

DOI numeris <https://doi.org/10.15388/vu.thesis.567>

<https://orcid.org/0000-0002-9611-5421>

VILNIAUS UNIVERSITETAS  
GAMTOS TYRIMŲ CENTRAS  
KLAIPĖDOS UNIVERSITETAS

Dalia Grendaitė

# Chlorofilo $\alpha$ koncentracijos ežeruose nustatymas, remiantis Sentinel-2 palydovų duomenimis

**DAKTARO DISERTACIJA**

Gamtos mokslai  
Fizinė geografija (N 006)

VILNIUS 2023

Disertacija rengta 2018–2023 metais Vilniaus universitete.

Mokslinius tyrimus rėmė Vilniaus universiteto mokslo skatinimo fondas (nr. MSF-JM-7/2021).

**Mokslinis vadovas** – prof. dr. Edvinas Stonevičius (Vilniaus universitetas, gamtos mokslai, fizinė geografija – N 006).

Gynimo taryba:

**Pirmininkas** – prof. dr. Donatas Pupienis (Vilniaus universitetas, gamtos mokslai, fizinė geografija – N006).

**Nariai:**

doc. dr. Krista Alikas (Tartu universitetas, technologijos mokslai, aplinkos inžinerija – T004),

doc. dr. Laurynas Jukna (Vilniaus universitetas, gamtos mokslai, fizinė geografija – N006),

akad. dr. Jūratė Kriaučiūnienė (Lietuvos energetikos institutas, technologijos mokslai, aplinkos inžinerija – T004),

dr. Rita Linkevičienė (Gamtos tyrimų centras, gamtos mokslai, fizinė geografija – N 006).

Disertacija ginama viešame Gynimo tarybos posėdyje 2023 m. gruodžio mėn. 21 d. 13 val. Vilniaus universiteto Chemijos ir geomokslų fakulteto Geomokslų instituto 313 auditorijoje. Adresas: M. K. Čiurlionio 21, Vilnius, Lietuva), tel. +370 5 239 8292.

Disertaciją galima peržiūrėti Gamtos tyrimų centro, Klaipėdos universiteto, Vilniaus universiteto bibliotekose ir VU interneto svetainėje adresu:

*<https://www.vu.lt/naujienos/ivykiu-kalendorius>*

DOI numeris <https://doi.org/10.15388/vu.thesis.567>

<https://orcid.org/0000-0002-9611-5421>

VILNIUS UNIVERSITY  
NATURE RESEARCH CENTER  
KLAIPĖDA UNIVERSITY

Dalia Grendaitė

# Chlorophyll $\alpha$ Concentration Estimation in Lakes Using Sentinel-2 Satellite Data

**DOCTORAL DISSERTATION**

Nature Sciences,  
Physical Geography (N 006)

VILNIUS 2023

The dissertation was prepared between 2018 and 2023 at Vilnius University. The research was supported by Vilnius University Research Promotion Fund (Grant no. MSF-JM-7/2021).

**Academic supervisor –**

Professor Dr. Edvinas Stonevičius (Vilnius University, Nature Sciences, Physical Geography – N 006).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

**Chairman** – Prof. Dr. Donatas Pupienis (Vilnius University, Nature Sciences, Physical Geography – N 006).

**Members:**

Assoc. Prof. Dr. Krista Alikas (Tartu university, Technologies, Environmental Engineering – T 004),

Assoc. Prof. Dr. Laurynas Jukna (Vilnius University, Nature Sciences, Physical Geography – N 006),

Acad. Dr. Jūratė Kriaučiūnienė (Lithuanian Energy Institute, Technologies, Environmental Engineering – T 004),

Dr. Rita Linkevičienė (Nature Research Center, Nature Sciences, Physical Geography – N 006).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 13:00 on the 21st of December 2023 in Room 313 of Faculty of Chemistry and Geosciences, Institute of Geosciences (Vilnius University).

Address: M. K. Čiurlionio str. 21, Room No. 313, Vilnius, Lithuania

Tel. +370 5 239 8292.

The text of this dissertation can be accessed at the libraries of Klaipėda University, the Nature Research Centre and Vilnius University, as well as on the website of Vilnius University:

*[www.vu.lt/lt/naujienos/ivykiu-kalendorius](http://www.vu.lt/lt/naujienos/ivykiu-kalendorius)*

## SANTRUMPOS

<b>Santrumpa</b>	<b>Anglų kalba</b>	<b>Lietuvių kalba</b>
AAA / EPA	Environmental Protection Agency	Aplinkos Apsaugos Agentūra
Ada	Adaptive Boosting – AdaBoost	Mašininio mokymosi adaptyvaus didinimo metodas, AdaBoost
AK / AC	Atmospheric Correction	Atmosferos korekcija
ANN	Artificial Neural Network	Dirbtinis neuroninis tinklas
C2X	Case 2 Extreme	C2X metodas, skirtas 2 tipo vandenims
C2RCC	Case 2 Regional Coast Colour	C2RCC metodas, skirtas 2 tipo vandenims
Chla	Chlorophyll $\alpha$ concentration	Chlorofilo $\alpha$ koncentracija
CDOM	Coloured Dissolved Organic Matter	Spalvotoji ištirpusi organinė medžiaga
CZCS	Coastal Zone Colour Scanner	Pakrančių zonų spalvos skanavimo instrumentas
DSF	Dark spectrum fitting	Tamsaus spektro pritaikymas
EKA	European Space Agency	Europos kosmoso agentūra
GEE	Google Earth Engine	Google Žemės platforma, skirta palydovinių duomenų apdorojimui, parsisiuntimui
F1	F1 score	F1 įvertis
LR	Logistic Regression	Logistinė regresija
MAPE	Mean Absolute Percentage Error	Vidutinė absoliuti procentinė paklaida
MERIS	MEDium Resolution Imaging Spectrometer	Vidutinės raiškos spektrometras
MODIS	Moderate Resolution Imaging Spectroradiometer	Vidutinės skiriamosios gebos spektrometras
MSI	Multispectral Instrument (onboard Sentinel-2)	Daugiaspektris jutiklis (naudojamas Sentinel-2 palydovų)
NIR	Near-infrared	Artimoji infraraudoji spinduliuotė
OAS	Optically active substances	Optiškai aktyvios medžiagos
POLYMER	Polynomial-based algorithm applied to MERIS	Polinominis atmosferos korekcijos algoritmas MERIS jutikliui
XGB	eXtreme Gradient Boosting – XGBoost	Mašininio mokymosi ekstremalių gradientų didinimo XGBoost metodas

R <sup>2</sup>	Coefficient of determination	Determinacijos koeficientas
RF	Random Forest	Mašininio mokymosi atsitiktinio miško klasifikacijos modelis
RFC	Random Forest Classification	Mašininio mokymosi atsitiktinio miško klasifikatorius
RFR	Random Forest Regressor	Mašininio mokymosi atsitiktinio miško regresijos modelis
rho <sub>w</sub>	Water-leaving reflectance	Vandens atspindys
RMSE	Root Mean Squared Error	Vidutinė kvadratinė paklaida
sd	Standard Deviation	Standartinis nuokrypis
SeaWiFS	Sea-viewing Wide Field-of-view Sensor	Jūrą stebintis plataus stebėjimo kampo jutiklis
SM	Suspended matter	Suspenduota medžiaga
SVM	Support Vector Machine	Mašininio mokymosi atraminių vektorių mašinos modelis
SWIR	Shortwave infrared	Trumpųjų bangų infraraudojoji spinduliuotė
TOA	Top of Atmosphere (reflectance)	Viršutinės atmosferos atspindys (atspindėtosios spinduliuotės santykis su krentančia, apimantis paviršiaus ir atmosferos signalą)
TSI	Trophic State Index	Trofinės būklės indeksas
L1C	Level 1C satellite data	Pirmo lygio palydovinė nuotrauka
L2A	Level 2A satellite data	Antro lygio palydovinė nuotrauka
	Reflectance	atspindys
	Image-based atmospheric correction products	Nuotrauka paremti atmosferos korekcijos produktai

## TURINYS

SANTRUMPOS .....	5
ĮVADAS.....	9
1. LITERATŪROS APŽVALGA.....	18
1.1. Optinių palydovinių jutiklių panaudojimo galimybės vandens kokybės stebėjimams.....	18
1.2. Optinių palydovinių duomenų neapibrėžtumas ir atmosferos korekcija .....	21
1.3. Vandens tipai ir klasifikacija .....	23
1.4. Vandens parametrų nustatymas iš palydovinių duomenų .....	24
1.5. <i>In situ</i> optines vandens savybes apibūdinančių rodiklių matavimai Lietuvos ežeruose .....	26
2. DUOMENYS IR METODIKA.....	28
2.1. Tyrimo eiga .....	28
2.2. Tyrimų teritorija.....	33
2.3. Aplinkos apsaugos agentūros valstybinio monitoringo duomenys.....	33
2.4. Palydoviniai duomenys .....	34
2.5. Atmosferos korekcijos algoritmų neapibrėžtumo vertinimas chlorofilo $\alpha$ koncentracijai .....	39
2.6. Mašininio mokymosi algoritmai.....	44
2.7. Mašininio mokymosi modelių įvertinimui naudotos metrikos .....	47
2.7.1. Ežerų klasifikacijos modelių vertinimo metrikos .....	47
2.7.2. Chlorofilo $\alpha$ regresijos modelių vertinimo metrikos.....	49
3. ATMOSFEROS KOREKCIJOS ALGORITMŲ NEAPIBRĖŽTUMO VERTINIMAS CHLOROFILO $\alpha$ KONCENTRACIJOS NUSTATYMO IŠ PALYDOVINIŲ DUOMENŲ .....	50
3.1. Atmosferos korekcijos produktų poveikis vandens spektrams .....	50
3.2. Chlorofilo $\alpha$ koncentracijos neapibrėžtumas .....	56
3.3. Chlorofilo $\alpha$ koncentracijos algoritmų įvertinimas .....	59
4. BIOFIZINĖ EŽERŲ KLASIFIKACIJA .....	66

4.1 Biofizinių ežerų klasių išskyrimas .....	66
4.2. Biofizinės ežerų klasės pagal <i>in situ</i> matavimus.....	69
4.3. Biofizinės klasės modelių sudarymas.....	69
4.4. Svarbūs ežerų spektriniai požymiai biofizinės klasės modelių sudarymui .....	71
4.5. Biofizinės ežero klasės modelių validavimo ir testavimo rezultatai ..	76
4.5.1. Binarinė klasifikacija .....	76
4.5.2. Daugelio klasių klasifikacija .....	78
4.6. Biofizinė ežerų klasifikacija naudojant atsitiktinio miško modelį.....	80
4.7. Rekomendacijos biofizinių klasių vertinimui.....	86
5. CHLOROFILO A KONCENTRACIJOS MODELIAI IR JOS KAITOS ANALIZĖ .....	89
5.1. Chlorofilo $\alpha$ koncentracijos modeliai, paremti atsitiktinio miško regresijos algoritmu .....	89
5.2. Atsitiktinio miško regresijos ir dirbtiniais neuroniniais tinklais paremtų modelių rezultatų palyginimas.....	99
5.3. Chlorofilo $\alpha$ koncentracijos išskirčių pašalinimas .....	102
5.4. Potencialiai probleminių vandens telkinių išskyrimas .....	103
5.5. Chlorofilo $\alpha$ kaita ežeruose.....	106
REKOMENDACIJOS .....	115
IŠVADOS .....	118
LITERATŪROS SĄRAŠAS.....	119
CURRICULUM VITAE.....	131
PRIEDAI.....	132
SUMMARY .....	144
PUBLIKACIJŲ SĄRAŠAS .....	172



## ĮVADAS

Ežerų ir tvenkinių vanduo yra naudojamas rekreacijai, pramonei, geriamam vandeniui, taip pat ežerai ir tvenkiniai yra svarbi pasaulinio anglies ciklo dalis, tad ir klimato sistemos sudedamoji (Tranvik ir kt., 2009). Ežerai ir tvenkiniai yra buveinė daugumai gyvųjų organizmų (Schindler & Scheuerell, 2002). Bet kokiam vandens naudojimui svarbi vandens kokybė, o dėl žmogaus veiklos, kai yra pakeičiami vandens telkiniai ar, kai į vandens telkinius patenka teršalai, maistinės medžiagos, iš kurių pagrindiniai – azoto ir fosforo junginiai, spartėja vandens telkinių eutrofikacija, vandens kokybė prastėja, o vandens organizmams tenka prisitaikyti prie pakitusių sąlygų.

Klimato kaita kelia vis daugiau iššūkių ekosistemoms, taip pat kyla ir vandens trūkumo ir kokybės problemų. Ežeruose ir tvenkiniuose aukštesnė temperatūra (Woolway ir kt., 2017), ilgesni stratifikacijos laikotarpiai (Woolway ir kt., 2021), ledo dangos mažėjimas lemia tinkamesnes sąlygas dumblių ir melsvabakterių žydėjimui. Dumblių žydėjimas yra paplitęs visame pasaulyje (Coffer ir kt., 2021; Donis ir kt., 2021; Kashulin ir kt., 2021; Ndebele-Murisa ir kt., 2010), o jo daromas neigiamas poveikis gali sutrikdyti vandens naudojimą. Dėl intensyvaus žydėjimo sumažėja vandens skaidrumas, tad ir šviesos patekimas į gilesnius vandens sluoksnius, dėl ko keičiasi ekosistemos struktūra, nyksta dugno augalija. Taip pat mažėja deguonies kiekis, kas gali būti pražūtinga žuvims. Melsvabakterių sukeliamas žydėjimas gali būti pavojingas dėl melsvabakterių išskiriamų toksinų, žalingų ir žmogaus sveikatai (Huisman ir kt., 2018).

Lietuvoje yra apie 6000 ežerų ir tvenkinių, kurie sudaro apie 1,4 % šalies ploto (Kilkus ir Stonevičius, 2012). Siekiant saugaus vandens naudojimo ir geros vandens sistemų būklės pagal Europos Sąjungos vandens pagrindų direktyvą (Europos Parlamentas ir Taryba, 2014), vandens telkiniuose atliekami stebėjimai. Tradiciniais metodais renkama informacija leidžia tiksliai įvertinti vandens telkinių ir maudyklų vandens kokybės parametrus, tačiau reikalauja daug laiko ir išteklių. Lietuvoje sistemingą valstybinę paviršinių vandens telkinių stebėseną vykdo Aplinkos Apsaugos Agentūra (AAA), kuri pateikia ataskaitas apie vandens telkinių būklę ir Europos Sąjungai. Fizikinių-cheminių, hidromorfologinių ir biologinių rodiklių visuma leidžia įvertinti natūralius ir žmogaus veiklos sukeltus pokyčius vandens telkiniuose. Tačiau stebėjimų tinklas bei matavimų dažnumas veikiausiai leidžia įvertinti ilgalaikį poveikį baseino mastu, tačiau informacijos apie trumpalaikius pokyčius vandens telkiniuose (vandens žydėjimo pradžią, intensyvumą ir jo trukmę) trūksta.

Optiniai Žemės stebėjimo palydovai išsprendžia kai kuriuos tradicinių metodų trūkumus. Palydovais vandens telkinius stebėti galima žymiai dažniau, be to, palydovais gaunama informacija apima visą vandens telkinį ir visus, nutolusius, sunkiai pasiekiamus vandens telkinius. Tad dažniausiai viename taške ir 4–6 kartus per metus tradiciniais metodais atliekami stebėjimai gali būti papildomi kas keletą dienų gaunamais palydoviniais stebėjimais. Visgi palydovinių duomenų panaudojimas vandenių būklės vertinimui ir informacijos apie juos perdavimui pagal Europos Sąjungos vandens pagrindų direktyvą kol kas ribotas, kadangi palydoviniai duomenys kaip duomenų šaltinis nėra įtraukti į stebėjimo politikos sistemą bei trūksta vieningos standartizuotos duomenų panaudojimo metodikos (Papathanasopoulou ir kt., 2019).

Optiniai palydoviniai duomenys jau nuo 1970-ųjų naudojami pirmiausia vandens spalvos stebėjimui (Hovis ir kt., 1980). Šiais palydovais galima stebėti vandens rodiklius, kurie susiję su vandens priemaisių ir šviesos sąveika. Dėl vandenyje esančių medžiagų išsklaidymo ir sugerties keičiasi palydovu užfiksuojamas spektras, tad spektriniuose duomenyse galima užfiksuoti fitoplanktono biomasę apibūdinančio pigmento chlorofilo  $\alpha$  (chl $\alpha$ ), suspenduotos medžiagos (SM) ir ištirpusios spalvotosios organinės medžiagos (CDOM) signalą. Šios medžiagos, dar vadinamos optiškai aktyviomis medžiagomis, apibūdina vandens telkinio ekologinę būklę, tad jų kaitos nustatymas gali būti siejamas su vandens telkinių eutrofikacija. Taip pat palydovais galima stebėti šiuo metu vis dar nepakankamai stebimus aplinkos būklės parametrus – vandens žydėjimų dažnumą, pradžią, trukmę ir išplitimą. Taip pat įtraukiant palydovinius duomenis ekologinės būklės įvertinimo patikimumas pagerėtų dėl didesnio vandens telkinių įvairovės reprezentavimo ir vandens kokybės kaitos juose nustatymo (Papathanasopoulou ir kt., 2019).

Vandens parametrų nustatymas iš palydovinių duomenų yra sudėtingas dėl vandens savybių ir optiškai aktyvių medžiagų mišinio sudėtinio signalo. Grynas vanduo sugeria didžiąją dalį spinduliuotės, tad palydovu užfiksuojamas signalas yra silpnas (McClain, 2009). Taip pat palydovu užfiksuojamas signalas yra stipriai veikiamas atmosferos. Atmosferos poveikio pašalinimas yra vienas iš pagrindinių nuotolinio stebėjimo iššūkių, trukdančių kokybišką biogeocheminių parametrų nustatymą vandenyje (Pahlevan ir kt., 2021). Nuotolinis stebėjimas yra paprastesnis, kai vandens telkinyje dominuoja viena iš optiškai aktyvių medžiagų – pvz., vandenynuose (I tipo, *angl.* – *Case I*), vyrauja chlorofilas, tad jo koncentraciją įvertinti iš palydovinių duomenų galima gana tiksliai, naudojant mėlynų ir žalių bangų juostas (O'Reilly ir kt., 1998). Sudėtinguose (II tipo, *angl.* – *Case II*) pakrančių ir sausumos vandenyse minėti algoritmai veikia prastai ir tada yra

naudojami ilgesnių bangų, pvz., dviejų (Gurlin ir kt., 2011) arba trijų (Dall’Olmo ir kt., 2003) raudonos ir raudonosios ribos spektrinių juostų algoritmai.

Palydovinių jutiklių radiometrinė, spektrinė, erdvinė ir laiko raiška lemia jų panaudojimo galimybes vandens telkiniams. Dėl skirtingų palydovinių jutiklių spektrinių juostų konfigūracijos ir raiškos nėra galimybės naudoti vienos metodikos visų jutiklių duomenims apdoroti. Tad kiekvieno jutiklio duomenims interpretuoti tenka sudaryti naujus algoritmus. Taip pat algoritmus kurti tenka ir dėl menko jų pritaikomumo kitose teritorijose, nulemta skirtingų vandens telkinių optinių savybių. Copernicus programos Sentinel-2 palydovas A orbitoje yra nuo 2015 m, o palydovas B – nuo 2017 m. Dėl šių palydovų tinkamų vandens telkiniams stebėti jutiklių savybių, programoje numatyto misijos tęstinumo (ESA, 2021) bei regioninių vandens telkinių savybių yra poreikis sudaryti algoritmus vandens kokybės parametru įvertinimui pagal šių jutiklių duomenis.

Naudojant optinius palydovinius duomenis svarbu susipažinti su neapibrėžtumu, kylančiu dėl būdingų palydovinių duomenų savybių ir dėl jų apdorojimo metodikos pasirinkimo. Dėl minėtų vandens savybių ir atmosferos poveikio algoritmo pasirinkimas gaunamiems rezultatams gali būti labai reikšmingas. Šiame tyrime naudoti prieinami 357 ežerų ir tvenkinių valstybinio monitoringo duomenys ir nemokamai prieinami vidutinės raiškos Sentinel-2 palydovų duomenys. Tyrime sukurti metodai gali būti pritaikyti kituose panašaus tipo vidutinių platumų ežeruose ir tvenkiniuose. Sukurti metodai padidins palydovinių duomenų prieinamumą vandens stebėjimams ir galėtų būti naudojami vandens telkinius stebinčių institucijų darbe.

### **Mokslinio tyrimo motyvacija**

Lietuvoje apie 40 % ežerų kategorijos vandens telkinių neatitinka geros būklės kriterijų, o dėl didelės pasklidusios taršos visa Lietuvos teritorija yra paskelbta jautria teritorija pagal nitratų direktyvą (91/676/EEB) (Aplinkos Apsaugos Agentūra, 2018). Didelė tarša spartina eutrofikacijos procesus ir lemia vandens telkinių užžėlimą ir intensyvų dumblių žydėjimą. Dumblių žydėjimas yra per mažai stebimas, didelį poveikį tiek ekosistemoms, tiek žmonėms galintis turėti reiškiny. Panaudojant palydovinius duomenis ši reiškinį būtų galima stebėti geriau – nustatyti, kada jis prasideda, įvertinti jo intensyvumą, erdvinį paplitimą ir informuoti žmones apie esamą situaciją.

Visgi palydovinių duomenų panaudojimas yra gana sudėtingas – reikalauja žinių apie duomenų pobūdį ir jų apdorojimą, o siekiant gauti biofizinius rodiklius, kitose teritorijose kurti algoritmai gali netikti tiriamiems vandens telkiniams dėl skirtingų baseino ir vandens telkinio savybių, tad dažniausiai

reikalingas algoritmų priderinimas tiriamiems vandens telkiniams. Taip pat algoritmai, sukurti naudojant vieno palydovinio jutiklio duomenis, be papildomų tyrimų negali būti tiesiogiai pritaikyti kito palydovinio jutiklio duomenims dėl skirtingos jutiklių spektrinių juostų konfigūracijos. Dėl šių priežasčių Lietuvoje palydovinių duomenų panaudojimas vandens telkinių tyrimuose yra ribotas ir iki šiol dažniausiai apėmė didžiuosius vandens telkinius – Kuršių marias (Bresciani ir kt., 2014; Vaičiūtė ir kt., 2012, 2021) ir Kauno marias.

Sentinel-2 palydovų jutiklis pasižymi gera spektrine, erdvine ir laiko raiška, tad jais gaunami duomenys yra potencialiai tinkami Lietuvos ežerams stebėti. Palydovai orbitoje skrenda palyginti neseniai – nuo 2015 m., tad sukurtų biofizinių parametrų įvertinimo algoritmų, padedančių panaudoti šiuos kas kelias dienas gaunamus duomenis, dar vis stinga. Dažnai naudojami paprasti spektrinių juostų santykio algoritmai chlorofilo  $\alpha$  koncentracijai gauti, tačiau taip panaudojama tik dalis palydovais gaunamos informacijos. Mašininio mokymosi algoritmai vis dažniau pasitelkiami dideliame palydovinių duomenų kiekiui apdoroti ir naudingos informacijos iš jų išgavimui. Šiame disertaciniame darbe sukurti mašininio mokymosi atsitiktinio miško algoritmu paremti modeliai chlorofilo  $\alpha$  koncentracijai nustatyti.

**Darbo objektas** – ežerų vandens kokybės rodiklių nustatymo metodika, remiantis palydoviniais duomenimis.

**Darbo tikslas** – chlorofilo  $\alpha$  koncentracijos nustatymo iš palydovinių duomenų metodikos kūrimas.

#### **Darbo uždaviniai:**

1. Įvertinti atmosferos korekcijos algoritmų neapibrėžtumą chlorofilo  $\alpha$  nustatymui Lietuvos ežeruose.
2. Sukurti ežerų optinių tipų klasifikaciją paremtą palydoviniais duomenimis ir ją pritaikyti tiriamiems ežerams.
3. Sukurti ir validuoti skirtingoms ežerų biofizinėms klasėms pritaikytus chlorofilo  $\alpha$  koncentracijos vertinimo algoritmus.
4. Pritaikyti sukurtus algoritmus tiriamiems ežerams ir pagal chlorofilo  $\alpha$  koncentraciją išskirti potencialiai probleminius ežerus.
5. Atlikti chlorofilo  $\alpha$  koncentracijos kaitos ežeruose analizę.

### **Ginamieji teiginiai:**

1. Chlorofilo  $\alpha$  koncentracijos nustatymas iš Sentinel-2 MSI duomenų yra jautrus atmosferos korekcijos produkto ir chlorofilo  $\alpha$  algoritmo pasirinkimui.
2. Biofizinės ežerų klasifikacijos modelis padidina chlorofilo  $\alpha$  koncentracijos modelių tikslumą.
3. Naudojant palydovinius duomenis ir chlorofilo  $\alpha$  koncentracijos modelius, paremtus atsitiktinio miško regresijos algoritmu, galima nustatyti erdvinį chlorofilo  $\alpha$  koncentracijos pasiskirstymą ežeruose ir jos trumpalaikius pokyčius ežeruose, siejamus su vandens žydėjimais.
4. Sukurti algoritmai leidžia identifikuoti potencialiai probleminius ežerus naudojant chlorofilo  $\alpha$  koncentracijos vidutines ir maksimalias vertes.

### **Darbo mokslinis naujumas**

Įvairiose šalyse kuriami algoritmai, vandens parametrus gauti iš palydovinių duomenų, tačiau neretai jie labiausiai tinka tos šalies ar regiono vandens telkiniams dėl tiems vandenims būdingų savybių. Pasaulinėmis duomenų bazėmis paremti tyrimai vis tik apima mažą dalį pasaulio vandens telkinių, o Lietuvos vandens ežerų ir tvenkinių šiuose tyrimuose beveik nėra (pvz., duomenų bazėje Lakes-CCI yra tik Rėkyvos ežeras ir Kuršių marios, <https://climate.esa.int/en/projects/lakes/>). Šia disertacija siekiama užpildyti vandens telkinių stebėjimų Lietuvoje trūkumą remiantis palydoviniais duomenimis. Pateikti rezultatai prisideda prie didesnio duomenų panaudojimo pasiūlant ekologinę ežero būklę apibūdinančių rodiklių modelius. Darbo mokslinė vertė:

1. Įvertintas atmosferos korekcijos algoritmų neapibrėžtumas chlorofilo  $\alpha$  koncentracijos gavimui iš Sentinel-2 MSI duomenų.
2. Sukurtas ežero biofizinės klasės, remiantis išvestiniais spektriniais parametrais ir atsitiktinio miško (RFC) algoritmu, modelis.
3. Sukurti atsitiktinio miško regresijos algoritmu paremti modeliai biofizinėms ežerų klasėms chlorofilo  $\alpha$  koncentracijai gauti.
4. Išskirti potencialiai probleminiai vandens telkiniai pagal hidrologinį kriterijų – chlorofilo  $\alpha$  koncentraciją. Nustatyti trumpalaikiai chlorofilo  $\alpha$  koncentracijos pokyčiai, siejami su vandens žydėjimais.

Atmosferos korekcijos algoritmų vertinimas atliktas remiantis dideliu skaičiumi (136) Lietuvos ežerų, kurie pasižymi įvairia ekologine būkle. Šis

vertinimas aktualus ne tik taikant palydovinius duomenis Lietuvoje, tačiau ir platesnei ekspertų bendruomenei, dirbančiai su vidutinių platumų ežerais, kadangi dėl atmosferos poveikio pašalinimo iš palydovinių duomenų kylantis neapibrėžtumas daro įtaką gaunamiems rezultatams. Darbe pasiūlyta originali biofizinė klasifikacija gali būti naudojama ekologinės būklės įvertinimui net ir tose vietovėse, kur turimi tik *in situ* vandens kokybės parametrų ir palydoviniai duomenys. Sukurti modeliai, apmokyti remiantis Lietuvos ežerų duomenimis, tačiau gali būti pritaikyti be adaptavimo panašaus tipo (vidutinių platumų ežerams, kuriuose dažniausiai dominuojanti optiškai aktyvi medžiaga yra chlorofilas  $\alpha$ ) vandens telkiniams. Biofizinė klasifikacija pritaikyta 357 Lietuvos ežerams ir tvenkiniams, didesniems nei 50 ha.

Sukurta metodika chlorofilo  $\alpha$  koncentracijos vertinimui, paremta mašininio mokymosi algoritmais, suteikia galimybę efektyviai surinkti informaciją apie didelį kiekį ežerų nenuvykus prie jų. Šiuo darbu prisidedama prie ežerų vandens kokybės tyrimų naudojant nuotolinius metodus vidutinių platumų klimatinėmis sąlygomis.

### **Darbo aktualumas ir pritaikomumas**

Tradiciniais metodais stebėjimai atliekami gana retai ir neleidžia įvertinti trumpalaikių vandens kokybės pokyčių ežeruose. Kasmet AAA vykdomi valstybinio ežerų ir tvenkinio monitoringo stebėjimai apima 70–75 ežerus ir tvenkinius, tad apie 20 % visų išskirtų reikšmingų vandens telkinių. Per tyrimo laikotarpį (2015–2021 m. laikotarpį) buvo stebėti 314 vandens telkinių, vadinasi, per minėtą laikotarpį nėra jokios informacijos apie 12 % reikšmingų vandens telkinių. Iš 7 kasmet stebimų vandens telkinių, tik Kauno mariose stebimas melsvabakterių žydėjimas, o kituose 53 vandens telkiniuose, kuriuose buvo fiksuojama aukšta chl  $a$  koncentracija ( $> 50 \text{ mg/m}^3$ ) per šį laikotarpį matavimai atlikti vos 1–3 metus.

Pritaikius šiame darbe sukurtus modelius nuo 2015 m. galima papildyti duomenų sekas dažnesniais stebėjimais. Taip pat gaunama daugiau informacijos apie vandens žydėjimo trukmę, galima nustatyti jo kaitos tendencijas. Sukurti modeliai gali būti naudojami ekspertų darbe vertinant ežero būklę beveik realiu laiku bei įtraukiant palydovais gautą informaciją į ilgalaikius ežerų būklės analizės vertinimus. Remiantis palydoviniais stebėjimais galima optimizuoti monitoringo programą – parinkti stebėjimų laiką, dažnį ar matavimo vietą atsižvelgiant į erdvinį chlorofilo  $\alpha$  koncentracijos pasiskirstymą. Naudojant darbe aprašytus modelius galima nustatyti ežero biofizinę klasę ir chlorofilo  $\alpha$  koncentraciją kas 2–3 dienas. Ši informacija gali padėti stebėti trumpalaikius vandens kokybės pokyčius

ežeruose ir gali būti siejama su meteorologiniais, žemės dangos ir žemės ūkio veiklos duomenimis nustatant vandens kokybės pokyčių priežastis.

Palydovais gaunamos informacijos kiekis kasdien didėja, o jų apdorojimui vis dažniau naudojami mašininio mokymosi metodai. Šiame darbe pasiūlytais mašininio mokymosi metodais galima apdoroti didelius duomenų kiekius ir greitai gauti informaciją apie vandens parametrus, aprėpiančią visą vandens telkinį bei apimančią daug vandens telkinių.

### **Darbo struktūra**

Šio darbo pagrindiniai rezultatų skyriai yra šie:

3-iaame skyriuje aptariamas atmosferos korekcijos produktų neapibrėžtumas chlorofilo  $\alpha$  koncentracijos gavimui iš palydovinių Sentinel-2 MSI duomenų. Tai atskleidžiama per palydovinių 136 ežerų iš rytų Lietuvos spektrų bei skirtingų chlorofilo  $\alpha$  koncentracijos algoritmų analizę.

4-ame skyriuje aprašomos biofizinės ežerų klasės ir joms nustatyti sudaryti mašininio mokymosi algoritmai. Taip pat šiame skyriuje pateikiama Lietuvos ežerų klasifikacija pagal biofizinės klases ir biofizinės klases panaudojimas ežero būklės laike stebėjimui.

5-ame skyriuje aprašyti sudaryti chlorofilo  $\alpha$  koncentracijos modeliai biofizinėms ežerų klasėms, jų rezultatų palyginimas su atsitiktinio miško modeliu ir dirbtinių neuroninių tinklų modeliu, kurie buvo sukurti neskirstant ežerų į klases. Išskirti neprobleminiai ir potencialiai probleminiai Lietuvos ežerai ir nagrinėjama chlorofilo  $\alpha$  kaita juose.

### **Rezultatų apibavimas**

Tyrimų rezultatai pristatyti 3-iose tarptautinėse ir 5-iose nacionalinėse mokslinėse konferencijose.

