkamumo funkcija, siejama su kiekviena chromosoma, aprašyta (3.6) lygtyje. Iš pradinės populiacijos į porų populiaciją yra atrenkamas lyginis chromosomų skaičius. Chromosomų atrinkimui į naują populiaciją yra naudojamas turnyrinis atsitiktinės atrankos metodas [151].

Porų populiacijoje visoms poroms vykdomas dviejų chromosomų kryžminimas. Šiame darbe naudojama vieno taško modifikuotas  $\beta$ -kryžminimo metodas, kurį pristatėme 2.3.1 skyrelyje. Beliko suderinti EA parametrus.

### 3.6. Evoliucinių algoritmų parametrų derinimas

Kaip ir 2.3.1 skyrelyje evoliuciniuose algoritmuose turi būti parinkti tokie parametrai: kryžminimo koeficientas  $\kappa$ , mutacijos koeficientas  $\mu$ , panašumo parametras  $\beta$  ir generacijų skaičius. Darbe pasinaudota klasikinio evoliucinio algoritmo konstravimo rekomendacijomis [151]. Kryžminimo koeficientas  $\kappa$  bus atrenkamas iš intervalo  $[0,6; 0,8]$ , o mutacijos parametras  $\mu$  iš intervalo  $[0; 0,15]$ . Kaip jau minėta anksčiau, kai panašumo parametras įgyja reikšmę  $\beta = \infty$  (arba  $\beta = 0$ ), evoliuciniai algoritmai atitinka klasikinį atvejį. Vietoje to čia tiriamas intervalas  $1 \leq \beta \leq 5$ . Kadangi nėra apibrėžta jokių metodų, kiek evoliuciniuose algoritmuose turi vykti generacijų, eksperimentuose buvo naudota 40 generacijų.

Įvykdžius evoliucinius algoritmus vieną kartą, rezultate gaunamas lokaliųjų korekcijų rinkinys  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$ . Rezultatas priklauso nuo sugeneruotos pradinės chromosomų populiacijos (tarp kitų atsitiktinių faktorių). Panašiai kaip PSO algoritme, EA vykdomas po 100 kartų, tai yra 100 kartų prognozuojame vieną reikšmę  $x_{15}$ , o tada skaičiuojama prognozės paklaidų RMSE. Prognozavimo tikslumas testuojamas keičiant parametrų  $\kappa$ ,  $\mu$  ir  $\beta$  reikšmes (3.4 lentelė). Geriausias rezultatas gaunamas kai  $\kappa = 0,7$ ;  $\mu = 0,15$  ir  $\beta = 2$ . Šis parametrų rinkinys bus naudojamas tolimesniems eksperimentiniams tyrimams.

**3.4 lentelė.** Dirbtinės laiko eilutės prognozavimo paklaidų RMSE gaunama, kai korekcijų  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$  paieškai naudojami EA.

$\kappa$	$\mu$	$\beta = 0$	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$	$\beta = 5$
0,6	0,001	0,1992	0,1974	0,1973	0,1962	0,1967	0,1964
0,6	0,005	0,1960	0,1956	0,1955	0,1955	0,1946	0,1961
0,6	0,01	0,1941	0,1943	0,1939	0,1932	0,1935	0,1937
0,6	0,05	0,1910	0,1893	0,1897	0,1904	0,1912	0,1914
0,6	0,1	0,1891	0,1894	0,1888	0,1889	0,1902	0,1900
0,6	0,15	0,1903	0,1889	0,1891	0,1893	0,1903	0,1893
0,7	0,001	0,1952	0,1965	0,1970	0,1951	0,1959	0,1966
0,7	0,005	0,1963	0,1950	0,1948	0,1949	0,1951	0,1960
0,7	0,01	0,1953	0,1938	0,1937	0,1941	0,1949	0,1956
0,7	0,05	0,1927	0,1891	<b>0,1882</b>	0,1904	0,1906	0,1908
0,7	0,1	0,1904	0,1894	0,1896	0,1894	0,1911	0,1901
0,7	0,15	0,1899	0,1890	0,1888	0,1901	0,1886	0,1901
0,8	0,001	0,2000	0,1958	0,1967	0,1957	0,1965	0,1955
0,8	0,005	0,1961	0,1944	0,1957	0,1952	0,1953	0,1960
0,8	0,01	0,1935	0,1942	0,1940	0,1931	0,1948	0,1949
0,8	0,05	0,1935	0,1893	0,1910	0,1902	0,1907	0,1904
0,8	0,1	0,1889	0,1890	0,1906	0,1900	0,1896	0,1897
0,8	0,15	0,1894	0,1894	0,1893	0,1900	0,1891	0,1906

Kaip jau minėta anksčiau, PSO funkcionalumas, pagrįstas prognozavimo metodu, gali būti paaiškintas parinkta tinkama parametro  $b_\lambda$  reikšme (3.6) lygtyje,

kuri aprašo tolerancijos koridorių lokaliųjų korekcijų aibei išmetant prognozavimo bandymų išskirtis. Panašūs patobulinimai taikomi ir EA grindžiamam prognozavimo modeliui, kuriuos toliau ir paaiškinsime. Skaitiniai eksperimentai su testine laiko eilute parodė, kad geriausi rezultatai gaunami, kai  $b_\lambda = 1$  (lentelė 3.1) ir kai išmetama 10% labiausiai išsiskiriančių prognozuojamų reikšmių (žr. 3.7 paveikslė).

Iš esmės algebrinės sekos atpažinimo procesas gali būti paaiškintas statistiniais metodais (pavyzdys 3.5 paveikslė). Tačiau svarbu pažymėti, kad čia iš signalo nefiltruojamas triukšmas. Triukšmo nuėmimas šiuo atveju yra visiškai skirtingas. Bendra šio uždavinio idėja yra rasti artimiausią algebrinę seką kuo mažiau deformuojant originalų signalą (tam naudojama (3.6) lygtis). Testinei laiko eilutei (kurioje algebrinė seka jau žinoma iš anksto) yra įmanomos kai kurios statistinės analizės rūšys. Tačiau triukšmo naikinimo statistinis įvertinimas tampa sudėtingu uždaviniu, kai dirbama su realiomis laiko eilutėmis, kadangi čia nėra žinoma kaip atrodo algebrinės sekos bazinis fragmentas, kai triukšmas dar nėra nuimtas. Be to, esant toms pačioms pradinėms sąlygoms galima gauti vieną rinkinį vienokios algebrinės sekos, o kitą rinkinį, jau visai kitokios sekos (pavaizduota 3.3 paveikslė). Šiame darbe vaizduojamos lokalsios korekcijos ir išmetami labiausiai išsiskiriantys vienos reikšmės prognozės bandymai anksčiau aprašytu būdu. Ši procedūra gali būti laikoma kaip apibendrinta statistinė analizė, nors ji nereikalauja nustatyti tam tikros tankio funkcijos, aprašančios prognozuojamų reikšmių pasiskirstymą. Procesas paliekamas evoliuciniam algoritmui ir derinami šio algoritmo parametrai, atsižvelgiant į tinkamumo funkciją aprašomą (3.6) lygtimi (taip pat tikrinamas prognozavimo paklaidų RMSE) ir stebint algoritmo našumą.

Galiausiai, siūlomo modelio bendro dizaino procedūra gali būti apibendrinta tokiu struktūriniu algoritmu:

1 žingsnis. Nustatyti skeletinės algebrinės sekos bazinio fragmento ilgį (parametrą  $n$ ).

2 žingsnis. Tikrinti, ar laiko eilutė yra ilgesnė už bazinį fragmentą (kitu atveju metodas negali būti taikomas).

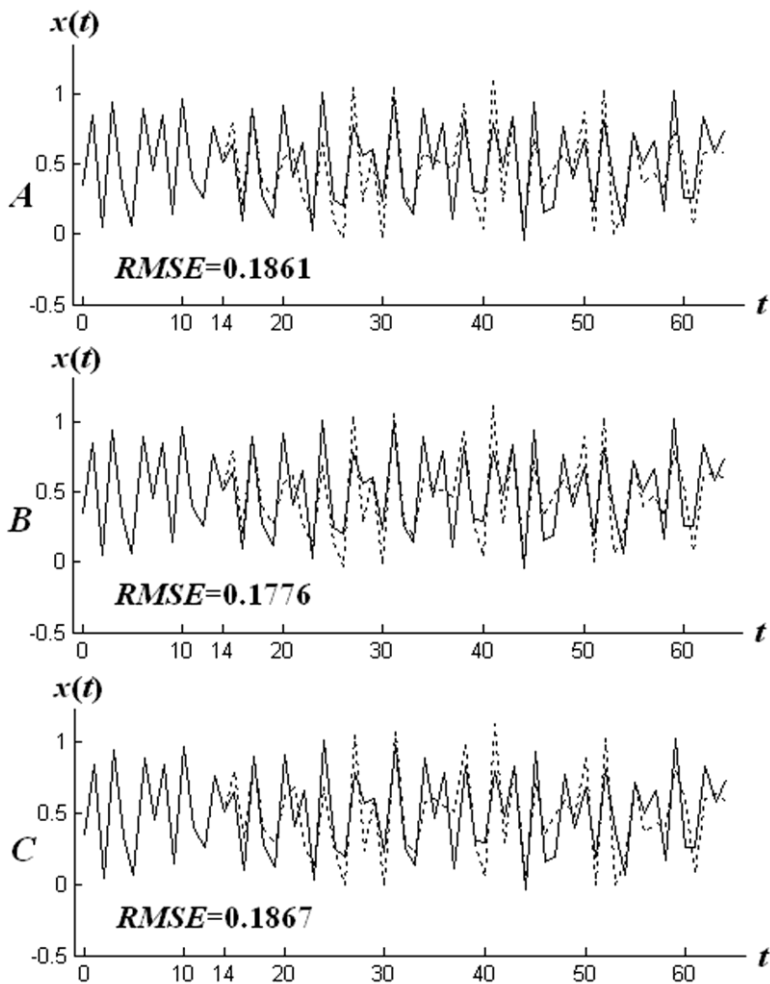
3 žingsnis. Nustatyti baudos dalį tinkamumo funkcijoje tarp Hankelio matricos determinanto reikšmės ir svorinių lokaliųjų korekcijų sumos (čia pasirenkama  $a_e = 1$ ).

4 žingsnis. Apibrėžti tolerancijos koridorių lokalsioms korekcijoms (pasirenkama  $b_\lambda = 1$ ).

5 žingsnis. Prognozuoti sekos sekantį elementą  $x_{2n+1}$ :

- algoritmą kartoti 100 kartų
  - a) Surasti vienintelę lokaliųjų korekcijų seką  $\{\varepsilon_0, \varepsilon_2, \dots, \varepsilon_{2n}\}$  (parenkami EA tokie parametrai:  $\kappa = 0,7$ ;  $\mu = 0,15$  ir  $\beta = 2$ ).
  - b) Sudaryti charakteristinę lygtį ir rasti jos šaknis.
  - c) Pagal (1.41) lygtį apskaičiuoti  $x_{2n+1}$  prognozuojamą reikšmę.
- Išmesti labiausiai nutolusias prognozuojamas reikšmes.
- Skaičiuoti likusių prognozės reikšmių  $x_{2n+1}$  aritmetinį vidurkį.

6 žingsnis. Per vieną žingsnį į priekį paslinkti stebėjimų langą ir grįžti į 5 žingsnį.



**3.7 pav.** Testinės laiko eilutės prognozė, kai korekcijų  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$  radimui panaudotas EA ( $n = 7$ ;  $\kappa = 0.7$ ,  $\beta = 2$ ,  $b = 1$ ) algoritmas: A – neišmetant nei vieno prognozės bandymo; B – išmetant 10% prognozės bandymų; C – išmetant 25% prognozės bandymų.

### 3.7. Skaitiniai eksperimentai su trumpomis laiko eilutėmis

#### 3.7.1. Testinė laiko eilutė su pridėtu tolygiuoju triukšmu

Tęsiami eksperimentiniai skaičiavimai su testine laiko eilute. Sukurto modelio prognozavimo funkcionalumas palyginamas su kitų prognozės metodų rezultatais. Eksperimentai buvo atlikti su Box-Jenkins laiko eilučių analizės procedūra ARIMA(4,1,3) (Autoregresinis integruotas slenkantis vidurkis). Bandymų keliu nustatoma, kad ARIMA modelis geriausiai dirba su testine laiko eilute, kai imama architektūra 4-1-3 (4 reiškia autoregresijos narių skaičių, 1 – nesezoninių skirtumų skaičių, 3 – slenkančio vidurkio prognozavimo lygties atsilikusių prognozuojamų paklaidų skaičius). Modelio adekvatumo testas pateiktas preide lentelėje Nr. 4. ARIMA modelio prognozavimo paklaidų RMSE lygi 0,1307 (3.5 lentelė, 3.8 paveikslas A dalis) ir maždaug 1,3 karto mažesnės nei mūsų modelio (palyginti su 7B

paveikslu). ARIMA modeliu laiko eilutę pradedama prognozuoti nuo 15 reikšmės, kad būtų galima adekvačiai palyginti su mūsų modelio rezultatais.

Akivaizdu, kad ARIMA(4,1,3) pranoksta sukurta modelį. Vis dėl to, ARIMA prognozavimas (pateiktas 3.8 paveikslo A dalyje) daugiau išreiškė slenkančio vidurkio modelio pobūdį. Sukurtas algebrinės sekos modelis, priešingai, grindžiamas ne kurio nors tipo statistiniais algoritmais, o kiekviename laiko žingsnyje vykdomu lokaliu individualiu skeletinės algebrinės progresijos atpažinimu. Turime pripažinti, kad demonstruojamo modelio prognozės rezultatai kai kuriuose taškuose yra gana nutolę nuo tikrosios laiko eilutės reikšmių (tai iš tikrųjų gadina vidutinį prognozavimo kokybės įvertinimą). Tačiau lokalieji šiuo modeliu prognozuotos laiko eilutės svyravimai duoda geresnį tikrosios laiko eilutės judesio pobūdžio reprezentavimą. Pavyzdžiui, šis modelis veiks geriau nei ARIMA, jei bus reikalinga nustatyti vienos dienos į priekį lokalų minimumą arba lokalų maksimumą [157].