Tarptautinės konferencijos:

1. **D. Grendaitė**, E. Stonevičius, J. Karosienė, K. Savadova, J. Kasperovičienė. „Seasonal Variation of Satellite Derived Chlorophyll-a Concentration in Eutrophic and Hypertrophic Lakes“. Living Planet Symposium 2019, 2019-05-16
2. **D. Grendaitė**, E. Stonevičius. „Atmospheric Correction Uncertainty in Mapping Chlorophyll-a Concentration in Lakes Using Sentinel-2 Data“ Baltic Sea Science Congress 2019, 2019-08-22
3. **D. Grendaitė**. „Lake Water Quality Assessment in Lithuania Using Remote Sensing And Machine Learning“. ESThub Vision Day 2021-05-20 (nuotoliniu būdu)

Nacionalinės konferencijos:

1. **D. Grendaitė**, E. Stonevičius, J. Karosienė, K. Savadova, J. Kasperovičienė. „Chlorophyll-a Concentration Retrieval From Sentinel-2 Data In Lakes In Lithuania“. Tarptautinė konferencija Ecobalt 2018. Aplinkos ministerija, Jakšto 4/9, Vilnius. 2018-10-26
2. **D. Grendaitė**, E. Stonevičius. „Lietuvos ežerų trofinės būklės įvertinimas naudojant palydovinius duomenis.“ 22-oji jaunųjų mokslininkų konferencija „Mokslas – Lietuvos ateitis, aplinkos inžinerija“, VGTU, 2019-03-20
3. **D. Grendaitė**, E. Stonevičius. „Sezoninė chlorofilo-a koncentracijos kaita eutrofiniuose ežeruose Lietuvoje naudojant palydovinius duomenis“. Konferencija „Klimato kaita Lietuvoje: globalūs ir nacionaliniai iššūkiai, stebėseną ir politikos gairės“, 2020-05-22
4. **D. Grendaitė**, E. Stonevičius. Vandens kokybės parametrų nustatymas Lietuvos ežeruose naudojant palydovinius duomenis ir mašininio mokymosi algoritmus. Nuotoliniai metodai gamtinių ekosistemų tyrimuose: privalumai ir iššūkiai. Gamtos tyrimų centras, Akademijos 2, Vilnius. 2021-10-12
5. **D. Grendaitė**, E. Stonevičius. Biofizinės klasės kaita Lietuvos ežeruose 2015–2021 m. Geographia juvena: VII-osios nacionalinės jaunųjų geografo konferencija. Vilniaus universiteto Geomokslų institutas, Čiurlionio 21, Vilnius. 2023-03-28

## Publikacijos

Publikacijos *Clarivate Analytics Web of Science* duomenų bazėje:

1. **Grendaitė D.** & Stonevičius E. Uncertainty of Atmospheric Correction Algorithms for Chlorophyll  $\alpha$  Concentration Retrieval in Lakes from Sentinel-2 Data. *Geocarto International* 37.23. 2022: 6867-6891. <https://doi.org/10.1080/10106049.2021.1958014>
2. **Grendaitė, D.** & Stonevičius, E. Machine Learning Algorithms for Biophysical Classification of Lithuanian Lakes Based on Remote Sensing Data. *Water* 2022, 14, 1732. <https://doi.org/10.3390/w14111732>

Kitos publikacijos *Clarivate Analytics Web of Science* duomenų bazėje:

1. Stonevičius, E., Uselis, G., **Grendaitė, D.** Ice Detection with Sentinel-1 SAR Backscatter Threshold in Long Sections of Temperate Climate Rivers. *Remote Sens.* 2022, 14, 1627. <https://doi.org/10.3390/rs14071627>



Kitos publikacijos:

1. Valskys V., Gulbinas Z., Stoyneva-Gärtner M., Uzunov B., Skorupskas R., Karosienė J., Kasperovičienė J., Rašomavičius V., Uogintas D., Audzijonytė A., Dainys J., Urbanavičius R., Urbanavičiūtė I., Vaičiūtė D., Bučas M., **Grendaitė D.**, Stonevičius E., Gedvilas A., Koreivienė J. Application of Remote Sensing in Environmental Studies: Advantages and Challenges. Annual of Sofia University “St. Kliment Ohridski”. Faculty of Biology. Book 2 – Botany, 106. 2022.

# 1. LITERATŪROS APŽVALGA

Šiame skyriuje apžvelgiami moksliniai tyrimai, nagrinėjantys palydovinių duomenų pritaikymo galimybes ir metodus vandens kokybės parametrų nustatymui iš palydovinių duomenų. Pirmoje dalyje pristatomi palydoviniai jutikliai, jų savybės, pritaikymo galimybės skirtingiems vandens telkiniams. Antroje dalyje aprašomi palydovinių duomenų neapibrėžtumo šaltiniai, koncentruojantis į atmosferos korekciją. Trečioje dalyje aptariami vandens tipai ir jų klasifikavimas pagal optinius duomenis. Ketvirtoje dalyje aptariami vandens parametrų nustatymo metodai.

## 1.1. Optinių palydovinių jutiklių panaudojimo galimybės vandens kokybės stebėjimams

Žemės stebėjimo palydovų era prasidėjo XX a. 7-ajame dešimtmetyje paleidus pirmąjį meteorologinį palydovą *TIROS-1* (Rao ir kt., 1990). Pirmasis vandens stebėjimams panaudotas palydovinis jutiklis buvo *Coastal Zone Colour Scanner (CZCS)* ant *Nimbus-7* palydovo ir teikė duomenis apie Žemę 1978–1986 m. (Hovis ir kt., 1980). Šiuo palydoviniu jutikliu gauti duomenys pirmąkart suteikė galimybę vandenynus stebėti globaliu mastu. Pirmasis stebimas parametras buvo vandens spalva, kuri buvo susieta su chla koncentracija, ir buvo įvertintas fitoplanktono žydėjimas.

Kitos vykusios misijos taip pat labai prisidėjusios prie vandenynų stebėjimo: *Sea-viewing Wide Field-of-view Sensor*, *SeaWiFS* (duomenis teikė nuo 1997 iki 2010 m.), *MODerate Imaging Spectroradiometer*, *MODIS* (Terra ir Aqua palydovai, nuo 1999) ir Europos kosmoso agentūros *Medium Resolution Imaging Spectrometer*, *MERIS* (McClain, 2009). Minėti jutikliai dėl savo žemesnės erdvinės raiškos (gardelės dydis nuo 300 m iki daugiau nei 1 km) labiau tinkami stebėti vandenynams ir dideliems vandens telkiniams (Philipson ir kt., 2016). Naudojant tokius duomenis nepavyktų stebėti ir erdvinio vandens parametrų pasiskirstymo mažesniuose ežeruose, kadangi ežerą apimtų vos kelios gardelės, kurioms didelę įtaką darytų sausumos paviršiai. Mažesniems vandens telkiniams, kokie yra ir Lietuvoje, labiau tinkami aukštesnės erdvinės raiškos palydovai, tokie kaip nacionalinės Amerikos kosmoso agentūros (*angl. – National Aeronautics and Space Administration, NASA*) Landsat serijos operatyvūs sausumos jutikliai (*angl. – Operational Land Imager, OLI*) (gardelės dydis 30 m) bei Europos Komisijos Copernicus programos Sentinel-2 palydovų daugiaspektris instrumentas (*angl. – MultiSpectral Instrument, MSI*) (gardelės dydis 10–20 m) (1.1 lentelė).

Naujesnių palydovinių jutiklių padidinta radiometrinė raiška (12-bit, pvz., Sentinel-2 MSI, Landsat 8 OLI ir 14-bit Landsat 9 OLI2) palyginus su ankstesnių jutiklių (pvz., CZCS, Landsat 5 TM, Landsat 7 ETM+) skiriamąja geba (8-bit), leidžia nustatyti mažus vandens parametrų skirtumus ir pokyčius vandenyje. Pagerintas signalo ir triukšmo santykis lemia didesnę duomenų tikslumą (Schott ir kt., 2016).

Viena iš svarbiausių palydovinių jutiklių charakteristikų – spektrinė raiška arba spektrinių juostų skaičius, jų vieta elektromagnetiniame spektre ir plotis. Šios charakteristikos tiesiogiai lemia, kokius objektus bus galima stebėti, pvz., Landsat palydovų jutikliai pasižymi palyginti plačiomis spektrinėmis juostomis, tad jais sunkiau užfiksuoti chla būdingus sugerties ir išsklaidymo spektrinius bruožus – absorbciją raudonų ir raudonosios ribos bangų dalyse (Kutser ir kt., 2016). Tam labiau tinka siauresnes spektrines juostas raudonosios ribos bangų ilgiuose turintis Sentinel-2 MSI (Drusch ir kt., 2012), dėl to šiuo jutikliu galima tiksliau užfiksuoti chla spektrines savybes (Kutser ir kt., 2016; Maier & Keller, 2018). Fitoplanktono grupėms išskirti irgi reikia daugiau siaurų spektrinių juostų, pvz., 620 nm centruotos spektrinės juostos, kuria galima įvertinti melsvabakterėms būdingo pigmento fikocianino absorbciją (Ogashawara, 2020). Šią spektrinę juostą turėjo MERIS jutiklis bei šiuo metu orbitoje esantys Sentinel-3 OLCI jutikliai, tačiau dėl 300 m gardelės dydžio šie jutikliai labiau tinka dideliems ežerams stebėti.

Minėti multispektriniai instrumentai spektrinėmis juostomis padengia tik nedidelę elektromagnetinio spektro dalį. Pastaraisiais metais daugėja hiperspektrinių misijų (pvz., *Hyperspectral Imager for the Coastal Ocean*, HICO, Italijos kosmoso agentūros *Precursore IperSpettrale della Missione Applicativa*, PRISMA (Loizzo ir kt., 2018)). Hiperspektriniai jutikliai turi kelis šimtus siaurų spektrinių juostų, tad jomis galima užfiksuoti spektrinius niuansus, kurių dauguma multispektrinių jutiklių neturi galimybės užfiksuoti. Vandens kokybės tyrimuose naudojant hiperspektrinius duomenis galima, pvz., nustatyti melsvabakterių biomą (O’Shea ir kt., 2021).

Dauguma Žemės stebėjimo palydovų skrenda poliarinėje orbitoje, 700–800 km atstumu nuo Žemės paviršiaus bei tą pačią teritoriją nufotografuoja kas 1–16 dienų. Regioniniam jūrų stebėjimui kai kurios šalys paleidžia palydovus, skrendančius geostacionarioje orbitoje ~36 000 km atstumu nuo Žemės paviršiaus, pvz., Pietų Korėjos palydovai GOCI ir GOCI II (1.1 lentelė). Šie palydovai pasižymi tuo, kad nuolat stebi tą pačią teritoriją ir nuotraukas užfiksuoja 8 kartus per dieną, tad jais galima stebėti vandens skaidrumo, chla pokyčius per dieną (Ding ir kt., 2022). Visgi geostacionarioje orbitoje skrendantys palydovai pasižymi prastesne erdvine raiška nei poliarinės orbitos palydovai, tad pastarieji yra naudojami dažniau.

**1.1 lentelė.** Optinių palydovinių jutiklių, naudojamų vandens tyrimuose, charakteristikos.

<b>Palydovas: jutiklis</b>	<b>Duomenų eilė, metai</b>	<b>Radio- metrinė raiška</b>	<b>Erdvinė raiška, m</b>	<b>Praskri- dimo dažnis</b>	<b>Spektrinių juostų skaičius ir diapazonas, nm</b>
Nimbus-7: CZCS*	1978–1986	8-bit	825	6 d.	6 juostos 433– 12500 nm
OrbView-2: SeaWiFS	1997–2011	10-bit	1100	1 d.	8 juostos 402– 885 nm
Aqua ir Terra: MODIS	Terra: 1999- Aqua: 2002-	12-bit	250, 500, 1000	1–2 d.	36 juostos 405–14385 nm
COMS*: GOCI* I	2010–2021	12-bit	500	1 h	8 juostos 400– 865 nm
GeoKompasat- 2B: GOCI II	2020–	12-bit	250	1 h	13 juostų 380– 900 nm
ENVISAT*: MERIS	2002–2012	12-bit	300, 1200	3 d.	15 juostų 412– 1050 nm ruože
Sentinel-3: OLCI*	A: 2016– B: 2018–	14-bit	300, 1200	1 d.	21 juosta 400– 1020 nm ruože
Landsat 5: TM*	1984–2012	8-bit	30, 120	16 d.	7 juostos, 450–12500 nm
Landsat 7: ETM+*	1999–2022	8-bit	15, 30, 100	16 d.	8 juostos 440– 12360 nm
Landsat 8 ir 9 OLI*	8: 2013– 9: 2021–	8: 12-bit 9: 14-bit	15, 30, 100	16 d., kartu – 8 d.	10 juostos 433–1390 nm ruože
Sentinel-2: MSI	A: 2015– B: 2017–	12-bit	10, 20, 60	2–5 d.	13 juostų 40– 2190 nm ruože

\* Naudojami trumpiniai: CZCS – Coastal Zone Colour Scanner, COMS – Communication, Ocean and Meteorological Satellite, GOCI – Geostationary Ocean Colour Imager, ENVISAT – Environmental Satellite, OLCI – Ocean and Land Colour Instrument, TM – Thematic Mapper, ETM+ – Enhanced Thematic Mapper, OLI – Operational Land Imager, MSI – MultiSpectral Instrument

Renkantis palydovinius duomenis pasirinktam tyrimo objektui svarbu atsižvelgti į minėtas charakteristikas, kadangi jos tiesiogiai lemia stebimo objekto vandens kokybės parametrų pasiskirstymo laike ir erdvėje užfiksavimo galimybes. Stebint mažesnius vandens telkinius, pvz., Lietuvos ežerus, pasirenkami poliarine orbita skrendantys palydovai, pasižymintys

geresne erdvine raiška, nufotografuojantys teritoriją kas 2–3 dienas bei pasižymintys pakankamai gera spektrine raiška (pvz., Sentinel-2 MSI).

## 1.2. Optinių palydovinių duomenų neapibrėžtumas ir atmosferos korekcija

Optinių palydovinių duomenų kokybė priklauso nuo daugelio faktorių, iš kurių pagrindiniai yra: paties palydovinio jutiklio stabilumas, dujų ir aerozolio dalelių pasiskirstymas atmosferoje (terpė tarp tiriamo objekto bei jutiklio), vėjas, bangos, vandens priemaišų vertikalus pasiskirstymas ir jų optinės savybės (IOCCG, 2019). Taip pat tiriamo objekto atspindžiui gali turėti įtakos objektą supanti aplinka, pvz., aukšti medžiai ar kalnai (topografinis efektas). Šie faktoriai bei silpnas signalas iš vandens paviršiaus lemia palydovinių duomenų neapibrėžtumą, kurį svarbu įvertinti.

Naudotojui pasiekiamiems palydoviniams duomenims yra pritaikytos kelios korekcijos: radiometrinė, geometrinė, atmosferos poveikio (pasirinktinai), topografinė (pasirinktinai). Pritaikant šias korekcijas yra pašalinamas aplinkos ir atmosferos poveikis bei duomenys pakoreguojami skirtingoms šviesos lauko sąlygoms. Tokie duomenys gali būti palyginami tarpusavyje, kadangi suteikia informacijos apie patį paviršių, šiuo atveju, apie vandens ir to kas yra jame optines savybes.

Nors visi paminėti faktoriai yra svarbūs, tačiau vandens kokybės tyrimuose naudojant palydovinius duomenis dažniausiai kyla klausimas, kurį atmosferos korekcijos (AK) – atmosferos poveikį palydovu užfiksuojamam signalui pašalinantį algoritmą – naudoti. Sentinel-2 MSI jutiklio duomenų korekcijai įvairiuose tyrimuose yra taikoma apie 10 įvairių AK algoritmų, tačiau nėra susitarimo, kuris algoritmas yra tinkamiausias. AK algoritmai, kaip ir visi modeliai, procesų aprašymą supaprastina, dėl to jų tikslumas labai priklauso nuo priimtų prielaidų, taip pat nuo duomenų rinkinių, kurie buvo naudojami algoritmų sudarymui, bei nuo pagalbinių duomenų tikslumo.

Yra išskiriamos dvi AK algoritmų grupės – nuotrauka paremtų (Chávez, 1996) bei paremtų spinduliuotės perdavimo modeliavimu (*angl. – radiative transfer modelling*) (Zhai ir kt., 2017). Nuotrauka paremti algoritmai priima prielaidą, kad sąlygos atmosferoje yra vienodos visoje nuotraukos aprėptyje bei reikalinga, kad nuotraukoje būtų tiek vandens, tiek sausumos paviršių (Warren ir kt., 2019). Kiti, pvz., optimizavimo metodai (Fan ir kt., 2017) yra pritaikomi kiekvienai vaizdo gardelei ir priklauso nuo mokymo duomenų aprėpties, kuriuos naudojant modelis buvo sukurtas (Brockmann ir kt., 2016).

Vieni algoritmai yra sukurti būtent vandens tyrimams, pvz., C2RCC (Case 2 Regional Coast Colour), Acolite, POLYMER ir iCOR, o kiti, pvz.,

Sen2Cor, buvo sukurti sausumos tyrimams, tačiau yra žymiai lengviau prieinami naudotojui. Šių AK algoritmų tikslumas priklauso nuo jų gebėjimo teisingai nustatyti aerolio daleles atmosferoje bei įvertinti aplinkos sąlygas. Atliktas AK palyginimo uždavinys (*angl. – The Atmospheric Correction Intercomparison eXercise, ACIX-Aqua*), paremtas Sentinel-2 ir Landsat 8 duomenimis atskleidė, kad dauguma AK algoritmų sunkiai tiksliai įvertina ir pašalina aerolio dalelių signalo indėlį į palydovu užfiksuotą signalą (Pahlevan ir kt., 2021).

Taip pat atmosferos korekcijos algoritmų tikslumas gali priklausyti nuo tiriamų vandens telkinių savybių. Pvz., daliai Estijos ežerų, kurie apibūdinami kaip tamsūs sugeriantys ežerai, C2RCC algoritmas veikia geriausiai. Taip pat buvo nustatyta, kad algoritmas neveikia taip gerai, kai duomenys yra konvertuojami į chla koncentraciją mažuose, siauruose ežeruose, kur yra stiprus topografiniai efektas (Ansper & Alikas, 2019) bei atspindys nuo vandens yra mažas. Topografiniai efektai nėra įtraukti į daugumą AK algoritmų ir taip pat gali lemti nemažas klaidas duomenyse (Paulino ir kt., 2022; Warren ir kt., 2019). Taip pat nustatyta, kad topografinis efektas yra stipresnis, kai atmosferoje gausu aerolio dalelių ir yra didesnis išsklaidymas atmosferoje (Minomura ir kt., 2001).

C2RCC algoritmas neveikia gerai chla absorbcijos raudonos spalvos bangos ilgiuose (665 nm) (Ansper & Alikas, 2019). Tačiau kito tyrimo metu C2RCC algoritmas pateikė realistiškus spektrus Baltijos jūrai ne žydėjimo metu. Visgi, kai vyksta melsvabakterių žydėjimas, C2RCC pateikia mažesnes atspindžio reikšmes (Toming ir kt., 2017). Nuotrauka paremti algoritmai, pvz., Sen2Cor ir iCOR, pateikia geresnius rezultatus vidaus vandens telkiniams nei priekrantės zonoms (Warren ir kt., 2019). Tad atmosferos korekcijos algoritmo pasirinkimas gali labai priklausyti nuo vandens charakteristikų. Tas pats algoritmas gali būti pakankamai geras vienomis sąlygomis, tačiau prastai veikti kitomis sąlygomis. Optinis vandens tipo nustatymas gali padėti sumažinti šiuos neapibrėžtumus (Pahlevan ir kt., 2021).

Atmosferos korekcijos algoritmo tikslumas yra įvertinamas pasitelkiant *in situ* radiometro duomenis, kurie laikomi suteikiantys tikrus vandens telkinių spektrus. Tačiau *in situ* spektrinių duomenų yra mažai, mažiau nei *in situ* vandens parametrų duomenų, kurie paprastai yra surenkami nacionalinių monitoringo programų metu. Dėl to atmosferos korekcijos algoritmai šiuo būdu gali būti įvertinami itin mažam skaičiui vandens telkinių, o esant poreikiui tirti didelį skaičių vandens telkinių, nėra galimybės atmosferos korekcijos algoritmus validuoti remiantis *in situ* spektriniais duomenimis.

### 1.3. Vandens tipai ir klasifikacija

Morel ir Prieur (1977) išskyrė du vandens tipus. I tipo vandenyse vyrauja fitoplanktonas, o autochtoninė detritinių ištirpusių bei kietųjų dalelių produkcija yra susijusi. II tipo vandenims būdingas išorinių kietųjų dalelių ir ištirpusios medžiagos patekimas į sistemą, su kuriomis fitoplanktono kiekis nekoreliuoja (Morel, 2001) ir fitoplanktono optinės savybės atsiskiria nuo tūrio optinių savybių (Spyrakos ir kt., 2018). Optinių savybių kompleksiskumas II tipo atveju apsunkina nuotolinį jų vandens kokybės nustatymą (Koponen ir kt., 2007).

Palydoviniai duomenys žymiai pagerino vandens telkinių erdvinį savybių stebėjimą, ypač fitoplanktono (per pagrindinio dumblių pigmento chlorofilo  $\alpha$  koncentraciją (Modabberi ir kt., 2020)), suspenduotų dalelių (Caballero ir kt., 2018; Saberioon ir kt., 2020; Sun ir kt., 2009) ir spalvotosios ištirpusios organinės medžiagos (Al-Kharusi ir kt., 2020). Optinės vandens telkinių savybės, nulemtos minėtų medžiagų, gali daryti neigiamą įtaką tiksliam chl $a$  koncentracijos nustatymui iš palydovinių duomenų, kai naudojami paprasti spektrinių juostų algoritmai (Kutser ir kt., 2016). Dėl šios priežasties vandens telkinių grupavimas pagal vyraujančias medžiagas ir sudarant algoritmus vandens telkinių grupėms leidžia gauti tikslesnius rezultatus (Soomets ir kt., 2020). Vandens telkinių grupavimas, klasterizavimas, klasifikavimas yra dažnai pasitaikantis būdas charakterizuoti skirtingus vandens telkinių tipus. Europos Sąjungoje ekologinė vandens telkinių būklė yra apibūdinama 5 klasėmis: labai bloga, bloga, vidutinė, gera ir labai gera. Būklė yra įvertinama Aplinkos apaugos agentūros atlikus kelis matavimus per metus, tačiau didžioji dalis vandens telkinių nėra stebimi kasmet.

Vandens telkinių grupavimui gali būti naudojami skirtingi metodai. Neprižiūrimo klasterizavimo metodas buvo pritaikytas sausumos vandens telkinių ir pakrančių vandenų sugrupavimui į optinius vandens tipus naudojant *in situ* hiperspektrinius duomenis ir buvo išskirta 13 skirtingų optinių vandens tipų (Spyrakos ir kt., 2018). Tačiau naudojant dabartinių ne hiperspektrinių palydovinių jutiklių duomenis gali būti sudėtinga išskirti tiek daug optinių vandens tipų. Paprastesnis metodas, kuriuo buvo išskirtos 5 klasės – skaidri, tarpinė, drumsta, labai drumsta ir ruda – paremtas k-vidurkių klasterizavimo metodu, buvo sukurtas Estijos mokslininkų (Reinart ir kt., 2003). Vandens telkinių tipų išskirimui buvo panaudotos optinės vandens telkinių savybės, pvz., silpimo dėl sklaidos (*angl. – diffuse attenuation*) koeficientas, išsklaidytas atspindys (*angl. – diffuse reflectance*), bei vandens kokybės parametrai – vandens skaidrumas, chl $a$  koncentracija, bendra suspenduota medžiaga ir spalvotosios ištirpusios organinės medžiagos. Šis 5 tipų metodas

buvo pritaikytas Estijos ežerams, Baltijos jūros pakrantėms, Vadeno jūrai ir Latvijos ežerams tirti (Soomets ir kt., 2020). Kitas optinių vandens tipų nustatymas buvo sudarytas Brazilijos vandenims, kur buvo išskirti aštuoni vandens tipai (da Silva ir kt., 2020). Vandens tipų išskyrimas itin priklauso nuo naudojamo duomenų rinkinio įvairovės, naudojamų požymių skaičiaus bei duomenų kiekio, kurie gali būti išskiriami kaip vienas tipas.

Vandens telkinių klasifikacija gali būti naudojama kaip duomenų šaltinis, apibūdinant ežero būklę (Soomets ir kt., 2020) arba pritaikyta vandens parametrų algoritmų kūrimui. Paprastus spektrinių juostų santykio algoritmus pastaruoju metu keičia sudėtingesni algoritmai, kurie gali sujungti kelių požymių informaciją (pvz., spektrinių duomenų išvestinius rodiklius) ir pateikti tikslesnius rezultatus. Klasifikavimui gali būti naudojami prižiūravimo mokymosi algoritmai, o vandens kokybės parametrų nustatymui – regresijos metodai. Chla koncentracijai išgauti sukurti atraminių vektorių metodu grindžiami regresijos algoritmai (Sun ir kt., 2009), dirbtiniai neuroniniai tinklai (Hafeez ir kt., 2019; Ioannou ir kt., 2013), Cubist (Saberioon ir kt., 2020) ir atsitiktinių miškų modeliai (Kim ir kt., 2014). Algoritmo pasirinkimas priklauso nuo duomenų pobūdžio ir taikymo srities. Kadangi universalus algoritmo nėra ir pirmenybė teikiama paprastesniam, o ne sudėtingam algoritmui, dažnai išbandomi keli algoritmai, siekiant parinkti konkrečiam taikymui geriausiai veikiančią algoritmą (Eskandari ir kt., 2022; Rezaei & Vadiati, 2020). Be to, gali būti naudojama automatinė modelio atranka pagrįsta sistema (Blix & Eltoft, 2018).

Daugeliu atvejų kuriant modelį naudojami *in situ* optiniai duomenys, tačiau, kaip anksčiau minėta, tokio tipo duomenys ne visada prieinami. Tačiau vis dėlto gerais rezultatais gali pasižymėti ir duomenimis pagrįsti (*angl.* – *data-driven*) algoritmai, kuriuose naudojami tik išmatuoti vandens parametrų duomenys ir palydoviniai duomenys (Saberioon ir kt., 2020).

#### 1.4. Vandens parametrų nustatymas iš palydovinių duomenų

Vandens parametrų nustatymas iš palydovinių duomenų yra paremtas jų fizikinėmis savybėmis – sugertimi (chl<sub>a</sub>, CDOM), atgaline sklaida (SM, chl<sub>a</sub>) ir fluorescencija (chl<sub>a</sub>). Bio-optiniai modeliai yra naudojami šviesos sąveikai su šiomis medžiagomis aprašyti per minėtus procesus (Morel, 2001). Biologinės chl<sub>a</sub> savybės lemia sąveiką su šviesa, tad iš to ir atsirado sąvoka vandenyno spalvos radiometrija (*angl.* – *ocean colour radiometry*), apibūdinanti mokslinę sritį, tiriančią vandenyno vandens spalvos nustatymą nuotoliniais metodais, kurios pirminis tikslas yra chl<sub>a</sub> koncentracijos nustatymas (Matthews, 2017). Kadangi dėl technologinio palydovinių jutiklių



progreso dabar stebimi ir maži sausumos vandens telkiniai, dažnai vartojama sąvoka – vandens spalvos radiometrija (*angl. – water colour radiometry*).

Chla koncentracija yra geras fitoplanktono biomasės indikatorius, randamas beveik visuose dumbliuose ir melsvabakterėse, tad tai yra dažnai į valstybinio monitoringo programas įtraukiamas rodiklis. Fitoplanktonas – mikroskopiniai, vienaląsčiai organizmai, plūduriuojantys vandenyje, žmogaus akimi matomi tik tada, kai jų yra didelis ląstelių kiekis. Esant didelei fitoplanktono biomasei pakinta vandens spektrinės savybės ir palydovais dažniausiai stebima žalia ar melsvai žalia vandens spalva. Palydovu fiksuojamas signalas priklauso nuo biomasės kiekio, fitoplanktono ląstelių dydžio bei formos, pigmentų, ląstelių struktūros ir atskirų ląstelių formavimosi į kolonijas ar siūlus. Chla sugeria mėlynos ir raudonos šviesos bangas. Sugerties maksimumai yra ties 440 nm ir 675 nm, o žalia šviesa atspindima (Matthews, 2017). Chla išsklaido šviesą 700–720 nm bangos ilgiuose, kas lemia padidėjusį atspindį (Schalles ir kt., 1998). Todėl jau seniai chla gavimui naudojamas atspindžio santykis 705 nm ir 670 nm bangos ilgiuose (Mittenzwey ir kt., 1992). Vandenynuose, kur chla yra vienintelė optiškai aktyvi medžiaga vandenyje (Gordon & Clark, 1980), algoritmai remiasi matomųjų bangų – žaliosios ir mėlynosios juostos duomenimis (O'Reilly ir kt., 1998), o pakrančių vandenyse ir gėlo vandens aplinkoje dėl kitų optiškai aktyvių medžiagų buvimo pirmenybė teikiama ilgesnėms bangoms (Yacobi ir kt., 1995).

Didelis spalvotas ištirpusių organinių medžiagų (CDOM) kiekis lemia mėlynos ir žalios bangos ilgių sugertį (Bricaud ir kt., 1981). Todėl daug CDOM turinčiuose vandenyse chla koncentracijai nustatyti naudojamos raudonosios ir artimosios infraraudonosios juostos (Gilerson ir kt., 2010; Kutser ir kt., 2016). Pastaruoju metu atlikta keletas tyrimų, kuriuose išbandyta daug algoritmų, skirtų sudėtingiems Baltijos jūros vandenims (Ligi ir kt., 2017) ir Estijos ežerams (Ansper ir Alikas 2019). Tyrimų išvadose teigiama, kad geriausias algoritmas chla koncentracijai drumstuose vandenyse gauti yra artimųjų infraraudonųjų (apie 705 nm) ir raudonųjų juostų santykis (Gilerson ir kt., 2010) arba trijų spektrinių juostų algoritmas, kuriame naudojamos raudonosios ir artimųjų infraraudonųjų spindulių 705 nm ir 740 nm juostos (Ansper & Alikas, 2019; Gitelson ir kt., 2009). Kai kuriais atvejais buvo padaryta išvada, kad modeliai, kuriuose naudojami TOA atspindžio duomenys, veikia panašiai kaip ir modeliai, kuriuose naudojami duomenys, kuriems pritaikyta AK (Ansper & Alikas, 2019; Soomets ir kt., 2020; Toming ir kt., 2017).

Naudojant sudėtingesnius mašininio mokymosi algoritmais pagrįstus modelius, galima panaudoti daugiau duomenų – spektrinių požymių – ir gauti

tikslesnius rezultatus. Čekijos ežerų tyrime iš Sentinel-2 MSI duomenų buvo apskaičiuota 11 skirtingų spektrinių indeksų, kurie buvo naudojami mašininio mokymosi algoritmuose. Normalizuotas augmenijos skirtumo indeksas (*angl.* – *normalized vegetation difference index*), paremtas raudonos ir NIR bangu spektrinėmis juostomis, raudonosios ir mėlynosios spektrinių juostų santykis ir raudonosios ribos (*angl.* – *red-edge*) juosta, B5, buvo svarbiausi požymiai chla koncentracijai gauti naudojant Cubist algoritimą (Saberioon ir kt., 2020). Mašininio mokymosi metodai padeda surasti tinkamiausius spektrinius požymius chla koncentracijai nustatyti ir leidžia panaudoti keletą jų, taip padidinant rezultatų tikslumą.

#### 1.5. *In situ* optines vandens savybes apibūdinančių rodiklių matavimai Lietuvos ežeruose

*In situ* tyrimai yra svarbūs siekiant suprasti vandens ekosistemų dinamiką. Šie tyrimai suteikia galimybę gauti išsamius ir tikslius duomenis apie vandens kokybės parametrus. Lietuvoje, kaip ir kitose Europos Sąjungos valstybėse yra vykdomas valstybinis ežerų ir tvenkinių monitoringas, kurio metu vykdomi reguliarūs *in situ* matavimai. Vandens skaidrumas, chla koncentracija ir suspenduotų medžiagų koncentracija – vandens kokybės indikatoriai, apibūdinantys limnosistemos ekologinę būklę.

*In situ* duomenys svarbūs ne tik nuolatiniam vandenių būklės stebėjimui, bet ir nuotoliniais būdais gautų duomenų validavimui ir modelių, naudojančių juos, kūrimui. Lietuvos ežeruose surinkti *in situ* duomenys gali būti naudojami modelių, paremtų palydoviniais duomenimis, kūrimui, tiek Lietuvos mokslininkų, tiek ir šalies ribas peržengiantiems tyrimams bei palydovinių duomenų pritaikymo gerinimui vidutinių platumų teritorijoje.

Valstybinis ežerų ir tvenkinių monitoringas yra sistemingas paviršinių vandens telkinių būklės, savaiminių pokyčių ir antropogeninio poveikio stebėjimas ir vertinimas. Monitoringo metu atliekama įvairių matavimų, tarp kurių ir skaidrumo, chla ir suspenduotos medžiagos koncentracijos matavimai. Chla koncentracija įvertinama spektrofotometriniu metodu po ekstrakcijos karštu 90 % etanolu pagal ISO 10260:1992 standartą (LAND 69-2005..., 2005). Pagal chlorofilo ekstrakto optinio tankio spektrofotometru matavimus 665 ir 750 nm bangos ilgiuose apskaičiuojama chla koncentracija. Suspenduotos medžiagos koncentracija, nustatoma gravimetriniu metodu (standartas LST EN 872:2005), ir vandens skaidrumas naudojant Secchi diską matavimai taip pat atitinka Europos standartizacijos komiteto ir Tarptautinės standartizacijos organizacijos išleistus standartus (Bendrieji reikalavimai..., 2003).

Valstybinio monitoringo programos apimtis bėgant metams keitėsi – minėti parametrai 2001–2002 m. buvo matuojami 13-oje ežerų, nuo 2003 m. įtrauktos Kauno marios, nuo 2005 m. matavimai atlikti jau 28-iuose ežeruose ir tvenkiniuose, 2006 m. – 37-iuose, o nuo 2007 m. kasmet stebima per 70 ežerų ir tvenkinių, didesnių nei 50 ha. 281 ežerų matavimai ir ekologinės būklės vertinimas apibendrinti AAA užsakytoje 2009 m. išleistoje Restauruotinių Lietuvos ežerų nustatymo ir preliminaraus restauravimo priemonių parinkimo šiems ežerams, siekiant pagerinti jų būklę ataskaitoje (Balevičienė ir kt., 2009).

Ekspedicinius vandens skaidrumo ir chla matavimus taip pat atlieka ir mokslo institucijos. Gamtos tyrimų centro mokslininkai chla nustatymui naudoja ir greitesnį fluorometrinį metodą (Savadova ir kt., 2018; Savadova-Ratkus ir kt., 2022). Paskutiniaisiais metais Jūrinių tyrimų centro mokslininkai renka duomenis palydovinių ir dronais surinktų duomenų validavimui (Tiškus ir kt., 2023).

Visgi *in situ* duomenų viešoji prieiga yra ribota, nėra oficialių duomenų rinkinių, dažniausiai pateikiamos vidutinės parametrų reikšmės, pvz., valstybinio monitoringo duomenyse, Balevičienės ir kt. (2009) ataskaitoje. Taip pat daugiausia duomenų yra apie didesniuosis (dažniausia didesni nei 50 ha dydžio) ežerus, o mažesniųjų ištirtumas yra ribotas, nors mažesni nei 50 ha (bet didesni nei 0,5 ha) ežerai sudaro beveik 70 % Lietuvos ežerų (Gailiūšis ir kt., 2001).

Palydovinių duomenų validavimui reikalingi konkrečių datų *in situ* duomenys, kurie taip pat turi ir matavimo vietos koordinatas, kadangi patikimiausia išsikirti centruotus ir artimus matavimo vietai duomenis iš palydovinių nuotraukų. Mašininio mokymosi algoritmais pagrįstiems modeliams kurti reikalinga surinkti kuo daugiau *in situ* duomenų, kurie reprezentuotų įvairias sąlygas ežeruose. Siekiant sukurti *in situ* duomenų bazę, kuri būtų naudinga palydovinių duomenų validavimui, reikia *in situ* matavimus vykdyti tuo pačiu metu, kada praskrenda palydovas.

## 2. DUOMENYS IR METODIKA

### 2.1. Tyrimo eiga

Šio tyrimo objektas – chla koncentracijos ežeruose nustatymas iš palydovinių duomenų Lietuvos ežerams. Siekiant įvertinti chla koncentraciją ežeruose, reikia atlikti keletą žingsnių, kad būtų įvertintas duomenų neapibrėžtumas, sukurti modeliai, patikrintas jų veikimas ir atlikta chla koncentracijos analizė laike ir erdvėje (2.1 pav.).

Vertinant chla pasirinkti vidutinės raiškos atviros prieigos Sentinel-2 misijos duomenys, kurie yra pakankamos erdvinės raiškos santykinai mažų Lietuvos ežerų stebėjimui (Grendaitė ir kt., 2018) bei pasižymi santykinai aukštu signalo ir triukšmo santykiu, tad yra galimybė užfiksuoti nedidelius vandens parametrų pokyčius. Sentinel-2 (palydovinio jutiklio MSI charakteristikos pateiktos 2.3 lentelėje) palydovų duomenimis pagrįstų algoritmų kūrimas labai reikalingas siekiant išnaudoti duomenų potencialą stebėjimams ir dėl planuojamo misijos tęstinumo keliems ateinantiems dešimtmečiams.

Palydovais užfiksuojamą signalą veikia ne tik tiriamo paviršiaus atspindys, bet ir atmosfera. Dėl šios priežasties paviršiaus duomenys pritaikius atmosferos poveikį pašalinančius algoritmus. Atmosferos korekcija – tai palydovinių duomenų apdorojimo žingsnis, kurį būtina atlikti siekiant gauti vandenį paliekantį atspindį. Šis atspindys naudojamas tiriant chla koncentraciją vandens telkiniuose. AK algoritmų yra sukurta ne vienas ir nėra susitarimo, kuris labiausiai tinka vidaus vandenims. Šiame tyrime įvertintas 7 atmosferos korekcijos produktais (Acolite, Acolite Rayleigh, C2RCC, C2X, iCOR, POLYMER ir Sen2Cor, detaliau aprašyti 2.5 skyriuje) gautų spektrų neapibrėžtumas chla koncentracijai įvertinti, pasitelkiant 10 radiometrinių chla algoritmų (CHLA1-CHLA10). Chla algoritmų rezultatai leidžia nustatyti, kurį atmosferos korekcijos algoritmą geriau naudoti, atsižvelgiant į apskaičiuotus mažiausius nuokrypius nuo išmatuotos chla koncentracijos. Atmosferos korekcijos neapibrėžtumui įvertinti buvo pasitelktos 7 Sentinel-2 nuotraukos (T35UMB, plačiau apie duomenis 2.4 skyriuje), apimančios 136 rytų Lietuvos ežerus (daugiau apie ežerus 2.2 skyriuje). Nuotraukos buvo darytos 2018–2019 m. Pirmiausia nagrinėta, kaip atmosferos korekcijos algoritmai sumažina ežerų spektrinius atspindžius, ypač atsižvelgiant į chla įvertinimui svarbias spektrines juostas (665 nm ir 705 nm) (rezultatai pateikti 3.1 skyriuje). Tada įvertinta 136 ežerams apskaičiuotos chla koncentracijos sklaida pasitelkus 10 algoritmų (rezultatai pateikti 3.2 skyriuje). Galiausiai nustatytas ryšys tarp 30 vienalaikių (ne daugiau nei dienos skirtumo,

duomenys aprašyti 2.5 skyriuje) išmatuotos bei iš palydovinių duomenų įvertintos chl<sub>a</sub> koncentracijos (rezultatai pateikti 3.3 skyriuje). *In situ* valstybinio monitoringo (daugiau apie duomenis 2.3 skyriuje) metu surinkti chl<sub>a</sub> duomenys buvo naudojami palydovinių duomenų validavimui.

Atmosferos korekcijos algoritmų neapibrėžtumas išryškėjo naudojant skirtingus radiometrinius chl<sub>a</sub> algoritmus, o paskaičiavus modelio veikimo metrikas buvo nustatyta geriausia atmosferos korekcijos ir chl<sub>a</sub> algoritmo kombinacija. Nuokrypiai tarp išmatuotų ir chl<sub>a</sub> algoritmu apskaičiuotų chl<sub>a</sub> reikšmių skirtinguose chl<sub>a</sub> intervaluose yra skirtingi, o tai implikuoja, kad chl<sub>a</sub> algoritmai, sukurti ežerų grupėms, kurie pasižymi panašiomis optinėmis savybėmis (šiuo atveju žema arba aukšta chl<sub>a</sub> koncentracija) galėtų veikti tiksliau nei imant visus stebėjimus jų neskirstant į grupes. Ežerų tipizacija pagal optines savybes dažnai nagrinėjama ir literatūroje, kuriamos klasifikacijos pagal spektrus (Spyrakos ir kt., 2018) ir sudaromi algoritmai siekiant nustatyti vandens parametrus ežerų optiniams tipams (Neil ir kt., 2019).

Įvertinus AK neapibrėžtumo poveikį chl<sub>a</sub> koncentracijai ir atsižvelgiant į tai, kad ežerų optines savybes lemia chl<sub>a</sub>, suspenduotos medžiagos (SM) ir ištirpusios spalvotosios organinės rūgštys (CDOM), buvo nuspręsta įvertinti ar sudarant stebėjimo sąlygų biofizinę klasifikaciją pagerėja chl<sub>a</sub> modelio rezultatai. Šiame tyrime ežero optinių savybių apibūdinimui buvo kuriamas biofizinės klasės modelis, kuriuo, naudojant spektrinius iš palydovo gautus duomenis, buvo galima nustatyti biofizinę ežero klasę kiekvieno palydovinio stebėjimo metu. Sukurti modeliai biofizinėms klasėms naudojami kaip chl<sub>a</sub> koncentracijos nustatymo metodikos dalis.

Biofizinės klasės modelių sudarymui buvo reikalinga paruošti vienalaikių *in situ* ir Sentinel-2 duomenų bazę (daugiau apie tai 2.3 ir 2.4 skyriuose). Kuriant duomenų bazę biofizinės klasės ir chl<sub>a</sub> modeliams apmokyti vienalaikiais stebėjimais buvo laikomos tokios duomenų poros, kai tarp *in situ* ir palydovinio stebėjimo buvo ne didesnis nei 3 dienų laiko tarpas. Ežerų stebėjimų suskirstymui į biofizines klases buvo pasitelkti Lietuvos ežeruose valstybinio ežerų ir tvenkinių monitoringo metu surinkti *in situ* duomenys ir ežerų tipologija pagal morfometrines savybes (LR Aplinkos ministerija, 2018). Buvo panaudota 2015–2020 m. išmatuota chl<sub>a</sub> koncentracija, SM koncentracija ir skaidrumas bei stebėjimai suskirstyti į biofizines klases (daugiau apie duomenis 2.3, 4.1 ir 4.2 skyriuose). Kiekvienam *in situ* stebėjimui buvo surastas artimiausias laike palydovinis stebėjimas, bet ne daugiau nei trijų dienų skirtumu esantys stebėjimai. Stebėjimų poros turi būti artimos siekiant užfiksuoti tas pačias sąlygas vandens telkinyje ir modelio apmokymas būtų tinkamas. Priešingu atveju, pvz., pasikeitus

meteorologinėms sąlygoms, vandens telkinyje pasikeistų ir vandens parametrų koncentracijos ir modelio, naudojančio palydovinius duomenis, apmokymas pagal išmatuotus *in situ* duomenis būtų nebetikslus, o tai reikšmingai neigiamai paveiktų modelio apmokymą ir modelio rezultatus. Trijų dienų skirtumas tarp *in situ* ir palydovinio stebėjimo naudojamas ir kituose tyrimuose (Ansper & Alikas, 2019). Galutiniai biofizinės klasės modeliai nustato vieną iš keturių biofizinių klasių: skaidrią, tarpinę, Chla-klasę arba drumstą klasę. Stebėjimai susiaurinti iki 4 klasių dėl nedidelio kiekio stebėjimų, pvz., klasės, kurioje vyrautų SM išskirti atskirai ir apmokyti modelio nebuvo galimybės dėl itin mažo duomenų kiekio. Tačiau surinkus daugiau reprezentatyvių duomenų būtų galima išskirti daugiau klasių naudojantis šio tyrimo pavyzdžiu ir darbo eiga.

Mašininio mokymosi modeliai leidžia panaudoti daugiau spektrinės informacijos nei vieną spektrinę juostą ar vieną radiometrinių indeksą, tad modelių sudarymui buvo pasirinkta naudoti mašininio mokymosi modelius. Modelių įvestis – šiuo atveju, spektriniai požymiai – apskaičiuoti iš palydovinių spektrų (atspindžių spektrinėse juostose, skirtumai, santykiai, taip pat kiti išvestiniai rodikliai, naudojantys kelias spektrines juostas ir apibūdinantys vandens telkinio spalvą) ir modeliu nustatomas rodiklis, šiuo atveju – biofizinė klasė. Tarp spektrinių požymių įtrauktos ir kai kurios spektrinės išraiškos (radiometriniai indeksai), naudotos CHLA1-CHLA10 modeliuose, tik CHLA1-CHLA10 modeliai yra empirinės lygtys, o čia buvo naudojami tik radiometriniai indeksai, kurie yra siejami su ežerų optinėmis savybėmis (pvz., ALG1 ir ALG2 iš 2.7 lentelės).

Siekiant surasti tinkamiausią algoritmą biofizinei klasifikacijai ir kadangi kiekvienu skirtingu modelių taikymo atveju turint konkrečią duomenų bazę geriau gali veikti vienas ar kitas algoritmas, buvo testuojami 6 skirtingi mašininio mokymosi klasifikacijos algoritmai:

1. logistinės regresijos (LR),
2. atraminių vektorių (SVM),
3. atsitiktinio miško (RFC),
4. adaptyvaus didinimo (AdaBoost),
5. ekstremalių gradientų didinimo (XGBoost),
6. dirbtinių neuroninių tinklų (ANN).

Algoritmas buvo pasirinktas pagal apskaičiuotas metrikas ir atsižvelgus į modelio sudėtingumą. Įprastai pasirenkamas mažiau sudėtingas modelis, kurio rezultatus taip pat lengviau interpretuoti.

Visas duomenų paruošimas vykdytas *Rstudio* aplinkoje (RStudio Team, 2020), o LR, SVM, Ada ir RFC mašininio mokymosi algoritmai buvo kuriami *python* aplinkoje naudojant *scikit-learn* modulį v1.0.2 (Pedregosa ir kt., 2011),

XGB – naudojant *XGBoost* modulį (Chen & Guestrin, 2016) ir ANN naudojant *keras* (Chollet ir kt., 2015).

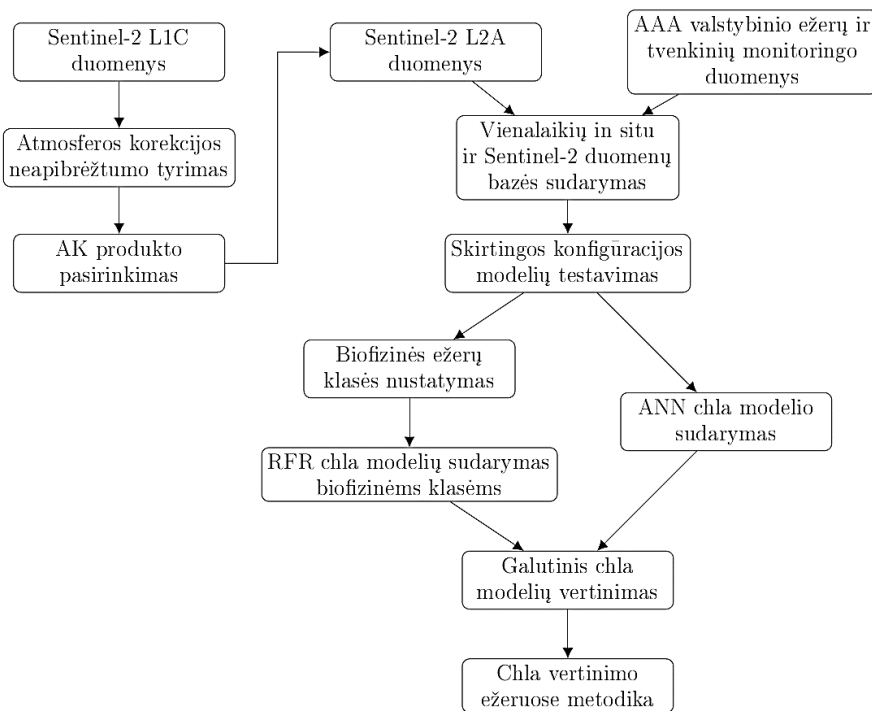
Konkrečiu metu ežere nustatyta biofizinė klasė parodo pradinę informaciją apie vandens telkinį ir gali būti naudojama kaip indikatorius apie vandens telkinio būklę. Tačiau siekiant tikslesnio chl<sub>a</sub> koncentracijos įvertinimo reikalingas chl<sub>a</sub> įvertinantis modelis. Biofizinė klasė leido atskirti skirtingus stebėjimus ir toliau buvo sukurti du modeliai chl<sub>a</sub> koncentracijai ežeruose nustatyti, pasiremiant turima vienašalių stebėjimų duomenų baze. Vienas modelis daugiausia apėmė žemas chl<sub>a</sub> koncentracijas (chl<sub>a</sub> < 7,2 mg/m<sup>3</sup>, skaidri biofizinė klasė), kitas – vidutines ir aukštas (chl<sub>a</sub> < 7,2 mg/m<sup>3</sup>, tarpinė ir Chl<sub>a</sub> biofizinė klasė). Tarpinės klasės stebėjimai buvo sudėti kartu su Chl<sub>a</sub>-klase dėl mažo duomenų kiekio. Ketvirtai – drumstai – klasei taip pat nebuvo kuriamas chl<sub>a</sub> modelis, paremtas mašininio mokymosi algoritmais, kadangi šiai klasei priskiriami ežerai, kuriuose vyrauja kitos optiškai aktyvios medžiagos (CDOM), kurios apsunkina chl<sub>a</sub> apskaičiavimą iš palydovinių duomenų. Taip pat šios klasės ežerų valstybinio monitoringo sistemoje yra palyginti nedaug (vienašalių duomenų bazėje ši klasė buvo priskirta 9 % duomenų) ir trūksta *in situ* duomenų modelių apmokymui.

Chl<sub>a</sub> modeliai buvo kuriami pasitelkiant atsitiktinio miško regresijos (RFR) algoritmą, kadangi RF parodė geriausias rezultatus jau kuriant biofizinės klasės modelius. Chl<sub>a</sub> modelių veikimas buvo įvertintas palyginant su modeliu, sukurtu naudojant visus duomenis (RFR), t.y., visą duomenų aibę. Taip pat jo veikimas buvo lygintas su sudėtingesniu ANN pagrįstu modeliu, taip atskleidžiant, kad skirstymas į klases leidžia nustatyti chl<sub>a</sub> koncentraciją tiksliausiai.

Sudarius ir įvertinus mašininio mokymosi algoritmu pagrįstus modelius modeliai buvo pritaikyti visai turimai 2015–2021 m. Sentinel-2 duomenų bazei. Nors duomenys buvo nufiltruoti, tačiau laiko eilutės chl<sub>a</sub> duomenyse buvo pastebėta staigių koncentracijos kritimų arba pakilimų, kurie sietini su debesimis ar jų šešėliais palydoviniuose duomenyse. Dėl to duomenys buvo papildomai filtruojami pašalinant išskirtis, apskaičiavus standartizuotą vidurkį (plačiau 5.3 skyriuje).

Po išskirčių pašalinimo duomenys buvo naudojami potencialiai probleminių ežerų nustatymui pagal chl<sub>a</sub> momentines (maksimalias) ir vidutines per metus bei tyrimo laikotarpį reikšmes. 2009 m. restauruotinių ežerų ataskaitoje potencialiai probleminiai ežerai buvo nustatyti pagal keletą kriterijų, vienas iš jų hidrobiologinis – chl<sub>a</sub> koncentracija. Remiantis 2009 m. išleista Restauruotinių ežerų ataskaita pagal hidrobiologinius rodiklius ežerai buvo priskirti potencialiai probleminiams, kai vidutinė chl<sub>a</sub> koncentracija sekliuose ir vidutiniškai giliuose ežeruose buvo didesnė už 8 mg/m<sup>3</sup>, o giliuose

–  $6 \text{ mg/m}^3$ , taip pat kai momentinė sekliuose ir vidutiniškai giliuose ežeruose viršijo  $15 \text{ mg/m}^3$ , o giliuose ežeruose buvo  $> 12 \text{ mg/m}^3$  (Balevičienė ir kt., 2009). Potencialiai probleminių ežerų nustatymas pagal chl<sub>a</sub> koncentraciją buvo atnaujintas šiame tyrime pasiremiant didesne duomenų baze – 7 metų laikotarpio duomenimis (2015–2021 m.) – bei dažnesniais stebėjimais, įvertintais iš palydovinių stebėjimų. Be to, dalyje ežerų (12 %) tyrimo metu nebuvo atlikta *in situ* tyrimų, tad palydovais ir tyrime sukurtais modeliais gauti duomenys leidžia įvertinti šių vandenų būklę bei priskirti juos neprobleminiams arba potencialiai probleminiams. Taip pat buvo atlikta chl<sub>a</sub> koncentracijos kaitos laike ir erdvėje analizė tiriamuose ežeruose ir tvenkiniuose.



**2.1 pav.** Tyrimo schema

Tyrime analizuoti vandens telkiniai apibūdinti 2.2 skyriuje. Tyrimui naudoti valstybinio monitoringo metu surinkti *in situ* duomenys (2.3 skyrius) ir palydoviniai Sentinel-2 duomenys (2.4 skyrius). Atmosferos korekcijos neapibrėžtumo vertinimo chl<sub>a</sub> koncentracijai metodika aprašyta 2.5 skyriuje, o neapibrėžtumo vertinimo rezultatai – 3 skyriuje. Biofizinės klasės klasifikacijos modeliai apibūdinti 2.6 skyriuje, o klasifikacijos ir regresijos modelių vertinimo metrikų formulės nurodytos 2.7 skyriuje. Biofizinės klasės



modelių sudarymas ir pritaikymas nagrinėjamas 4 skyriuje, o chla modelių sudarymas ir pritaikymas aprašytas 5 skyriuje.

## 2.2. Tyrimų teritorija

Tyrime nagrinėti 357 vandens telkiniai, esantys Lietuvoje ir įtraukti į valstybinio monitoringo programą, kuriuose kas 1–7 metus atliekami *in situ* stebėjimai. Šie ežerai ir tvenkiniai priklauso Baltijos jūros ekoregionui, o jų absoliutinis aukštis nesiekia 200 m virš jūros lygio. Lietuva pagal Koppeno klimato klasifikaciją yra drėgno žemyninio klimato su šiltomis vasaromis zonoje (Dbf). Vidutinė metinė oro temperatūra žemyninėse meteorologijos stotyse 7,9 C, o jūrinėse 5,8 C (Jaagus ir kt., 2014). Lietuva yra vakarinėje Rytų Europos plynaukštės dalyje. Pasklidoji tarša iš dirbamųjų laukų yra pagrindinis vandens telkinių taršos šaltinis (Aplinkos Apsaugos Agentūra, 2018).

Ežerų dydis yra nuo 0,35 iki 46,8 km<sup>2</sup> (vidurkis 2,24 km<sup>2</sup>, standartinis nuokrypis 4,1 km<sup>2</sup>). Remiantis Paviršinių vandens telkinių tipų aprašu pagal vidutinį ir maksimalų gylį ežerai ir tvenkiniai skirstomi į tris tipus (2.1 lentelė) (LR Aplinkos ministerija, 2018). Biofizinėje klasifikacijoje skirtingo gylio ežerams taikomos skirtingos skaidrumo slenkstinės vertės (4.1 lentelė), o išskiriant potencialiai probleminius ežerus skirtingo gylio ežerams taikomos skirtingos chla slenkstinės vertės (2.1 skyrius). Iš jų 289 yra natūralūs ežerai, 67 yra labai pakeisti vandens telkiniai (tvenkiniai) ir 1 yra dirbtinės kilmės vandens telkinys (Lampėdžių karjeras) (2.2 pav.).

**2.1 lentelė.** Ežerų tipai (LR Aplinkos ministerija, 2018) ir jų apibūdinimas.

Tipas	Gylio tipas	Vidutinis gylis, m	Didžiausias gylis, m	Ežerų ir tvenkinių skaičius
1	Seklūs	≤ 3 m   > 3 m	–   < 11 m	190
2	Vidutiniškai gilūs	> 3 m	11–30 m	121
3	Gilūs	–	> 30 m	46

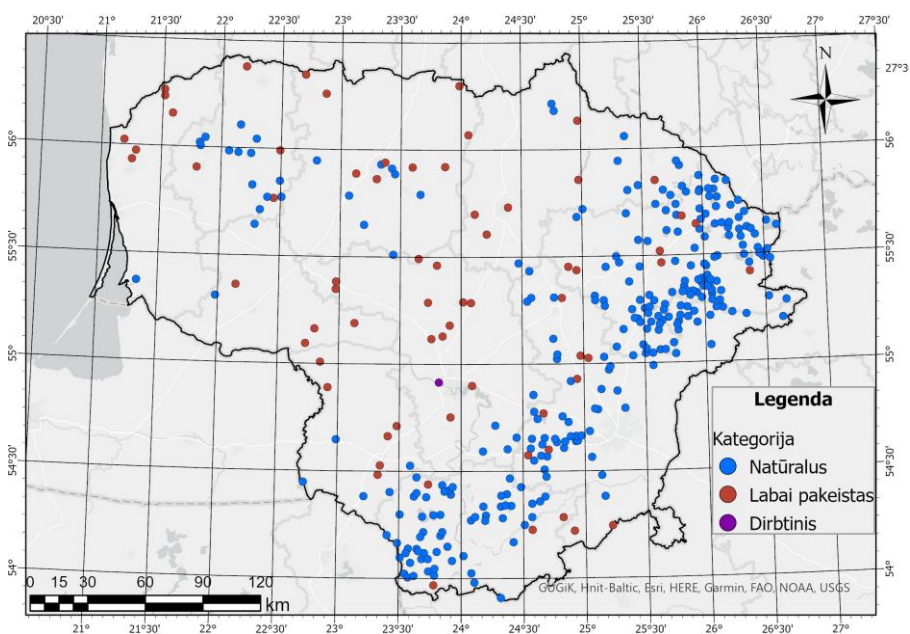
## 2.3. Aplinkos apsaugos agentūros valstybinio monitoringo duomenys

Tyrime palydovinių duomenų validavimui buvo naudoti valstybinio ežerų ir tvenkinių monitoringo duomenys – chla koncentracija, skaidrumas, suspenduota medžiaga. Duomenys gauti iš monitoringą vykdančios institucijos – Aplinkos Apsaugos Agentūros prie Lietuvos Respublikos Aplinkos ministerijos. Per 2015–2021 m. laikotarpį AAA atliko matavimus

314 iš 357 ežerų ir tvenkinių. Dauguma matavimų atliekami šiltuoju sezonu – balandžio–spalio mėnesiais.

Dažniausiai viename vandens telkinyje atliekami 4 stebėjimai per nurodytą laikotarpį (1 pavasarį, 2 – vasarą ir 1 rudenį) ir 13-oje vandens telkinių kai kuriais metais buvo atlikti 6 stebėjimai. Daugumoje vandens telkinių atliekami vienalaikiai skaidrumo ir chla koncentracijos matavimai, tačiau Drūkšių ežere (LTL52) chla matavimų nebuvo, nors skaidrumas buvo nustatytas 2015–2018 ir 2020, 2021 metais (24 stebėjimai). Krokų lankoje (LTL277) atvirkščiai – nėra skaidrumo matavimų.

Per metus AAA atlieka apie 300 matavimų, tačiau tik dalis jų dėl debesuotumo ir nutolimo laike gali būti panaudoti palydovinių duomenų validacijai (2.2 lentelė).



**2.2 pav.** Vandens telkinių kategorijos: natūralūs vandens telkiniai – ežerai, labai pakeisti vandens telkiniai – tvenkiniai, dirbtiniai vandens telkiniai – karjerai.

#### 2.4. Palydoviniai duomenys

Tyrimė buvo naudojami atviri vidutinės erdvinės raiškos Europos Kosmoso Agentūros (EKA) Sentinel-2 misijos palydovų daugiaspektrio jutiklio (*angl. – MultiSpectral Instrument*) optiniai duomenys. Sentinel-2 MSI stebi Žemę 13-oje spektrinių juostų, kurių gardelės dydis yra 10 m, 20 m ir 60 m (2.3 lentelė), o stebima elektromagnetinio spektro dalis apima nuo

matomųjų bangų (440 nm) iki trumposios infraraudonosios spinduliuotės (2190 nm) spektro dalį. Radiometrinė jutiklio raiška 12 bitų.

**2.2 lentelė.** Aplinkos apsaugos agentūros atliekamų matavimų skaičius 2015–2021 m. balandžio–spalio mėnesiais ir vienalaikių *in situ* ir Sentinel-2 stebėjimų skaičius.

Metai	AAA atlikta matavimų	Vienalaikių <i>in situ</i> ir palydovinių stebėjimų skaičius 2015–2020 m.	Papildytas vienalaikių <i>in situ</i> ir palydovinių stebėjimų skaičius 2015–2021 m.
2015	310	54	54
2016	298	73	73
2017	303	74	96
2018	304	107	164
2019	309	147	210
2020	318	108	194
2021	318		193

Sentinel-2 duomenys yra prieinami 100 km x 100 km (plotas 10 000 km<sup>2</sup>) dydžio žemėlapių regionais (*angl.* – *tiles*). Duomenys yra ortorektifikuoti ir pateikiami WGS84 koordinatų sistemoje UTM projekcijoje. Duomenyse yra bedimensė atspindžio reikšmė (*angl.* – *reflectance*) kiekvienoje gardelėje. Reikšmės kinta nuo 0 iki 1, kur 0 – atspindžio nuo paviršiaus nėra, o 1 – atspindimas maksimalus energijos kiekis. Pirmo lygio duomenys (Level 1C) yra dar apibūdinami kaip viršutinės atmosferos (*angl.* – *Top of Atmosphere, TOA*) atspindžiai, kadangi šiuose duomenyse užfiksuotas signalas apima ne tik atspindį nuo paviršiaus, bet ir signalą iš atmosferos.

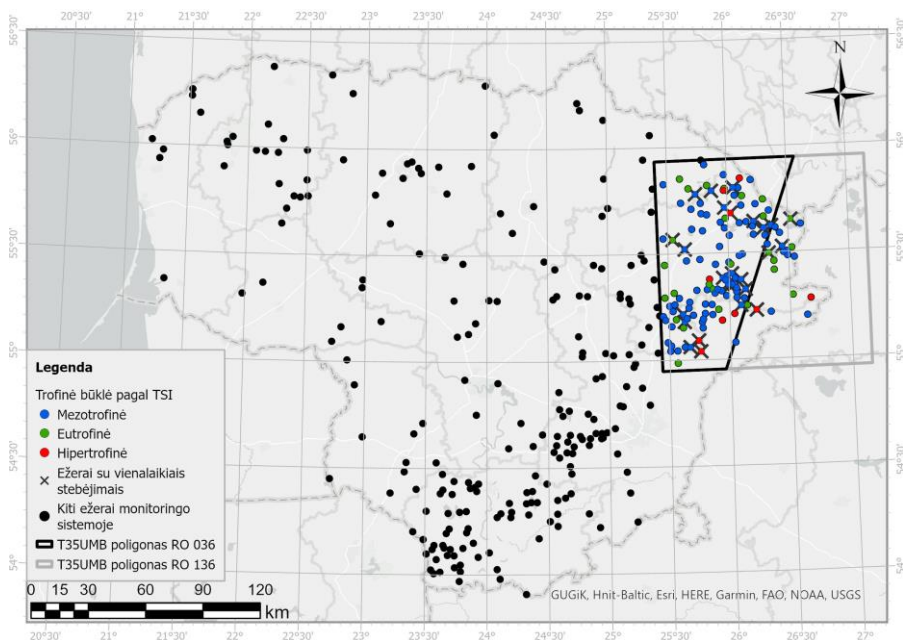
Atmosferos korekcijos algoritmų neapibrėžtumui įvertinti iš EKA Copernicus Open Access Hub buvo parsisiųstos pirmo lygio (L1C) Sentinel-2 MSI žemėlapių regiono T35UMB (2.3 pav.) 2018 ir 2019 m. septynios nuotraukos (2.4 lentelė), reprezentuojančios skirtingus sezonus – pavasarį, vasarą, rudenį. Nuotraukoms pritaikyti 7 atmosferos korekcijos algoritmai (2.5 skyrius). Spektrai buvo paimti iš 7 Sentinel-2 nuotraukų. Keturios nuotraukos buvo gautos palydovui skrendant santykinė orbita 136, kai TUMB35 regionas apimamas pilnai, o kitos trys nuotraukos buvo gautos palydovui leidžiantis 036 orbita, kuri apima ne visą T35UMB dalį, tačiau į ją vis tiek patenka didžioji dalis – 115 ežerų (2.4 lentelė).

Ežerų biofizinės klasifikacijos modeliui sudaryti buvo naudojamos 2015–2020 m. balandžio–spalio mėn. didžiąją Lietuvos dalį apimančios pirmo lygio nuotraukos (T34UEG, T34UFG, T35ULB, T35UMB, T34UFF, T35ULA),

kurių debesuotumas buvo mažesnis nei 30 %. 2015–2016 m. duomenims buvo pritaikytas Sen2Cor v2.5.5 atmosferos korekcijos algoritmas, kadangi naujesnis algoritmas nebuvo suderinamas su senesniais duomenimis. 2017–2020 m. duomenims pritaikytas Sen2Cor v2.8.0 atmosferos korekcijos algoritmas. Vėliau chla modelių sudarymui naudotas duomenų rinkinys, papildytas ežerų spektrais, apimančiais ir likusią Lietuvos dalį (šiaurėje – T34VEH, T34VFH, T35VLC ir pietuose – T34UFE, T34UGE, T35ULV), didesnio debesuotumo (< 60 %) nuotraukomis ir 2021 m. duomenimis. Duomenų rinkinys buvo papildytas naudojantis Google Earth Engine (GEE) duomenų pasiekimo paslauga (Gorelick ir kt., 2017), kai parsisiunčiami tik iškirpti duomenys, o ne visa nuotrauka. Duomenys buvo paimti iš S2\_SR kolekcijos, kur yra prieinami antro lygio, t. y., atėmus atmosferos poveikį, duomenys (L2A). Duomenys prieinami nuo 2017-03-28. Duomenims pritaikytas Sen2Cor atmosferos korekcijos algoritmas.

**2.3 lentelė.** Sentinel-2 daugiaspektrio jutiklio (MSI) spektrinių juostų erdvinė ir spektrinė raiška (Drusch ir kt., 2012).