Iš kitos pusės, galima pažymėti, kad modelio pobūdis grindžiamas algebrinės sekos atpažinimu ir jo ekstrapoliacija ateityje apima daug sudėtingesnę analizę, nei paprastas lokalaus maksimumo ar lokalaus minimumo radimas. Sukurtas modelis negali būti lyginamas su primityviais metodais, grindžiamais paprastu laiko eilutės praeities didžiausios ir mažiausios reikšmės ieškojimu (po to kai yra sumažinamas triukšmas). Toks požiūris, kuris grindžiamas praeities minimaliomis ir maksimaliomis reikšmėmis, gerai veikia su periodinėmis sekomis, kurių periodas yra gana trumpas. Akivaizdu, kad periodinės sekos yra algebrinės sekos. Tačiau algebrinių sekų aibė yra žymiai platesnė, o periodinės sekos yra tiesiog trivialis algebrinės sekos pavyzdys.

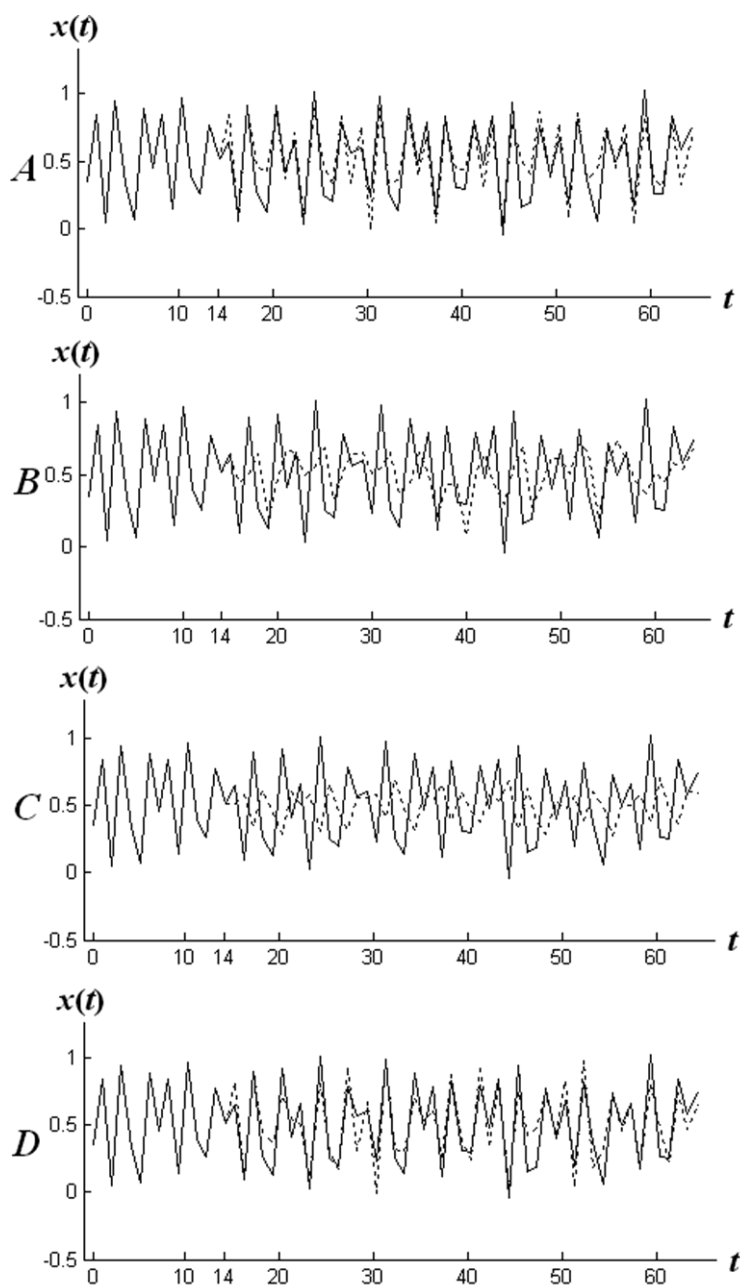
Gali būti įmanoma suderinti ARIMA(4,1,3) slenkančio vidurkio funkciją ir mūsų modelio lokalų kintamumą skaičiuojant šių modelių prognozuojamų reikšmių aritmetinį vidurkį kiekvienu atitinkamu laiko momentu. Tokio prognozavimo rezultatas vaizduojamas 3.8 paveikslo D dalyje. Tokio junginio prognozės paklaidų RMSE yra geresnis nei ARIMA(4,1,3) modelio (3.5 lentelė). Tačiau kas svarbiausia, prognozuojamos laiko eilutės sklaida yra mažesnė, artimesnė realiai laiko eilutei.

Sukonstruoto modelio funkcionalumas taip pat palyginamas su slenkančio vidurkio (MA) ir eksponentinio glodinimo (SES) prognozavimo rezultatais. Iš duomenų šalinant būdingus atsitiktinius pokyčius dažnai naudojamas paprastas slenkančio vidurkio glodinimo industrinis metodas (MA), kuris yra aprašytas 1.1 skyrelyje (1.4) lygtimi. Eksperimentuojant buvo nustatyta, kad MA modelio prognozavimo paklaidų RMSE yra mažiausia, kai prognozei imamos  $k = 2$  praeities reikšmės. Pirma prognozuojama reikšmė  $x_2$  suskaičiuojama iš elementų  $(x_0, x_1)$ . Tada pakopomis perkeliama stebėjimų langas ir skaičiuojama prognozuojamų reikšmių seka pagal (1.4) lygtį. MA modeliu prognozavimo paklaidų RMSE lygi 0,2602 (3.5 lentelė). Rezultatai pateikti 3.8 paveikslo B dalyje. Akivaizdu, kad šioje situacijoje paprasčiausias glodinimo negali veikti geriau, nei darbe sukurta modelis.

**3.5 lentelė.** Testinės laiko eilutės prognozavimo paklaidų RMSE

Prognozės modelis	Algebrinių sekų	ARIMA(4,1,3)	MA	SES	Jungtinis ARIMA(4,1,3) ir algebrinių sekų
<b>RMSE</b>	0,1776	0,1307	0,2602	0,3482	0,1290





**3.8 pav.** Testinės laiko eilutės prognozė: A – ARIMA(4, 1, 3) modeliu; B – slenkančio vidurkio modeliu; C – SES modeliu; D – kombinuotu modeliu.

Eksponentinis glodinimo modelis yra paprastas ir pragmatiškas požiūris į prognozavimą, kuriuo prognozavimas konstruojamas iš eksponentiniu svoriu paveikto praeities duomenų vidurkio [11, 12]. Testinėje laiko eilutėje nėra aiškiai išreikštas trendas arba sezonškumas, todėl čia naudojamas viengubo eksponentinio glodinimo (SES) modelis, kuriame pagal nutylėjimą pasirenkama reikšmė  $\alpha = 0,5$ .

Šiuo modeliu gaunama prognozavimo paklaidų RMSE yra 0,3482(3.5 lentelė); rezultatai vaizduojami 3.8 paveikslo C dalyje.

Šiame skyrelyje buvo pateiktas prognozavimo pavyzdys, kai prie sudarytosios sekos buvo pridėtas triukšmas, tolygiai pasiskirstęs intervale  $[-0,15; 0,15]$ . 3.6 lentelėje yra pateiktos prognozės paklaidos metrikos gautos pristatytų algebrinių sekų bei ARIMA modeliais. Galima pastebėti, kad sukonstruotas algebrinių sekų laiko eilučių modelis gerai prognozuoja kai pridėdamas triukšmas yra pasiskirstęs gana siaurame intervale. Modelių ARIMA(3,0,1) ir ARIMA(1,0,3) adekvatumo testai yra pateikiami priede lentelėse Nr. 5 ir Nr. 6.

**3.6. lentelė.** Paklaidų RMSE palyginimas, kai prie algebrinės sekos yra pridėdamas triukšmas tolygiai pasiskirstęs nurodytuose intervaluose.

Pridėtas tolygusis triukšmas	Algebrinių sekų RMSE	ARIMA : RMSE
Kai $[-0,15; 0,15]$	0,1776	(4,1,3) : 0,1307
Kai $[-0,5; 0,5]$	3,3421	(3,0,1) : 2,851
Kai $[-1; 1]$	3,2173	(1,0,3) : 4,503

### 3.7.2. Testinė laiko eilutė su pridėtu Gauso triukšmu

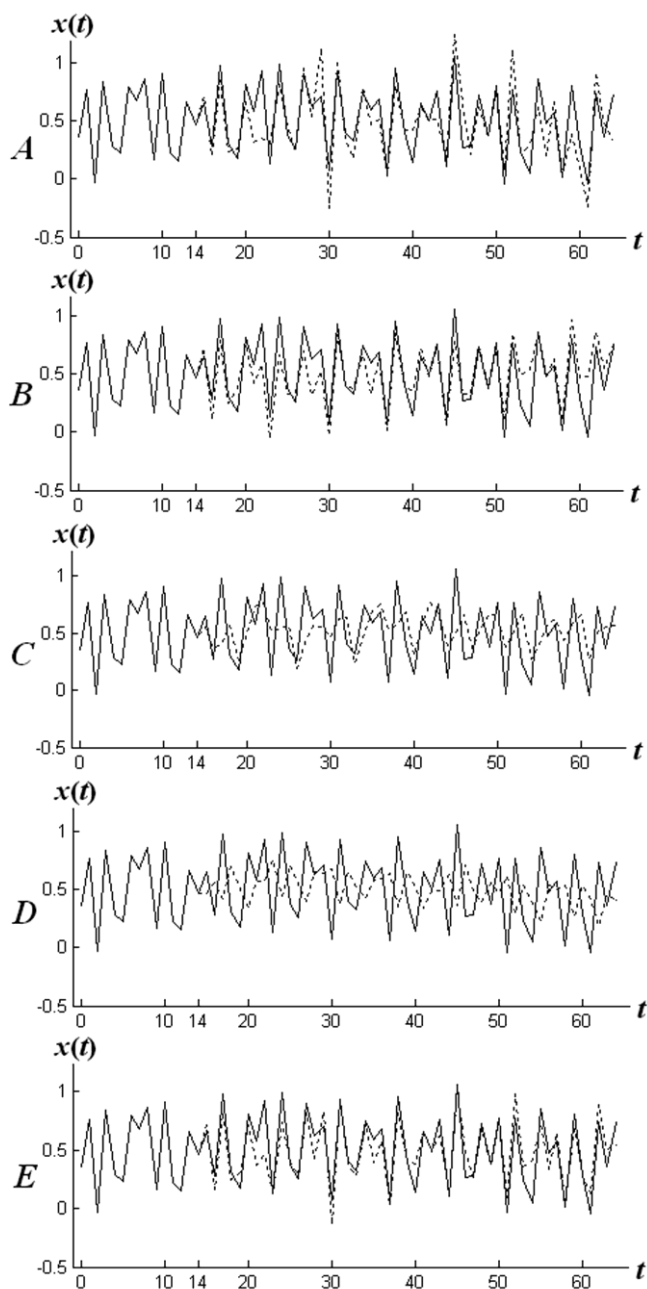
Bovo atlikti skaitiniai eksperimentai su testine laiko eilute, tačiau periodinei sekai vietoje tolygaus triukšmo dabar pridėdamas Gauso triukšmas (su vidurkiu 0 ir dispersija 0,1). Naujos testinės laiko eilutės nustatomas H-rangas yra 8; bazinio fragmento ilgis 15 ( $n=7$ ,  $2n+1=15$ ). Skeletinei algebrinei sekai nustatyti yra naudojamas evoliucinis algoritmas su tokiais pačiais parametrais, kurie buvo nustatyti praeitame eksperimente.

Sukonstruoto modelio funkcionalumas palyginamas su ARIMA, slenkančio vidurkio, eksponentinio glodinimo ir jungtiniu modeliais (3.9 paveikslas). Bandymų keliu randama geriausia ARIMA architektūra (5, 0, 3) (autoregresijos dalies eilė 5, integruotosios dalies – 0, slenkančio vidurkio – 3). Modelio adekvatumo testas pateiktas preide lentelėje Nr. 7. Slenkančio vidurkio modelis geriausiai prognozuoja, kai parametras  $k=2$ . Skirtingų metodų RMSE pateikti 3.7 lentelėje. Šiame pavyzdyje vėl, ARIMA(5,0,3) modelis veikia geriau nei siūlomas, bet duomenų kintamumas geriau matomas būtent algebrinių sekų modelio atveju. Galiausiai galima pažymėti, kad būtų pageidaujama naudoti jungtinį sukurto ir ARIMA modelį (iš 5 modelių pateiktų 3.9 paveiksle), jei reikalinga palaikyti vienos dienos į priekį prognozavimą.

**3.7 lentelė.** Laiko eilutės su Gauso triukšmu prognozės paklaidų RMSE

Prognozės modelis	Algebrinių sekų	ARIMA(5,0,3)	MA	SES	Jungtinis ARIMA(4,0,2) ir algebrinių sekų
<b>RMSE</b>	0,1771	0,1544	0,2523	0,3475	0,1288

Toliau taip pat kaip ir praeitame skyrelyje, atliekamas tyrimas dėl Gauso triukšmo įtakos į prognozės rezultatus. 3.8 lentelėje yra pateiktos prognozės paklaidų metrikos, gautos naudojant pristatytą algebrinių sekų bei ARIMA modelius. Modelių ARIMA(2,0,3) ir ARIMA(1,1,2) adekvatumo testai yra pateikiami priede lentelėse Nr. 8 ir Nr. 9.



**3.9 pav.** Laiko eilutės su Gauso triukšmu prognozė: A – pristatytu modeliui; B – ARIMA(5, 0, 3) modeliui; C – slenkančio vidurkio modeliui; D – SES modeliui; E – kombinuotu modeliui.