Spektrinės juostos numeris	Centrinis bangos ilgis, nm	Spektrinės juostos plotis, nm	Erdvinė raiška, m
B2	490 (mėlyna)	65	10
B3	560 (žalia)	35	
B4	665 (raudona)	30	
B8	842 (NIR)	115	
B5	705 (raudonosios ribos)	15	20
B6	740 (raudonosios ribos)	15	
B7	783 (raudonosios ribos)	20	
B8A	865 (NIR)	20	
B11	1610 (SWIR)	90	
B12	2190 (SWIR)	180	
B1	443 (mėlyna)	20	60
B9	945 (NIR)	20	
B10	1380 (NIR)	30	



**2.3 pav.** Į valstybinę monitoringo sistemą įtraukti ežerai ir tvenkiniai. Pažymėti šie regionai: palydovo Sentinel-2 pilnas žemėlapio regionas T35UMB iš santykinės orbitos RO 136 ir dalis žemėlapio regiono, kai palydovas praskrenda orbita RO 036. T35UMB regiono nuotraukos, naudotos AK neapibrėžtumo vertinime. Ežerų trofinė būseną buvo gauta pagal valstybinio monitoringo duomenis. Ežerai, naudoti lyginamojoje chl<sub>a</sub> koncentracijos analizėje, pažymėti kryžiu.

**2.4 lentelė.** Sentinel-2 nuotraukų, naudotų atmosferos korekcijos algoritmu neapibrėžtumui įvertinti, charakteristikos. Nuotraukos apima tą patį regioną – Sentinel-2 TUMB35 – Rytų Lietuvą. Nuotraukos darytos palydovui praskendant santykinę orbita 036 (dalinis T35UMB padengimas) ir 136 (pilnas T35UMB padengimas).

Data	Santykinė orbita	Nuotraukos debesuotumas, %	<i>In situ</i> ir palydovinių chl <sub>a</sub> koncentracijos stebėjimų skaičius, naudotų CHLA1-CHLA10 algoritmuose
2018-05-02	136	3,19	3
2018-05-07	136	0,07	4
2019-07-21	136	2,1	2
2019-07-24	036	3,3	4
2019-08-20	136	10,4	4
2019-08-28	036	5,2	8
2019-09-27	036	0,0	5

Vienalaikių *in situ* ir palydovinių duomenų bazei sudaryti palydoviniai duomenys buvo paimti devyniose 20 m raiškos gardelėse (3 x 3 makrogardelė), centruotose nacionalinio monitoringo matavimų atlikimo vietose. Matavimo vietos dažniausia yra centrinėje arba giliausioje ežero vietoje. 3 x 3 makrogardelė pasirinkta, kadangi tai gali padidinti signalo ir triukšmo santykį ir dažnai naudojama kituose tyrimuose (Ansper & Alikas, 2019; Kratzer ir kt., 2019). 20 m raiška naudota siekiant išlaikyti originalią 705 nm spektrinės raiškos juostos raišką, kuri yra svarbi chla koncentracijos nustatymui. Remiantis Sen2Cor nuotraukos klasifikacija buvo palikti tik tie spektrai, kuriems buvo priskirta vandens klasė (vandens kaukės reikšmė – 6). Taip pat buvo pašalinti spektrai, kurių atspindys 1610 nm centruotoje spektrinėje juostoje (B11) buvo didesnis nei 0,0215. Ši reikšmė yra dažnai naudojama vandens gardelių atskyrimui nuo ne vandens gardelių, kadangi šiame bangos ilgyje vandenį paliekanti spinduliuotė yra lygi 0 (Wang, 2007). Didesni atspindžiai šiuose bangos ilgiuose būtų siejami su atmosferiniu išsklaidymu, saulės atspindžiu, migla ar debesimis (Vanhellemont, 2019). Taip pat likę galimai klaidingi spektrai buvo šalinami pasitelkiant atspindžio variaciją makrogardelėje 783 nm spektrinėje juostoje (B7), kadangi didelės variacijos makrogardelės galėjo būti paveiktos debesų ar debesų šešėlių. Makrogardelės, kur standartinis nuokrypis buvo didesnis nei 0,002, buvo pašalintos.

Iš GEE buvo parinktos nuotraukos, kur debesuotumas buvo mažesnis nei 60 %, o tada naudojant GEE prieinamą duomenų rinkinį su debesų tikimybe (kolekcija „Sentinel-2: Cloud Probability) buvo nufiltruotos visos gardelės, kur debesų tikimybė buvo didesnė nei 10 %. Taip pat pritaikyti kiti jau paminėti filtravimo metodai – paliktos tik vandens kaukę turinčios gardelės, pritaikytas filtravimas pagal 1610 nm ir pagal 783 nm spektrinės juostos variaciją.

Ežerų biofizinės klasifikacijos ir chla modeliams sudaryti buvo naudojama vienalaikių palydovinių ir *in situ* duomenų bazė. Vienalaikių stebėjimų 2015–2020 m. buvo 563 (226 skirtingi vandens telkiniai), kurie buvo naudoti biofizinės klasės modeliams kurti. Prijungus daugiau duomenų ir prailginus tyrimo laikotarpį pasitelkus GEE, duomenų rinkinys apėmė 2015–2021 m. ir išaugo iki 984, tačiau po vizualios palydovinių duomenų RGB nuotraukų patikros liko 749 stebėjimai, kurie buvo panaudoti chla modelių kūrimui.

## 2.5. Atmosferos korekcijos algoritmų neapibrėžtumo vertinimas chlorofilo $\alpha$ koncentracijai

Siekiant įvertinti atmosferos korekcijos algoritmų neapibrėžtumą chl  $a$  koncentracijai buvo naudojami 7 dažnai praktikoje taikomi atmosferos korekcijos procesoriai: 3 nuotrauka paremti produktai – Acolite, Acolite Rayleigh (v.20190326.0) ir iCOR (v2.0: 2019-02-05), vienas dalinai nuotrauka paremtas produktas Sen2Cor (v2.8.0), 2 paremti dirbtiniais neuroniniais tinklais – C2RCC (v.2.0, 20150319) ir C2RCC, skirtas itin drumstiems ir itin sugeriantiems vandens telkiniams – C2X, bei 1 paremtas polinominiu spektro derinimo metodu – POLYMER (v.4.12). Nuotrauka paremti algoritmai nenaudoja papildomos informacijos, o atmosferos savybės suskaičiuojamos iš nuotraukos informacijos (pvz., atspindžių ilgųjų bangų spektrinėse juostose). 7 AK procesoriai buvo pritaikyti 7 nuotraukoms, apimančioms tą pačią teritoriją (2.2 pav.). Algoritmai buvo paleidžiami su numatytais nustatymais parenkant 20 m raišką taip siekiant išvengti raiškos perskaičiavimo 705 nm spektrinėje juostoje. Paleidžiant C2RCC ir C2X druskingumo reikšmė buvo nustatyta į mažiausią galimą – 0,0001 PSU (*angl. – practical salinity units*), nes buvo tiriami tik gėlieji vandenys. Sen2Cor, iCOR, C2RCC ir C2X produktai buvo pritaikyti SNAP (pilnas pavadinimas *angl. – SeNtinel Application Platform*) programoje, Acolite paleidžiamas atskiroje programoje, o POLYMER – Linux aplinkoje.

Acolite (*angl. – atmospheric correction for OLI*) buvo sukurtas koreguoti nuotraukas priekrantės ir vidaus vandenims. Šis algoritmas yra paremtas tik nuotrauka. Acolite yra integruoti du metodai, iš kurių numatytasis yra tamsaus spektro parinkimas (*angl. – dark spectrum fitting, DSF*) (Vanhellemont, 2019; Vanhellemont & Ruddick, 2018). Algoritmas atlieka aerozolio dalelių korekciją. Algoritme taip pat gali būti atlikta saulės atspindžio (*angl. – sun glint*) korekcija, tačiau pagal numatytuosius nustatymus ji nėra atliekama. DSF suskaičiuoja kelio atmosferoje atspindį (*angl. – atmospheric path reflectance*) remiantis daug tamsių objektų, esančių nuotraukoje, be išankstinės informacijos apie nustatytą tamsią spektrinę juostą. Tamsūs objektai yra nustatomi pagal nuokrypį nuo mažiausių kvadratų metodu sudarytos tiesės pagal pirmą kiekvienos spektrinės juostos gardelių tūkstantį, paimtą iš gardelių histogramos. Tada yra sukuriamas tamsusis spektras (Vanhellemont, 2019). Pagal numatytuosius nustatymus, kelio atspindys (*angl. – path reflectance*) yra skaičiuojamas visai nuotraukai. Šis metodas yra paremtas dviem prielaidomis: kelio atmosferoje atspindys yra pastovus visoje nuotraukoje bei nuotraukoje yra gardelių, kurių atspindys yra artimas nuliui bent vienoje iš spektrinių juostų. Remiantis šiomis prielaidomis yra

apskaičiuojamas kelio atmosferoje atspindys (Vanhellemont, 2019). Acolite pateikia dvi vandens atspindžio reikšmes: atspindį, pakoreguotą atimant Reilėjaus išsklaidymą (be dujų ar dangaus korekcijos, toliau tekste vadinama Acolite Rayleigh) ir atspindį, pataisytą atimant Reilėjaus išsklaidymą bei aerozolio dalelių sugėrimą (toliau tekste bus vadinama Acolite).

iCOR – taip pat nuotrauka paremtas AK produktas. Atmosferos korekcija naudojant iCOR yra sudaryta iš kelių žingsnių: sausumos ir vandens gardelių nustatymas pagal artimosios infraraudonosios spinduliuotės spektrinę 842 nm juostą, sausumos gardelės yra panaudojamos aerozolio dalelių optiniam storiui gauti (naudojant adaptuotą (Guanter, 2006) sukurtą metodą), atliekama artimos aplinkos (*angl. – adjacency*) korekcija naudojant SIMEC (Sterckx ir kt., 2015) vandeniui ir galiausiai išsprendžiama spinduliuotės perdavimo lygtis (De Keukelaere ir kt., 2018). Aerozolio dalelių storio apskaičiavimo algoritmas remiasi nuotraukos spektrine sausumos gardelių variacija (VITO, 2019). Nuotrauka yra padalinama į 15 km x 15 km dalis ir šiose dalyse priimama atmosferos vientisumo prielaida. Šios dalys taip pat yra pakankamai didelės, kad jose spektrinė variacija būtų didelė. Kiekvienai nuotraukos daliai žemiausia spinduliuotė (*angl. – radiance*) yra nustatoma kiekvienai spektrinei juostai ir šiam aproksimuotam tamsaus objekto spektrui iš MODTRAN5 lentelių (*angl. – look-up tables*) yra gaunama atitinkama atmosferos kelio spinduliuotė. iCOR pagal numatytuosius nustatymus naudoja nekintamą neurbanizuotą (*angl. – fixed rural*) teritorijų modelį (De Keukelaere ir kt., 2018).

Sen2Cor (SENtinel-2 CORection) – atmosferos korekcijos produktas, sukurtas sausumos taikymams. Šis produktas buvo įtrauktas į tyrimą, nes tai yra standartinis dažniausiai naudojamas atmosferos korekcijos produktas Sentinel-2 duomenims apdoroti. Sen2Cor atlieka atmosferos (aerozolio dalelių optinio storio nustatymą, vandens garų nustatymą), vietovės ir plunksninių debesų korekciją, taip pat pirmo lygio (Level 1C) nuotraukos klasifikavimą (Mueller-Wilm ir kt., 2019). Aerozolio dalelių optinis storis nustatomas naudojant tankios tamsios augalijos (*angl. – dense dark vegetation*) algoritmą. Tinkamam algoritmo pritaikymui nuotraukoje turi būti charakteringų tamsios tankios augalijos, tamsios dirvos arba vandens telkinių paviršių (Mueller-Wilm ir kt., 2019).

C2RCC (*angl. – Case 2 Regional Coast Colour*) atmosferos korekcijos algoritmas yra pritaikomas kiekvienai gardelei. Pasitelkus dirbtinius neuroninius tinklus iš viršutinės atmosferos atspindžio (TOA) yra gaunamas vandenį paliekantis atspindys (*angl. – water-leaving reflectance*) ir būdingos (*angl. – inherent*) optinės savybės iš vandenį paliekančio atspindžio. C2RCC yra sudarytas iš devynių dirbtinių neuroninių tinklų: ANN tikrinimui ar TOA



patenka į ANN mokymo duomenų rinkinio diapazoną, atmosferos ANN, ANN normalizuoto atspindžio suskaičiavimui, ANN būdingų vandens optinių savybių nustatymui, ANN vandens atspindžio nustatymui iš būdingų optinių savybių, ANN neapibrėžtumo apskaičiavimui bei ANN spinduliuotės silpimo koeficientui 490 nm bangos ilgyje nustatyti (Doerffer, 2015). C2X (*angl. – Case 2 Extreme*) sudarytas ekstremalioms išsklaidymo ir sugerties sąlygoms (Brockmann ir kt., 2016). C2X duomenų rinkinys buvo sudarytas EKA finansuoto „Case 2 Extreme“ (C2X) projekto metu (Hieronymi ir kt., 2016). Šiame duomenų rinkinyje yra matavimų iš Šiaurės ir Baltijos jūrų bei *Coastcolour* duomenų rinkinio (Doerffer, 2015).

POLYMER – atmosferos korekcijos produktas, sukurtas MERIS jutikliui ir vėliau pritaikytas kitiems palydoviniams jutikliams, pvz., Sentinel-2 MSI (Steinmetz & Ramon, 2018). POLYMER buvo naudojamas sudarant *Ocean Colour Climate Change Initiative* duomenų rinkinį (Plymouth Marine Laboratory, 2012). TOA signalas yra išskaidomas į ozono sugerties, azoto dioksido sugerties ir Reilėjaus sklaidos (865 nm bangos ilgyje) komponentus. Ozono ir azoto dioksido praleidimo faktorius (*angl. – transmittance*) suskaičiuojamas iš ozono ir azoto dioksido stulpelio duomenų, gautų iš pagalbinių duomenų šaltinių (ERA-Interim duomenų bazės iš Europos vidutinės trukmės prognozių centro (*angl. – ECMWF*)). POLYMER tikslas yra panaudoti atmosferos ir saulės atspindžio modelį bei vandens atspindžio modelį likutinio saulės atspindžio, aerolio dalelių sklaidos ir signalo iš vandens atskyrimui. Siekiant geriausio spektrinio atitikimo spektro derinimą sudaro atmosferos modelio ir vandens atspindžio modelio parametrų optimizavimas (Steinmetz ir kt., 2011).

Atmosferos korekcijos neapibrėžtumo vertinimui buvo naudoti spektrai iš 136 ežerų ir tvenkinių esančių Rytų Lietuvoje ir patenkančių į Sentinel-2 MSI T35UMB regiono aprėptį. Šie ežerai sudaro 38 % ežerų, įtrauktų į valstybinio monitoringo programą. Pagal valstybinio monitoringo 2018–2019 m. duomenis chl<sub>a</sub> koncentracija šiuose ežeruose kito intervale nuo 0,2 iki 148,7 mg/m<sup>3</sup>. Pagal Carlsono trofinės būklės indeksą (TSI) 74 % šių ežerų yra mezotrofiniai ir pasižymi žemomis chl<sub>a</sub> koncentracijomis, o kiti yra eutrofiniai ir hipertrofiniai (2.5 lentelė). TSI indeksas apskaičiuotas pagal vidutinę *in situ* išmatuotą chl<sub>a</sub> koncentraciją. 50-iai ežerų TSI apskaičiuotas pagal 2007–2014 m. monitoringo duomenis, 79-iems ežerams naudoti duomenys iš 2015–2018 m. Septyniuose ežeruose AAA matavimų nebuvo, tad trofinei būklei nustatyti buvo pasitelkti literatūros šaltiniai (Balevičienė ir kt., 2009).

**2.5 lentelė.** Trofinė analizuojamų ežerų būklė (Carlson, 1977) pagal valstybinio monitoringo chlorofilo  $\alpha$  koncentraciją 2007–2018 m.

<b>Carlson trofinės būklės indeksas, TSI</b>	<b>Ežerų skaičius</b>	<b>Vidutinė chlorofilo <math>\alpha</math> koncentracija, mg/m<sup>3</sup></b>
Mezotrofinis	100	5.1
Eutrofinis	26	12.7
Hipertrofinis	10	34.0

Atmosferos korekcijos neapibrėžtumo vertinime nagrinėjami atspindžiai 10 m raiškos matomųjų bangų spektrinėse juostose 490 nm (B2), 560 nm (B3), 665 nm (B4) ir 20 m raiškos artimosios infraraudonosios spinduliuotės bangų juostose – 705 nm (B5), 740 nm (B6) ir 783 nm (B7). Palydovinių duomenų raiška buvo suvienodinta į 20 m, kadangi buvo norėta išvengti 705 nm (B5) spektrinės juostos, kurios raiška yra 20 m, raiškos keitimo, kadangi ši spektrinė juosta yra naudojama visuose chla algoritmuose.

7 Sentinel-2 MSI pirmo lygio nuotraukoms buvo pritaikyti 7 atmosferos korekcijos algoritmai (2.4 pav.). Duomenys buvo nufiltruoti remiantis IDEPIX (*angl. – pixel identification*) nuotraukos klasifikatoriumi, pasirenkant tik vandens (gardelės pažymėtos Idepix\_Water, Idepix\_Clear\_Water buvo pasiliktos) gardeles ir kitais 2.4 skyriuje aprašytais metodais. Dauguma pašalintų spektrų, kurie sudarė 36 % visų spektrų (2.6 lentelė), buvo paveikti plunksninių debesų arba priskirti sausumos klasei. Dėl filtravimo pagal B11 spektrinę juostą buvo pašalinti dar 0,87 % TOA spektrų. Dėl atspindžio variacijos filtravimo buvo pašalinti 6,37 % spektrų.

Atskiruose atmosferos korekcijos produktuose buvo pritaikyti papildomi žingsniai galimai klaidingų spektrų nufiltravimui. Acolite ir Acolite Rayleigh buvo pašalinti spektrai, kuriuose pasitaikė neigiamų reikšmių mėlynoje spektrinėje juostoje (B1). POLYMER produkte buvo palikti tie spektrai, kuriems nebuvo priskirtas joks žymuo arba buvo priskirtas *Case2* arba *Inconsistency* žymuo, kaip rekomenduojama produkto aprašyme (Plymouth Marine Laboratory, 2012). Kiti duomenims priskirti žymenys (*Land, Cloud, L1\_invalid, Negative\_BB, Out\_of\_Bounds, Exception, Thick\_aerosol ir High\_Air\_Mass*) lėmė 0,57 % spektrų pašalinimą. Taip pat POLYMER atveju buvo gauta daug spektrų su neigiamomis reikšmėmis, daugiausia neigiamų atspindžio reikšmių gauta 705 nm spektrinėje juostoje (B5) – 80 % spektrų. Taip pat buvo spektrų, kur neigiamos reikšmės buvo 665 nm (B4) ar 740 nm (B6) spektrinėse juostose. Šie spektrai buvo filtruojami pagal šiuos kriterijus: jei vienos ar dviejų gardelių spektruose nustatyta neigiamų reikšmių, buvo pašalinami tų gardelių spektrai, o kiti to ežero ir datos spektrai paliekami. Tais atvejais, kai daugiau nei dviejų gardelių spektrai iš 9 ežero gardelių

pasižymėjo neigiamomis reikšmėmis bet kurioje iš spektrinių juostų, visų 9 gardelių spektrai to ežero ir datos buvo pašalinami. Pasitaikius tokiems atvejams buvo pašalinti 89 % spektrų (2.6 lentelė). Daugiausiai spektrų (19 %) liko 2018-05-07 dienai, o kitoms dienoms liko 6–14 % duomenų. Literatūros šaltiniuose minima, kad naudojantis POLYMER gaunamos neigiamos atspindžių reikšmės – tai gali būti siejama su slėgio pokyčiu, kai vandens telkinio altitudė yra didesnė nei jūros lygis arba dėl algoritmo įvedamos klaidingos vėjo greičio pataisos, o tai lemia sumažintą atspindį (Pereira-Sandoval ir kt., 2019).

C2RCC ir C2X duomenys gauna žymenis, kai spektriniai duomenys nepatenka į algoritmo apmokymo duomenų rinkinio diapazoną arba, kai algoritmu gautos vandens atspindžio reikšmės nepatenka į ANN būdingų vandens savybių duomenų rinkinio diapazoną. Naudojant 7 nuotraukų duomenų rinkinyje tokių spektrų nepasitaikė. iCOR ir Sen2Cor produktuose nėra specifinių filtravimo pasirinkimų.

**2.6 lentelė.** Pradinių spektrų skaičius 136 ežerams (orbita 136) ir 115 ežerų (orbita 036) kiekvienai nuotraukai, taip pat TOA spektrų skaičius, likęs po spektrų filtravimo remiantis IDEPIX nuotraukos klasifikatoriumi ir spektrų skaičius likęs pritaikius kiekvieno atmosferos korekcijos algoritmo filtravimo žingsnius.

Data	Pradinis spektrų skaičius	TOA	Aco-lite	Aco-lite Rayleigh	C2R CC	C2X	iCOR	POLYMER	Sen2Cor
2018-05-02	1224	538	538	538	538	538	538	62	538
2018-05-07	1224	1086	1086	1086	1086	1086	1086	216	1086
2019-07-21	1224	781	781	781	781	781	781	97	781
2019-07-24	1035	357	357	357	357	357	357	9	357
2019-08-20	1224	685	667	667	685	685	685	123	685
2019-08-28	1035	234	234	234	234	234	234	0	234
2019-09-27	1035	972	972	972	972	972	972	0	972

Nufiltravus duomenis buvo naudojami suvidurkinti makrogardelės (3 x 3) spektrai. Pritaikius atmosferos korekcijos algoritmus gauti paviršiaus (vandens) spektrai – atspindžiai spektrinėse juostose. Naudojantis atmosferos korekcijos algoritmais gautais vandens atspindžiais buvo įvertintas atmosferos korekcijos produktų poveikis vandens spektrui. Apskaičiavus chla koncentraciją remiantis 10 chla algoritmų (2.7 lentelė) buvo vertinama, kaip TOA signalo transformacija į vandens atspindį paveikė gautą chla

koncentraciją. Buvo parinkti 10 skirtingos struktūros chla algoritmai siekiant atskleisti jų jautrumą atmosferos korekcijos algoritmo pasirinkimui (2.4 pav.).

Chla koncentracijai, apskaičiuotai iš palydovinių duomenų, validuoti buvo naudojama AAA išmatuota chla koncentracija. Buvo naudojami 30 vienalaikių stebėjimų duomenys, kur tarp *in situ* matavimo ir palydovo praskridimo buvo ne didesnis nei vienos dienos tarpas. Pagal TSI indeksą 21 matavimas buvo užfiksuotas mezotrofiniuose ežeruose, 3 – eutrofiniuose ir likę 6 hipertrofiniuose ežeruose. Chla koncentracija buvo 1,2–148,7 mg/m<sup>3</sup> (vidutinė 18,1 mg/m<sup>3</sup>, standartinis nuokrypis 29,9 mg/m<sup>3</sup>).

Chla algoritmai buvo vertinami pasitelkiant determinacijos koeficientą, vidutinę kvadratinę paklaidą (RMSE), vidutinę kvadratinę logaritmuotą paklaidą (RMSLE), medianinį nuokrypį, ir medianinę absoliutinę procentinę paklaidą (MAPE).

## 2.6. Mašininio mokymosi algoritmai

Ežero klasei nustatyti, remiantis spektriniais požymiais, buvo testuojami 6 prižiūrimo mašininio mokymosi algoritmai. Mašininio mokymosi algoritmų taikymo eigoje testuojami keli skirtingomis savybėmis pasižymintys algoritmai ir parenkamas labiausiai tinkantis konkrečiu atveju konkrečiam duomenų rinkiniui. Dėl to šiame tyrime buvo išbandyti 6 skirtingo sudėtingumo algoritmai.

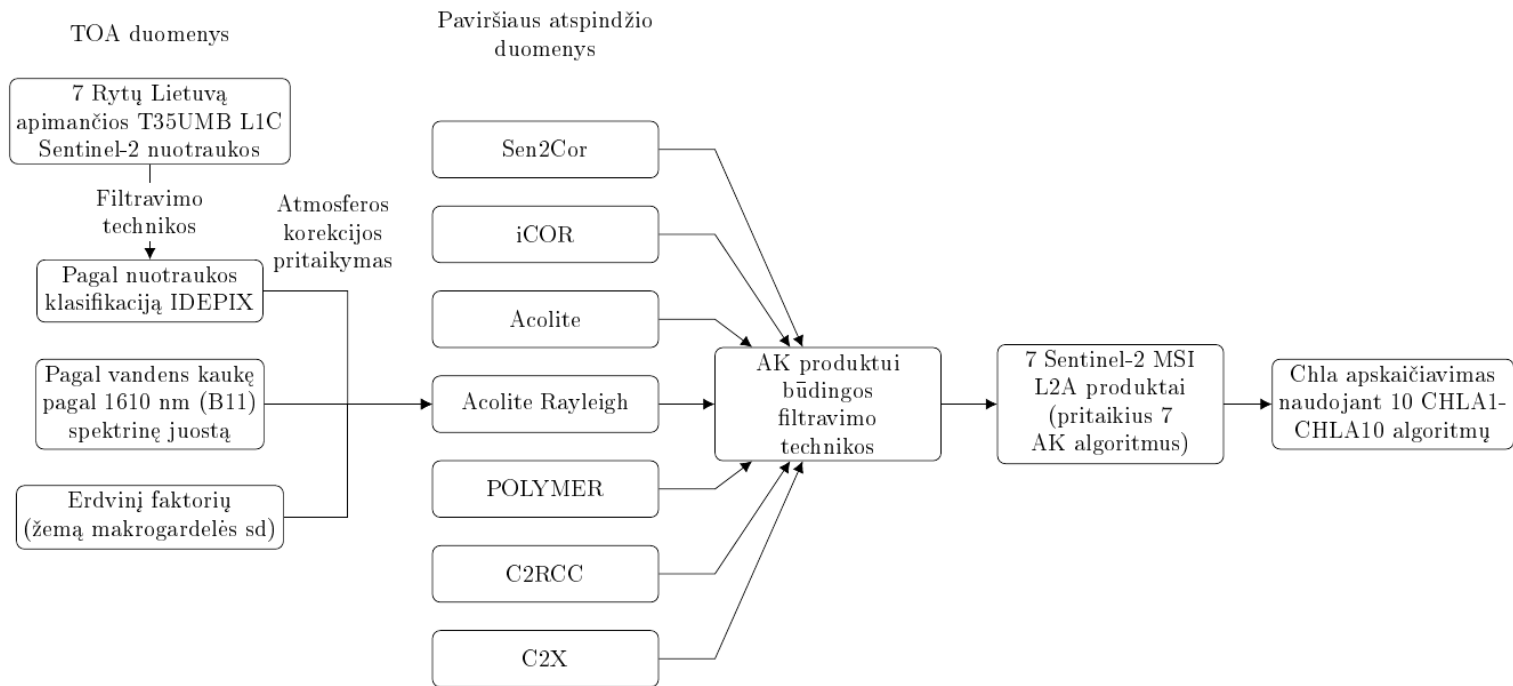
Pirmas iš testuotų algoritmų – logistinė regresija (LR) – tai parametrinis tiesinis modelis, naudojamas tikimybės modeliavimui diskrečiam rezultatų skaičiui. Pirmiausia apskaičiuojama svertinė įvesties (požymių) suma, tada ji paduodama sigmoidinei funkcijai ir grąžinama tikimybė. Po to ji paverčiama binarine išvestimi – 0 arba 1.

Antro iš testuotų algoritmų – atraminių vektorių klasifikatoriaus (SVM) tikslas – rasti hiperplokštumą, geriausiai padalijančią duomenis. Sprendimo funkcija priklauso nuo duomenų (atraminių vektorių), kurie yra arčiau dvi klases skiriančios hiperplokštumos, poaibio. Duomenys gali būti transformuojami naudojant tiesinę, radialinę, polinomine, sigmoidinę ar kitą funkciją. SVM dažnai būna tikslesnis, kai duomenų aibės yra mažos ir kai yra daug požymių (Mountrakis ir kt., 2011).

Atsitiktinio miško klasifikatorius (RFC) yra ansamblinio tipo modelis, kuris naudoja daug sprendimų medžių nuspręsti, kuriai klasei pavyzdys priklauso. Atskiri sprendimų medžiai suskirsto duomenis pagal sprendimų taisykles, pagrįstas požymių duomenimis ir pasirinktomis jų ribinėmis vertėmis. Pavyzdžiui yra priskiriama ta klasė, už kurią balsavo daugiausia sprendimų medžių. Atsitiktiniame miške naudojamas perkėlimo (*angl.* –

**2.7 lentelė.** Chlorofilo  $\alpha$  algoritmai koncentracijos nustatymui pagal Sentinel-2 MSI duomenis. R665 – atspindys 665 nm spektrinėje juostoje, R705 – atspindys 705 nm, R740 – 740 nm spektrinėje juostoje.

Radiometrinio indekso kodas	Radiometrinio indekso lygtis	Literatūros šaltinis	Chlorofilo $\alpha$ algoritmo kodas	Empirinės lygties literatūros šaltinis	Empirinė lygtis	Palydovas, jutiklis	Atmosferos korekcijos algoritmas, naudotas sudarant lygtį	Chlorofilo koncentracija, pagal kurią buvo sudaryta empirinė lygtis, mg/m <sup>3</sup>
ALG1	$R705 - (R665 + R740) / 2$	Toming ir kt., 2016)	CHLA1	Toming ir kt., 2016	$-2231 * ALG1 + 12.7$	Sentinel-2 MSI	Sen2Cor	4–73
ALG2	R705/R665	Mittenzwey ir kt., 1992	CHLA2	Moses ir kt., 2009	$61.32 * ALG2 - 37.94$	MERIS	Ryškos gardelės AK	0–65
			CHLA3	Watanabe ir kt., 2019	$185.34 * ALG2 - 125.9$	Sentinel-2 MSI	Sen2Cor	0–800
ALG3	$(R705 / R665) / (R705 - R665)$	Mishra & Mishra, 2010	CHLA4	Watanabe ir kt., 2019	$0.000001 * ALG3 + 66.038$	Sentinel-2 MSI	Sen2or	0–800
ALG4	R665/R705	Soomets ir kt., 2020	CHLA5	Soomets ir kt., 2020	$ALG4 * (-105.3) + 140.6$	Sentinel-2 MSI	TOA duomenys	0–110
ALG5	$(R665^{(-1)} - R705^{(-1)}) * R740$	Dall'Olmo & Gitelson, 2005	CHLA6	Moses ir kt., 2009	$-232.29 * ALG5 + 23.174$	MERIS	Ryškos gardelės AK	0–65
			CHLA7	Watanabe ir kt., 2019	$474.69 * ALG5 + 67.502$	Sentinel-2 MSI	Sen2Cor	0–800
ALG6	$(R740 / R705) - (R740 / R665)$	Soomets ir kt., 2020	CHLA8	Soomets ir kt., 2020	$-368.5 * ALG6 + 39.1$	Sentinel-2 MSI	TOA duomenys	0–110
ALG7	$(R705 - R665) / (R705 + R665)$	Mishra & Mishra, 2012	CHLA9	Watanabe ir kt., 2019	$((ALG7)^2) * (-1093.2) + 283.47 * ALG7 + 25.947$	Sentinel-2 MSI	Sen2Cor	0–800
ALG8	R705–R665	Grendaitė, 2018	CHLA10	Grendaitė, 2018	$2054 * R705 - 2154 * R665 + 17$	Sentinel-2 MSI	Sen2Cor	2–57



**2.4 pav.** Atmosferos korekcijos produktų neapibrėžtumo vertinimo chlorofilo  $\alpha$  koncentracijai schema.

*bootstrap*) agregavimas, leidžiantis sprendimų medžiams atrinkti duomenis ir taip sukurti skirtingus medžius. Be to, atsitiktiniame miške medžiams leidžiama naudoti ne visus požymius, o tik atsitiktinį jų poaibį, kuris sumažina koreliaciją tarp medžių (Breiman, 2001).

AdaBoost klasifikatorius (Ada) yra adaptyvusis stiprinimo (*angl.* – *boosting*) metodas iš ansamblinių modelių šeimos. Pavadinimas „adaptyvusis“ susijęs su svorių priskyrimu – svoriai priskiriami kiekvienam pavyzdžiui, o neteisingai klasifikuotiems pavyzdžiams priskiriami didesni svoriai (Freund & Schapire, 1996). AdaBoost viduje galima naudoti bet kurią mašininio mokymosi algoritmą, tačiau tyrime naudota numatytoji versija, kurioje naudojami sprendimų medžiai.

XGBoost klasifikatorius (XGB). XGBoost reiškia ekstremalų gradiento didinimą (*angl.* – *eXtreme Gradient Boosting*). Tai sprendimų medžių pagrindu sudarytas ansamblinis modelis, kuriame naudojamas gradientinio stiprinimo metodas (Chen & Guestrin, 2016). XGB algoritmas palaipsniui į sprendimų medį prideda vis daugiau šakų (sąlygų), kad sukurtų geresnį modelį. Apskritai tai greitai ir gerai veikiantis algoritmas.

Dirbtiniai neuroniniai tinklai (ANN) yra sudėtingas algoritmas, kurį sunkiau interpretuoti nei kitus mašininio mokymosi algoritmus, kadangi ANN dažnai turi bent kelis paslėptus sluoksnius (Wang, 2003). Be to, ANN reikia optimizuoti daugybę hiperparametrų. Tačiau daugeliu atvejų ANN pateikia tiksliausius rezultatus, todėl ANN plačiai taikomi daugelyje sričių, įskaitant nuotolinio stebėjimo duomenų analizę (Brockmann ir kt., 2016; Hieronymi ir kt., 2017; Maxwell ir kt., 2018).

## 2.7. Mašininio mokymosi modelių įvertinimui naudotos metrikos

### 2.7.1. Ežerų klasifikacijos modelių vertinimo metrikos

Siekiant įvertinti modelių veikimą ir palyginti juos tarpusavyje, buvo naudotos 7 modelio vertinimo metrikos.

Klaidų matricoje nurodomas tikrų atvejų bei teisingų ir klaidingų modelių nustatytų atvejų skaičius (2.8 lentelė). Šiame tyrime teisingi teigiami (*angl.* – *true positives*) rezultatai – tai teisingai suklasifikuoti skaidrios klasės stebėjimai, teisingi neigiami (*angl.* – *true negatives*) rezultatai – teisingai suklasifikuoti OAS-klasės stebėjimai, klaidingi teigiami rezultatai – tai neteisingai suklasifikuoti atvejai iš skaidrios klasės, o klaidingi neigiami rezultatai – klaidingai klasifikuoti stebėjimai iš OAS-klasės stebėjimų.

**2.8 lentelė.** Klaidų matricos pavyzdys binarinės klasifikacijos atveju.

		Modeliu nustatomas žymuo	
		0	1
Tikras žymuo	0	Teisingi neigiami (TN)	Klaidingi teigiami (FP)
	1	Klaidingi neigiami (FN)	Teisingi teigiami (TP)

Tikslumas (1 lygtis) – visų klasių teisingai nustatytų atvejų dalis.

$$tikslumas = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Preciziškumas (*angl. – precision*) (2 lygtis) – teisingų teigiamų rezultatų ir teigiamai nustatytų modelių atvejų skaičiaus santykis. Tikslumas apibrėžia klasifikatoriaus gebėjimą nepažymėti neigiamų atvejų kaip teigiamų. Pavyzdžiui, šio tyrimo atveju jis parodo, kokia dalis visų modelio rezultatų, pažymėtų kaip skaidri klasė, buvo teisingai nustatyta.

$$preciziškumas = \frac{TP}{TP + FP} \quad (2)$$

Jautrumas (*angl. – recall*) (3 lygtis) – teisingų neigiamų ir neigiamai nustatytų modelių atvejų santykis. Šiuo atveju jautrumas parodo klasifikatoriaus gebėjimą klasifikuoti OAS-klasės pavyzdžius. Geriausias rezultatas yra 1, o blogiausias – 0.

$$jautrumas = \frac{TN}{TN + FN} \quad (3)$$

F1 įvertis (4 lygtis) – tai svertinis preciziškumo ir jautrumo vidurkis. Santykinis preciziškumo ir jautrumo indėlis yra vienodas. Kelių klasių klasifikacijos atveju tai yra kiekvienos klasės F1 įverčio vidurkis, kurio svoris priklauso nuo vidutinio parametro.

$$F1 = 2 \cdot \frac{preciziškumas \cdot jautrumas}{preciziškumas + jautrumas} \quad (4)$$

AUC (*angl. – Area Under Curve*) – plotas po ROC (*angl. – receiver operating characteristic*) kreive, apskaičiuotas pagal modelio rezultatų



įverčius. ROC kreivė vaizduoja du parametrus – teigiamų teisingai nustatytų atvejų dalį (TP/(TP+FN)) ir klaidingų klaidingai nustatytų atvejų dalį (FP/(FP+TN)). AUC rodiklis yra integralas po visa ROC kreive. Reikšmė 0,0 apibūdina modelį, neturintį jokių prognostinių įgūdžių, 0,5 – modelį, kurio rezultatai tiek pat geri, kaip ir atsitiktinis spėjimas (binarinėje klasifikacijoje), o reikšmė 1,0 rodo, kad modelio rezultatai yra 100 % teisingi. AUC rodiklio reikšmė 0,7 rodo, kad yra 70 % tikimybė, jog modelis atskirs dvi klases.

Log\_loss arba kryžminės entropijos nuostolis yra nuostolio funkcija, apibrėžiama kaip neigiama logaritminė tikimybė modelio, kuris gražina mokymo duomenų prognozių tikimybes. Mažesnis log\_loss rodiklis reiškia geresnius modelio rezultatus.

### 2.7.2. Chlorofilo $\alpha$ regresijos modelių vertinimo metrikos

Vertinant chla koncentracijos algoritmus buvo naudojami dažnai regresijos modeliams įvertinti taikomi rodikliai: determinacijos koeficientas ( $R^2$ ), vidutinė kvadratinė paklaida (RMSE), vidutinė logaritmuota kvadratinė paklaida (RMSLE), medianinis nuokrypis (*angl. – median bias*), medianinė absoliuti procentinė paklaida (MAPE). Šių rodiklių skaičiavimo formulės pateiktos 5–9 lygtyse.

$$R^2 = \frac{(\sum_{i=1}^n (m_i - \bar{m})(p_i - \bar{p}))^2}{\sum_{i=1}^n (m_i - \bar{m})^2 (p_i - \bar{p})^2} \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(p_i - m_i)^2}{n}} \quad (6)$$

$$RMSLE = \sqrt{\sum_{i=1}^n \frac{(\log_{10}(p_i) - \log_{10}(m_i))^2}{n}} \quad (7)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (p_i - m_i) \quad (8)$$

$$MAPE (\%) = 100 \times \bar{r} \quad \text{kur } \bar{r} \text{ yra } \frac{|p_i - m_i|}{m_i} \text{ mediana } i = 1, \dots, n \quad (9)$$

Kur  $m_i$  yra išmatuota/tikra parametro reikšmė,  $\bar{m}$  – išmatuotų reikšmių vidurkis,  $p_i$  – modeliu nustatyta reikšmė,  $\bar{p}$  – modeliu nustatytų reikšmių vidurkis,  $n$  – stebėjimų skaičius.

### 3. ATMOSFEROS KOREKCIJOS ALGORITMŲ NEAPIBRĖŽTUMO VERTINIMAS CHLOROFILO A KONCENTRACIJOS NUSTATYMOI IŠ PALYDOVINIŲ DUOMENŲ

Pirmasis rezultatų skyrius padalintas į tris dalis: pirmoje dalyje aprašomas atmosferos korekcijos algoritmų poveikis vandens spektrui (3.1 skyrius), antroje – chl<sub>a</sub> algoritmų neapibrėžtumas pristatomas pagal 136 ežerų duomenis (3.2 skyrius). Trečiame poskyryje chl<sub>a</sub> koncentracija, apskaičiuota panaudojant skirtingais atmosferos korekcijos produktais gautus vandens spektrus, įvertinama pagal 30 vienalaikių *in situ* ir palydovinių stebėjimų duomenis (3.3 skyrius).

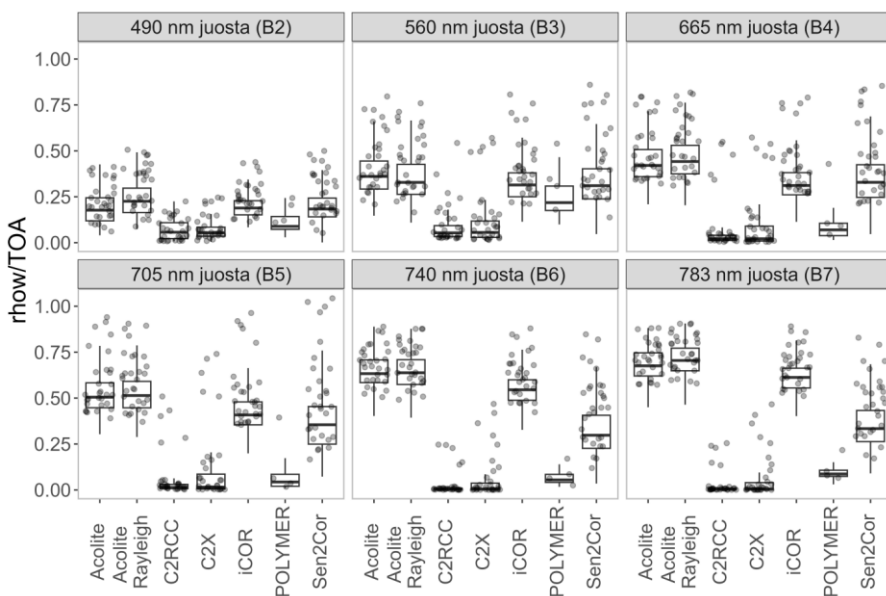
#### 3.1. Atmosferos korekcijos produktų poveikis vandens spektrams

Atmosferos korekcijos (AK) produktai naudoja įvairius metodus atmosferos poveikio pašalinimui iš palydovinių duomenų. AK produktų poveikiui įvertinti pasirinktiems 136 ežerų spektrams, patenkantiems į T35UMB Sentinel-2 MSI regiono aprėptį, buvo apskaičiuoti vandens atspindžio ( $\rho_{\text{H}_2\text{O}}$ ) ir viršutinės atmosferos atspindžio (TOA) duomenų santykiai. Palydoviniai spektrai paimti iš 7 nuotraukų iš 2018 ir 2019 m. Duomenų variaciją lėmė AK produktų poveikis, būdingos optinės ežerų savybės ir jų kaita stebėjimų laikotarpiu.

Nuotrauka paremti produktai (Acolite, Acolite Rayleigh ir iCOR) parodė mažiausius pokyčius trumpesniuose bangų ilgiuose palyginus su kitais AK produktais (3.1 pav.). Pagal šiuos tris produktus apskaičiuotas  $\rho_{\text{H}_2\text{O}}$ /TOA santykis parodė panašų transformacijos pobūdį ir dydį: didėjant bangos ilgiui, visų ežerų ir datų atspindžio santykio mediana palaipsniui didėjo nuo 0,18–0,23 TOA 490 nm spektrinėje juostoje iki 0,61–0,7 TOA 783 nm spektrinėje juostoje. Nuotrauka paremti produktai, panašiai kaip Sen2Cor, sumažina atspindį 490–705 nm juostos diapazone, tačiau 740 ir 783 nm juostose atspindžio reikšmes sumažina kiek mažiau. Sen2Cor pateikia nuoseklią transformaciją visuose bangų ilgiuose.

Naudojant kitus AK produktus, atspindžio pokyčiai pasižymėjo skirtingu transformacijos pobūdžiu ir dydžiu. POLYMER ir C2RCC produktai sumažino atspindį 490 nm ir 560 nm spektrinėse juostose mažiau nei 665–783 nm juostose, tačiau šių dviejų produktų transformacijos dydis labai skiriasi. Naudojant C2RCC produktą gautos itin mažos vandens atspindžio vertės, palyginti su kitais tyrime naudotais AK produktais. Vandens atspindžio mediana ilgujų bangų ruože (740 nm ir 783 nm spektrinėse juostose) sudaro

tik 0,003 TOA signalo. Transformacija 490 nm ir 560 nm juostose yra mažesnė, o mediana sudaro atitinkamai 0,06 ir 0,05 TOA signalo. Palyginus su kitais produktais pagal POLYMER produktą apskaičiuoto  $\rho_{\text{h}}/\text{TOA}$  ir TOA santykio mediana yra maža, tačiau gerokai didesnė nei apskaičiuota naudojant C2RCC. POLYMER produkto  $\rho_{\text{h}}/\text{TOA}$  mediana svyruoja nuo 0,04 iki 0,09 490 nm ir 665–783 nm spektrinėse juostose, tačiau santykio mediana daug didesnė 560 nm spektrinėje juostoje (0,22). Signalas, transformuojamas C2X produktu, pasiskirsto tolygiai –  $\rho_{\text{h}}/\text{TOA}$  mediana svyruoja nuo 0,004 740 nm ir 783 nm juostose iki 0,06 560–705 nm spektrinėse juostose.



**3.1 pav.** Stačiakampė diagrama parodo viršutinės atmosferos (TOA) atspindžio transformaciją į atspindį nuo vandens paviršiaus ( $\rho_{\text{h}}$ ). Apatiniai ir viršutiniai stačiakampio kraštai parodo pirmą ir trečią kvartilius, linija stačiakampio viduje žymi medianą, vertikalūs brūkšniai už stačiakampių kraštų žymi didesnes ir mažesnes reikšmes, bet ne daugiau nei 1,5 tarpkvartilinio intervalo. Taškai reprezentuoja ežerus, kurie buvo įtraukti į chla koncentracijos *in situ* ir palydovinių duomenų lyginamąją analizę.

Mažiausi analizuotų AK produktų atspindžio skirtumai yra 490 nm ir 560 nm spektrinėse juostose (3.1 pav.). Ilgesnių bangų spektrinėse juostose C2RCC, C2X ir POLYMER produktai sumažina TOA signalą daug labiau nei naudojant kitus keturis produktus (3.1 pav.).

665-705 nm spektrinės juostos yra svarbios chla koncentracijai nustatyti vandens telkiniuose. Dažniausiai naudojamos spektrinės juostos yra susijusios

su chla sugertimi raudonųjų bangų ilgiuose ir sklaida raudonosios ribos bangų ilgiuose apie 700–720 nm (Sentinel-2 MSI spektrinė juosta B5, 705 nm). Atspindžio pikas 705 nm spektrinėje juostoje įvertinamas lyginant su atspindžiu gretimose spektrinėse juostose. Sentinel-2 MSI atveju atspindžio pikas 705 nm juostoje paprastai lyginamas su atspindžiu 665 nm ir 740 nm juostose (Ansper & Alikas, 2019; Toming ir kt., 2016).

Vandens gardelių atspindys 665–740 nm bangos ilgių spektrinėse juostose pasižymi dideliu neapibrėžtumu dėl didelių taikomų AK produktų spektro transformacijų skirtumų (3.1 pav.). Naudojant C2RCC ir POLYMER, 75 % ar daugiau  $\rho_{\text{h}}/\text{TOA}$  verčių 665–740 nm spektrinėse juostose yra mažesnės nei 0,11, o daugiau nei 75 %  $\rho_{\text{h}}/\text{TOA}$  verčių, apskaičiuotų naudojant Acolite, Acolite Rayleigh ir iCOR, yra didesnės nei 0,26. Naudojant C2X produktą apskaičiuotų  $\rho_{\text{h}}/\text{TOA}$  verčių intervalas yra platesnis nei naudojant C2RCC ir POLYMER, tačiau  $\rho_{\text{h}}/\text{TOA}$  vertės yra gerokai mažesnės nei apskaičiuotos naudojant nuotrauka paremtus produktus ir Sen2Cor.

7 AK produktų skirtingas atspindžio transformacijas lemia skirtingos AK produktų architektūros. Nuotrauka paremti produktai labiau transformuoja atspindį trumpesnių bangų ilgiuose (490–665 nm) ir mažiau pakeičia atspindį ilgesnių bangų ilgiuose (705–783 nm). Atspindžiai trumpesnių bangų ilgiuose labiau koreguojami dėl Reilėjaus sklaidos, o visi atspindžiai visuose bangos ilgiuose koreguojami dėl aerolio dalelių sugerties. Korekcijos dydžio skirtumus tarp nuotrauka paremtų produktų lemia produktų aerolio dalelių modeliai. Dėl to Acolite, Acolite Rayleigh, iCOR ir Sen2Cor pasižymėjo panašiomis transformacijomis trumpesnių bangų ilgių spektrinėse juostose, o didesni skirtumai pastebimi ilgesnių bangų spektrinėse juostose. Produktai, kuriuose naudojami neuroniniai tinklai – C2RCC ir C2X – bei spektrinio derinimo metodas – POLYMER – pasižymi daug didesnėmis korekcijomis nei nuotrauka paremti produktai. C2RCC ir C2X produktų struktūrą sudaro keli neuroniniai tinklai, o gauti spektrai labai priklauso nuo ANN mokymo duomenų rinkinio. Naudojant duomenims C2RCC ir C2X produktai nepriskyrė ne mokymo intervalo duomenų žymens, tad tyrinėti spektrai buvo panašūs į tuos, kurie buvo naudoti apmokant C2RCC ir C2X ANN. Tačiau, kadangi C2RCC daugiausia buvo mokomas pagal vandenynų ir pakrančių vandenų spektrus, o ne vidaus vandenų, tikėtina, kad analizuotų ežerų optinės sąlygos skyrėsi nuo tų, kurios buvo C2RCC ir C2X mokymo duomenų rinkinyje, o tai lėmė, kad tyrinėti spektrai buvo už algoritmo apibrėžties ribų.

Dauguma analizuojamų ežerų, remiantis valstybinio monitoringo duomenimis, yra mezotrofiniai (100), o visų mezotrofinių ežerų spektriniai profiliai, gauti kiekvienu AK produktu kiekvienai nuotraukai, yra panašios formos ir dydžio (3.2 pav.). Jeigu AK produktai vienodai pašalintų atmosferos

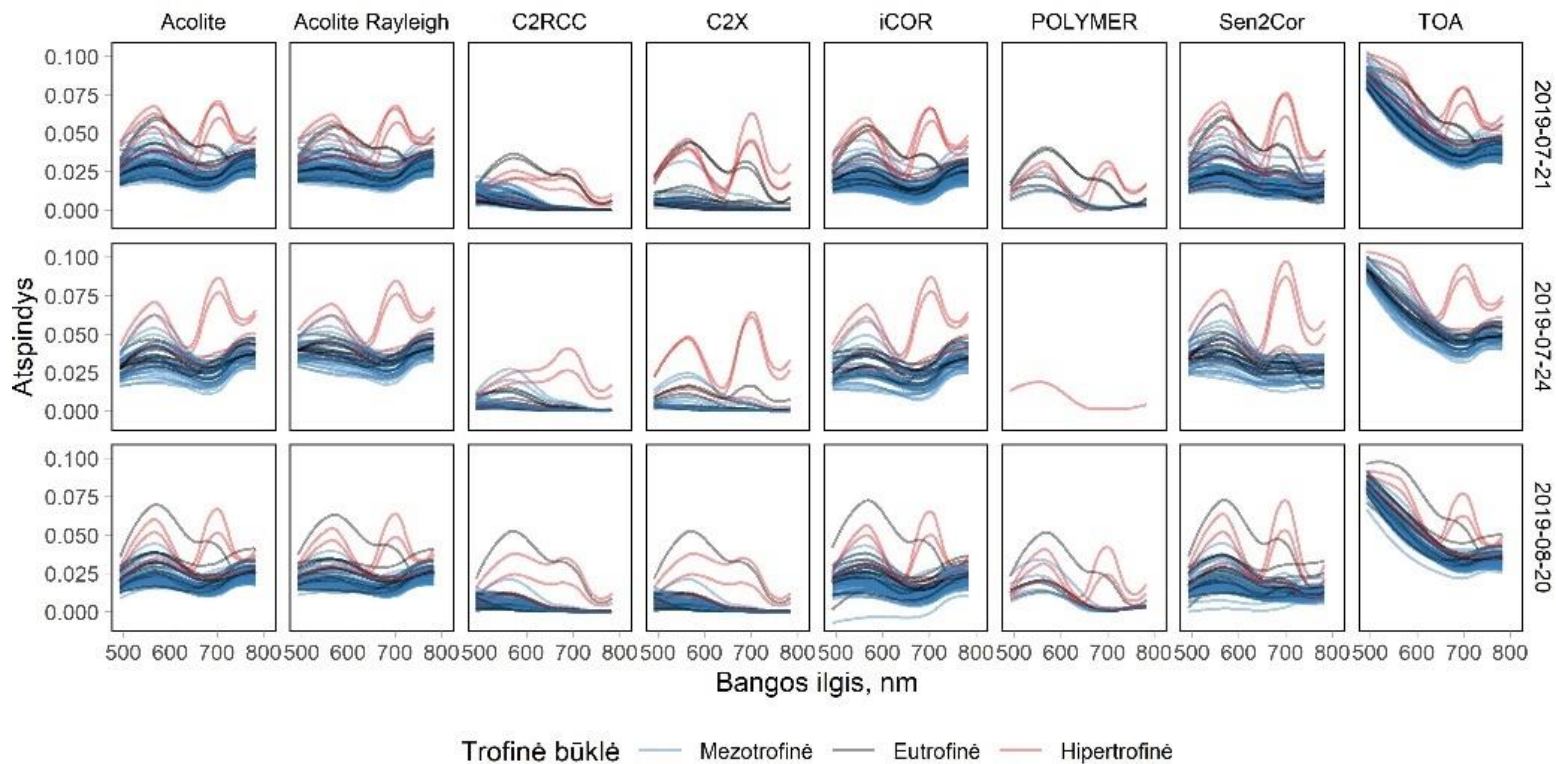
poveikį, ežerų spektrų forma atrodytų taip pat. Visų mezotrofinių ežerų spektrų forma yra panaši, tik atskirų ežerų atveju keičiasi atspindžio dydis. Eutrofiniuose (26) ir hipertrofiniuose (10) ežeruose dumblių žydėjimo metu chl<sub>a</sub> koncentracija gali būti daug didesnė. Tokiu atveju vandens optinės savybės būtų labiau išreikštos ir atspindžiai būtų didesni. 705 nm atspindžio pikas matomas tik kai kuriuose eutrofiniuose ir daugumoje hipertrofinių ežerų (3.2 pav.).

Antroje vasaros pusėje kelių ežerų, daugiausia eutrofinių ir hipertrofinių, 705 nm ir 560 nm spektrinėse juostose stebimas padidėjęs atspindys (3.2 pav.). Aiškus pikas 705 nm juostoje matomas spektriniuose profiliuose, apskaičiuotuose naudojant visus AK produktus. C2RCC atveju atspindys 705 nm juostoje yra daug didesnis nei 740 nm juostoje, tačiau jis panašus į atspindį 665 nm juostoje, kuri, palyginti su spektrais, gautais naudojant kitus AK produktus, gali būti pervertinta.

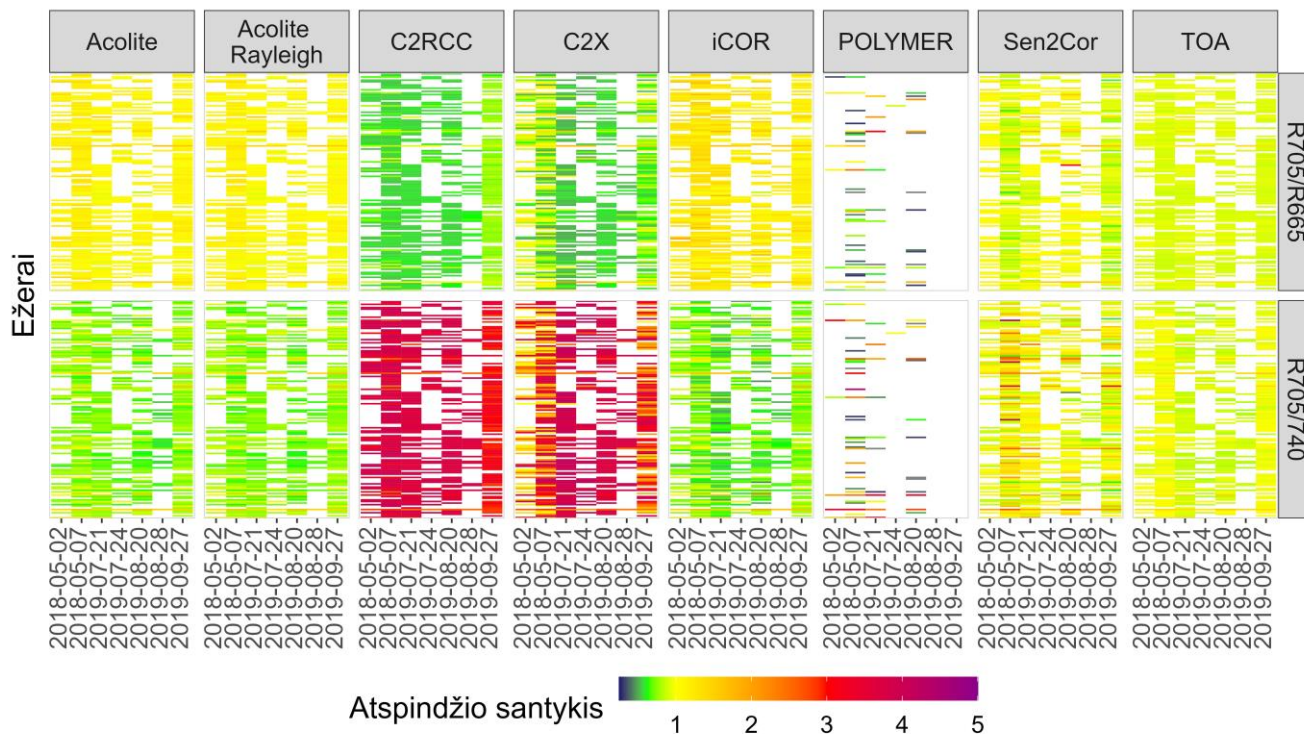
Spektrinių profilių, apskaičiuotų naudojant C2RCC, formos skiriasi nuo profilių gautų naudojant kitus AK produktus. Keliuose eutrofiniuose ir hipertrofiniuose ežeruose C2RCC atspindys 705 nm juostoje yra didelis, tačiau šiuose ežeruose atspindys 490–665 nm juostose yra panašaus dydžio. Kita vertus, šiuose ežeruose atspindžio intensyvumas 705 nm juostoje yra didesnis nei 740 nm juostoje (3.2 pav.).

Skirtumai tarp TOA į rh<sub>ow</sub> transformacijų, gautų pritaikius skirtingus AK produktus, lėmė skirtingą vandens atspindžio spektrinio profilio formą. Siekiant apibūdinti spektrinį profilį chl<sub>a</sub> koncentracijai nustatyti reikšminguose bangos ilgiuose, buvo panaudoti spektrinių juostų santykiai R<sub>705</sub>/R<sub>665</sub> (atspindys 705 nm ir 665 nm juostose) ir R<sub>705</sub>/R<sub>740</sub> (atspindys 740 nm juostoje). Siekiant vizualizuoti vandens atspindžio spektrų formos neapibrėžtumų dydį, susijusį su pasirinktu AK produktu buvo nubrėžti visų analizuotų ežerų bei visų datų R<sub>705</sub>/R<sub>665</sub> ir R<sub>705</sub>/R<sub>740</sub> verčių grafikai (3.3 pav.).

TOA atspindys 665 nm juostoje yra vidutiniškai didesnis nei 705 nm juostoje daugumoje analizuotų ežerų. Atspindys 740 nm juostoje 75 % atveju yra didesnis nei 705 nm juostoje. Sen2Cor produktu gauti 705 nm juostos atspindžiai paprastai yra didesni nei 740 nm juostoje (mediana R<sub>705</sub>/R<sub>740</sub> = 1,03) ir mažesni nei 665 nm juostoje (mediana R<sub>705</sub>/R<sub>665</sub> = 0,97).



**3.2 pav.** 136 ežerų vandens atspindys ( $\rho_{\text{row}}$ ), gautas pritaikius 7 atmosferos korekcijos algoritmus, 3 dienas 2019 m. vasarą, kai didesnio produktyvumo ežeruose vyksta dumblių žydėjimai.



**3.3 pav.** Sentinel-2 R705/R665 (B5/B4) ir R705/R740 (B5/B6) spektrinių juostų santykiai viršutinės atmosferos (TOA) ir vandens atspindžio ( $\rho_{\text{row}}$ ), gauto pritaikius 7 atmosferos korekcijos algoritmus 136 ežerams, naudojant 7 dienų 2018–2019 m. nuotraukas. Kiekviena horizontali linija parodo tam tikro ežero atspindžio santykį nurodytą dieną. Balti plotai parodo trūkstamus duomenis dėl Sentinel-2 padengimo (orbita RO 036) arba dėl pritaikyto duomenų filtravimo.

Nuotrauka parenti produktai transformuoja atspindžius trumpųjų bangų spektro dalyje žymiai labiau nei ilgųjų bangų spektro dalyje. Todėl taikant Acolite, Acolite Rayleigh ir iCOR gautas vandens atspindys 740 nm juostoje yra didesnis nei 705 nm juostoje (mediana R705/R740 = 0,71–0,75), o 665 nm juostoje – mažesnis nei 705 nm juostoje (mediana R705/R665 = 1,07–1,21).

Didžiausias R705/R740 santykis gaunamas naudojant C2RCC ir C2X produktus. Mažiausias juostos santykis R705/R665 taip pat apskaičiuotas naudojant C2RCC ir C2X produktus. C2RCC produkto atveju atspindys 705 nm juostoje yra didesnis (> 2,4 karto) nei 740 nm juostoje ir mažesnis nei 665 nm juostoje visų ežerų ir datų atveju.

R705/R740 ir R705/R665 santykio kitimas skirtinguose ežeruose ir skirtingomis datomis yra didžiausias, kai vandens atspindys apskaičiuojamas POLYMER produktu. Santykiai svyruoja nuo mažų iki labai didelių, o konkrečių ežerų ar skirtingomis datomis gauti atspindžiai nepasižymi nuoseklumu (3.3 pav.). Didelė dalis POLYMER spektrų prieš analizę buvo pašalinta dėl neigiamų atspindžio verčių (2.5 skyrius). Nufiltravus POLYMER duomenis iš pradinio duomenų rinkinio liko tik 11 % duomenų, dėl to turint mažesnę duomenų rinkinį, sunkiau daryti išvadas apie POLYMER rezultatus. Dauguma atvejų POLYMER pateikė neigiamas atspindžio vertes, o atspindžiai buvo teigiami tais atvejais, kai TOA signalas buvo palyginti didelis. Tai rodo, kad POLYMER veikė prastai ežeruose, kuriuose chl<sub>a</sub> koncentracija yra maža. Neigiamus POLYMER rezultatus galėjo lemti neteisingas vėjo greičio įvertinimas koreguojant saulės atspindį, kaip pastebėjo ir Pereira-Sandoval ir kt. (2019). Ansper ir Alikas (2019) nustatė, kad daugeliu atvejų POLYMER nepateikė rezultatų mažų ir siaurų ežerų atveju. Tačiau POLYMER gerai veikia tais atvejais, kai tikimasi didesnio saulės atspindžio (*angl.* – *sun glint*), pavyzdžiui, Ispanijos vandenyse (Pereira-Sandoval ir kt., 2019).

Daugeliu nagrinėtų atvejų spektrinio profilio forma 665–740 nm bangų diapazone labiausiai priklauso nuo pasirinkto AK produkto. Vidutiniškai R705/R740 ir R705/R665 variaciją labiau lemia pasirinktas AK produktas, o ne ežerui būdinga optiškai aktyvių medžiagų, tokių kaip spalvotos ištirpusios organinės medžiagos, bendros suspenduotos medžiagos ir chlorofilo *a*, dinamika.

### 3.2. Chlorofilo *a* koncentracijos neapibrėžtumas

Siekiant patikrinti kaip AK produkto pasirinkimas veikia išvestą chl<sub>a</sub> koncentraciją, visų analizuotų ežerų koncentracija buvo apskaičiuota pagal

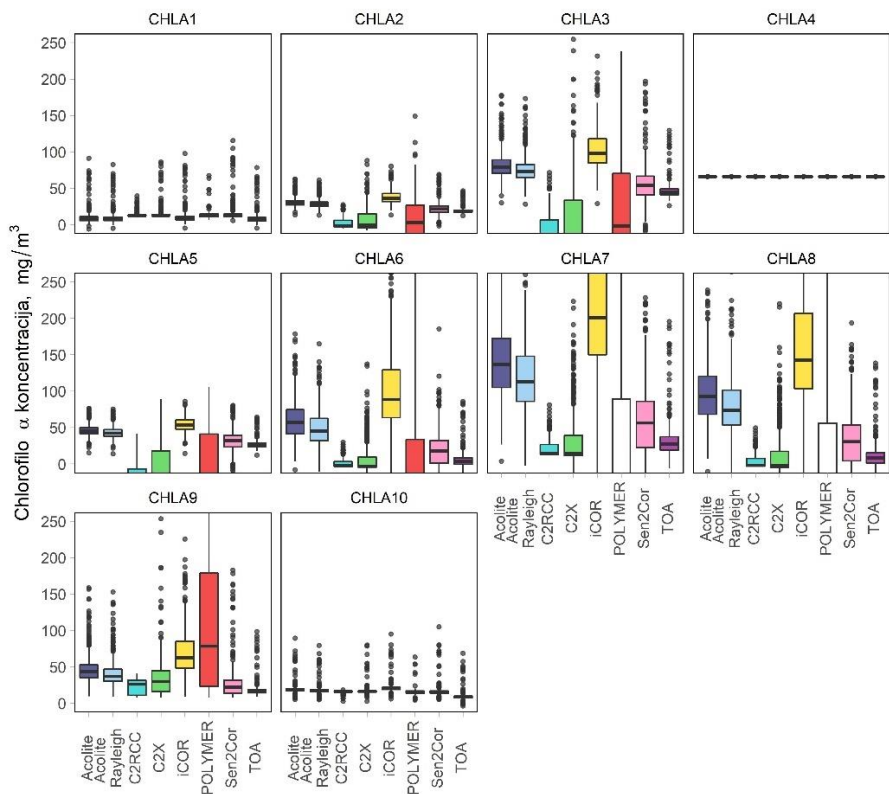


10 literatūroje pateiktų chla koncentracijos nustatymo algoritmų (2.7 lentelė). Algoritmai yra skirtingos struktūros ir paremti dviem (665 nm ir 705 nm) arba trimis (665 nm, 705 nm ir 740 nm) spektrinėmis juostomis. R705/R665 ir R705/R740 neapibrėžtumai persidavė į chla koncentracijos neapibrėžtumą.