**3.8. lentelė.** Paklaidų RMSE palyginimas, kai prie algebrinės sekos yra pridedamas skirtingas Gauso triukšmas.

Pridėtas Gauso triukšmas	Algebrinių sekų RMSE	ARIMA : RMSE
Kai dispersija 0,1	0,1771	(5,0,3) : 0,1544
Kai dispersija 0,5	4,9854	(2,0,3) : 1,780
Kai dispersija 1	5,3844	(1,1,2) : 4,497

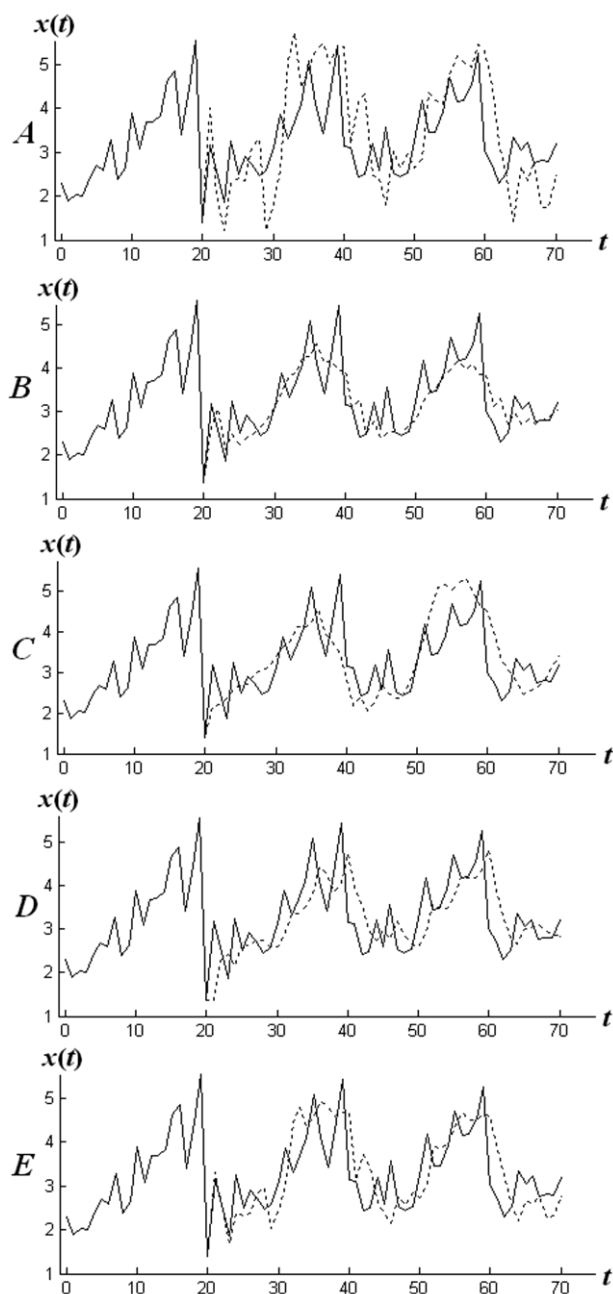
### 3.7.3. Andersono sausumos termometro temperatūros laiko eilutė

Pradžioje reikia rasti bazinio fragmento ilgį Andersono sausumos termometro temperatūros laiko eilutei [142]. Hankelio matricos determinanto pirmasis minimumas gaunamas, kai matricos formatas yra  $11 \times 11$  ( $n=10$ , laiko eilutės bazinio fragmento ilgis yra  $2n+1=21$ ). Bazinio fragmento ilgis fiksuojamas visai laiko eilutei. Vykdomas vieno žingsnio į priekį prognozavimas pasinaudojant evoliuciniu algoritmu, kurio parametrai paliekami tokie patys kaip ir ankstesniuose dviejuose eksperimentuose (žiūrėti 3.10 paveikslo A dalį).

Eksperimentiniu būdu buvo nustatyta, kad ARIMA modelio prognozė geriausia, kai imama architektūra (3,0,3) (modelio adekvatumo tyrimas pateiktas preid lentelėje Nr. 10), o slenkančio vidurkio, kai pasirenkamas parametras  $k=5$ . Prognozavimo rezultatai vaizduojami 3.10 paveiksle, o visų modelių prognozavimo paklaidų RMSE pateikiami 3.9 lentelėje. Šiuo atveju ir vėl ARIMA(3,0,3) prognozuoja tiksliau nei sukonstruotas modelis. Tačiau ARIMA(3,0,3) modelis išreiškė slenkančio vidurkio modelio pobūdį ir akivaizdžiai suvidurkino visas Andersono laiko eilutės smailias viršūnes. Eksponentinio glodinimo modelis taip pat duoda žymiai geresnius rezultatus, bet galima pastebėti tam tikrą fazės vėlavimą tarp tikrosios laiko eilutės ir prognozuotos (kas greičiausiai gali būti laikoma neigiamu faktoriumi prognozuojant vienu žingsniu į priekį). Nors jungtinio modelio prognozavimo paklaidų RMSE yra prastesnis nei ARIMA(3,0,3), tačiau prognozuojamos laiko eilutės šiuo atveju kintamumas yra žymiai artimesnis tikrajai.

**3.9 lentelė.** Andersono sausumos termometro temperatūros laiko eilutės prognozės paklaidų RMSE

Prognozės modelis	Algebrinių sekų	ARIMA(3,0,3)	MA	SES	Jungtinis ARIMA(3,0,3) ir algebrinių sekų
<b>RMSE</b>	0,9204	0,4556	0,6117	0,6321	0,6079



**3.10 pav.** Andersono sausumos termometro temperatūros laiko eilutės prognozė: A – pristatytu modeliui; B – ARIMA(3, 0, 3) modeliui; C – slenkančio vidurkio modeliui; D – SES modeliui; E – kombinuotu modeliui.

#### 3.7.4. Australijos bazinės mėnesinės geležies gamybos laiko eilutė nuo 1956 m. sausio iki 1995 m. rugpjūčio

Australijos bazinės mėnesinės geležies gamybos (tūkstančiais tonų) laiko eilutės sudarytos nuo 1956 m. sausio iki 1995 m. rugpjūčio [142], Hankelio matricos determinanto pirmasis minimumas pasiekiamas kai rangas lygus 6 ( $n=5$ , laiko

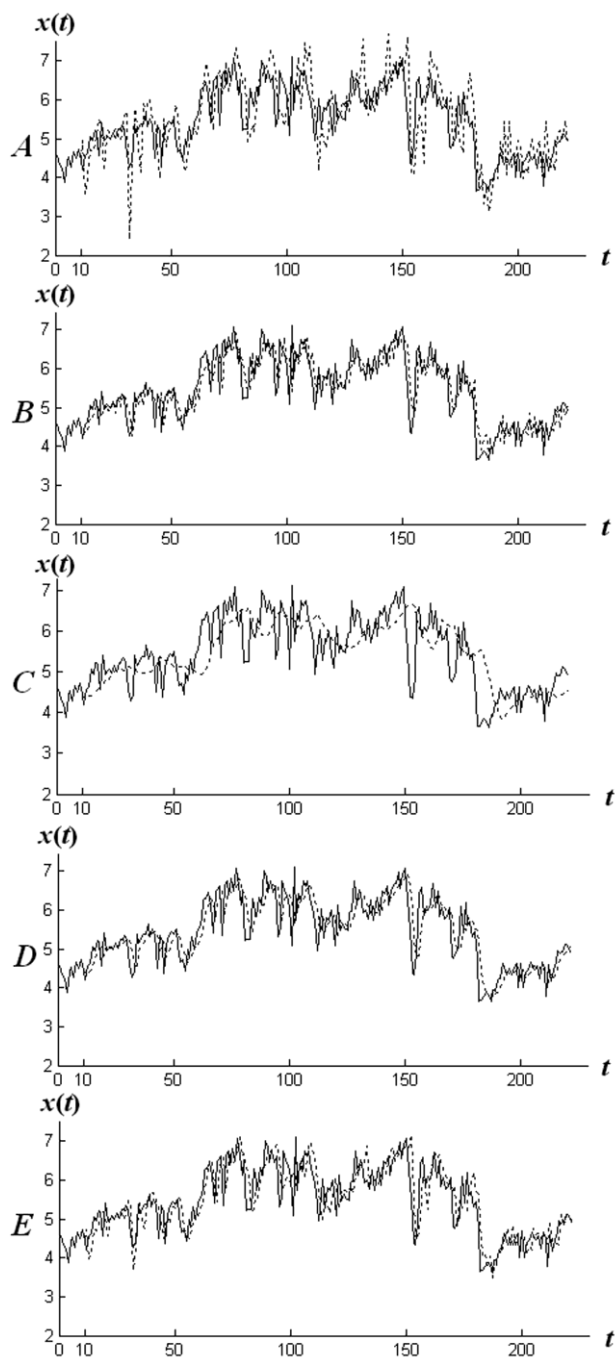
eilutės bazinio fragmento ilgis  $2n+1=11$ ). Evoliucinio algoritmo parametrai lieka nepakeisti. Gauti vieno žingsnio į priekį prognozavimo rezultatai vaizduojami 3.11 paveikslo A dalyje (prieš prognozuojant laiko eilutę, mes visi duomenys dalijami iš 100).

Eksperimentiniu būdu nustatyta, kad ARIMA modelio prognozavimas geriausias, kai jo architektūra (5,1,4) (modelio adekvatumo tyrimas pateiktas preide lentelėje Nr. 11), o slenkančio vidurkio, kai pasirenkamas parametras  $k=9$ . Prognozavimo rezultatai vaizduojami 3.11 paveiksle, o paklaidų RMSE palyginami 3.10 lentelėje. Bendru atveju algebrinių sekų metodas gerai dirba tam tikruose pradinės laiko eilutės intervaluose. Tai reiškia, kad algebriniai sąryšiai, valdantys laiko eilutės evoliuciją, egzistuoja tik tam tikrose duomenų srityse. Čia yra keletas vietų, kuriose sukonstruoto modelio prognozavimo paklaidos yra labai didelės. ARIMA(5,1,4) vykdymas šiuo atžvilgiu yra geresnis (nors šis modelis neįvertina lokaliųjų minimumų ir maksimumų). Jungtinis modelis (mūsų siūlomo ir ARIMA(5,1,4) modelių vidurkis) rodo gerą kintamumą, o prognozuojama laiko eilutė mėgdžioja tikrosios laiko eilutės dinamiką. Reikia pažymėti, kad eksponentinio glodinimo metodo ir jungtinio metodo RMSE 3.10 lentelėje yra panašūs, tačiau vis vien prastesnės nei ARIMA(5,1,4) prognozavimo paklaidų RMSE. Todėl konstruoti siūlomo ir eksponentinio glodinimo modelių aritmetinį vidurkį nebūtų naudinga.

Bendru atveju, prognozavimo rezultatai gali būti dar labiau pagerinti, jei keletas momentinių klaidų gautų algebrinių sekų modeliu, būtų sumažintos. Kaip jau minėjome anksčiau, viena iš priežasčių tokių momentinių paklaidų gali būti staigus algebrinės taisyklės apibūdinančios pradinės laiko eilutės dinamiką, pasikeitimas (dėl įvairių sezoninių ir ciklinių komponentų).

**3.10 lentelė.** Australijos bazinės mėnesinės geležies gamybos laiko eilutės prognozės paklaidų RMSE

Prognozės modelis	Algebrinių sekų	ARIMA(5,1,4)	MA	SES	ARIMA(5,1,4) ir algebrinių sekų
RMSE	0,6727	0,4181	0,6403	0,4702	0,4759



**3.11 pav.** Australijos bazinės mėnesinės geležies gamybos laiko eilutės prognozė: A – pristatytu modelių; B – ARIMA(5, 1, 4) modelių; C – slenkančio vidurkio modelių; D – SES modelių; E – kombinuotu modelių.

### 3.8. Adaptyvi rango nustatymo strategija

Ankstesniuose skyreliuose ieškant laiko eilutės bazinio fragmento buvo laikoma, kad surastas sekos rangas yra vienodas visai laiko eilutei. Tačiau tiriant realaus pasaulio laiko eilutes dažnai susiduriama su duomenyse esančiu sezoniškumu, trendu, triukšmu. Todėl ieškant skeletinės sekos reikia atsižvelgti į laiko eilutėje vykstančius staigius stebimojo proceso pokyčius, t. y. bus pritaikyta adaptyvi rango nustatymo strategija. Teoriškai sekos minorų  $H_0^{(m)}$  determinantų skaičiavimas stebimas tol, kol su tam tikra  $m$  reikšme gaunamas  $\det(H_0^{(m)})=0$ . Deja realios laiko eilutės dažnai būna chaotinės, kurioms rangas praktiškai neegzistuoja (įrodyta [113],  $m$  artėja į begalybę). Todėl [156] straipsnyje siūloma sekos rangui suteikti viršutinį rėžį  $\bar{m}$ . Jei sekos Hankelio minorų determinantai neartėja į nulį iki  $\bar{m}$ , tuomet rango paieškos procesas užbaigiamas priimant, kad  $HrS_0 \geq \bar{m}$ .

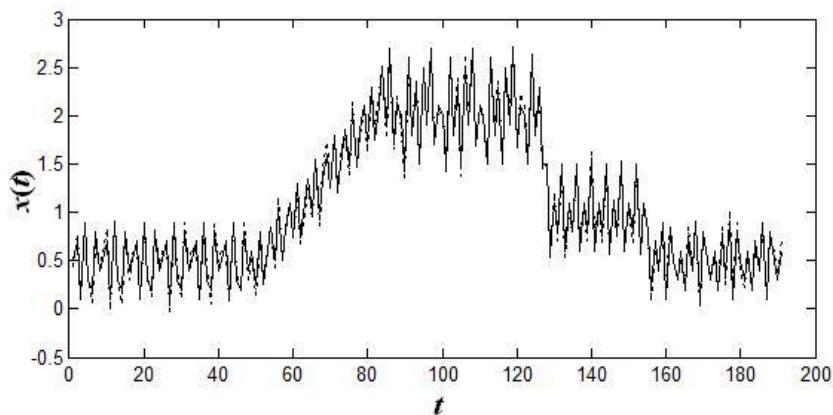
Nors teoriškai reikia surasti tokio matavimo Hankelio matricą, kurios determinantas yra lygus nuliui, praktiškai užtenka surasti determinantą iki užsibrėžto tikslumo, kaip šiuo atveju  $\varepsilon_{(r)}$ . Taigi, determinantų skaičiavimas yra tęsiamas iki  $|\det(H_0^{(m)})| < \varepsilon_{(r)}$  arba  $m > \bar{m}$ . Šiuo atžvilgiu skaičiuojamas ne sekos rangas, o pseudo rangas (analogiškai kaip tiesiniuose operatoriuose skaičiuojamas pseudo spektras [156]). Tikslas – rasti adaptyvią strategiją optimalaus rango parinkimui, su kuriuo Hankelio matricos determinantas artėtų į optimalią  $\varepsilon_{(r)}$  reikšmę.