Chla algoritmai parodė skirtingą gebėjimą įvertinti chla koncentraciją 136 analizuotuose ežeruose. Siekiant įvertinti chla algoritmų jautrumą AK produkto pasirinkimui buvo palyginta kiekvieno algoritmo chla koncentracijos variacija naudojant skirtingus AK produktus. Mažiausias chla koncentracijos kitimo intervalas, gautas naudojant skirtingus AK produktus, buvo nustatytas naudojant spektrinių juostų skirtumo algoritmus (CHLA1 ir CHLA10). CHLA1 algoritmu gautos koncentracijos mediana (visiems ežerams ir visoms datoms) svyravo nuo 8,1 mg/m<sup>3</sup>, gautos naudojant TOA, iki 12,7 mg/m<sup>3</sup>, gautos naudojant C2RCC, C2X ir Sen2Cor AK produktus (3.4 pav.). CHLA10 gautos koncentracijos medianos buvo didesnės nei naudojant CHLA1 su bet kuriuo AK produktu. Naudojant TOA duomenis, CHLA10 gautos koncentracijos mediana buvo didesnė už gautą naudojant CHLA1 tik 0,5 mg/m<sup>3</sup>. Didžiausias skirtumas (9,1–10,5 mg/m<sup>3</sup>) tarp koncentracijos medianų, gautų naudojant šiuos du chla algoritmus, buvo pastebėtas naudojant nuotrauka paremtus AK produktus – Acolite, Acolite Rayleigh ir iCOR. Chla koncentracijos variacija tarp AK produktų buvo šiek tiek mažesnė naudojant CHLA10 algoritmą (standartinis nuokrypis nuo 1,3 iki 10,7 mg/m<sup>3</sup>) nei naudojant CHLA1 algoritmą (standartinis nuokrypis nuo 2,6 iki 12,4 mg/m<sup>3</sup>).

Labiausiai jautrūs AK produkto parinkimui pagal koncentracijos amplitudės medianą buvo trijų spektrinių juostų CHLA6-CHLA8 algoritmai. Šiais algoritmais ir 7 AK produktais bei TOA duomenimis gautos koncentracijos medianų amplitudės buvo 136,1 mg/m<sup>3</sup> CHLA6 algoritmui ir 276,0 mg/m<sup>3</sup> CHLA7 algoritmui. Šie algoritmai naudojo 740 nm spektrinę juostą ir R705/R740 juostų santykį. AK produktai pasižymėjo didesniu neapibrėžtumu šioje spektrinėje juostoje ir santykyje, o tai atsispindėjo gautoje chla koncentracijoje. 740 nm juostos įtaka išvestai chla koncentracijai yra mažesnė, kai naudojamas 665 nm ir 740 nm spektrinių juostų vidurkis, kaip CHLA1 algoritmo atveju.

Naudojant CHLA6 algoritmą gauta daug neigiamos koncentracijos atvejų: naudojant Sen2Cor ir TOA 24–28 % duomenų buvo neigiami, o naudojant šį algoritmą ir C2X, POLYMER bei C2RCC duomenis – 62–66 % koncentracijų buvo neigiamos. Taip pat 21–62 % duomenų buvo neigiami naudojant CHLA8 duomenis. Mažiau neigiamų verčių gauta naudojant CHLA7 algoritmą – 13 % chla koncentracijų, gautų naudojant Sen2Cor duomenis ir 56 % koncentracijų naudojant POLYMER duomenis.



**3.4 pav.** Chlorofilo  $\alpha$  koncentracijos, apskaičiuotos naudojant viršutinės atmosferos (TOA) ir vandens atspindžio (rho<sub>w</sub>) duomenis ir 10 chlorofilo  $\alpha$  algoritmų. Grafikas gautas naudojant 136 ežerų duomenis ir 7 nuotraukas. Y ašis apima 0–250 mg/m<sup>3</sup> koncentracijos intervalą, neigiamos ir ekstremalios reikšmės neapimtos.

CHLA2 ir CHLA3 algoritmais, į kuriuos įtrauktas dažniausiai naudojamas juostų santykis R705/R665, gautos koncentracijos buvo jautrios naudojamam AK produktui. Naudojant CHLA2 algoritimą gautų koncentracijų medianų amplitudė buvo 38,1 mg/m<sup>3</sup>, o naudojant CHLA3 algoritimą – 115,2 mg/m<sup>3</sup>. CHLA3 algoritmu nustatytas didesnis koncentracijos medianos intervalas nei CHLA2, naudojant skirtingus AK produktus; kadangi CHLA3 lygties koeficientų vertės yra didesnės nei CHLA2, todėl šių koeficientų įtaka chla koncentracijai yra didesnė. Naudojant CHLA2 ir C2RCC bei C2X gauta koncentracijos mediana buvo neigiama. Naudojant C2RCC, C2X ir POLYMER duomenis buvo gauta daug neigiamų chla koncentracijų (atitinkamai 63 %, 52 % ir 49 % koncentracijų). Naudojant CHLA3 algoritimą, su POLYMER duomenimis gauta koncentracijos mediana taip pat buvo

neigiama – 51 % koncentracijų buvo neigiamos, naudojant C2X – 60 % koncentracijų, o naudojant C2RCC – 71 %. Neigiamos koncentracijos buvo gautos, kai R705/R665 santykis CHLA2 algoritmo atveju buvo mažesnis nei 0,62, o CHLA3 algoritmo atveju – mažesnis nei 0,68.

CHLA4 algoritmas neparodė jokio atsako į bet kokio AK produkto naudojimą – standartinis chla koncentracijos, gautos taikant šį algoritmą, nuokrypis buvo lygus 0, naudojant bet kurį AK produktą. Taip nutiko dėl algoritmo konfigūracijos – R705/R665 santykis buvo padalintas iš šių spektrinių juostų bangų ilgių skirtumo, o vėliau padaugintas iš 0,000001 koeficiento ir pridėta 66,038 koeficiento reikšmė. Taikant šį algoritmą gautos vertės buvo lygios pastarajam koeficientui, kadangi juostų R705/R665 santykis tirtuose ežeruose buvo gana mažas. Galima prasto algoritmo veikimo priežastis galėjo būti ta, kad empirinė lygtis buvo sukurta Brazilijos vandens telkiniams, kuriuose chla koncentracija siekė iki  $800 \text{ mg/m}^3$  (Watanabe ir kt., 2019), o šiame tyrime analizuojamuose ežeruose išmatuota koncentracija siekė tik  $150 \text{ mg/m}^3$ .

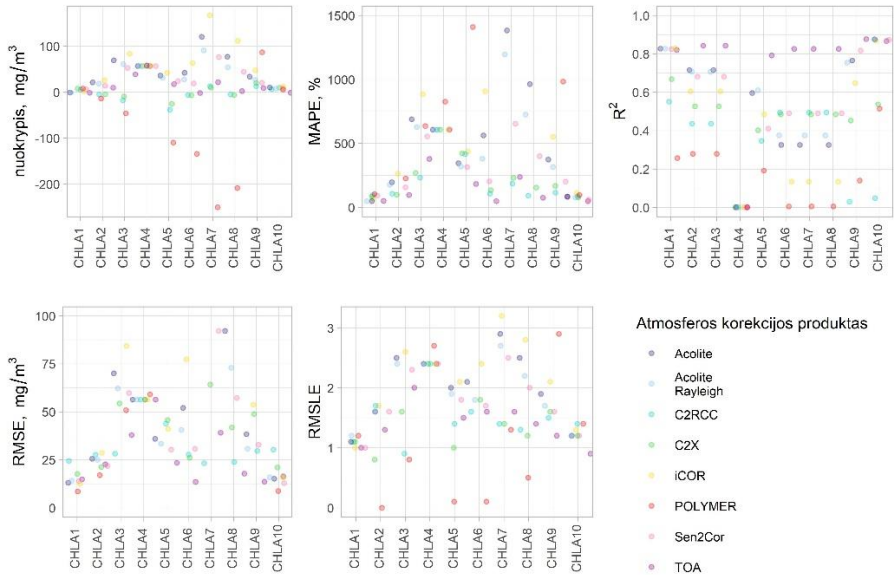
CHLA5 algoritmas, kuriame naudojamas R665/R705 santykis, ir CHLA9 algoritmas, paremtas normalizuotu chlorofilo  $\alpha$  skirtumu, pasižymėjo didele variacija tarp naudojamų AK produktų. CHLA5 algoritmo koncentracija mediana buvo  $92,4 \text{ mg/m}^3$ , o CHLA9 algoritmo –  $62,4 \text{ mg/m}^3$ . Kaip ir CHLA3 atveju, taikant CHLA5 algoritmą taip pat gauta daug neigiamų koncentracijų – naudojant C2RCC, C2X ir POLYMER produktus gautų koncentracijų mediana buvo neigiama. Nustatyta daugiau neigiamų koncentracijų nei naudojant CHLA3 algoritmą: C2RCC – 80 %, C2X – 66 %, POLYMER – 57 % ir Sen2Cor – 3 %. Naudojant CHLA9 algoritmą neigiamų koncentracijų negauta.

Atmosferos korekcijos algoritmai sumažino nedidelį atspindį iš ežerų, o naudojant skirtingus radiometrinius chla algoritmus, neapibrėžtumas išaugo. Dažniau buvo gauta neigiamų reikšmių tais atvejais, kai ežerai mezotrofiniai ir pasižymi mažomis optiškai aktyvių medžiagų koncentracijomis, todėl iš jų TOA atspindys yra mažas. Šiuose ežeruose AK algoritmais gauti R665/R705 ir R705/R665 santykiai maži, gali būti ir neigiami. Tokiu atveju priklausomai nuo chla algoritmo gaunamos ir neigiamos koncentracijos.

### 3.3. Chlorofilo $\alpha$ koncentracijos algoritmų įvertinimas

Chla algoritmai buvo lyginti pasitelkiant 30 *in situ* chla koncentracijos matavimų analizuotuose ežeruose, kurie buvo artimi laike su palydovo stebėjimais (skirtumas tarp duomenų gavimo buvo ne daugiau vienos dienos, duomenys aprašyti 2.5 skyriuje). Chla algoritmai buvo įvertinti naudojant

keletą dažniausiai naudojamų modelio vertinimo rodiklių (3.5 pav.) bei nuokrypio medianą dviejuose chla koncentracijos intervaluose:  $chla < 10 \text{ mg/m}^3$  ir  $chla > 10 \text{ mg/m}^3$ . Nuokrypio mediana buvo pasirinkta, nes tai lengvai interpretuojama metrika ir ją galima pateikti sprendimų priėmėjams, kurie gali būti nesusipažinę su kitomis modelių vertinimo metrikomis.



**3.5 pav.** Chlorofilo  $\alpha$  algoritmų įvertinimo metrikos: medianinis nuokrypis ( $\text{mg/m}^3$ ), vidutinė absoliuti procentinė klaida (%), determinacijos koeficientas,  $R^2$ , vidutinis kvadratinis nuokrypis, RMSE ( $\text{mg/m}^3$ ), vidutinis kvadratinis logaritmuotas nuokrypis, RMSLE.

Pagal modelio vertinimo rodiklius (3.5 pav.) ryšio stiprumas tarp chla koncentracijos, gautos iš palydovinių duomenų ir *in situ* koncentracijos, skyrėsi tarp chla algoritmų ir AK produktų. 17-os modelių determinacijos koeficientas buvo didesnis nei 0,8, tačiau šių algoritmų nuokrypis svyravo nuo  $-0,9 \text{ mg/m}^3$  (CHLA1 su Acolite) iki  $38,6 \text{ mg/m}^3$  (CHLA3 su TOA duomenimis). 15-os algoritmų  $R^2$  vertės svyravo tarp 0,6 ir 0,8, o šių algoritmų nuokrypio mediana buvo nuo  $7,7$  (CHLA1 su C2RCC) iki  $83,5 \text{ mg/m}^3$  (CHLA3 su iCOR). Kitų algoritmų  $R^2$  reikšmės buvo mažesnės (0,0–0,55), o jų nuokrypių mediana kito nuo  $-250,7$  iki  $167 \text{ mg/m}^3$ .

Trijų spektrinių juostų algoritmas (CHLA1) pasižymėjo nedideliu nuokrypiu (iki  $7,8 \text{ mg/m}^3$ ), o keturių AK produktų (Sen2Cor, iCOR, Acolite, Acolite Rayleigh) ir TOA determinacijos koeficientas buvo didesnis nei 0,8.

Mažesnės  $R^2$  vertės buvo stebimos C2RCC ir C2X algoritmams, o POLYMER  $R^2$  buvo mažiausias ( $R^2 = 0,26$ ) ir pasižymėjo didžiausiu nuokrypiu ( $7,8 \text{ mg/m}^3$ ). Kitas dviejų juostų skirtumu pagrįstas algoritmas (CHLA10) parodė panašius rezultatus, tačiau  $R^2$  buvo didesnis ( $0,87\text{--}0,88$ ). Nuokrypis daugeliu atvejų buvo didesnis (iki  $11,9 \text{ mg/m}^3$ ). Tik tuo atveju, kai algoritmas buvo naudojamas su TOA duomenimis,  $R^2$  buvo didesnis, o nuokrypis šiek tiek mažesnis nei naudojant AK produktus ( $-1,6 \text{ mg/m}^3$ ).

Iš dviejų juostų santykio algoritmų (CHLA2-CHLA5) geriausi rezultatai buvo užfiksuoti CHLA2 su TOA duomenimis ( $R^2 = 0,84$ , nuokrypis =  $9,6 \text{ mg/m}^3$ ). Nuokrypis buvo didesnis ( $13,6\text{--}25,8 \text{ mg/m}^3$ ), o  $R^2$  kito  $0,6\text{--}0,71$  intervale, kai CHLA2 algoritmas buvo naudojamas su nuotrauka paremtais AK produktais (iCOR, Sen2Cor, Acolite ir Acolite Rayleigh). Mažesnės  $R^2$  vertės nustatytos ANN paremtiems algoritmams C2RCC ir C2X ( $R^2 = 0,44\text{--}0,53$ ), tačiau juos naudojant gauti mažesni nuokrypiai ( $4,7\text{--}5,3 \text{ mg/m}^3$ ). CHLA3 algoritmas dažniausiai pasižymėjo prastesniais rezultatais, nes visų AK produktų atveju nuokrypiai buvo didesni. CHLA5 atveju, kai buvo naudojamas R665/R705 santykis, geriausias algoritmo veikimas buvo stebimas naudojant TOA duomenis ( $R^2 = 0,79$ , nuokrypis =  $17,6 \text{ mg/m}^3$ ). Nuotrauka paremtų iCOR, Acolite ir Acolite Rayleigh rezultatai buvo prastesni ( $R^2 = 0,48\text{--}0,61$ ), o su Sen2Cor duomenimis algoritmas veikė panašiai kaip su C2X ( $R^2 = 0,40\text{--}0,41$ ), Sen2Cor atveju nuokrypis buvo teigiamas (nuokrypis =  $23,9 \text{ mg/m}^3$ ), o C2X atveju – neigiamas (nuokrypis =  $25,3 \text{ mg/m}^3$ ).

Trijų spektrinių juostų santykio algoritmai (CHLA6 ir CHLA7) pasižymėjo mažesniu determinacijos koeficientu, naudojant 5 AK produktų duomenis: Acolite, Acolite Rayleigh, C2X, iCOR ir POLYMER. Naudojant TOA duomenis CHLA6 ir CHLA7 algoritmai parodė geresnius rezultatus nei dviejų juostų santykio algoritmas CHLA5, tačiau prastesnius nei CHLA2, CHLA3, CHLA9 ir CHLA10. Naudojant visus AK produktus nuokrypis buvo mažesnis algoritmo CHLA6 atveju, o TOA duomenų naudojimas lėmė mažiausią nuokrypį (nuokrypis =  $-2,2 \text{ mg/m}^3$ ).

Kitas trijų juostų algoritmas, CHLA8, pasižymėjo panašiais rezultatais kaip ir CHLA6 ir CHLA7, tačiau nuokrypis dažniausiai buvo didesnis nei CHLA7 algoritmo atveju ir mažesnis CHLA6 atveju, išskyrus tada, kai jis buvo naudojamas su C2RCC ir C2X AK produktais – nuokrypis buvo mažesnis. Naudojant TOA duomenis nuokrypis buvo panašus, tačiau teigiamas, o CHLA6 algoritmo atveju nuokrypis buvo neigiamas.

Panašiai kaip ir CHLA1 su Sen2Cor duomenimis ( $R^2 = 0,82$ ), veikė normalizuoto chla skirtumo algoritmas CHLA9, tačiau nuokrypis buvo mažesnis naudojant CHLA1 algoritmą (nuokrypis =  $6,4 \text{ mg/m}^3$ ) nei CHLA9

algoritmą (nuokrypis = 20,1 mg/m<sup>3</sup>). Naudojant kitus AK produktus, CHLA9 taip pat veikė prasčiau – R<sup>2</sup> reikšmės buvo mažesnės, o nuokrypis – didesnis.

Tik kai kuriais atvejais išmatuotos ir apskaičiuotos remiantis palydoviniais duomenimis koncentracijos gerai sutapo (3.6 pav.). Grafike pavaizduotas *in situ* išmatuotų koncentracijų intervalas 0–150 mg/m<sup>3</sup>. Kai kuriais atvejais matoma mažiau nei 30 taškų, kadangi algoritmai pateikė mažesnes arba didesnes koncentracijas. Kai kuriais atvejais, kai buvo stebimas aukštas determinacijos koeficientas (R<sup>2</sup> > 0,8) (pavyzdžiui, CHLA2 ir CHLA3 su TOA duomenimis, CHLA9 su Sen2Cor bei CHLA10 su Acolite, Acolite Rayleigh, iCOR ir Sen2Cor), iš palydovinių duomenų gautos chla koncentracijos atitikimas su *in situ* matavimais buvo prastas mažų chla koncentracijų intervale (chla < 10 mg/m<sup>3</sup>). Tikėtina, kad aukštą determinacijos koeficientą lėmė palyginti geras duomenų atitikimas didesnių koncentracijų diapazone (chla > 10 mg/m<sup>3</sup>). Be to, kai kurie algoritmai linkę pervertinti visas koncentracijas (pavyzdžiui, CHLA3 ir CHLA7), o kiti negebėjo įvertinti mažesnių koncentracijų (pavyzdžiui, CHLA2 naudojant didžiąją dalį AK produktų). Mažiausias vidutinis nuokrypis – 1,7 mg/m<sup>3</sup> – mažesnių nei 10 mg/m<sup>3</sup> koncentracijų intervale nustatytas naudojant CHLA1 ir TOA duomenis, o didesnėms nei 10 mg/m<sup>3</sup> koncentracijoms mažiausias vidutinis nuokrypis – 0,0 mg/m<sup>3</sup> – nustatytas naudojant CHLA1 algoritmą ir Sen2Cor duomenis (3.1 lentelė).

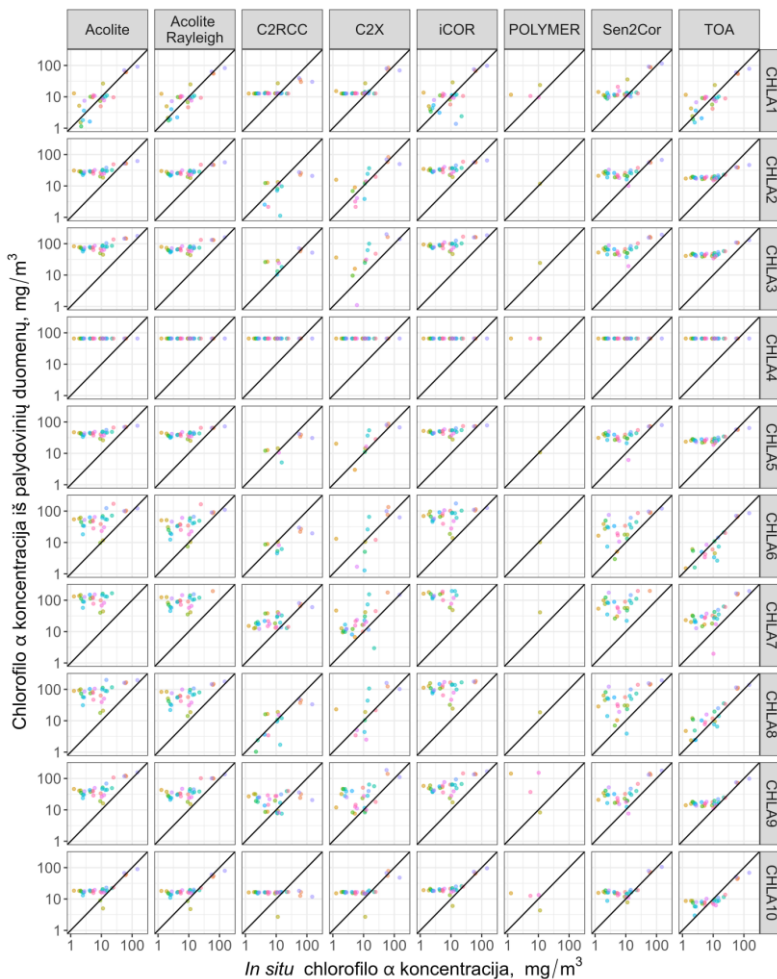
Kadangi AK produktai sumažino atspindį trumpųjų bangų spektro dalyje labiau nei ilgųjų bangų spektro dalyje, skirtumas tarp atspindžio 665 nm juostoje ir 705 nm juostoje padidėjo. 705 nm ir 665 nm spektrinėmis juostomis pagrįsti algoritmai (CHLA2 ir CHLA3) pervertina chla koncentraciją, ypač kai naudojami nuotrauka paremti AK produktai, kuriuose naudojama interpoliacija tarp tamsiausių ir šviesiausių nuotraukos gardelių (Vanhellemont, 2019). Tyrimo rezultatai parodė, kad algoritmai, į kuriuos įtraukti santykiai su 740 nm juosta (CHLA6-CHLA8), pasižymėjo didžiausiu neapibrėžtumu, nes jie pateikė daug neigiamų chla koncentracijos verčių ir pasižymėjo didžiausiu koncentracijos svyravimu. AK produktai pateikė gana skirtingas 740 nm juostos transformacijas ir šios juostos įtaka yra daug didesnė, kai skaičiuojamas santykis, nei tada, kai 740 nm juostos atspindys naudojamas vidurkiui su 665 nm juosta apskaičiuoti, kaip CHLA1 algoritmo atveju. Nustatyta, kad trijų juostų santykio algoritmai veikia prasčiau nei dviejų juostų santykio algoritmai, kai buvo naudojami didesni duomenų rinkiniai (Neil ir kt., 2019).

Dauguma tyrime išbandytų chla algoritmų buvo gauti naudojant Sentinel-2 duomenis (išskyrus CHLA2 ir CHLA6) ir Sen2Cor AK produkto duomenis

**3.1 lentelė.** Apskaičiuotas medianinis nuokrypis remiantis 30 matavimų dviem koncentracijos intervalams  $chl\alpha < 10 \text{ mg/m}^3$  ir  $chl\alpha > 10 \text{ mg/m}^3$ ) 10 chlorofilo  $\alpha$  algoritmų. Pilkai pažymėti laukeliai žymi žemiausius medianinius nuokrypius kiekvienam chlorofilo  $\alpha$  (CHLA1-CHLA10) algoritmui.

Atmosferos korekcijos produktas	Chlorofilo $\alpha$ koncentracijos intervalas	CHLA1	CHLA2	CHLA3	CHLA4	CHLA5	CHLA6	CHLA7	CHLA8	CHLA9	CHLA10
Acolite	Chl $\alpha$ < 10	2,1	23,6	69,3	62,5	38,5	42,1	110,1	71,4	33,5	13,5
Acolite Rayleigh		2,2	21,1	61,0	62,5	34,4	28,7	82,3	49,7	26,5	11,9
C2RCC		9,5	-4,1	-18,7	62,5	-38,7	-4,4	14,2	-2,7	17,4	12,3
C2X		9,3	-5,1	-22,1	62,5	-43,9	-7,2	8,6	-7,2	23,9	12,9
iCOR		2,9	27,9	82,5	62,5	44,5	65,9	157,8	108,8	46,3	15,4
POLYMER		7,8	-13,9	-46,3	62,6	-109,9	-134,3	-250,7	-208,7	86,8	10,6
Sen2Cor		7,5	18,5	54,5	62,5	30,0	19,8	64,2	35,7	21,2	11,7
TOA		1,7	14,6	39,9	62,5	22,3	-1,9	21,3	2,3	12,2	3,2
Acolite		Chl $\alpha$ > 10	-3,2	12,9	69,1	51,2	25,9	41,1	140,2	91,9	33,0
Acolite Rayleigh	-4,0		11,6	61,3	51,2	23,1	27,2	104,6	64,3	25,6	2,9
C2RCC	-2,0		-15,1	-2,0	51,2	-25,6	-15,8	8,6	-12,2	-2,9	-2,0
C2X	2,0		-1,0	18,3	51,2	3,3	-7,7	19,8	-1,9	16,0	2,8
iCOR	-1,5		15,8	86,0	51,2	32,1	52,6	183,3	120,9	50,3	5,0
POLYMER	5,9		-13,4	-30,1	55,3	-90,3	-178,9	-334,2	-275,2	69,5	-1,9
Sen2Cor	0,0		6,1	51,7	51,2	17,1	18,7	86,5	46,6	19,3	2,6
TOA	-4,6		4,3	32,9	51,2	10,6	-5,7	33,1	3,7	2,5	-6,1

(išskyrus CHLA5 ir CHLA8, kurie buvo gauti naudojant TOA duomenis). Algoritmai, kurie iš pradžių buvo išvesti naudojant MERIS duomenis, pasižymėjo gerais rezultatais, kai buvo naudojami su TOA duomenimis (3.5 pav.). CHLA8 algoritmas (Soomets ir kt., 2020) parodė gerus rezultatus naudojant TOA duomenis, o naudojant AK produktus koncentracijos dažniausiai buvo pervertintos (3.6 pav.). Geriausiai veikė juostų skirtumu pagrįsti algoritmai (CHLA1 ir CHLA10), nes jie gerai atitiko *in situ* duomenis naudojant TOA, Sen2Cor ir nuotrauka paremtus Acolite, Acolite Rayleigh ir iCOR AK produktus (3.5 ir 3.6 pav.). Šie algoritmai parodė stabiliausius rezultatus, susijusius su AK produktų parinkimu, išskyrus tuos atvejus, kai naudojami ANN paremti AK produktai C2RCC ir C2X.



**3.6 pav.** Išmatuotos ir algoritmais gautos chlorofilo  $\alpha$  koncentracijos ryšio grafikai (skalės logaritmuotos).



Tyrimo rezultatai parodė, kad naudojant chla algoritmus su TOA duomenimis apskaičiuota chla koncentracija gerai atitiko *in situ* chla duomenis ( $R^2 > 0,8$  iš 10 algoritmų ir nuokrypis  $< 2,5 \text{ mg/m}^3$  naudojant CHLA1, CHLA6, CHLA8 ir CHLA10 algoritmus). Šie rezultatai rodo, kad taikant AK produktą gali nepavykti gauti tikslesnių atspindžio duomenų, todėl tais atvejais, kai AK produkto validavimui nėra *in situ* spektrų, pasirinkimas naudoti TOA duomenis gali būti patikima išeitis siekiant sumažinti išvestinių biofizinių duomenų neapibrėžtumą. Rezultatai rodo, kad santykinai didesnės paklaidos pastebimos esant mažesnei chla koncentracijai, todėl patikimesnę chla koncentraciją iš nuotolinio stebėjimo duomenų būtų galima išvesti taikant algoritmus ežerų, turinčių panašų optiškai aktyvių medžiagų kiekį, grupei (Spyrakos ir kt., 2018), kaip tai buvo padaryta Neil ir kt. (2019), Soomets ir kt. (2020) ir Pahlevan ir kt. (2021) tyrimuose.

## 4. BIOFIZINĖ EŽERŲ KLASIFIKACIJA

### 4.1 Biofizinių ežerų klasių išskyrimas

Siekiant išskirti ežerų biofizines klases, kurios vėliau naudojamos chla koncentracijos algoritmų kūrimui, buvo naudojama AAA išmatuota chla koncentracija, vandens skaidrumas ir suspenduotos medžiagos koncentracija. Kiekvienam stebėjimui buvo priskirta biofizinė klasė, kuri vėliau buvo naudojama mašininio mokymosi algoritmams apmokyti. Vandens skaidrumas apibūdina ekologinę vandens telkinio būklę ir yra tiesiogiai susijęs su optiškai aktyviomis medžiagomis – chla, suspenduota medžiaga ir ištirpusia spalvotąja organine medžiaga. Atsižvelgiant į ežero tipą pagal gylį yra išskiriamos 5 ekologinės būklės klasės pagal skaidrumą. Ekologinė seklaus ežero būklė (1 tipas, 2.1 lentelė) yra gera arba labai gera, kai skaidrumas yra didesnis nei 1,3 m. Vidutinio gilumo (2 tipas) ir gilių (3 tipas) ekologinė būklė laikoma gera arba labai gera, kai skaidrumas yra didesnis nei 2 m. Atitinkamai mažesnis nei 1,3 m skaidrumas sekliuose ežeruose ir mažesnis nei 2 m skaidrumas yra siejamas su vidutine, bloga arba labai bloga ežero ekologinė būkle (LR Aplinkos Ministerija, 2019). Pigmento chla koncentracija yra geras trofinės būklės indikatorius, todėl klasių sudarymui buvo pasirinktos dvi chla slenkstinės reikšmės ir stebėjimai suskirstyti į tris grupes. Naudotos slenkstinės reikšmės yra siejamos su skirtinga ežero trofine būkle pagal Carlsono trofinės būklės indeksą (Carlson, 1977): oligotrofiniai ir mezotrofiniai ežerai (mažo biologinio produktyvumo ežerai pasižymintys žema chla koncentracija  $< 7,2 \text{ mg/m}^3$ , eutrofiniai ežerai (didesnio produktyvumo ežerai) siejami su  $7,2\text{--}20 \text{ mg/m}^3$  chla koncentracija, o hipertrofiniai ežerai (didelio produktyvumo ežerai) su aukštesne nei  $20 \text{ mg/m}^3$  chla koncentracija (4.1 lentelė). Taip pat buvo atsižvelgta į kitos optiškai aktyvios medžiagos – suspenduotos medžiagos koncentraciją, pagal kurią stebėjimai buvo dalinami į dvi grupes –  $< 10 \text{ mg/m}^3$  ir  $> 10 \text{ mg/m}^3$ . Suskirsčius stebėjimus į grupes buvo nustatyta 12 klasių (4.1 lentelė).

1\_skaidri klasė, apibūdinama mažomis chla ir SM koncentracijomis ir dideliu skaidrumu, buvo dažniausiai pasitaikanti klasė ir jai priskirta 46 % stebėjimų. Kita klasė – 2\_chla\_skaidri – pasižymėjo didesnėmis chla koncentracijomis, geros būklės skaidrumu ir žemomis SM koncentracijomis. Šios klasės, kuri apėmė 21 % stebėjimų, vandenyse pagrindinė optiškai aktyvi medžiaga yra chla. Dar viena klasė, kurioje vyrauja chla, priskirta 9 % stebėjimų – 3\_chla\_drumsta. Ši klasė pasižymi prastesniu skaidrumu – jų būklė siejama su vidutine, bloga arba labai bloga ekologine klase pagal skaidrumą. Likusios klasės buvo retai pasitaikančios – jos sudarė nuo 0 iki 7

% stebėjimų. Taip pat pabrėžtina, kad stebėjimų, kur vyrauja SM (1\_SM\_drumsta, 2\_chla\_SM\_drumsta) buvo vos keli. Dėl nepakankamo reprezentatyvumo šie stebėjimai nebuvo įtraukti į mokymo-testavimo duomenų rinkinį, naudojamą klasifikacijos algoritmams apmokyti.

Pirmame žingsnyje mašininio mokymosi algoritmais buvo siekiama atskirti geru skaidrumu ir mažomis optiškai aktyvių medžiagų koncentracijoms pasižyminčius matavimus (skaidri klasė, paremta 1\_skaidri klase, 4.1 lentelė) nuo kitų (OAS-klasė). Skaidrių ežerų atskyrimas nuo kitų suteikia galimybę nustatyti matavimus, kurie nesiejami su probleminiais vandenimis. Dėl to binarinėje klasifikacijoje buvo naudojamos dvi klasės – skaidri klasė ir OAS-klasė. Klasės, kuriose matavimų buvo vos keli, nebuvo įtrauktos į algoritmus, kadangi mašininio mokymosi algoritmai paprastai veikia geriau, kai jų apmokymui naudojamas didelis duomenų kiekis. O mažas duomenų kiekis šiose klasėse lemtų netinkamą jų klasifikavimą. Binariniam klasifikacijos uždaviniui naudotos klasės buvo skirtingo dydžio – 46 % stebėjimų skaidrioje klasėje ir 54 % stebėjimų OAS-klasėje.

Antrame klasifikavimo žingsnyje OAS-klasė buvo dalinama į tris klases (4.1 lentelė):

- Geros arba labai geros ekologinės skaidrumo klasės ežerai nepaisant kiek didesnės chla koncentracijos ( $7,2\text{--}20\text{ mg/m}^3$ ). Chla yra vyraujanti optiškai aktyvi medžiaga šiuose vandenyse. Priskirta tarpinė klasė.
- Didžioji dalis matavimų pasižymi dideliu drumstumu dėl išaugusios dumblių, o kai kuriais atvejais dėl išaugusios dumblių ir SM koncentracijos. Priskirta klasė – Chla-klasė.
- Didesnio drumstumo ežerai dėl kitų priežasčių nei didelė chla ar SM koncentracija. Tikėtina, kad mažas skaidrumas šiuose ežeruose yra nulemtas didesnio ištirpusios spalvotosios organinės medžiagos kiekio. Priskirta klasė – drumsta klasė.