Vėl konstruojama dirbtinė algebrinė seka, kuri susideda iš kelių periodinių sekų. Pavyzdys pateiktas 3.12 paveiksle: ištisine linija atitinka seką, kurioje pirmą periodinių elementų grupę  $\{0,5; 0,7; 0,1; 0,9; 0,3; 0,2; 0,8; 0,4\}$  (elementai periodiškai kartojasi šešis kartus), o jos H-rangas lygus 9, antra geometrinių greičių auganti elementų grupė periodiškai kartojasi kas šeštas elementas septynis kartus, trečios grupės H-rangas lygus 11 (ši elementų grupė kartojama keturis kartus), ketvirtos – 5 (elementų grupė kartojama septynis kartus) ir paskutinės grupės – 10 (laiko eilutėje elementų grupė kartojama keturis kartus). Toliau kiekvienam sekos elementui pridedami atsitiktiniai skaičiai, pasiskirstę pagal tolygųjį skirstinį intervale  $[-0,15; 0,15]$  (testinė eilutė demonstruojama 3.12 paveiksle: punktyrine linija). Testinė laiko eilutė bus naudojama adaptyvaus rango nustatymo strategijai parinkti.

Pirmiausia reikia panagrinėti, kokį optimalų  $\varepsilon_{(r)}$  reikia parinkti, kad ieškomas Hankelio matricos pseudo rangas būtų optimalus. Akivaizdu, jeigu skeletinės sekos evoliucija žinoma ir ji nesikeičia, Hankelio matricos determinantas turėtų būti lygus nuliui. Tokiu atveju gaunama prognozė tiksliai atkartoja stebimos laiko eilutės duomenis. Tačiau tuo atveju, kai stebimo proceso evoliucija pasikeičia, prieš tai nustatytas skeletinės sekos bazinis fragmentas savyje saugo tik tokius duomenis kurie buvo nustatyti iki tam tikro proceso šuolio. Tokiu atveju, Hankelio matricos rangas jau nebebus lygus nuliui. Jeigu laiko eilutėje vyksta dažni evoliucijos pokyčiai, jie turi įtakos optimalaus rango nustatymui, ir dažnai Hankelio matricos determinanto reikšmė yra žymiai didesnė už nulį. Taigi, tegu pradžioje ieškomas Hankelio matricos determinantas, kuris būtų mažesnis už 0,5 (pradinė  $\varepsilon_{(r)}$  reikšmė).



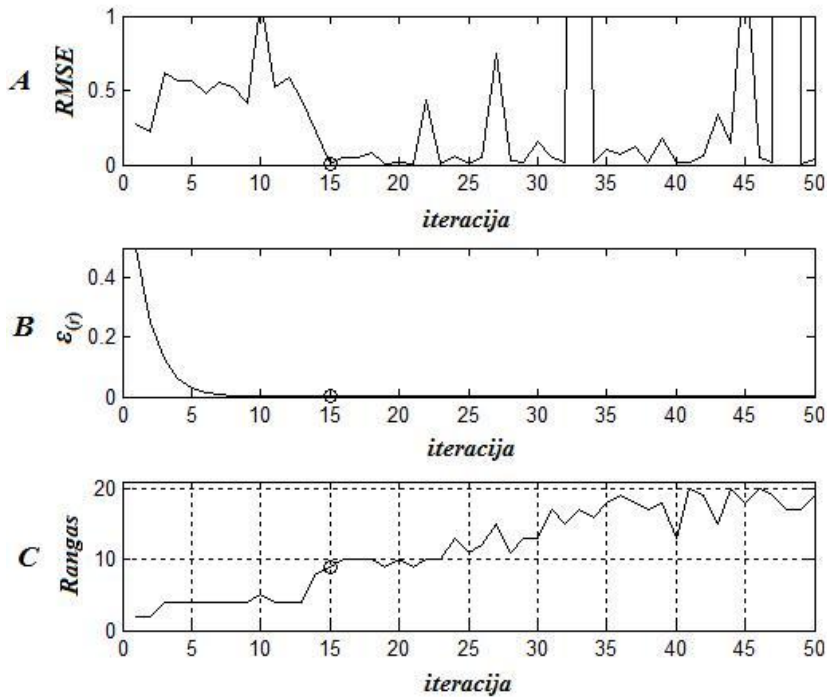
Tada ieškomas laiko eilutės toks sekos bazinis fragmentas, su kuriu ne tik Hankelio matricos determinantas būtų mažesnis už  $\varepsilon_{(r)}$ , bet ir laiko eilutės prognozuojama reikšmė būtų tiksliausia. Tokiu atveju tikrinamas prognozės paklaidų RMSE. Suradus šį rinkinį jis fiksuojamas ir pereinama į kitą iteraciją sumažinant  $\varepsilon_{(r)}$  per pusę. Paieškos procesas tęsiamas tol, kol iteracijų skaičius pasieks 50. Tyrimo ieškomasis rangas bus laikomas optimaliu, jei stebimos reikšmės nuo ekstrapoliuotos reikšmės paklaidos RMSE yra mažiausia. Vienu metu skaičiavimuose stebimi prognozuojamos sekos RMSE, sekos Hankelio matricos determinantas ir rangas. Tokio skaitinio eksperimento rezultatai pateikiami 3.13 paveiksle. Šiame paveiksle galima pastebėti, kad prognozės paklaidų RMSE (3.13 paveikslo A dalis) priklauso nuo  $\varepsilon_{(r)}$  reikšmės (3.13 paveikslo B dalis). Sekos geriausia prognozė gaunama, kai Hankelio matricos determinantas mažesnis  $\varepsilon_{(r)} = 3 \cdot 10^{-5}$ . Tada sekos rangas yra 9 (3.13 paveikslo C dalis). Šiuo pavyzdžiu galima pastebėti, kad nebūtinai mažesnė  $\varepsilon_{(r)}$  reikšmė atitinka geresnę laiko eilutės prognozę, t. y. tokiu būdu, buvo sukonstruota adaptyvaus rango parinkimo strategija.



**3.12 pav.** Testinės laiko eilutės (ištininė linija) ir kelių dalių periodinės laiko eilutės (punktyrinė linija) grafinis vaizdas

Dabar kai jau žinoma, su koku rangu geriausia susidaryti skeletinės sekos bazinį fragmentą, evoliucinių algoritmų pagalba galima prognozuoti tiriamą laiko eilutę. Pasiekama prie tos pačios 3.1 skyrelyje aprašytos prognozės strategijos. Evoliucinių algoritmų parametrai yra jau suderinti 3.6 skyrelyje. Šiuo atveju prognozuojant laiko eilutės vieną reikšmę į priekį, stebima, ar einamuoju momentu prognozuojama reikšmė nenukrypsta nuo tikrosios laiko eilutės reikšmės daugiau nei nusistatyta riba. Čia galima reikalauti, pavyzdžiui, kad prognozuojamos reikšmės skirtumo nuo tikrosios reikšmės modulis neviršytų 0,75. Jei užsibrėžta riba viršijama, tolesnei laiko eilutės prognozei atlikti iš naujo nusistatomas rangas jau minėtu būdu. Tokiu būdu adaptyviai parinkinėjant rangą yra prognozuojama laiko eilutė atsižvelgiant į joje vykstančius staigius pokyčius. Tokios prognozės rezultatas pateiktas 3.14 paveikslo A dalyje punktyrine linija. Šio paveikslo B dalyje pavaizduota, kuriose vietose prognozės absoliučiosios paklaidos viršija nusistatytą ribą. Tuomet, atitinkamai yra keičiamas sekos rangas, kuris demonstruojamas 3.14

paveikslo C dalyje. Keičiantis rangui reikalinga, kad procesas nusistovėtų, todėl tose vietose, kuriose rangas žymimas punktyrine linija, prognozė gaunama gana netiksli. Kai sekos fragmentas yra pakankamo ilgio, prognozuojamos ateities reikšmės gaunamos tikslesnės.

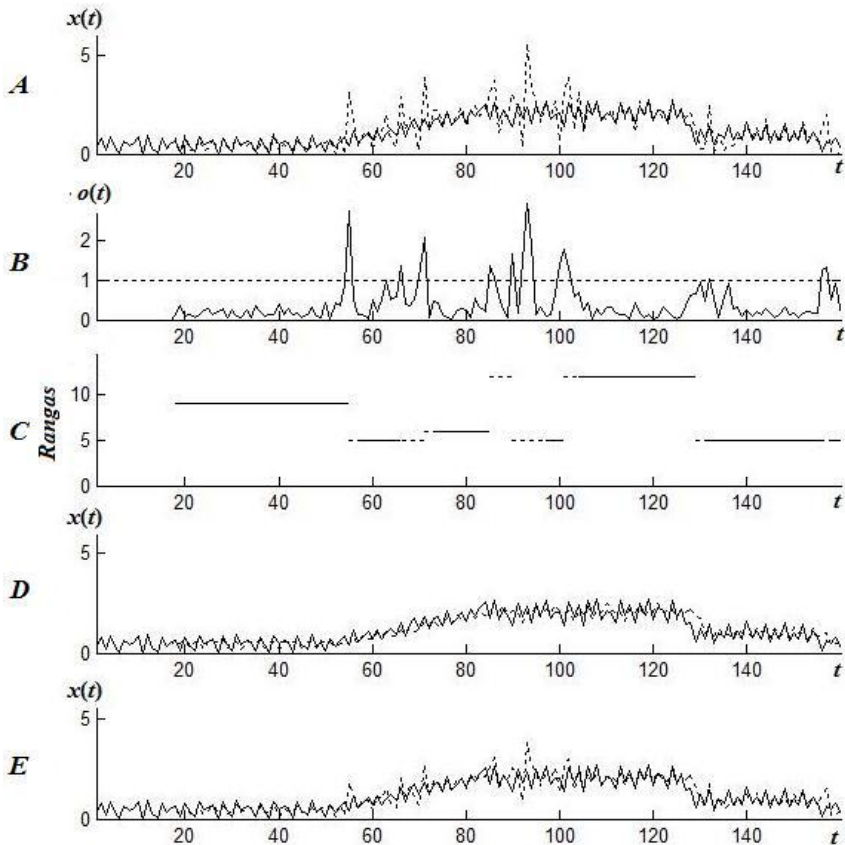


**3.13 pav.** Sekos rango priklausomybė nuo  $\varepsilon_{(r)}$ . Po kiekvienos iteracijos fiksuojama: A – prognozės paklaidos RMSE; B –  $\varepsilon_{(r)}$  mažėjimas; C – optimalus sekos Hankelio matricos rangas.

Sudaryto modelio funkcionalumas palyginamas su ARIMA, slenkančio vidurkio, eksponentinio glodinimo ir jungtinių modeliais (3.11 lentelė). ARIMA modelis geriausiai prognozuoja, kai naudojama architektūra (3, 1, 2) (autoregresijos dalies eilė 3, integruotosios dalies – 1, slenkančio vidurkio – 2). Modelio adekvatumo tyrimas pateiktas preide lentelėje Nr. 12. Šiuo modeliu gaunamų prognozės paklaidų RMSE lygi 0,3273 (prognozės rezultatai pateikti 3.14 paveikslo D dalyje). Slenkančio vidurkio modelis geriausiai prognozuoja, kai parametras  $k = 3$ . Šiame pavyzdyje nors sukonstruoto modelio paklaidos yra didžiausios, tačiau geriausius rezultatus duodančio ARIMA(3,1,2) modelio prognozės reikšmės yra labiau sugludintos. Taigi norint pademonstruoti triukšmingo signalo geresnę prognozę, nei sukonstruoto modelio, bet ne sugludinti, galima pasiūlyti naudoti jungtinių sukonstruoto ir ARIMA modelių prognozavimą. Tokiu būdu paklaidų RMSE sumažėja nuo 0,6280 iki 0,3794. Tokio modelio prognozės rezultatai pateikti 3.14 paveikslo E dalyje punktyrine linija.

**3.11 lentelė.** Testinės laiko eilutės su prognozės adaptyviu rango parinkimu paklaidų RMSE

Prognozės modelis	Kintamo rango algebrinių sekų	ARIMA(3, 1, 2)	MA	SES	Jungtinis ARIMA(3,1,2) ir kintamo rango algebrinių sekų
RMSE	0,6280	0,3273	0,3956	0,4194	0,3794



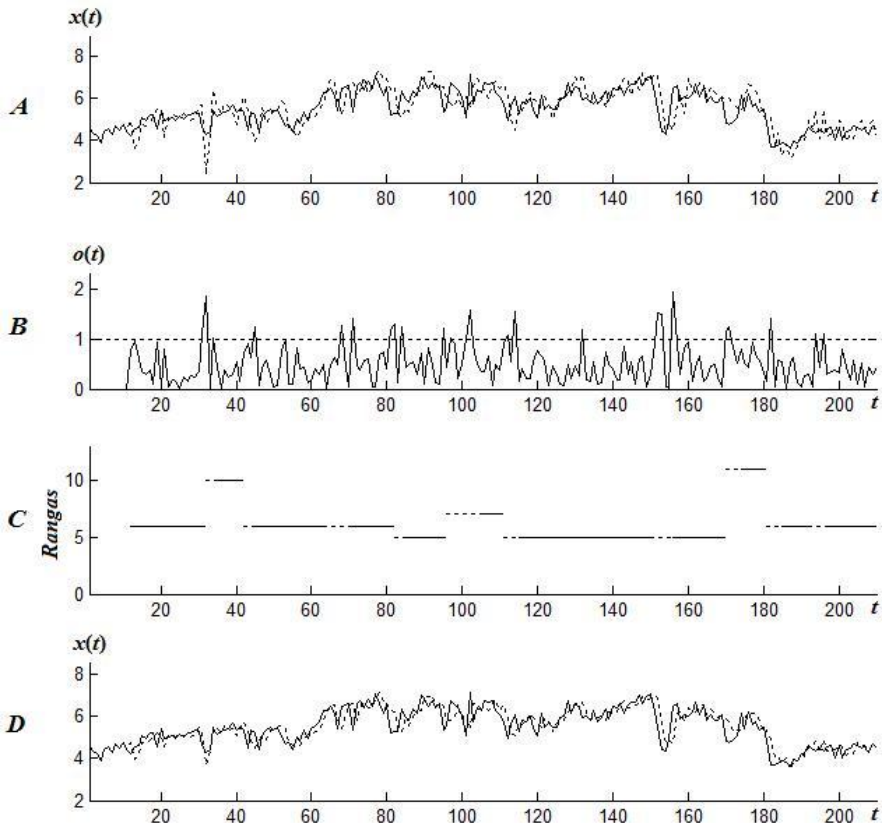
**3.14 pav.** Kelių periodų testinės laiko eilutės prognozė: A – mūsų siūlomo modelio prognozė (punkttyrinė linija) su adaptyviu rango parinkimu, B – modelio prognozės absoliučiosios paklaidos, C – rango kitimas, D – ARIMA(3, 1, 2) prognozė (punkttyrinė linija), E – Jungtinė ARIMA ir mūsų siūlomo modelio prognozė (punkttyrinė linija).

**3.9. Australijos bazinės mėnesinės geležies gamybos laiko eilutės prognozė adaptyvaus rango parinkimo modeliu**

Iš naujo nagrinėjamas Australijos bazinės mėnesinės geležies gamybos laiko eilutės skeletinės sekos bazinio fragmento detekcijos uždavinys, tik šiuo atveju Hankelio matricos rango ieškomas adaptyviai pagal 3.8 skyrelyje aprašytą metodiką. Pradinei sekos daliai, kaip ir 3.7.4 skyrelyje, bazinio fragmento ilgis paliekamas 11, t. y. rangas 6. Nekeičiant evoliucinių algoritmo parametrų prognozės vieno žingsnio į priekį rezultatai vaizduojami 3.15 paveikslo A dalyje. Šiuo atveju reikalaujama, kad prognozės absoliučiosios paklaidos neviršytų vieneto (kaip demonstruojama

3.15 paveikslo B dalyje). Tokiu būdu rango kitimas priklausomai nuo prognozės paklaidų vaizduojamas 3.15 paveikslo C dalyje. Tokios prognozės RMSE lygi 0,6239 (kai su pastoviu rango konstruojant skeletinę seką prognozės paklaidų RMSE buvo 0,6727). Tačiau 3.7.4 skyrelyje gauta ARIMA(5,1,4) prognozė vis vien lieka geresnė, nes šio modelio prognozės paklaidų RMSE lygi 0,4235 (žiūrėti 3.10 lentelę).

Jungtinis modelis (sukonstruoto ir ARIMA(5,1,4) modelių vidurkis) demonstruojamas 3.15 paveikslo D dalyje rodo gerą kintamumą ir pakankamai tikslų laiko eilutės atkartojimą. Tokio modelio RMSE lygi 0,4235, kuri mažai skiriasi nuo ARIMA modelio RMSE, tačiau prognozės vaizdas yra mažiau suglotnintas.



**3.15 pav.** Australijos bazinės mėnesinės geležies gamybos laiko eilutės prognozė: A – mūsų siūlomo modelio prognozė (punktyninė linija) su adaptyviu rango parinkimu, B – modelio prognozės absoliučiosios paklaidos, C – rango kitimas, D – Jungtinė ARIMA(5, 1, 4) ir mūsų siūlomo modelio prognozė (punktyninė linija).

### 3.10. Skyriaus išvados

Šiame skyriuje buvo pristatytas laiko eilučių prognozės modelis, kuris pagrįstas skeletinės sekos identifikavimu. Šis modelis efektyviai veikia su trumpomis laiko eilutėmis. Tokia algebrinės sekos detekcija leidžia iš pakankamai mažo duomenų kiekio ištraukti kiek įmanoma daugiau informacijos apie stebimą procesą ir tada panaudoti šią informaciją praeities duomenims ekstrapoliuoti į ateitį.

Šiame modelyje rastas algebrinės sekos bazinis fragmentas yra taisyklė, kuri rekonstruojama į kiekvieną laiko eilutės momentą. Prognozės strategijai vykdyti buvo sukurta tinkamumo funkcija, kurioje pagrindinis akcentas dedamas laiko eilutės lokalsioms korekcijoms rasti. Šios lokalsios korekcijos buvo panaudotos sekos pseudo rangui konstruoti. Kiekvienas laiko eilutės prognozuojamas taškas buvo gaunamas po 100 kartų vykdant evoliucinius algoritmus. Kadangi evoliuciniais algoritmais prognozuojant laiko eilutės atitinkamas reikšmės nepavykdavo išvengti tam tikrų išskirčių, dėl kurių nukentėdavo prognozės kokybė, buvo suskurtas būdas, kaip atsikratyti tam tikru procentu labiausiai nutolusių prognozės reikšmių.

Laiko eilutės prognozės funkcionalumui patikrinti panaudotos dirbtinės algebrinės sekos su jas veikiančiu tam tikru triukšmo lygiu bei keletas realių laiko eilučių. Sukurto modelio prognozės paklaidų RMSE visą laiką buvo didesnės nei ARIMA, o kartais nei slenkančio vidurkio ar viengubo eksponentinio glodinimo, tačiau pastarieji modeliai daugiau suglotnina laiko eilutes ir neparodo laiko eilutės atitinkamo kintamumo. Algebrinių sekų modelis, priešingai, labai jautriai reaguoja į laiko eilutės variabilumą. Todėl buvo pasiūlyta naudoti jungtinį algebrinių sekų modelio ir ARIMA modelio prognozavimą.

Norint pagerinti sukonstruotą laiko eilučių prognozės modelį, buvo pademonstruota, kad galima laiko eilutės bazinį fragmentą konstruoti adaptyviai parenkant Hankelio matricos rangą. Darbe parodyta, kad konstruojant skeletinės sekos pseudo rangą reikia atsižvelgti, kiek nykstamai maža jo reikšmė gali būti. Taigi stebint laiko eilutės prognozės paklaidas einamuoju momentu, priimamas sprendimas, ar sekančio momento reikšmei prognozuoti reikia keisti skeletinės sekos rangą.

## IŠVADOS

1. Sukurta nauja laiko eilučių nereguliaraus rekonstravimo į laiko vėlinimų erdvę metodika, skirta optimalaus laiko vėlinimų rinkiniui sudaryti, kuris būtinas ilgų laiko eilučių prognozavimui. Sukurta laiko eilučių prognozavimo metodika, pagrįsta neraiškių logikų sistemų tinklais ir nereguliariu atraktoriaus rekonstravimu į vėlinimų erdvę. Pademonstruoti šios prognozavimo metodikos pranašumai prieš kai kurias geriausias žinomas metodikas (savaime apsimokantys ANFIS modeliai, hibridiniai ANN ir ARMA modeliai ir kt.)
2. Konstruojant evoliucinius algoritmus, kryžminimo procese įdiegtas naujas panašumo parametras, ypač tinkamas sprendžiant diskrečiuosius daugiamačius optimizavimo uždavinius.
3. Sukurtas prognozavimo modelis ypač efektyvus, dirbant su ergodinėmis chaotinėmis laiko eilutėmis (Mackey-Glass, Rosstler, kintamos žvaigždės kasdienio ryškumo laiko eilutės), nors ir negali prognozuoti netikėtų impulsų, šuolių ir sezoniškumų – jei tokių nepasitaikė modelio mokymo etape (tai parodė eksperimentai su ECG bei S ir P laiko eilutėmis).
4. Trumpų laiko eilučių prognozei sukurtas modelis, kuriame panaudota algebrinių sekų bazinio fragmento identifikacijos koncepcija. Algebrinės sekos baziniam fragmentui nustatyti sudaryta lokaliųjų korekcijų, kurios laiko eilutę veikia kaip adaptyvus triukšmo valymo filtras, tinkamumo funkcija. Atlikus eksperimentus įsitikinta, kad prognozavimas efektyvus, kai laiko eilutė nėra trumpesnė nei mažiausia reikalingo ilgio skeletinė algebrinė seka. Tyrime parodyta, kad minėtas trumpų laiko eilučių modelis gerai dirba su eilutėmis, kurias veikia tik gana siauro intervalo triukšmas.
5. Sudarytas originalus labiausiai nukrypusių algebrinių skeletinių sekų atmetimo algoritmas leido dar pagerinti prognozuojamų reikšmių paklaidas.
6. Sukurto trumpų laiko eilučių prognozės modelio funkcionalumui patikrinti vertinta prognozės paklaidų RMSE metrika ir lyginta su ARIMA, slenkančio vidurkio ir viengubo eksponentinio glodinimo prognozės modeliais. Nors algebrinių sekų prognozės rezultatai daug geriau parodo duomenų sklaidą, kartais gautos net beveik 2 kartus didesnės paklaidos, nei taikant ARIMA modelį. Eksperimentais įrodyta, jog mūsų sukurto modelio ir ARIMA modelio apjungimas leidžia išlaikyti gerą duomenų sklaidą ir priimtinas paklaidas.

## 7. LITERATŪRA

1. *NIST/SEMATECH e-Handbook of Statistical Methods*, 2003, [interaktyvus] [žiūrėta 2010-06-29]. Prieiga per internetą: <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc42.htm>.
2. R. Harris, *Applied Time Series Modeling and Forecasting*, John Wiley & Sons, Chichester, 2004, 302 p., ISBN 0-470-84443-4.
3. Hyndman, R. J. *Moving Averages*, 2008 [interaktyvus] [žiūrėta 2011-03-21]. Prieiga per internetą: <http://robjhyndman.com/papers/movingaverage.pdf>.
4. Mills, T. C. *The econometric modeling of financial time series*, Cambridge University Press, Cambridge, 2003, 372 p., ISBN: 0-521-624134.
5. Sitte, R.; Sitte, J. Neural network approach to the random walk dilemma of financial time series, *Applied Intelligence*, 2002, vol. 16, no. 3, p. 163-171.
6. Araujo, R. de A. Swarm-based translation-invariant morphological prediction method for financial time series forecasting, *Information Science*, 2010, vol. 180, p. 4784-4805.
7. Chatfield, C. *Time series Forecasting*, Chapman&Hall/CRC, 2001. 280 p., ISBN 978-1-58488-063-9.
8. J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Kluwer Academic Publishers, Norwell, 2001).
9. E. M. Azoff, *Neural Network Time Series: Forecasting of Financial Markets* (John Wiley & Sons, Chichester, 1994).
10. S. Makridakis, C. C. Wheelwright and R. J. Hyndman, *Forecasting Methods and Applications* (Wiley, New York, 1998).
11. Taylor, J. W. Smooth transition exponential smoothing, *Journal of Forecasting*, 2004, vol. 23, p. 385-394.
12. Gardner, E. S. Jr. Exponential smoothing: The state of the art – Part II, 2005, [interaktyvus], [žiūrėta 2009-12-13]. Prieiga per internetą: <http://www.bauer.uh.edu/gardner/docs/pdf/Exponential-Smoothing.pdf>.
13. Kunst, R. M. Econometric Forecasting, 2009, [interaktyvus] [žiūrėta 2011-02-11]. Prieiga per internetą: <http://homepage.univie.ac.at/robert.kunst/prognos.pdf>.
14. Costa, M.; Alpuim, T. Parameter estimation of state space models for univariate observations, *Journal of Statistical Planning and Inference*, 2010, vol. 140, p. 1889-1902.
15. De Gooijer, G. J.; Hyndman, R. J. 25 years of time series forecasting, *Internacional Journal of Forecasting*, 2006, vol. 22, p. 443-473.
16. Ferkingstad, E.; Loland, A.; Wilhelmsen, M. Causal modelling and inference for electricity markets, *Energy Economics*, 2011, vol. 33, p. 404-412.
17. Paul, P. S. Predictors of work injury in underground mines – an application of a logistic regression model, *Mining Science and Technology*, 2009, vol. 19, p. 282-289.
18. Brandimarte, P.; Zotteri, G. *Introduction to distribution logistics*. John Wiley & Sons. Hoboken, New Jersey, 2007. 587 p. ISBN 978-0-471-75044-4.
19. Dablemont, S. et al. Time series forecasting with SOM and local non-linear models – application to the DAX30 index prediction, *Proceedings of the Fourth Workshop on Self-Organizing Maps, (WSOM) 03*, 2003.
20. Coyle, D.; Prasad, G.; McGinnity, T. M. A time-series prediction approach for feature extraction in a brain-computer interface, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2005, vol. 13, no. 4, p. 461-467.
21. Leu, Y. H.; Lee, C. P.; Jou, Y. Z. A distance-based fuzzy time series model for exchange rates forecasting. *Expert Systems with Applications*, 2009, vol. 36, no. 4, p. 8107-8114.
22. Chen, S. M.; Hwan, J. R. Temperature prediction using fuzzy time series. *IEEE Transactions on System Man and Cybernetics B: Cybernetics*, 2000, vol. 30, no. 2, p. 263-275.