Šiame žingsnyje stebėtas didesnis klasių disbalansas – 39 % atvejų buvo priskirta tarpinei klasei, 43 % – Chla-klasei, o likę 18 % – drumstai klasei. Klasių disbalansas – skirtingas duomenų skaičius vienoje klasėje – galintis turėti įtakos rezultatams, kai modelis dažniau prognozuoja dažniau pasitaikančią klasę. Vienas iš būdų klasių disbalansui kompensuoti yra klasių svorių naudojimas (4.2 lentelė). Pritaikius klasių svorius daugiau svorio suteikiama rečiau pasitaikančioms klasėms.

**4.1 lentelė.** Klasių apibrėžimas pagal vandens kokybės parametrus. Binarinei biofizinei klasifikacijos problemai spręsti buvo priskirtos klasės: skaidri klasė arba OAS-klasė, kuriai priskirti stebėjimai su didesnėmis optiškai aktyvių medžiagų koncentracijomis. Klasės, pažymėtos „–“, dėl mažo duomenų kiekio nebuvo įtrauktos į mokymo-testavimo duomenų rinkinį.

Klasė	Skaidrumas*	Chlorofilo $\alpha$ klasė	Suspenduotos medžiagos klasė	Atvejų skaičius	Žymuo binarinėje klasifikacijoje	Žymuo daugianarėje klasifikacijoje
1_skaidri	Skaidru*	< 7,2 mg/m <sup>3</sup>	< 10 mg/m <sup>3</sup>	260	Skaidri	–
1_drumsta	Drumsta			15	OAS-klasė	Drumsta
1_SM_skaidri	Skaidru		≥ 10 mg/m <sup>3</sup>	0	–	–
1_SM_drumsta	Drumsta			1	–	–
2_chla_skaidri	Skaidru	7,2–20 mg/m <sup>3</sup>	< 10 mg/m <sup>3</sup>	119	OAS-klasė	Tarpinė
2_chla_drumsta	Drumsta			38	OAS-klasė	Drumsta
2_chla_SM_skaidri	Skaidru		≥ 10 mg/m <sup>3</sup>	3	–	–
2_chla_SM_drumsta	Drumsta			10	OAS-klasė	Chla-klasė
3_chla_skaidri	Skaidru	> 20 mg/m <sup>3</sup>	< 10 mg/m <sup>3</sup>	31	OAS-klasė	Chla-klasė
3_chla_drumsta	Drumsta			49	OAS-klasė	Chla-klasė
3_chla_SM_skaidri	Skaidru		≥ 10 mg/m <sup>3</sup>	2	OAS-klasė	Chla-klasė
3_chla_SM_drumsta	Drumsta			38	OAS-klasė	Chla-klasė

\* Skaidri klasė apibūdinama > 1,3 m skaidrumu sekliems ežerams ir > 2 m skaidrumu vidutinio gilumo ir giliems ežerams. Drumsta apibūdinama priešingai – skaidrumas < 1,3 ir < 2 m.

#### 4.2 lentelė. Naudoti klasių svoriai biofizinių klasių modelių kūrimo metu.

Binarinė		Daugelio klasių	
Klasės žymuo	Svoris	Klasės žymuo	Svoris
OAS-klasė	0,93	Tarpinė klasė	1,27
Skaidri klasė	1,08	Chla-klasė	1,16
		Drumsta klasė	2,86

#### 4.2. Biofizinės ežerų klasės pagal *in situ* matavimus

Biofizinė ežero klasė buvo nustatyta kiekvieno stebėjimo metu ir galėjo kisti per tyrimo laikotarpį ir skirtingais sezonais. Ežeruose atlikti matavimai buvo suskirstyti į keturias biofizines klases pagal chla koncentraciją, SM koncentraciją ir skaidrumą atsižvelgiant į gylį. Chla ir SM yra optiškai aktyvios medžiagos, o skaidrumas apibūdina bendrą vandens telkinio ekologinę būklę. Informacija apie gylį taip pat prisideda prie vandens telkinio ekologinės būklės apibrėžimo ir gali padėti atskirti natūraliai senus ir eutrofikuosius ežerus.

Skaidrios (*angl. – Clear*) klasės vidutinė chla koncentracija buvo  $3,8 \text{ mg/m}^3$ , vidutinis skaidrumas – 4,5 m (4.3 lentelė). Tarpinė (*angl. – Moderate*) klasė pasžymėjo didesne chla koncentracija (vidutiniškai  $11,7 \text{ mg/m}^3$ ) ir mažesniu nei skaidrios klasės vidutiniu skaidrumu (2,5 m). Kitų dviejų klasių ekologinė būklė dažniausiai buvo vidutinė, bloga arba bloga. Chla-klasei buvo būdingas dumblių žydėjimas, o drumstos klasės vandens telkinių skaidrumas buvo mažas, nepaisant neaukštos chla koncentracijos, kaip Chla-klasėje. Šios klasės buvo naudojamos kaip žymenys (*angl. – target label*) mašininio mokymosi algoritmuose modelių kūrimo metu, t. y., žymuo yra tai, ką modeliu siekiama nustatyti.

#### 4.3. Biofizinės klasės modelių sudarymas

Biofizinės klasės modeliai buvo sudaromi naudojant mašininio mokymosi klasifikacijos algoritmus. Algoritmai pagal jiems pateiktus požymius skaičiuoja ir nustato žymenis, šiuo atveju – biofizines klases. Modelio mokymui buvo naudoti įvairūs spektriniai požymiai, apibūdinantys ežero vandens atspindžio spektrą: atspindžio amplitudė – didžiausio atspindžio 490–865 nm bangų ilgiuose skirtumas su mažiausiu atspindžiu 490–865 nm bangų ilgiuose, spektrinių juostų santykiai R705/R665, R560/R490, R560/R665 ir R560/R705, spektrinių juostų skirtumai BD1 ir BD2, empirinė lygtis, kurioje naudojamas BD2, matomasis regimasis bangos ilgis (*angl. – Apparent Visible Wavelength, AVW*), atspalvio kampas (*angl. – hue angle*), spalva pagal

Forelio-Ulės spalvų skalę (FU) ir mėnuo, kuris galėtų padėti atskirti dumblių žydėjimus, kurie dažniau pasitaiko vasarą (4.4 lentelė).

**4.3 lentelė.** Biofizinių ežerų klasių vandens parametrų (chlorofilo  $\alpha$  koncentracijos, skaidrumo, suspenduotos medžiagos (SM)) reikšmės pagal 2015–2020 m. duomenis.

Klasė	N	Chla intervalas, mg/m <sup>3</sup>	Vidutinė chla koncentracija, mg/m <sup>3</sup>	Skaidrumo intervalas, m	Vidutinis skaidrumas, m	SM intervalas, g/m <sup>3</sup>	Vidutinė SM koncentracija, g/m <sup>3</sup>
Skaidri	260	0,2–7,1	3,8	1,4–11,0	4,5	0,9–7,5	2,5
OAS-klasė:	303						
Tarpinė	119	7,3–19,8	11,7	1,3–6,8	2,7	0,9–8,0	3,7
Chla-klasė	131	11,3–148,1	42,4	0,3–5,0	1,1	1,0–56,0	9,3
Drumsta	53	0,8–19,4	11,6	0,5–1,9	1,3	1,7–9,6	4,9

Modelio kūrimo metu turima duomenų aibė padalinama į mokymo aibę (80 %) ir testavimo aibę (20 %) pagal ežerus (visi vieno ežero stebėjimai patenka į mokymo arba į testavimo aibę, kad modelis neišmokytų konkrečiam ežerui būdingo požymių derinio). Duomenų padalinimas ir testavimo aibės atskyrimas suteikia galimybę patikrinti modelio veikimą ant modeliui kurti nenaudotų duomenų. Testavimo aibė atskiriama ir paliekama iki modelio validavimo etapo. Binariniam uždaviniui spręsti mokymo aibėje buvo 440 stebėjimų, o testavimo aibėje – 123 stebėjimai (iš viso 563). Trijų klasių uždaviniui spręsti buvo naudojamas duomenų rinkinys iš 303 stebėjimų – 245 stebėjimai priskirti mokymo aibei ir 58 stebėjimai testavimo aibei.

Požymiai buvo transformuojami dviem būdais – kad požymių pasiskirstymas būtų priartintas prie Gauso (normaliojo) pasiskirstymo buvo pritaikyta *PowerTransform()* funkcija iš *scikit-learn* paketo ir atliktas duomenų standartizavimas (pagal vidurkį ir standartinį nuokrypį) pagal mokymo duomenų aibę naudojant *scikit-learn* funkciją *StandardScaler()*. Duomenų standartizavimas rekomenduojamas naudojant bet kurį mašininio mokymosi algoritmą.

Kuriant modelius buvo apmokyti šeši algoritmai naudojant tą pačią duomenų padalijimą, paliekant numatytuosius hiperparametrus. Buvo testuojami 6 klasifikacijos algoritmai, kadangi skirtinguose taikymuose geriau veikti gali vienas ar kitas algoritmas. Siekiant geriausio rezultato patikrinami keli variantai ir pagal apskaičiuotas modelių vertinimo metrikas parenkamas

tinkamiausias modelis. Modeliams įvertinti naudoti rodikliai: tikslumas (*angl. – accuracy*), preciziškumas (*angl. – precision*), jautrumas (*angl. – recall*), F1 įvertis, ploto po kreive (AUC) įvertis ir log\_loss įvertis, apskaičiuoti naudojant *scikit-learn* modulį (2.7.1 skyrius).

Modelių įvertinimas vykdytas naudojant stratifikuotą trijų dalių kryžminį validavimą, kuris tinka modelių įvertinimui naudojant mažas mokymo aibes, ir testavimo aibę. Kryžminis validavimas leidžia įvertinti kaip modelio tikslumas priklauso nuo duomenų padalinimo. Metrikų svyravimas skirtingo padalinimo metu parodo modelio jautrumą duomenų padalinimui. Naudojant stratifikuoto daugiasluoksnio validavimo metodą mokymo aibės duomenys padalijami į 3 dalis pagal klasių žymenis. Trijų dalių atveju mokymo duomenų aibė padalijama į tris vienodo dydžio dalis [A, B, C]. Pirmą kartą A ir B dalys naudojamos kaip mokymo duomenų aibė, o C dalis – kaip validavimo aibė, pagal kurią apskaičiuojami modelio įverčiai. Antrą ir trečią kartą kitos dalys naudojamos kaip validavimo rinkiniai.

Sudarius modelius naudojant numatytuosius hiperparametrus buvo atrinktas geriausiai veikiantis modelis. Siekiant pagerinti rezultatus buvo atliktas modelio hiperparametrų optimizavimas, įprastai pagerinantis modelio generalizavimo galimybes. Optimizavimui naudotas mokymo duomenų rinkinys ir stratifikuotas 3 dalių kryžminis validavimas. Hiperparametrų paieška atlikta naudojant python modulį *optuna*. *Optuna* yra greitas automatizuotas hiperparametrų optimizavimo metodas (Akiba ir kt., 2019).

#### 4.4. Svarbūs ežerų spektriniai požymiai biofizinės klasės modelių sudarymui

Modeliams sudaryti naudojamų požymių atranka yra svarbus mašininio mokymosi modelių kūrimo žingsnis. Kuriant modelius pirmiausia buvo panaudoti visi požymiai (4.4 lentelė), kurie atskirai yra dažnai naudojami kituose tyrimuose. Dalis jų yra panašūs (pvz., R705/R665 ir BD2) ir tik atskirais atvejais vieni gali suteikti daugiau informacijos nei kiti. Kuriant mašininio mokymosi modelius svarbu, kad požymiai nebūtų itin panašūs tarpusavyje – nepasižymėtų aukšta tarpusavio koreliacija, kitaip dar vadinama multikolinearumu (*angl. – multi-collinearity*). Multikolinearumo nebuvimas tarp požymių yra svarbi mašininio mokymosi modeliams priimama prielaida. Multikolinearumo nebuvimas padeda gauti prasmingus ir interpretuojamus rezultatus.

**4.4 lentelė.** Mašininio mokymosi algoritmuose ežero klasei nustatyti naudoti požymiai.

Požymio pavadinimas	Požymio išraiška arba naudoti bangos ilgiai, parametrai	Naudotos Sentinel-2 MSI spektrinės juostos	Literatūros šaltinis
Atspindžio amplitudė	$Max(R_{490-865}) - Min(R_{490-865})$	B2–B8A	–
R705/R665	$R_{705}/R_{665}$	B5/B4	Ammenberg ir kt., 2002
R560/R490	$R_{560}/R_{490}$	B3/B2	Hussein & Assaf, 2020
R560/R665	$R_{560}/R_{665}$	B3/B4	Toming ir kt., 2016
R560/R705	$R_{560}/R_{705}$	B3/B5	Sòria-Perpinyà ir kt., 2021
BD1	$R_{705} - \left(\frac{R_{665} + R_{740}}{2}\right)$	$B5 - (B4 + B6)/2$	Toming ir kt., 2016
BD2	$R_{705} - 1.05 \times R_{665}$	$B5 - 1.05 \times B4$	Grendaitė & Stonevičius, 2021
Lygtis_BD2	$2054 * BD2 + 17$	$B5 - 1.05 \times B4$	Grendaitė & Stonevičius, 2021
Matomas regimasis bangos ilgis, AVW	$R_{490}, R_{560}, R_{665}$	B2, B3 ir B4	Vandermeulen ir kt., 2020
Atspalvio kampas	$R_{490}, R_{560}, R_{665}, R_{705}$	B2, B3, B4 ir B5	van der Woerd & Wernand, 2018
FU (Forelio-Ulės skalė)	Atspalvio kampas	–	van der Woerd & Wernand, 2018
Mėnuo	Mėnesiai (nuo 04 iki 10)	–	–

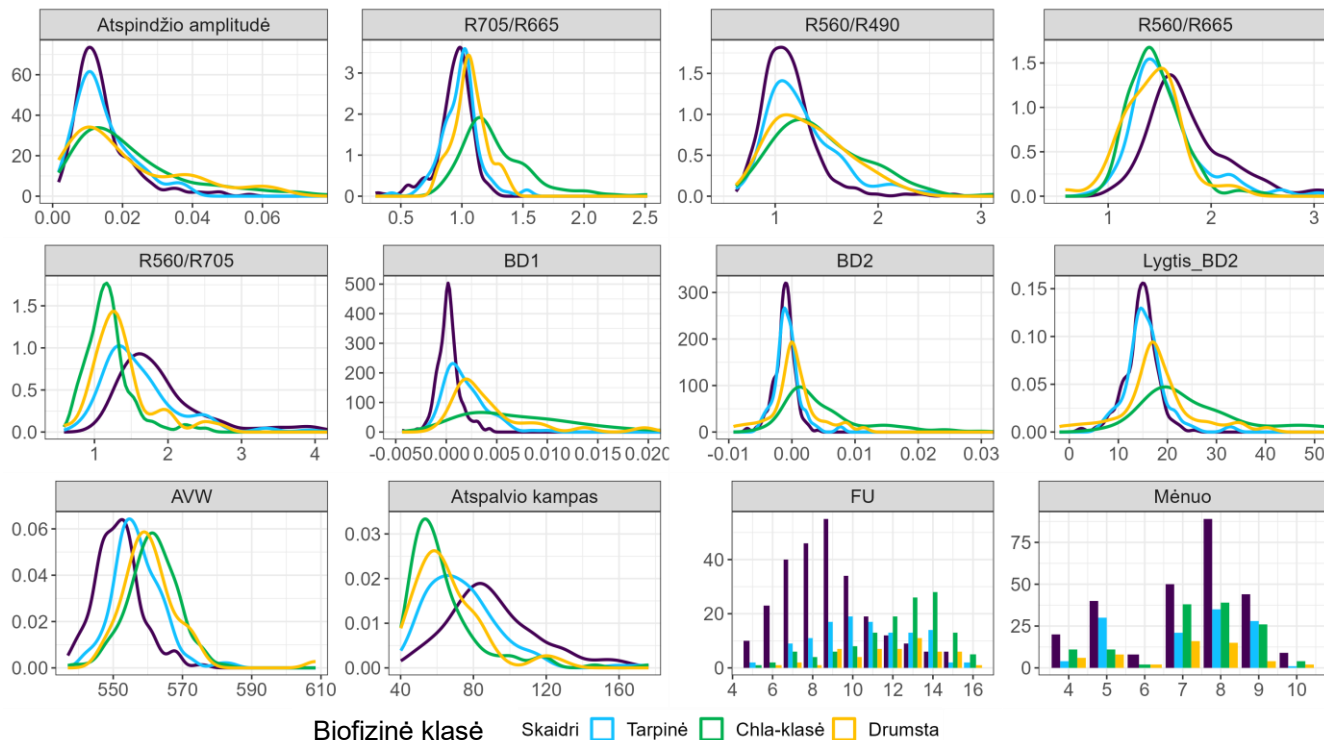


Siekiant atrinkti labiau modelio rezultatus lemiančius požymius ir išvengti multikolinearumo buvo apskaičiuota koreliacija tarp požymių bei tarp požymių ir biofizinės klasės (4.5 lentelė). Naudotus požymius galima suskirstyti į dvi grupes: požymius, kurie yra tiesiogiai apskaičiuoti iš atspindžio duomenų – spektrinių juostų skirtumas, arba juostų santykis – ir kitą grupę – sudėtingesnius požymius, išvestinius iš kelių spektrinių juostų atspindžių (AVW, atspalvio kampas ir FU). Kadangi AVW, atspalvio kampas ir FU tarpusavyje labai koreliuoja ( $R = -0,94, 0,9$  ir  $0,94$ ), buvo nuspręsta naudoti vieną iš jų – AVW, kadangi AVW koreliacija su biofizine klase buvo didžiausia. Be to, požymių pasiskirstymas parodė geresnę klasių atskiriamumą naudojant AVW, nes dominuojantis spalvą apibūdinantis bangos ilgis pereina prie ilgesnių bangų ilgių, kai klasė keičiasi nuo skaidrios per tarpinę į Chla-klasę (4.1 pav.).

Iš kitos požymių grupės buvo pasirinkti BD1 ir R560/R705 požymiai, kadangi jie tarpusavyje mažiau susiję nei kiti požymiai ( $R = -0,23$ ) ir pakankamai gerai koreliuoja su biofizine klase (atitinkamai  $R = -0,38$  ir  $R = 0,3$ ). Renkantis požymius taip pat buvo atsižvelgta į preliminarius rezultatus, gautus taikant logistinę regresiją, nes multikolinearumas blogina šio algoritmo veikimą. Atsižvelgiant į dvinarę logistinę regresiją remiantis pasirinktais požymiais galima paaiškinti gaunamus rezultatus. Padidinus R560/R705 požymio vertę vienu vienetu, tikimybė, kad ežeras bus priskirtas skaidriųjų ežerų klasei, padidėja 1,7. BD1 ir AVW reikšmes padidinus vienu vienetu, tikimybė, kad ežeras bus priskirtas skaidrių klasei, sumažėtų (4.6 lentelė).

**4.6 lentelė.** Į dvinarės logistinės regresijos modelį įtrauktų požymių koeficientai ir šansai nustatyti skaidrią klase.

Požymis	Koeficientas	Šansai
R560/R705	0,56	1,7
BD1	-2,04	0,1
AVW	-0,46	0,6



**4.1 pav.** 12-os požymių pasiskirstymas pagal biofizines klases (skaidri, tarpinė, Chla-klasė, drumsta). Grafikas parodo tam tikros reikšmės tikimybę. Pirmoje ir antroje eilutėje esantys požymiai yra bedimensiniai, išskyrus Lygtis\_BD2 – tai chlorofilo  $\alpha$  koncentracija  $\text{mg}/\text{m}^3$ . Matomo regimojo bangos ilgio (AVW) vienetas yra nm, atspalvio kampo vienetas – laipsniai, FU – klasė pagal Forelio-Ulės spalvos skalę, o paskutinis – stebėjimo mėnesio numeris.

**4.5 lentelė.** Koreliacijos koeficientai tarp požymių ir požymių bei požymių ir klasių žymenų. Mėlyni atspalviai žymi neigiamą koreliaciją, o oranžinės/raudonos spalvos – teigiamą koreliaciją. Didesnės nei 0,08 ir mažesnės nei –0,08 koreliacijos buvo statistiškai reikšmingos ( $p$  reikšmė = 0,05).

	Atspindžio amplitudė	R705/R665	R560/R490	R560/R665	R560/R705	BD1	BD2	Lygtis_BD2	AVW	Atspalvio kampas	FU	Mėnuo	2 klasių žymuo	3 klasių žymuo
Atspindžio amplitudė	1,00	0,30	0,35	0,26	0,04	0,70	0,49	0,49	0,08	-0,10	0,15	0,12	-0,19	0,25
R705/R665	0,30	1,00	0,29	-0,39	-0,58	0,73	0,82	0,82	0,44	-0,38	0,43	0,05	-0,40	0,44
R560/R490	0,35	0,29	1,00	0,04	0,00	0,44	0,34	0,34	0,77	-0,69	0,75	0,15	-0,31	0,32
R560/R665	0,26	-0,39	0,04	1,00	0,83	-0,10	-0,12	-0,12	-0,53	0,37	-0,40	0,23	0,35	-0,35
R560/R705	0,04	-0,58	0,00	0,83	1,00	-0,23	-0,27	-0,27	-0,41	0,28	-0,31	0,15	0,30	-0,31
BD1	0,70	0,73	0,44	-0,10	-0,23	1,00	0,92	0,92	0,39	-0,35	0,42	0,07	-0,38	0,42
BD2	0,49	0,82	0,34	-0,12	-0,27	0,92	1,00	1,00	0,32	-0,28	0,34	0,08	-0,28	0,31
Lygtis_BD2	0,49	0,82	0,34	-0,12	-0,27	0,92	1,00	1,00	0,32	-0,28	0,34	0,08	-0,28	0,31
AVW	0,08	0,44	0,77	-0,53	-0,41	0,39	0,32	0,32	1,00	-0,90	0,94	-0,03	-0,48	0,50
Atspalvio kampas	-0,10	-0,38	-0,69	0,37	0,28	-0,35	-0,28	-0,28	-0,90	1,00	-0,94	-0,02	0,44	-0,43
FU	0,15	0,43	0,75	-0,40	-0,31	0,42	0,34	0,34	0,94	-0,94	1,00	0,02	-0,49	0,49
Mėnuo	0,12	0,05	0,15	0,23	0,15	0,07	0,08	0,08	-0,03	-0,02	0,02	1,00	0,00	-0,02
2 klasių žymuo	-0,19	-0,40	-0,31	0,35	0,30	-0,38	-0,28	-0,28	-0,48	0,44	-0,49	0,00	1,00	-0,86
3 klasių žymuo	0,25	0,44	0,32	-0,35	-0,31	0,42	0,31	0,31	0,50	-0,43	0,49	-0,02	-0,86	1,00

## 4.5. Biofizinės ežero klasės modelių validavimo ir testavimo rezultatai

### 4.5.1. Binarinė klasifikacija

Pirmajame etape buvo išsikeltas uždavinys atskirti ežerus, kuriuose yra mažai optiškai aktyvių medžiagų (chla ir SM), dėl ko jie pasižymi dideliu vandens skaidrumu (skaidri klasė) nuo tų ežerų, kuriuose fiksuotos didesnės optiškai aktyvių medžiagų koncentracijos (OAS-klasė). Toks skirstymas iškart leidžia atskirti ežerus su didesniais optiškai aktyvių medžiagų kiekiais, kurie gali būti potencialiai probleminiai.

Klasifikacijos modelių tikslumas siekė 79–81% naudojant mokymo duomenų rinkinį (4.7 lentelė). Buvo naudojamas trijų dalių kryžminis validavimas, kad būtų sumažinta atskirų modelio paleidimų įtaka ir modelis pasižymėtų geresnėmis generalizavimo galimybėmis. Kryžminio validavimo rezultatai parodė, kad modelių veikimas skirtingo duomenų padalinimo metu skiriasi (standartinis nuokrypis tarp skirtingų duomenų padalinimų siekia 0,05), tačiau standartinis nuokrypis tarp skirtingų duomenų padalinimų AUC įverčiui siekė iki 0,03 SVM ir ANN modeliams ir 0,02 AdaBoost modeliui.

Modelių veikimas iš esmės buvo gana panašus, tačiau, remiantis AUC įverčiu, geriausiai su testavimo duomenimis veikė RFC modelis (4.8 lentelė). Be to, RFC užtikrino mažiausią  $\log\_loss$  reikšmę, o tai rodo gerą modelio veikimą. Binarinės klasifikacijos uždavinyje svarbu ne tik teisingas stebėjimo priskyrimas skaidriai klasei (TP), bet ir tai, kad ežerui, kuriame yra tam tikrų optiškai aktyvių medžiagų, nebūtų priskirta skaidri klasė. Taigi pageidautina, kad klaidingai skaidriai klasei priskiriamų atvejų (FN) būtų mažai. Palyginus modelių veikimą, buvo nuspręsta naudoti RFC modelį, nes jis užtikrino mažiausią klaidingai klasifikuotų ežerų skaičių ( $n = 19$  iš visų 123), taip pat palyginti nedidelį klaidingai priskirtų OAS-klasei atvejų skaičių (FN,  $n = 7$ ). Be to, RFC yra mažiau jautrus požymių multikolinearumui, todėl jis yra geresnis pasirinkimas nei LR, kuris šiame duomenų rinkinyje pasižymėjo panašiais rezultatais, tačiau naudojant LR modelį multikolinearumas gali turėti įtakos. Buvo atlikti bandymai optimizuoti RFC modelio hiperparametrus, tačiau tai nepagerino modelio tikslumo, todėl buvo nuspręsta naudoti pradinę jo versiją. RFC klasifikatoriuje svarbiausias požymis buvo BD1 (santykinė svarba = 0,4), o R560/R705 ir AVW pasižymėjo vienoda svarba (0,30).

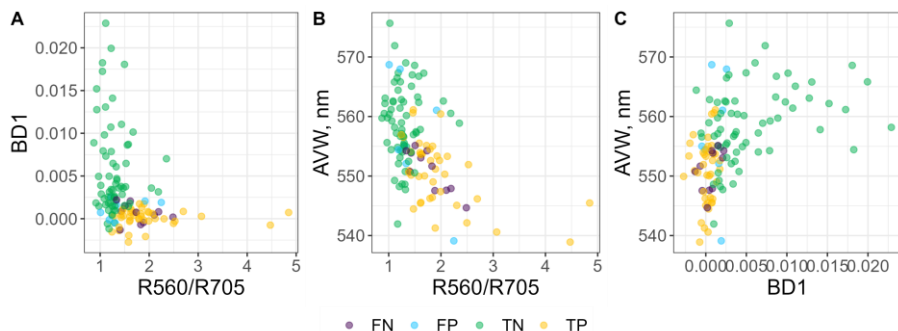
**4.7 lentelė.** Modelio įvertinimo rodikliai be hiperparametrų optimizavimo validavimo etape (naudojant 3 dalių kryžminį validavimą). Pateikiamos vidutinės  $\pm$  standartinio nuokrypio reikšmės.

Klasifikatorius	Validavimo tikslumas	Validavimo preciziškumas	Validavimo jautrumas	Validavimo F1 įvertis	Validavimo AUC	Validavimo Log_Loss
<b>LR</b>	0,81 $\pm$ 0,02	0,76 $\pm$ 0,03	0,88 $\pm$ 0,04	0,82 $\pm$ 0,01	0,90 $\pm$ 0,00	6,0
<b>SVM</b>	0,81 $\pm$ 0,03	0,78 $\pm$ 0,05	0,86 $\pm$ 0,03	0,82 $\pm$ 0,02	0,89 $\pm$ 0,01	5,9
<b>RFC</b>	0,81 $\pm$ 0,00	0,80 $\pm$ 0,02	0,81 $\pm$ 0,03	0,80 $\pm$ 0,01	0,89 $\pm$ 0,01	$9,99 \times 10^{-16}$
<b>Ada</b>	0,79 $\pm$ 0,01	0,77 $\pm$ 0,02	0,83 $\pm$ 0,01	0,79 $\pm$ 0,01	0,87 $\pm$ 0,02	4,8
<b>XGB</b>	0,79 $\pm$ 0,01	0,78 $\pm$ 0,01	0,81 $\pm$ 0,04	0,79 $\pm$ 0,02	0,89 $\pm$ 0,01	3,0
<b>ANN</b>	0,79 $\pm$ 0,03	0,77 $\pm$ 0,02	0,80 $\pm$ 0,12	0,78 $\pm$ 0,05	0,89 $\pm$ 0,01	3,0

**4.8 lentelė.** Modelio įvertinimo rodikliai naudojant testavimo duomenis.

Klasifikatorius	Testavimo tikslumas	Testavimo preciziškumas	Testavimo jautrumas	Testavimo F1 įvertis	Testavimo AUC	Testavimo Log_Loss	Teisingai priskirta OAS-klasė (TN)	Klaidingai priskirta skaidri klasė (FP)	Klaidingai priskirta OAS-klasė (FN)	Teisingai priskirta skaidri klasė (TP)
<b>LR</b>	0,83	0,73	0,87	0,80	0,84	5,9	61	15	6	41
<b>SVM</b>	0,83	0,72	0,89	0,80	0,84	5,9	60	16	5	42
<b>RFC</b>	0,85	0,77	0,85	0,81	0,85	5,1	64	12	7	40
<b>Ada</b>	0,81	0,79	0,70	0,74	0,79	6,5	67	9	14	33
<b>XGB</b>	0,82	0,75	0,81	0,78	0,82	6,2	63	13	9	38
<b>ANN</b>	0,83	0,76	0,81	0,78	0,83	6,0	64	12	9	38

Naudotoje duomenų aibėje buvo 123 duomenys, iš kurių 19 buvo klaidingai suklasifikuoti RFC modelio. Dauguma klaidingai klasifikuotų atvejų buvo iš OAS-klasės (12) ir jiems buvo suteiktas skaidrių ežerų klasės žymuo (klaidingi teigiami) (4.8 lentelė). Svarbiausio požymio BD1 reikšmių pasiskirstymas tarp teisingai klasifikuotų skaidrių ežerų klasės stebėjimų (teisingų teigiamų) rodo, kad reikšmės pasiskirsto apie nulį (mediana = 0,0003) (4.2 pav.), o daugumos teisingai priskirtų OAS-klasės stebėjimų (TN) požymių reikšmės yra didesnės (mediana = 0,004). BD1 klaidingai OAS-klasei priskirtų atvejų (FN) reikšmės (mediana = 0,0008) yra panašios į teisingai skaidriai klasei priskirtų atvejų (TP) reikšmes, o klaidingai priskirtų skaidriai klasei atvejų (FP) reikšmės (mediana = 0,002) yra artimesnės teisingai priskirtų OAS-klasei atvejų (TN) reikšmėms. Panaši situacija stebima ir su kitais požymiais (4.2 pav.). Šių neteisingai suklasifikuotų atvejų chla koncentracijos mediana buvo 10,3 mg/m<sup>3</sup>, o tai rodo, kad daugumai atvejų su didele chla koncentracija, klasė buvo priskirta teisingai (OAS-klasei).



**4.2 pav.** R560/R705 ir BD1 (A), R560/R705 ir AVW (B), BD1 ir AVW (C) požymių ryšys klaidingai priskirtų OAS-klasei (FN), klaidingai priskirtų skaidriai klasei (FP), teisingai priskirtų OAS-klasei (TN) ir teisingai skaidriai klasei priskirtų (TP) stebėjimų ežeruose.

#### 4.5.2. Daugelio klasių klasifikacija

Siekiant atskirti skirtingus ežerus OAS-klasėje toliau buvo sudaromas modelis, paremtas atsitiktinio miško algoritmu, OAS-klasei, kurią sudarė matavimai iš ežerų, turinčių didelį kiekį optiškai aktyvių medžiagų – chla ir SM. Ežerai buvo suskirstyti į tris grupes: tarpinė klasė (turintys šiek tiek chla, tačiau gero ir labai gero skaidrumo), Chla-klasę (drumstus dėl chla ir/ar SM ir pasižyminčius nedideliu skaidrumu) ir drumstus (drumstus dėl kitų priežasčių,

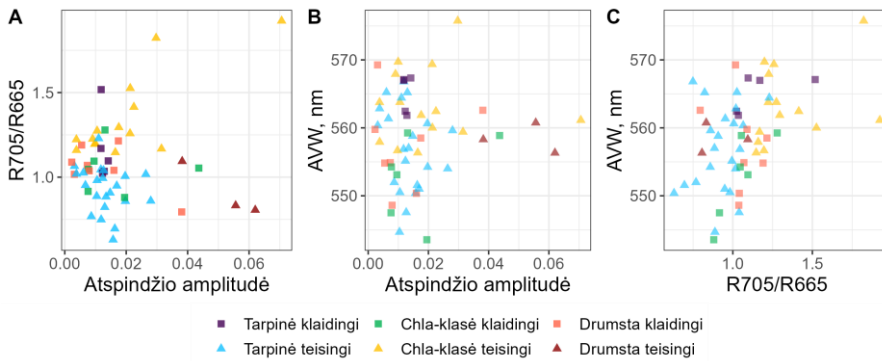
tikėtina dėl didesnio kiekio CDOM). Tokia biofizinė klasifikacija leidžia atskirti silpnus ir stiprius dumblių žydėjimus bei drumstus vandenius.