23. Chen, W.-S.; Du, Y.-K. Using neural network and data mining techniques for the financial distress prediction model, *Expert Systems with Application*, 2009, vol. 36, p. 4075-4086.
24. Zounemat-Kermani, M.; Teshnehlab, M. Using adaptive neuro-fuzzy inference system for hydrological time series prediction, *Applied Soft Computing*, 2008, vol. 8, 928-936.
25. Harvey, A. C. Forecasting structural time series models, and the Kalman filter, Cambridge University Press, U.K., 1989, p. 554. ISBN 0-521-32196-4.
26. Corberan-Vallet, A.; Bermudez, J. D.; Vercher, E. Forecasting correlated time series with exponential smoothing models, *International Journal of Forecasting*, 2011, vol. 27, p. 252-265.
27. Bermudez, J. D.; Segura, J. V.; Vercher, E. SIOPRED: a prediction and optimization integrated system for demand, *TOP*, 2008, vol. 16, 258-271.
28. Hyndman, R. J. et al. *Forecasting with exponential smoothing: the state space approach*. New York, NY: Springer, 2008, [interaktyvus] [žiūrėta 2011-02-12]. Prieiga per internetą: <http://www.exponentialsMOOTHING.net/>.
29. Kalekar, P. S. *Time series forecasting using Holt-Winters exponential smoothing*, 2004, [interaktyvus] [žiūrėta 2011-03-21]. Prieiga per internetą: [http://www.it.itb.ac.in/~praj/acads/seminar/04329008\\_ExponentialSmoothing.pdf](http://www.it.itb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf).
30. Hahn, H.; Meyer-Neiberg, S.; Pickl, S. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, 2009, vol. 199, p. 902-907.
31. El-Talbany, M.; El-Karmi, F. Short-term forecasting of Jordanian electricity demand using particle swarm optimization. *Electric Power Systems Research*, 2008, vol. 78, p. 425-433.
32. Lee, Ch. M.; Ko, Ch. N. Short-term load forecasting using lifting scheme and ARIMA models, *Expert Systems with Applications*, 2011, vol. 38, p. 5902-5911.
33. Khashei, M.; Bijari, M.; Ardali, G. A. R. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs), *Neurocomputing*, 2009, vol. 72, p. 956-967.
34. Aznarte, M. J. L. et al. Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models, *Expert Systems with Applications*, 2007, vol. 32, no. 4, p. 1218-1225.
35. Ture, M.; Kurt, I. Comparison of four different time series methods to forecast hepatitis A virus infection, *Expert Systems with Applications*, 2006, vol. 31, no. 1, p. 41-46.
36. Rojas, I. et al. Soft-computing techniques and ARMA model for time series prediction, *Neurocomputing*, 2008, vol. 71, p. 519-537.
37. Valenzuela, O. et al. Hybridization of intelligent techniques and ARIMA models for time series prediction, *Fuzzy Sets and Systems*, 2009, vol. 159, p. 821-845.
38. Kaminskas, V. Идентификация динамических систем по дискретным наблюдениямб (Mokslas, Vilnius, 1982).
39. Nagpul P.S. Time series analysis in winIDAMS, 2005, India, [interaktyvus] [žiūrėta 2011-12-20]. Prieiga per internetą: <http://portal.unesco.org/ci/en/files/18650/11133194701TimeSeriesAnal.pdf/TimeSeriesAnal.pdf>.
40. Fisher, T. J. Testing Adequacy of ARMA Models using a Weighted Portmanteau Test on the Residual Autocorrelations, [interaktyvus] [žiūrėta 2012-01-05]. Prieiga per internetą: <http://support.sas.com/resources/papers/proceedings11/327-2011.pdf>.
41. Porporato, A.; Ridolfi, L. Multivariate nonlinear prediction of river flows, *Journal of Hydrology*, 2001, vol. 248, p. 109-122.
42. Kato, H.; Naniwa, S.; Ishiguro, M. A bayesian multivariate nonstationary time series model for estimating mutual relationships among variables, *Journal of Econometrics*, 1996, vol. 75, p. 147-161.



43. Swift, S.; Liu, X. Predicting glaucomatous visual field deterioration through short multivariate time series modeling, *Artificial Intelligence in Medicine*, 2002, vol. 24, p. 5-24.
44. Preez, J.; Witt, S. F. Univariate versus multivariate time series forecasting: an application to international tourism demand, *International Journal of Forecasting*, 2003, vol. 19, p. 435-451.
45. Kunst, R. M. *Multivariate forecasting methods*, 2004, [interaktyvus] [žiūrėta 2011-01-21]. Prieiga per internetą: <http://homepage.univie.ac.at/robert.kunst/prognos4.pdf>.
46. Melard, G.; Roy, R.; Saidi, A. Exact maximum likelihood estimation of structured or unit root multivariate time series models, *Computational Statistics & Data Analysis*, 2006, vol. 50, p. 2958-2986.
47. Kascha, C.; Mertens, K. Business cycle analysis and VARMA models, *Journal of Economic Dynamics and Control*, 2009, vol. 33, issue 2, p. 267-282.
48. Thomakos, D. D.; Guerard, Jr. J. B. Naïve, ARIMA, nonparametric, transfer function and VAR models: A comparison of forecasting performance, *Internacional Journal of Forecasting*, 2004, vol. 20, p. 53-67.
49. Rueda, C.; Rodriguez, P. State space models for estimating and forecasting fertility, *International Journal of Forecasting*, 2010, vol. 26, p. 712-724.
50. Commandeur, J. J. F.; Koopman, S. J. *An introduction to state space time series analysis*. Oxford University Press, 2007, 192 p., ISBN: 978-0-19-922887-4.
51. Haung, M. J.; Chen, M. Y.; Lee, S. C. Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis, *Expert Systems with Applications*, 2007, vol. 32(3), p. 856-867.
52. Kirkos, E.; Spathis, C.; Manolopoulos, Y. Data mining techniques for the detection of fraudulent financial statements, *Expert Systems with Applications*, 2007, vol. 32, p. 995-1003.
53. Huang, Y. P.; Hsu, C. C.; Wang, S. H. Pattern recognition in time series database: A case study on financial database, *Expert Systems with Applications*, 2007, vol. 33, p. 199-205.
54. Weigend, A.; Gershenfeld, N. *Time series prediction: Forecasting the future and understanding the past*, Addison-Wesley, Reading, 1994, 643 pp., ISBN 0-201-62.
55. Menezes, J. M. P. Jr.; Barreto, G. P. Long-term time series prediction with NARX network: an empirical evaluation, *Neurocomputing*, 2008, vol. 71, p. 3335-3343.
56. Zhang, G.; Patuwo, B. E.; Hu, M. Y. Forecasting with artificial neural networks: the state of the art, *International Journal of Forecasting*, 1998, vol. 14, no. 1, p. 35-62.
57. Khashei, M.; Bijari, M. An artificial neural network (p,d,q) model for time series forecasting, *Expert Systems with Applications*, 2010, vol. 37, p. 479-489.
58. Yeh, J. P. Short-term chaotic time series forecasting, [interaktyvus] [žiūrėta 2011-01-12]. Prieiga per internetą: [http://www.intechopen.com/source/pdfs/13502/InTech-Short\\_term\\_chaotic\\_time\\_series\\_forecast.pdf](http://www.intechopen.com/source/pdfs/13502/InTech-Short_term_chaotic_time_series_forecast.pdf).
59. Tenti, P. Forecasting foreign Exchange rates using recurrent neural networks, *Applied Artificial Intelligence*, 1996, vol. 10, p. 567-581.
60. Ma, Q. et al. Chaotic time series prediction based on evolving recurrent neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, Montreal, Canada, 2005.
61. Assaad, M.; Bone, R.; Cardot, H. Predicting chaotic time series by boosted recurrent neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, Hong Kong, 2006, vol. 4233, p. 831-840.
62. Avci, E. Forecasting daily and sessional returns of the ISE-100 index with neural network models, *Journal of Dogus University*, 2007, vol. 8, no. 2, p. 128-142.
63. Armano, G.; Marchesi, M.; Murru, A. A hybrid genetic-neural architecture for stock indexes forecasting, *Information Science*, 2005, vol. 170, p. 3-33.

64. Huang, C. L.; Tsai, C. Y. A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting, *Expert Systems with Applications*, 2009, vol. 36, no. 2, p. 1529-1539.
65. Cheng, C. H.; Teoh, H. J.; Chen, T. L. High-order fuzzy time series based on rough set for forecasting TAIEX, *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, Hong Kong, 2007, no.3, p. 1354-1358.
66. Cho, S. Y. *et al.* HebbR<sup>2</sup>-traffic: A novel application of neuro-fuzzy network traffic monitoring system, *Expert Systems with Application*, 2009, vol. 36, no. 3, Part 2, p. 6343-6356.
67. Cagdas, H. Aladag , et al. Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations, *Expert Systems with Applications*, 2009, vol. 36, p. 4228-4231.
68. Qiu, W.; Liu, X.; Li, H. A generalized method for forecasting based on fuzzy time series, *Expert Systems with Applications*, 2011, vol. 38, p. 10446-10453.
69. Boyacioglu, M. A., Avci, D. An adaptive network-based fuzzy inference system (FIS) fot the prediction of stock market return: the case of the Istanbul stock Exchange, *Expert Systems and Application*, 2010, vol. 37, p. 7908-7912.
70. Wolkenhauer, O. *Fuzzy mathematics in systems theory and data analysis*. John Wiley & Sons, Inc, NY, 2001. 263 p. ISBN 0-471-41656-8.
71. Wang, F. K.; Chang, K. K.; Tzeng, Ch. W. Using adaptive network-based fuzzy inference systems to forecast automobile sales, *Expert Systems with Application*, 2009, vol. 38, p. 10587-10593.
72. Talei, A.; Chua, L. H. Ch.; Quek, Ch. A novel application of a neuro-fuzzy computational technique in event-based rainfall-runoff modeling, *Expert Systems with Application*, 2010, vol. 37, p. 7456-7468.
73. Shiri, J.; Kasi, O. Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model, *Journal of Hydrology*, 2010, vol. 394, p. 486-493.
74. Atsalakis, G. S.; Dimitrakakis, E. M.; Zopounidis, C. D. Elliott wave Theory and neuro-fuzzy systems, in stock marcet prediction: the WASP system, *Expert Systems with Applications*, 2011, vol. 38, p. 9196-9206.
75. Chen, T. A hybrid fuzzy and neural approach for DRAM price forecasting, *Computer in Industry*, 2011, vol. 62, p. 196-204.
76. Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 2003, vol. 50, p. 159-175.
77. Zou, H. F.; et al. An investigation and comparison of artificial neural network and time series models for chinese food grain price forecasting, *Neurocomputing*, 2007, vol. 70, p. 2913-2923.
78. Yu, T. H. K.; Huarng, K. H. A neural network-based fuzzy time series model to improve forecasting, *Expert Systems with Applications*, 2010, vol. 37, p. 3366-3372.
79. Jang, J. R. ANFIS: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, and Cybernetics*, 1993, vol 23, no. 3, p. 665-685.
80. Abraham, A. Neuro-Fuzzy Systems: State-of-the-Art Modeling Techniques, Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence, *Lecture Notes in Computer Science* in: Mira, J. et al (Eds.), 2001, Germany, Springer, p. 269-276.
81. Kim, J.; Kasabov, N. HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical system, *Neural Networks*, 1999, vol. 12, p. 1301-1319.

82. Abdelrahim, E. M.; Yahagi, T. A new transformed input-domain ANFIS for highly nonlinear system modeling and prediction, in: *Canadian Conference on Electrical and Computer Engineering*, 2001, vol. 1, p. 655-660.
83. Jovanovic, R. *et al.* Modified ANFIS architecture – improving efficiency of ANFIS technique, in: *7th Seminar on Neural Network Applications in Electrical Engineering, NEUREL 2004*, 2004, p. 215-220.
84. Maguire, L. P. *et al.* Predicting chaotic time series using a fuzzy neural network, *Information Science*, 1998, vol. 112, p. 125-136.
85. Yilmaz, N. A. *et al.* ANFIS unfolded in time for multivariate time series forecasting, *Neurocomputing*, 2004, vol. 61, p. 139-168.
86. Chen, D. W.; Zhang, J. P. Time series prediction based on ensemble ANFIS, in: *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, 2005, vol. 6, p. 3552-3556.
87. Li, H. *et al.* Hybrid model of WT and ANFIS and its application on time series prediction of ship roll motion, in: *IMACS Multiconference on Computational Engineering in Systems Applications*, 2006, vol. 1, p. 333-337.
88. Vairappan, C. *et al.* Batch type local search-based adaptive neuro-fuzzy inference system (ANFIS) with self-feedbacks for time series prediction, *Neurocomputing*, 2009, vol. 72, p. 1870-1877.
89. Gershenfeld, N. A.; Weigend, A. S. The future of time series, eds. Reading, MA: Addison-Wesley 1993, [interaktyvus] [žiūrėta 2012-01-29]. Prieiga per internetą: <http://www.cs.colorado.edu/departments/publications/reports/docs/CU-CS-670-93.pdf>.
90. Wang, J. *et al.* Combined modeling for electric load forecasting with adaptive particle swarm optimization, *Energy*, 2010, vol. 35, p. 1671-1678.
91. Zhi-Peng, L. *et al.* An improved adaptive exponential smoothing model for short-term travel time forecasting of urban arterial street, *Acta Automatica Sinica*, 2008, vol. 34, no. 11, p. 1404 – 1409.
92. Hilborn, R. C. *Chaos and nonlinear Dynamics: An introduction for scientists and engineers*, Second Edition, Oxford University Press, New York, 2000. 650 p. ISBN 0-19-850723-2.
93. Feng, H.; Shu, Y.; Yang, O. W. W. Research on characterization of wireless LANs traffic, *Commun Nonlinear Sci Numer Simulat*, 2011, vol. 16, p. 3179-3187.
94. Wang, J., *et al.* Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting, *Expert Systems with Application*, 2011, vol. 38, p. 8419 – 8429.
95. Packard, N. H.; Crutchfield, J. P.; Shaw, R. S. Geometry from a time series, *Physical Review Letters*, 1980, vol. 45, p. 712-716.
96. Berliner, L. M. Statistics, probability and chaos, *Statistical Science*, 1992, vol. 7, p. 69-122.
97. Eckman, J. P.; Ruelle, D. Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems, *Physica D*, 1992, vol. 56, p. 185-187.
98. Kennel, M. B.; Brown, R.; Abarbanel, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review A*, 1992, vol. 15, p. 3403-3411.
99. Takens, F. Detecting strange attractor in turbulence. In *D. A. Rand and L. S. Young. Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, 1981 vol. 898. Springer Verlag. p. 366–381.
100. Mane, R. On the dimension of the compact invariant sets of certain nonlinear maps. In *D. A. Rand and L.S. Young. Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, 1981 vol. 898. Springer Verlag. p. 230–242.
101. Bourbaki, N.; Eggleston, H. G.; Madan, S. *Topological vector spaces: Chapters 1- 5*. Springer-Verlag, New York, 1987. 368 p. ISBN 3-540-13627-4.

102. Judd, K.; Mees, A. Embedding as a modeling problem, *Physica D: Nonlinear Phenomena*, 1998, vol. 120, p. 273-286.
103. Rosenstein, M. T.; Collin, J. J.; De Luca, C. J. A practical method for calculating largest Lyapunov exponents from small data sets, *Physica D*, 1993, vol. 65, p. 117-134.
104. Kodba, S.; Marhl, M. Detecting chaos from a time series, *European Journal of Physics*, 2005, vol. 26, p. 205-215.
105. B. L. Ho, R. E. Kalman, Effective construction of linear state-variable models from input-output functions, *Regelungstechnik*, 1966, vol. 14, p. 545-592.
106. L. S. de Jong, Numerical aspects of recursive realization algorithms, *SIAM Journal on Control and Optimization*, 1978, vol. 16, p. 646-59.
107. H. Akaike, Stochastic theory of minimal realization, *IEEE Transactions on Automatic Control*, 1974, vol. 19, p. 667-674.
108. M. Gevers, T. Kailath, An innovations approach to least-squares estimation, part vi: Discrete-time innovations representations and recursive estimation, *IEEE Transactions on Automatic Control*, 1973, vol. 18, p. 588-600.
109. J. Juang, R. Pappa, An eigensystem realization algorithm for modal parameter identification and model reduction, *Journal of Guidance, Control and Dynamics*, 1985, vol. 8, p. 620-627.
110. R. Carniel, F. Barazza, M. Tarraga, R. Ortiz, On the singular values decoupling in the Singular Spectrum Analysis of volcanic tremor at Stromboli, *Natural Hazards and Earth System Sciences*, 2006, vol. 6, p. 903-909.
111. J. W. Layman, The Hankel Transform and Some of its Properties, *Journal of Integer Sequences*, 2001, vol. 4, p. 1-11.
112. Lai, D.; Chen, G. Dynamical systems identification from time time-series data: a Hankel matrix approach, *Mathematical and Computer Modelling*, 1996, vol. 24, p. 1-10.
113. Navickas, Z.; Bikulciene, L. Expressions of solutions of ordinary differential equations by standard functions, *Mathematical Modeling and Analysis*, 2006, vol. 11, p. 399-412.
114. Navickas, Z.; Ragulskis, M. How far one can go with the Exp-function method? *Applied Mathematics and Computation*, 2009, vol. 211, p. 522-530.
115. Navickas, Z.; Bikulciene, L.; Ragulskis, M. Generalization of Exp-function and other standard function methods, *Applied Mathematics and Computation*, 2010, vol. 216, p. 2380-2393.
116. Navickas, Z.; Ragulskis, M.; Bikulciene, L. Be careful with the Exp-function method – Additional remarks, *Communications in Nonlinear Science and Numerical Simulations*, 2010, vol. 15, p. 3874-3886.
117. Youness, A. *et. al.* Adapting genetic algorithms for combinatorial optimization problems in dynamic environments. In *Kosinski, W. Advances in Evolutionary Algorithms, I-Tech Education and Publishing*, Vienna, Austria, 2008, p. 207-230.
118. Sun, Z. L. *et. al.* Optimal selection of time lags for TDSEP based on genetic algorithm, *Neurocomputing*, 2006, vol. 69, p. 884-887.
119. Ferreira, T. A. E.; Vasconcelos, G. C.; Adeodato, P. J. L. A new intelligent system methodology for time series forecasting with artificial neural networks, *Neural Processing Letters*, 2008, vol. 28, p. 113-129.
120. Araujo, R. de A.; Ferreira, T. A. E. An intelligent hybrid morphological-rank-linear method for financial time series predictio, *Neurocomputing*, 2009, vol. 72, p. 2507-2524.
121. Liu, B. *et. al.* Permutation genetic algorithm for stacking sequence design of composite laminates, *Computer Methods in Applied Mechanics and Engineering*, 2000, vol.186, p. 357-372.

122. Goldberg, D. E.; *Genetic algorithms in search, optimization and machine learning*. Addison Wesley, Reading, MA, 1989. p. 412, ISBN: 0201157675.
123. Cheng, C.-H.; Chen, T. -L.; Wei, L.-Y. A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting, *Information Science*, 2010, vol. 180, p. 1610-1629.
124. Shimodaira, H. A new genetic algorithm using large mutation rates and population-elitist selection (GALME). In *Proceedings of 8th Conference on Tools with Artificial Intelligence*, 1996, pp. 25-32.
125. Eberhart, R. C.; Kennedy, J. A new optimizer using particle swarm theory. In *Proceedings of Sixth International Symposium on Micro Machine and Human Science*, Nagoya, Japan, IEEE Service Center, Piscataway, NJ, 1995, p. 39-43.
126. Lin, S. W., *et al.* Particle swarm optimization for parametre determination and feature selection of support vector machines, *Expert Systems with Applications*, 2008, vol. 35, p. 1817-1824.
127. Zhao, L.; Yang, Y. PSO-based single multiplicative neuron model for time series production, *Expert Systems with Applications*, 2009, vol. 36, p. 2805-2812.
128. Coelcho, L. S. Gaussian Quantum-behaved particle swarm optimization approaches for constrained engineering design problems, *Expert Systems with Applications*, 2010, vol. 37, p. 1676-1683.
129. Eberhart, R. C.; Kennedy, J. Particle swarm optimization: developments, applications and resources. In *Proceedings of IEEE Congress on Evolutionary Computation*, Seoul, Korea, IEEE Service Center, Piscataway, NJ, 2000, p. 81-86.
130. Shi, Y. H.; Eberhart, R. C. Parameter selection in particle swarm optimization. In *Proceedings of Seventh Annual Conference on Evolutionary Programming*, San Diego, CA, Springer-Verlag, New York, 1998, p. 591-600.
131. Clerc, M. The swarm and the queen: towards a deterministic and adaptive particle swarm optimization. In *Proceedings of IEEE Congress on Evolutionary Computation*, Washington, DC, IEEE Service Center, Piscataway, NJ, 1999, p. 1951-1957.
132. Trelea, I. C. The particle swarm optimization algorithm: convergence analysis and parameter selection, *Information Processing Letters*, 2003, vol. 85, no. 6, p. 317-325.
133. Kennedy, J.; Eberhart, R. C.; Shi, Y. H. *Swarm Intelligence*, Morgan Kaufman, San Francisco, CA 2001.
134. Kirkpatrick, S.; Gelatt C.D. Jr.; Vecchi, M. P. Optimization by simulated annealing, *Science*, 1983, vol. 220, p. 671-680.
135. Haupt, R.; Haupt, S.E. *Practical genetic algorithms*, Second Edition, John Wiley & Sons, Inc., Canada, 2004, 253 p. ISBN: 0-471-45565-2.
136. Behnamian, J.; Fatemi Ghomi, S. M. T. Developement of PSO-SA hybrid metaheuristic for a new comprehensive regression model to time-series forecasting, *Expert Systems with Applications*, 2010, vol. 37, p. 974-984.
137. Tian, P.; Ma, J.; Zhang, D.-M. Application of the simulated annealing algorithm to the combinatorial optimization problem with permutation property: An investigation of generation mechanizm, *European Journal of Operational Research*, 1999, vol. 118, p. 81-94.
138. Ben-Ameue, W. Computing the initial temperature of simulated annealing, *Computational Optimization and Application*, 2004, vol. 29, p. 369-385.
139. Schreiber, T. Interdisciplinary application of nonlinear time series methods, *Physics Reports – Review Section of Physics Letters*, 1999, vol. 308, p. 2-64.
140. Manabe, Y.; Chakrabortya, B. A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network, *Neurocomputing*, 2007, vol. 70, p. 1360-1371.
141. Cao, L. Practical method for determining the minimum embedding dimensijon of scalar time series, *Physica D: Nonlinear Phenomena*, 1997, vol. 110, p. 43-50.

142. Hyndman, R. J. (n.d.) *Time Series Data Library* [interaktyvus] [žiūrėta 2011-03-27]. Prieiga per internetą: <http://robjhyndman.com/TSDL>.
143. Gu, H.; Wang, H. Fuzzy prediction of chaotic time series based on singular value decomposition, *Applied Mathematics and Computation*, 2007, vol.185, p. 1171-1185.
144. Harpharm, C.; Dawson, C. W. The effect of different basis functions on a radial basis function network for time series prediction: A comparative study, *Neurocomputing*, 2006, vol. 69, p. 2161-2170.
145. Zunxiong, L.; Jianhui, L. Chaotic time series multi-step direct prediction with partial least squares regression, *Journal of Systems Engineering and Electronics*, 2007, vol. 18, p. 611-615.
146. MathWorks MATLAB [interaktyvus] [žiūrėta 2011-03-27]. Prieiga per internetą: <http://www.mathworks.com/products/fuzzylogic/demos.html?file=/products/demos/shipping/fuzzy/mgtsdemo.html>
147. Rössler, O. E. An Equation for Continuous Chaos, *Physics Letters A*, 1976 vol. 57, p. 397–398.
148. Strogatz, S. H. *Nonlinear Dynamics and Chaos*, Perseus publishing, Massachusetts, 1994, 498 p., ISBN 0-201-54344-3.
149. Gedeon, T.; Holzer, M.; Pernarowski, M. Attractor reconstruction from interspike intervals is incomplete, *Physica D: Nonlinear Phenomena*, 2003, vol. 178, p. 149-172.
150. Lian, J.; Shuai, J.; Durand, D. M. Control of phase synchronization of neuronal activity in the rat hippocampus, *Neural Engineering*, 2004, vol. 1, p. 46-54.
151. Holland, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control, and Artificial Intelligence*, The MIT Press, 1992, 228 p., 978-0-262-58111-0.
152. Whitley, D. genetic algorithm tutorial, *Statistics and Computing*, 1994, vol. 4, p. 65-85.
153. Du, H.; Zhang, N. Time series prediction using evolving radial basis function networks with new encoding scheme, *Neurocomputing*, 2008, vol. 71, p. 1388-1400.
154. Newcastle Engineering Design Centr, Merz Court, Newcastle University. [interaktyvus] [žiūrėta 2011-05-10]. Prieiga per internetą: <http://www.edc.ncl.ac.uk/highlight/rhjanuary2007g01.php>.
155. Keogh, E.; *et. al.* Finding the most unusual time series subsequence: algorithms and applications, *Knowledge and Information Systems*, 2006, vol. 11, p. 1-27.
156. Ragulskis, M.; Navickas, Z. The rank of a sequence as an indicator of chaos in discrete nonlinear dynamical systems, *Common Nonlinear SCi Numer Simulat*, 2011, vol. 16, p. 2894-2906.
157. Satpathy, H. P.; Liew, A. C. A real-time short-term peak and average load forecasting system using a self-organasing fuzzy neural network, *Engineering Applications of Artificial Intelligence*, 1998, vol. 11, p. 307-316.

## **MOKSLINIŲ PUBLIKACIJŲ DISERTACIJOS TEMA SĄRAŠAS**

### **STRAIPSNAI**

#### **Mokslinės informacijos instituto (ISI) pagrindinio sąrašo leidiniuose su citavimo indeksais**

1. Ragulskis M.; Lukoševičiūtė K. Non-Uniform Attractor Embedding for Time Series Forecasting by Fuzzy Inference System // Neurocomputing, Amsterdam: Elsevier Science. ISSN 0925-2312. 2009, vol. 72, Issues 10-12, p. 2618-2626. [ISI Web of Science].
2. Lukoševičiūtė K.; Ragulskis M. Evolutionary Algorithms for the Selection of Time Lags for Time Series Forecasting by Fuzzy Inference Systems // Neurocomputing, Amsterdam: Elsevier Science. ISSN 0925-2312. 2010, vol. 73, Issues 10-12, p. 2077-2088. [ISI Web of Science].
3. Ragulskis M.; Lukoševičiūtė K.; Navickas Z.; Palivonaitė R. Short-term time series forecasting based on the identification of skeleton algebraic sequences, // Neurocomputing, Amsterdam: Elsevier Science. ISSN 0925-2312. 2011, vol. 74, Issues 10-12, p. 2077-2088. [ISI Web of Science].

#### **Kituose Mokslinės informacijos instituto (ISI) duomenų bazėse referuojamuose leidiniuose [Proceedings ir kt.]**

1. J. Ragulskiene, K. Lukoseviciute, M. Ragulskis. Near-Optimal Embedding by Genetic Algorithms for Time Series Prediction. International Conference Numerical Analysis and Applied Mathematics, Rethymno, Greece 2009. American Institute of Physics Proceedings, Vol. 1168, p.148-151. [ISI Proceedings].
2. M. Ragulskis, A. Fedaravičius, K. Lukoševičiūtė, Cryptographic applications based on time averaged geometric Moire // Proceedings of the 2009 SEM Annual Conference and Exposition on Experimental and Applied Mechanics [elektroninis išteklius: June 1-4, 2009, Albuquerque, New Mexico, USA. ISBN 9789896740009 p. 63-69.

#### **Kituose recenzuojamuose mokslo leidiniuose**

1. Ragulskis, Minvydas; Lukoševičiūtė, Kristina; Ragulskienė, Jūratė; Nonlinear embedding for chaotic time series prediction by fuzzy inference systems // The Improvement of the Quality, Reliability and Long Usage of Technical Systems and Technological Processes, December 12-19, 2010, Sharm el Sheikh, Egypt : IX International Conference. ISBN 978966 p. 74-76.

## **Konferencijų pranešimų medžiagoje**

1. Construction of Cryptographic Hash Functions Based on Time Average Chaotic Map/ M.Ragulskis, Z. Navickas, L. Saunoriene, K. Lukoševičiūtė// Second International Interdisciplinary Chaos Symposium on CHAOS and COMPLEXSYSTEMS, CCS2008 List of Abstracts, Istanbul, Turkey, May 07-10, 2008, 30-31p.



## PRIEDAS

**1 lentelė.** ECG signalui nustatyto ARIMA(4,1,3) modelio adekvatumo tikrinimas:

$$x_t = 1,2995314x_{t-1} - 0,5899878x_{t-2} - 0,3178376x_{t-3} + 0,303846x_{t-4} + \\ + e_t - 0,4257694e_{t-1} + 0,3421525e_{t-2} + 0,2142127e_{t-3};$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	0,002	0,032	0,006	0,940
2	0,001	0,032	0,006	0,997
3	-0,011	0,032	0,130	0,988
4	0,009	0,032	0,218	0,994
5	0,008	0,032	0,284	0,998
6	-0,023	0,032	0,839	0,991
7	0,029	0,031	3,274	0,859
8	-0,028	0,031	4,049	0,853
9	0,007	0,031	4,101	0,905
10	0,030	0,031	6,021	0,814
11	-0,031	0,031	10,583	0,479
12	-0,024	0,031	11,724	0,468
13	-0,030	0,031	13,330	0,423
14	0,013	0,031	13,512	0,487
15	0,001	0,031	13,512	0,563
16	-0,026	0,031	16,700	0,505

**2 lentelė.** S&P500 laiko eilutei parinkto modelio ARIMA(1,0,1) adekvatumo tikrinimas:

$$x_t = -0,8677643x_{t-1} + e_t + 0,90023942e_{t-1} + 0,00651048;$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,017	0,047	0,134	0,714
2	0,005	0,047	0,147	0,929
3	0,033	0,047	0,654	0,884
4	0,009	0,047	4,805	0,308
5	0,005	0,047	6,004	0,306
6	-0,026	0,047	7,998	0,238
7	-0,033	0,047	8,498	0,291
8	0,004	0,047	8,504	0,386
9	0,035	0,047	9,062	0,432
10	-0,039	0,047	12,746	0,238
11	-0,002	0,046	12,748	0,310
12	0,043	0,046	15,927	0,195
13	-0,029	0,046	17,554	0,175
14	-0,042	0,046	19,950	0,132
15	0,020	0,046	20,139	0,167
16	0,026	0,046	20,508	0,249

**3 lentelė.** Rossler eilutei parinkto modelio ARIMA(2,0,3) adekvatumo tikrinimas:

$$x_t = 1,9241144x_{t-1} - 0,9833252x_{t-2} + u_t - 0,1924666e_{t-1} - 0,6412270e_{t-2} - 0,1650730e_{t-3} + 0,1274409$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,026	0,032	7,469	0,053
2	0,007	0,032	7,513	0,058
3	-0,014	0,032	7,705	0,063
4	-0,020	0,032	8,095	0,088
5	-0,005	0,032	8,119	0,150
6	-0,007	0,031	8,173	0,226
7	0,000	0,031	8,173	0,318
8	0,002	0,031	8,177	0,416
9	0,006	0,031	8,211	0,513
10	0,007	0,031	8,264	0,603
11	0,008	0,031	8,328	0,684
12	0,008	0,031	8,386	0,754
13	0,006	0,031	8,427	0,815
14	0,004	0,031	8,441	0,865
15	0,000	0,031	8,441	0,905
16	-0,004	0,031	8,454	0,934

**4 lentelė.** Testinės sekos su pridėtu triukšmo lygiu [-0,15; 0,15] ARIMA(4,1,3) modelio adekvatumo tikrinimas:

$$x_t = -1,9321306x_{t-1} - 1,8566398x_{t-2} - 1,2510344x_{t-3} - 0,6218348x_{t-4} + e_t + 0,7559998e_{t-1} + 0,7892625e_{t-2} - 0,9292909e_{t-3}$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,006	0,057	0,012	0,912
2	0,010	0,057	0,045	0,978
3	0,037	0,057	0,429	0,512
4	-0,022	0,057	0,574	0,751
5	-0,055	0,057	1,109	0,775
6	0,057	0,057	2,159	0,340
7	-0,024	0,056	2,813	0,421
8	0,034	0,056	3,311	0,346
9	-0,050	0,056	4,469	0,346
10	-0,050	0,056	6,220	0,285
11	-0,046	0,056	6,743	0,326
12	-0,054	0,056	7,178	0,327
13	0,048	0,056	9,692	0,138
14	0,037	0,056	10,502	0,075
15	-0,041	0,056	14,502	0,326
16	0,042	0,056	19,340	0,412

**5 lentelė.** Testinės sekos su pridėtu triukšmo lygiu  $[-0,5; 0,5]$  ARIMA(3,0,1) modelio adekvatumo tikrinimas:

$$x_t = -0,30103513x_{t-1} - 0,07514834x_{t-2} + 0,2446108x_{t-3} + e_t - 0,99781799e_{t-1}$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,011	0,045	0,065	0,806
2	0,000	0,045	0,065	0,970
3	0,029	0,045	0,482	0,923
4	0,038	0,044	1,563	0,725
5	0,032	0,044	1,912	0,684
6	-0,042	0,044	2,173	0,704
7	0,036	0,044	2,678	0,749
8	0,024	0,044	3,309	0,469
9	-0,027	0,044	3,996	0,361
10	-0,021	0,044	5,921	0,302
11	-0,043	0,044	6,843	0,285
12	-0,044	0,044	7,749	0,197
13	-0,029	0,044	9,824	0,148
14	0,038	0,044	11,372	0,177
15	-0,041	0,044	15,439	0,253
16	-0,043	0,044	18,103	0,336

**6 lentelė.** Testinės sekos su pridėtu triukšmo lygiu  $[-1; 1]$  ARIMA(1,0,3) modelio adekvatumo tikrinimas:

$$x_t = -0,96593131x_{t-1} + e_t + 0,82672497e_{t-1} - 0,19147574e_{t-2} - 0,01828734e_{t-3} + 0,50751892$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	0,012	0,045	0,068	0,795
2	-0,008	0,045	0,098	0,952
3	-0,030	0,044	0,560	0,905
4	0,026	0,044	0,899	0,925
5	-0,041	0,044	1,725	0,829
6	-0,037	0,044	2,800	0,731
7	0,029	0,044	2,925	0,403
8	-0,044	0,044	3,553	0,298
9	0,022	0,044	4,704	0,317
10	0,008	0,044	5,174	0,348
11	-0,022	0,044	6,952	0,374
12	0,028	0,044	7,988	0,299
13	-0,041	0,044	9,227	0,315
14	0,027	0,044	10,618	0,194
15	-0,039	0,044	13,397	0,249
16	-0,043	0,044	19,449	0,189

**7 lentelė.** Testinės sekos su pridėtu Gauso triukšmo lygiu (kai dispersija lygi 0,1) ARIMA(4,1,3) modelio adekvatumo tikrinimas:

$$x_t = -1,6672038x_{t-1} - 1,3861353x_{t-2} - 0,5585565x_{t-3} + 0,1477121x_{t-4} + \\ + 0,4082957x_{t-5} + e_t + 1,5832645e_{t-1} + 1,0540409e_{t-2} + 0,2318670e_{t-3} + 0,4401441$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,017	0,044	0,152	0,697
2	0,034	0,044	0,775	0,679
3	-0,022	0,044	1,030	0,794
4	-0,033	0,044	1,295	0,862
5	-0,009	0,044	1,942	0,585
6	0,023	0,043	2,228	0,694
7	-0,011	0,043	2,297	0,807
8	0,043	0,043	3,260	0,776
9	-0,018	0,043	3,430	0,843
10	-0,041	0,043	4,284	0,292
11	-0,004	0,043	4,292	0,368
12	-0,013	0,043	4,379	0,625
13	0,042	0,043	7,374	0,497
14	-0,016	0,043	7,518	0,583
15	0,040	0,043	8,941	0,177
16	0,037	0,043	11,510	0,319

**8 lentelė.** Testinės sekos su pridėtu triukšmo lygiu [-0,5; 0,5] ARIMA(3,0,1) modelio adekvatumo tikrinimas:

$$x_t = -0,45361172x_{t-1} + 0,53025315x_{t-2} + e_t + 0,37640176e_{t-1} - 0,59914754e_{t-2} + \\ + 0,2383523e_{t-3} + 0,527736$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	0,003	0,045	0,004	0,950
2	0,007	0,045	0,025	0,987
3	-0,014	0,044	0,118	0,990
4	-0,014	0,044	0,224	0,994
5	0,003	0,044	0,229	0,999
6	-0,008	0,044	0,261	1,000
7	0,011	0,044	6,401	0,494
8	-0,043	0,044	9,087	0,335
9	-0,001	0,044	9,087	0,429
10	0,039	0,044	12,610	0,246
11	0,023	0,044	12,886	0,301
12	0,035	0,044	13,505	0,333
13	-0,044	0,044	14,486	0,341
14	0,035	0,044	15,119	0,370
15	-0,043	0,044	18,870	0,220
16	0,040	0,044	19,709	0,234

**9 lentelė.** Testinės sekos su pridėtu Gauso triukšmo lygiu (kai dispersija lygi 1) ARIMA(1,1,2) modelio adekvatumo tikrinimas:

$$x_t = -0,76059445x_{t-1} + e_t - 0,17645255e_{t-1} - 0,79662183e_{t-2} - 0,00058137$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	0,004	0,045	0,007	0,932
2	-0,001	0,045	0,008	0,996
3	-0,032	0,045	0,520	0,914
4	-0,021	0,044	0,737	0,947
5	-0,012	0,044	0,815	0,976
6	-0,015	0,044	0,934	0,988
7	0,041	0,044	2,814	0,902
8	-0,004	0,044	2,821	0,945
9	-0,015	0,044	2,938	0,967
10	0,034	0,044	3,533	0,966
11	0,044	0,044	4,640	0,947
12	0,038	0,044	6,994	0,858
13	-0,035	0,044	7,622	0,867
14	0,038	0,044	10,720	0,708
15	0,023	0,044	10,988	0,753
16	0,002	0,044	10,991	0,810

**10 lentelė.** Andersono sausumos termometro temperatūros laiko eilutei parinkto modelio ARIMA(3,0,3) adekvatumo tikrinimas:

$$x_t = 0,70616477x_{t-1} - 0,99721786x_{t-2} + 0,70929x_{t-3} + e_t - 0,14935523e_{t-1} + 0,98610752e_{t-2} - 0,14330598e_{t-3} + 0,59159925$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,037	0,057	7,469	0,053
2	-0,022	0,057	7,513	0,058
3	0,094	0,057	7,705	0,063
4	0,061	0,057	8,095	0,088
5	-0,075	0,057	8,119	0,150
6	0,105	0,057	8,173	0,226
7	-0,124	0,056	8,173	0,318
8	0,124	0,056	8,177	0,416
9	-0,070	0,056	8,211	0,513
10	-0,090	0,056	8,264	0,603
11	-0,106	0,056	8,328	0,684
12	-0,054	0,056	8,386	0,754
13	0,048	0,056	8,427	0,815
14	0,077	0,056	8,441	0,865
15	-0,041	0,056	31,135	0,905
16	0,042	0,056	31,171	0,011

**11 lentelė.** Australijos bazinės mėnesinės geležies gamybos (tūkstančiais tonų) laiko eilutei parinkto modelio ARIMA(5,1,4) adekvatumo tikrinimas:

$$x_t = -0,4953726x_{t-1} + 0,7470315x_{t-2} + 0,5439009x_{t-3} - 0,4153680x_{t-4} - \\ -0,0140179x_{t-5} + e_t + 0,2405519e_{t-1} - 1,0448383e_{t-2} - \\ -0,5306933e_{t-3} - 0,5512426e_{t-4}$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,009	0,066	0,018	0,892
2	0,035	0,065	0,304	0,859
3	0,033	0,065	0,560	0,905
4	0,056	0,065	1,308	0,860
5	-0,018	0,065	1,387	0,926
6	0,006	0,065	1,397	0,966
7	-0,064	0,065	4,101	0,768
8	0,021	0,065	4,210	0,838
9	-0,056	0,064	6,943	0,643
10	0,034	0,064	7,229	0,704
11	-0,003	0,064	7,231	0,780
12	0,057	0,064	10,041	0,612
13	-0,037	0,064	10,375	0,663
14	0,059	0,064	12,331	0,580
15	-0,044	0,063	12,815	0,617
16	0,058	0,063	14,764	0,542

**12 lentelė.** Kintančio rango algebrinės sekos indentifikuoto ARIMA(3,1,2) modelio adekvatumo tikrinimas:

$$x_t = -1,1225587x_{t-1} - 0,2634742x_{t-2} + 0,1048423x_{t-3} + e_t - 0,271095e_{t-1} - \\ -0,2515050e_{t-2}$$

Laiko vėlinimas	Autokoreliacija	Paklaidų standartas	Box-Ljung statistika	Tikimybė
1	-0,004	0,072	0,004	0,953
2	-0,013	0,072	0,036	0,982
3	-0,031	0,072	0,225	0,973
4	0,047	0,071	0,666	0,955
5	0,060	0,071	1,369	0,928
6	-0,004	0,071	1,372	0,968
7	0,053	0,071	1,923	0,964
8	0,050	0,071	2,424	0,965
9	0,034	0,070	2,657	0,976
10	-0,089	0,070	4,263	0,935
11	0,146	0,070	8,635	0,656
12	0,059	0,070	9,337	0,674
13	0,012	0,070	9,367	0,745
14	-0,071	0,069	10,416	0,731
15	0,048	0,069	10,897	0,760
16	0,048	0,069	11,381	0,785