Kadangi geriausi rezultatai sprendžiant dviejų klasių problemą buvo pastebėti naudojant RFC modelį, buvo nuspręsta RFC pritaikyti ir trijų klasių problemai spręsti. RFC nėra mažiau jautrus požymių multikolinearumui nei kiti algoritmai, o kadangi duomenų rinkinys yra nedidelis, buvo pabandyta RFC naudoti su visais 12 požymių. Tačiau tai nedavė laukiamų rezultatų, tad buvo parinkti požymiai, kurių pasiskirstymas pagal klases šiek tiek skyrėsi (4.1 pav.), ypač atkreipiant dėmesį į skirtumus tarp tarpinės ir Chla-klasių spektrinių požymių pasiskirstymo, kadangi drumsta klasė duomenų rinkinyje nebuvo dažna. Buvo naudojami trys požymiai – atspindžio amplitudė, R705/R665 santykis ir AVW. Didžiausią santykinę svarbą turėjo R705/R665 požymis – 0,38, o kitų požymių santykinė svarba buvo po 0,31. Modelį pritaikius naudojant testavimo duomenis tikslumas tarp klasių svyravo nuo 27 % iki 81 % (4.9 lentelė). Hiperparametrų optimizavimas bendro tikslumo nepakeitė, tačiau pagerino drumstos klasės tikslumą, vis dėlto padidėjo ir kitų klasių klaidingai klasifikuojamų rezultatų skaičius, dėl to buvo nuspręsta naudoti RFC modelio versiją su pradiniais (numatytais) hiperparametrais, kad būtų išlaikytas aukštesnis svarbesnių tarpinės ir Chla-klasės tikslumas.

**4.9 lentelė.** Atsitiktinio miško klasifikatoriaus (RFC) klaidų matrica pagal testavimo (nematyto) duomenų rinkinį, kurį sudaro 58 stebėjimai.

Tikra biofizinė klasė	Modeliu nustatyta biofizinė klasė					Klasifikavimo tikslumas	
	Klasė	Tarpinė	Chla-klasė	Drumsta	Iš viso		
	Tarpinė	22	5	0	27		81 %
	Chla-klasė	5	14	1	20		70 %
	Drumsta	6	2	3	11		27 %

Kai kuriais atvejais klaidingai klasifikuotų stebėjimų spektrinių požymių reikšmės sutampa su kitų klasių požymių reikšmėmis (4.3 pav.). Pvz., klaidingai suklasifikuotų Chla-klasės stebėjimų požymių reikšmės yra artimos teisingai suklasifikuotoms tarpinės klasės stebėjimų reikšmėms. Taip gali atsitikti dėl to, kad klasių riba yra griežta, o panašios požymių reikšmės gali būti susijusios su skirtingomis klasėmis.



**4.3 pav.** Atspindžio amplitudės ir R705/R665 (A), atspindžio amplitudės ir matomo regimojo bangos ilgio (AVW) (B), R705/R665 ir AVW (C) požymių sklaidos diagramos. Spalvos žymi teisingai suklasifikuotus (tarpinės klasės teisingų, Chla-klasės teisingų ir drumstos klasės teisingų) ir neteisingai suklasifikuotus (tarpinės klaidingų, Chla-klasės klaidingų ir drumstos klasės klaidingų) stebėjimus, gautus pagal atsitiktinio miško (RF) klasifikatorių. Atspindžio amplitudė ir R705/R665 yra bedimensiai dydžiai, o AVW išreiškiamas nanometrais.

Tarpinė ir drumsta klasės buvo panašios pagal chla ir panašios į skaidrią klasę pagal SM (4.3 lentelė). Tačiau drumsta klasė buvo panašesnė į Chla-klasę pagal skaidrumo reikšmes. Kai kurie panašumai galėjo lemti klaidingą stebėjimų klasifikavimą, nes klasės buvo formuojamos remiantis griežtomis vandens kokybės parametru ribinėmis vertėmis. Spektrinis drumstos klasės panašumas į skaidrią ir tarpinę klases, kaip matyti iš apskaičiuotų spektrinių požymių (4.1–4.3 pav.), prisidėjo prie mažo klasifikacijos tikslumo (naudojant testavimo duomenis), kuris drumstos klasės atveju buvo 27 %, o skaidrios klasės atveju tikslumas buvo daug didesnis – 85 %. Be to, drumstos klasės stebėjimų buvo mažiau (iš viso 53), o skaidrios klasės stebėjimų buvo 260.

#### 4.6. Biofizinė ežerų klasifikacija naudojant atsitiktinio miško modelį

Biofizinės klasės modelis buvo sudarytas skirtingų stebėjimų ir ežerų atskyrimui, kurie klasės viduje pasižymi panašesnėmis optinėmis savybėmis negu atskiros klasės. Biofizinė klasifikacija naudojama vandens parametru, pvz., chla, tikslesniam nustatymui. Biofizinės ežero klasės nustatymo modelis buvo sudarytas naudojant atsitiktinio miško klasifikacijos modelį, kuris pasižymėjo geriausiomis modelio veikimo metrikomis. Modeliu gaunama,



kuriai iš keturių biofizinės klasių (skaidriai, tarpinei, Chla-klasei ar drumstai) priskiriamas stebėjimas.

Dviejų ežerų klasių modelis buvo pritaikytas 43 022 ežerų spektrams, gautiems iš 357 ežerų septynerių metų 2015–2021 m. laikotarpio. Vienam ežerui teko nuo 36 iki 160 stebėjimų (vidurkis = 121, standartinis nuokrypis,  $sd = 21$ ). Mažiau debesuotais metais ir kai orbitoje buvo jau abu Sentinel-2 palydovai (nuo 2017 m.), vidutinis stebėjimų kiekis per metus buvo 26 ( $sd = 6$ ) 2018 metais ir 29 ( $sd = 5$ ) 2019 metais, kiek daugiau debesuotais metais – 20–21 (2020–2021 m.), o itin debesuotais – 11 stebėjimų per metus ( $sd = 4$ ) (4.10 lentelė).

**4.10 lentelė.** Vidutinis palydovinių stebėjimų skaičius 357 ežeruose ir tvenkiniuose per metus.

Metai	2015	2016	2017	2018	2019	2020	2021
Vidutinis stebėjimų skaičius $\pm$ standartinis nuokrypis	$5 \pm 1$	$10 \pm 3$	$11 \pm 4$	$26 \pm 6$	$29 \pm 5$	$20 \pm 5$	$20 \pm 4$

Biofizinė klasė nustatoma kiekvieno stebėjimo metu. Skaidri klasė buvo nustatyta 58,4 % ežerų spektrų, likusiai daliai spektrų (17 903) buvo pritaikytas trijų klasių modelis ir 18,1 % stebėjimų buvo priskirti tarpinei klasei, 21,6 % – Chla-klasei, o 1,9 % – drumstai klasei (4.11 lentelė). Atskirais metais nustatytas, kuri klasė tais metais vyrauja, buvo įvertinta, kuri biofizinė klasė vyraavo ežere per tyrimo laikotarpį. Skaidri klasė buvo nustatyta kaip vyraujanti 242 ežeruose, iš kurių 161 ežere skaidri klasė vyraavo per visą 2015–2021 m. laikotarpį (4.4 pav.). Šiuose ežeruose pavienių stebėjimų metu buvo fiksuojamos ir kitos klasės. Su vandens žydėjimais siejamos tarpinė, Chla-klasė ir drumsta klasė buvo užfiksuotos atitinkamai vidutiniškai 19,4 %, 14,7 % ir 6,8 % stebėjimų ežere per tyrimo laikotarpį.

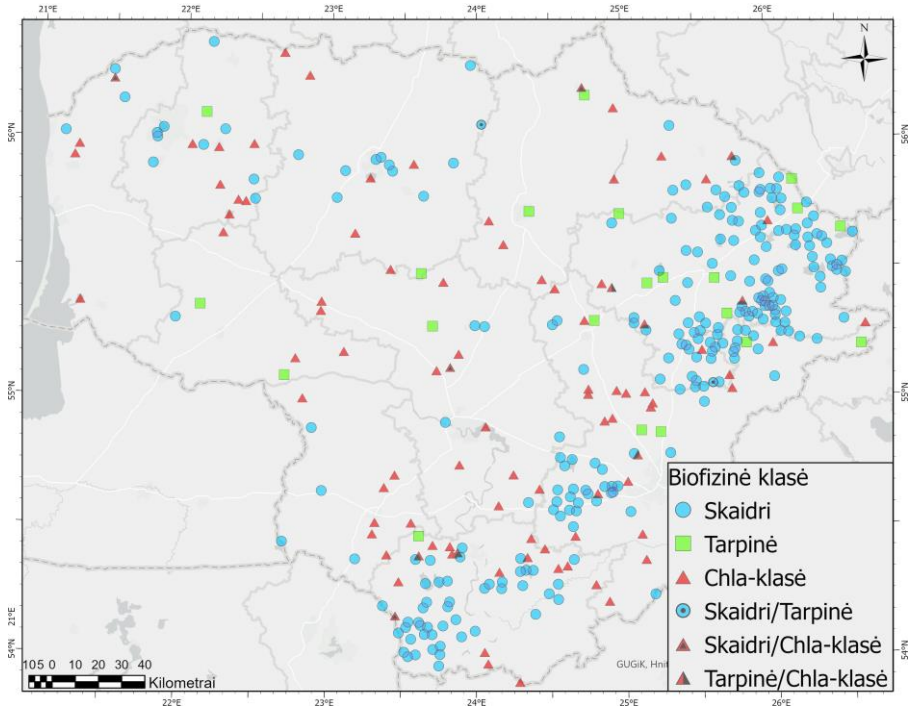
Kai kuriuose ežeruose, kuriuose skaidri klasė nustatyta kaip vyraujanti, atskirais vienais ar dvejais metais vyraavo tarpinė arba Chla-klasė. 81-ame ežere pastebėta klasių kaita tarp skirtingų metų. Dažniausiai biofizinės klasės pasikeitimas buvo stebimas tarp skaidrios ir tarpinės klasių (54 % ežerų). Kituose ežeruose klasė keisdavosi tarp skaidrios ir Chla-klasės (30 % ežerų) arba tarp trijų klasių – skaidrios, tarpinės ir Chla-klasės (16 % ežerų).

Didžiausia vidutinė metinė skaidraus tipo klasės stebėjimų dalis (93,4 %,  $sd = 14,7$  %) buvo 2015 m., o mažiausia (63,3 %,  $sd = 22,3$  %) – 2018 m. Chla-klasė bent kartą stebėta 182 ežeruose 2018 m., taigi ši klasė vidutiniškai sudarė 17,1 % ( $sd = 13,4$  %) stebėjimų, ežeruose, kur per 2015–2021 m. laikotarpį skaidri klasė buvo priskirta kaip vyraujanti. Iš ežerų, kur skaidri

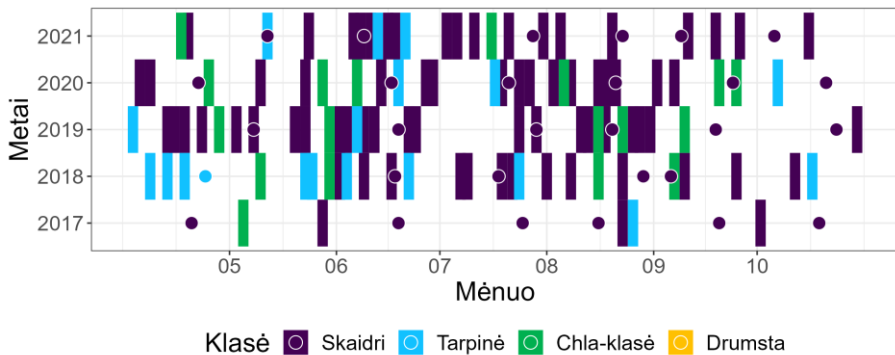
klasė visais metais buvo vyraujanti, 41-ame ežere Chla-klasė užfiksuota iš eilės 2 kartus. Kai kuriais atvejais tai gali pasitaikyti dėl atmosferos drumstumo, vėjo sukkelto vandens nelygumo, kai palydovu užfiksuojamas didesnis atspindys, kas lemia, kad nustatoma Chla, o ne skaidri klasė. Kita vertus, esama skaidrių ežerų, kuriuose yra fiksuojami epizodiniai vandens žydėjimai, pvz., Platelių ežere (Žemaitijos nacionalinis parkas, 2019). Tokių trumpų žydėjimų, kurie gali būti užfiksuoti vos viena palydovine nuotrauka (4.5 pav.), validavimui būtų galima panaudoti papildomą informaciją, surinktą vykdant piliečių mokslo iniciatyvas, žinutes ir nuotraukas iš naujienų portalų, socialinių tinklų. Platelių ežere užfiksuotas vandens žydėjimas 2018 m. birželio mėnesį (Žemaitijos nacionalinis parkas, 2019), kuris matosi ir iš biofizinės klasės erdvinio pasiskirstymo ežere (4.6 pav.). Iš 2018 m. birželio 17 d. užfiksuotos nuotraukos matyti, kad žydėjimas vyko centrinėje dalyje netoli salų (4.6 pav.), tačiau atkreiptinas dėmesys, kad užfiksavimas priklauso nuo naudojamos duomenų raiškos. Skirtingos Sentinel-2 spektrinės juostos yra 10, 20 ir 60 m raiškos, aukščiausios raiškos duomenys leidžia užfiksuoti smulkiausias parametrų pasiskirstymo niuansus, kurie žemesnės raiškos (60 m) duomenyse gali būti ne tokie ryškūs. Dėl to geriau rinktis bent jau 20 m raišką taip išnaudojant palydovinių vaizdų galimybes.

**4.11 lentelė.** Biofizinių klasių pasikartojimo 2015–2021 m. tirtuose 357 ežeruose ypatumai.

Biofizinė klasė		Skaidri	Tarpinė	Chla-klasė	Drumsta
Nustatyta ežerų spektrų, skaičius, %		25119, 58,4 %	7774, 18,1 %	9306, 21,6 %	823, 1,9 %
Klasė nustatyta kaip vyraujanti 2015–2021 m., ežerai (visais metais)		242 (161)	21 (2)	87 (40)	0 (0)
Vyraujanti biofizinė klasė		Skaidri	Tarpinė	Chla-klasė	Drumsta
Vidutinis klasės nustatytų stebėjimų skaičius ežeruose, %	Skaidrios	78,4±20,7	29,8±14,7	24,0±17,9	–
	Tarpinės	19,4±14,5	60,6±13,8	23,5±16,1	–
	Chla-klasės	14,7±14,0	15,8±8,7	64,9±23,4	–
	Drumstos klasės	6,8±4,8	11,0±5,6	10,9±9,4	–

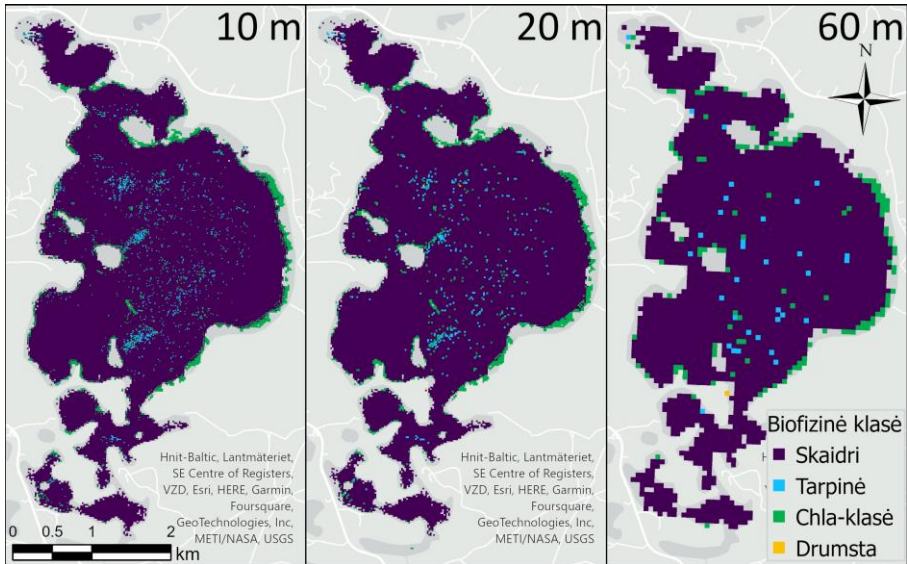


**4.4 pav.** Vyraujanti biofizinė klasė tirtuose ežeruose 2015–2021 m. laikotarpiu.



**4.5 pav.** Biofizinės klasės kaita Platelių ežere 2015–2021 m.

Ežeruose, kuriuose tarpinė klasė nustatyta kaip vyraujanti 2015–2021 m., biofizinė klasė keitėsi labiau ir buvo tik du ežerai (Galuonis ir Vaitiekūnų tvenkinys), kuriuose visais metais 2015–2021 m. vyravo tarpinė klasė (4.11 lentelė). Šiuose ežeruose dažnai buvo nustatoma ir skaidri klasė, rečiau Chla-klasė ir drumsta klasė. Kituose 19 ežerų, kuriuose tarpinė klasė nustatyta kaip vyraujanti 4–5 metus, o kitais metais dažniausiai daugiausia stebėjimų nustatyta Chla-klasė.

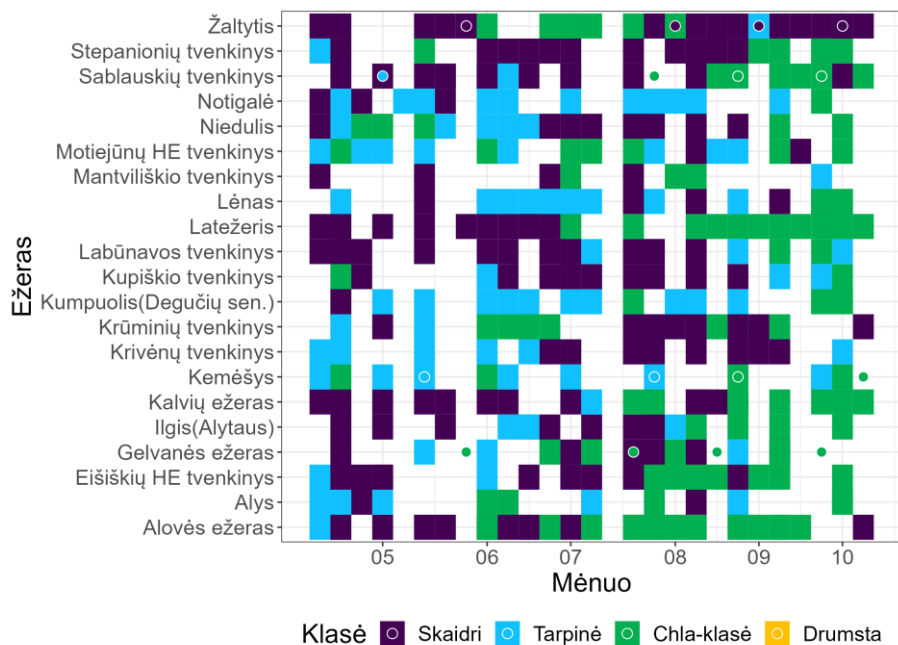


**4.6 pav.** Biofizinės klasės pasiskirstymas Platielių ežere 2018 m. birželio 17 d., nustatytas naudojant skirtingos raiškos (10 m, 20 m ir 60 m) duomenis.

Chla-klasė nustatyta vyraujanti 87 ežerų, o iš jų 40 ežerų Chla-klasė buvo vyraujanti visais septyneriais 2015–2021 m. Šiuose ežeruose kiekvienais metais Chla-klasei buvo priskirta nuo 33 iki 100 % stebėjimų, o tai rodo skirtingą dumblių žydėjimo trukmę ežeruose. Keliuose ežeruose klasė keitėsi labai nežymiai (Ūdrijos ežeras, Amalvas, Petraičių tvenkinys, Papis ir Pravalas (Vidugiris)) – visais metais nuo pavasario iki rudens šiuose ežeruose vyko dumblių žydėjimas ir 81–100 % stebėjimų buvo nustatyta Chla-klasė. Didesnė biofizinės klasės kaita buvo užfiksuota kituose Chla ežeruose, iš kurių 45 % ežerų 2020 m. buvo stebimos kitos klasės nei dominuojanti Chla-klasė. Šiuose ežeruose 2020 m. buvo nuo 8 iki 29 stebėjimų (4.7 pav.). Kai kuriuose ežeruose pavasarį ir rudenį vyravo Chla-klasė, o vasarą stebėtos skaidri ir tarpinė klasės (Niedulio ežeras), kurios kai kuriais atvejais neatitiko *in situ* matavimų (Gelvanės ežeras).

Atlikus vizualinę palydovinių duomenų natūralių spalvų RGB vaizdų patikrą paaiškėjo, kad iš 2015–2020 m. duomenų 43 % klaidingai klasifikuotų atvejų buvo susiję su plonų debesų ar debesų šešėlių poveikiu palydoviniams duomenims. Sen2Cor vaizdų klasifikavimas, kuris buvo pritaikytas šiame tyrime ne vandens gardelėms nufiltruoti, šiais atvejais neteisingai identifikavo plonus debesis ir debesų šešėlius. Nors ši problema yra žinoma, ją išspręsti virš vandens telkinių, kurių signalas yra silpnas, tebėra sudėtinga, o norint padidinti nuotraukos klasifikavimo tikslumą galima naudoti kelis nuotraukos

klasifikavimo algoritmus (Tarrío ir kt., 2020). Tokiais atvejais, kai duomenis paveikė ploni debesys ar ant ežero krintantys debesų šešėliai (pavyzdžiui, Gelvanės ežeras 2020 m., 4.7 pav.), spektrinis signalas buvo sumažėjęs ir chla signalas palydovo nebuvo užregistruotas. Šiame tyrime kuriant modelius buvo naudotos Sentinel-2 MSI nuotraukos, kurių debesuotumas neviršijo 30 %. Naudojant didesnio debesuotumo nuotraukas galima išplėsti mokymo ir testavimo duomenų aibę, tačiau duomenų kokybei užtikrinti būtina atlikti kitokią arba papildomą duomenų kokybės patikrą. Kitais atvejais (43 % atvejų) dažnai tarp *in situ* matavimų ir palydovinio stebėjimo buvo dvejų–trijų dienų laiko skirtumas, todėl sąlygos ežere galėjo pasikeisti, o stebėjimas buvo priskirtas tarpinei klasei.



**4.7 pav.** Klasės pokyčiai pagal *in situ* (apskritimai) ir modelio (kvadratai) duomenis ežeruose, kuriuose per visą tiriamą laikotarpį dominavo Chla-klasė, o grafike pateikta 2020 m. klasės kitimo laiko eilutė, kai šiuose ežeruose buvo stebimos skaidri arba tarpinė klasė.

Nebuvo nustatyta ežerų, kuriuose drumsta klasė būtų buvusi dažniausia bent vienais metais, remiantis palydoviniais duomenimis, nors pavieniais atvejais ši klasė ežeruose buvo identifikuota (817 spektrų). Taip pat buvo 7 ežerai, kuriuose nebuvo išskirta vyraujanti klasė, nes vienodą metų skaičių buvo dvi arba trys klasės. Kai kuriais atvejais metinės klasės parinkimui įtakos turėjo mažas stebėjimų skaičius skirtingais metų laikais, todėl galėjo atsitikti

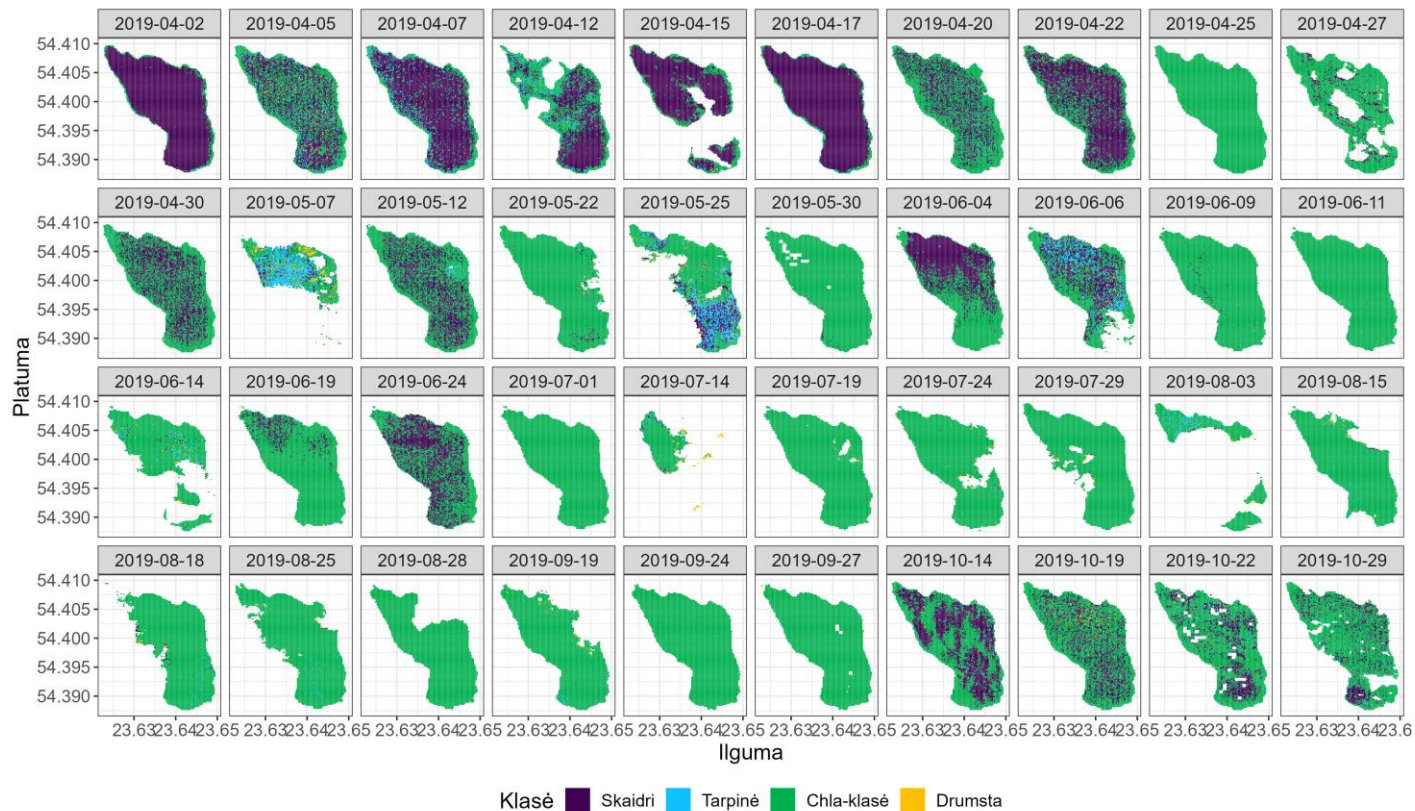
taip, kad vyraujanti klasė buvo nustatyta remiantis pavasario ir rudens stebėjimais, tačiau trūko vasaros stebėjimų, kada kai kuriuose ežeruose tikėtinas dumblių žydėjimas. Be to, buvo atvejų, kai modelis neteisingai klasifikavo klasę, kai buvo prieinami *in situ* duomenys.

Biofizinės klasės pasiskirstymas gali parodyti kaip laike ir erdvėje ežere vystosi vandens žydėjimas. Pvz., Simno ežere per stebėjimų laikotarpį (2015–2021 m.) vyravo Chla-klasė. 2019 m. 69 % stebėjimų buvo nustatyta Chla-klasė, tačiau šie duomenys buvo paremti taškiniais, AAA matavimo vietoje gautais duomenimis. Erdvinė biofizinės klasės analizė leidžia stebėti, kokia ežero dalis buvo apimta žydėjimo. Simno atveju buvo užfiksuota, kad balandžio pradžioje (2019-04-02) visa ežero akvatorija buvo priskirta skaidriai klasei, o iki balandžio vidurio ežere galėjo būti dumblių, kadangi šiaurinėje ežero dalyje 8–36 % gardelių nustatyta Chla-klasė (4.8 pav.). Link balandžio pabaigos (2019-04-25) visa ežero akvatorija buvo paveikta žydėjimo. Kai kurių stebėjimų metu (pvz., 2019-04-12 ir 2019-06-04) tik ežero dalyje vyko dumblių žydėjimas, tad tokiu atveju taškinė monitoringo vieta gali neatspindėti viso ežero situacijos. Gegužės ir birželio mėnesiais keičiantis meteorologinėms sąlygoms žydėjimo intensyvumas svyravo. O nuo liepos pradžios iki rugsėjo pabaigos pagal turimus duomenis visoje ežero akvatorijoje vyko dumblių žydėjimas.

#### 4.7. Rekomendacijos biofizinių klasių vertinimui

Nepaisant įvardintų RFC biofizinės klasės modelio trūkumų, modelis gali būti naudojamas vyraujančiai biofizinei ežero klasei nustatyti (4.4 pav.) ir klasių kaitai per sezoną identifikuoti (4.5, 4.7 ir 4.8 pav.), nustatant silpną (tarpinę klasę) ir stiprų dumblių žydėjimą (Chla-klasę). Biofizinės klasės pritaikymas erdviniam duomenis leidžia stebėti dumblių žydėjimo raidą ežero akvatorijoje ir nustatyti kokia ežero dalis yra veikiama žydėjimo. Biofizinė klasė, įvertinta iš palydovinių duomenų, gali padėti nustatyti dumblių žydėjimo pradžią ir dinamiką, kai stebima klasių kaita, taigi tai gali būti papildomas informacijos šaltinis vandens kokybės tyrėjams.

Tai pačiai problemai spręsti galima panaudoti įvairius mašininio mokymosi algoritmus, tačiau patartina naudoti kuo paprastesnį metodą, kuris pateikia geriausias rezultatus. Šiame tyrime sukurti modeliai, paremti atsitiktinio miško klasifikacijos (RFC) algoritmu, kuris dažnai naudojamas dėl savo paprastumo ir gebėjimo išskirti svarbius požymius. Šiame tyrime, kaip ir kituose chla nustatymo tyrimuose (Gómez ir kt., 2021; Song ir kt., 2015), geriausias rezultatus pateikė RFC algoritmu paremtas modelis, o sudėtingesni



**4.8 pav.** Biofizinēs klasēs pasiskirstymas Simno (LTL196) ežere 2019 m. Kai kuriomis datomis (pvz., 2019-06-14) dalis duomenų buvo nufiltruota dėl debesuotumo.

algoritmai, tokie kaip ANN, pateikia tik šiek tiek geresnius rezultatus (Maier & Keller, 2019).

Sukurtą biofizinės klasės modelį galima taikyti be adaptavimo kituose regionuose, kuriuose vandenys pasižymi panašiomis optinėmis savybėmis. Modelis buvo sukurtas remiantis 226 ežerų stebėjimais ir daugiausia dėmesio skiriama vandenims, kuriuose dominuoja fitoplanktonas. Taip pat į klasifikaciją buvo įtraukta ir drumstų ežerų (dėl kitų medžiagų) klasė, tačiau norint pagerinti šios klasės klasifikavimą, reikalingas geresnis klasės apibūdinimas, pavyzdžiui, naudojant sugerties koeficientą 440 nm bangos ilgyje, kuris gerai apibūdina CDOM kiekį vandenyje (Nima ir kt., 2019). Kituose regionuose 4.1 lentelėje aprašytą klasių apibrėžimą būtų galima papildyti šiuo parametru ir atnaujinti klasių grupavimą mašininio mokymosi modelių apmokymui. Sudarytą modelį būtų galima papildyti daugiau klasių, pavyzdžiui, klase, kurioje dominuoja SM (pvz., 1\_SM\_drumsta klasė, kaip nurodyta 4.1 lentelėje). Tokiu atveju modelį reikėtų iš naujo apmokyti ir apsvarstyti naujų spektrinių požymių įtraukimą.

Biofizinė ežerų klasifikacija gali būti gera pagalbinė priemonė vandenių būklės vertinimo ekspertams. Ji gali padėti nuotoliniu būdu stebėti staigius pokyčius ežere ir padėti priimant sprendimus atlikti matavimus vietoje. Klasifikacija gali pasitarnauti kaip priemonė stebėti dumblių žydėjimo pradžią ir dinamiką. Taip ji gali informuoti apie ežero būklę tarp in situ matavimų. Taip pats vienas svarbesnių biofizinės klasės pritaikymo būdų, naudojamų ir šiame tyrime – biofizinės klasės naudojamos kuriant vandens parametrų nustatymo iš palydovinių duomenų algoritmus, kadangi panašių optinių savybių vandenims algoritmus sukurti galima tikslesnius (Neil ir kt., 2019).



## 5. CHLOROFILO A KONCENTRACIJOS MODELIAI IR JOS KAITOS ANALIZĖ

Trečiame skyriuje buvo pristatytas paprastų chla algoritmų įvertinimas naudojant vandens atspindį, gautą apdorojus palydovinius duomenis įvairiais atmosferos korekcijos algoritmais. Atskleista, kad skirtingais AK algoritmais apdorojus duomenis galima gauti pakankamai skirtingus rezultatus, todėl chla algoritmus gali tekti adaptuoti priklausomai nuo naudojamo AK algoritmo. Taip pat naudojant skirtingus chla algoritmus chla koncentracijos sklaida gali būti žymiai didesnė nei išmatuotų duomenų, o dažniausiai sunku tiksliai įvertinti mažas chla reikšmes dėl vandens telkinių ypatybių bei silpno palydovu užfiksuojamo signalo dėl mažo kiekio optiškai aktyvių medžiagų vandenyje. Dėl to ketvirtame skyriuje buvo aprašyti sukurti klasių modeliai, leidžiantys pagal vandens spektro savybes sugrupuoti ežerų stebėjimus. Biofizinės klasės modeliai priskiria ežerui klasę tam tikru laiko momentu. Galima stebėti biofizinės klasės pokytį, kuris siejamas su optinių vandens savybių kitimu. Klasė geriausiai nustatoma vandens telkiniuose, kuriuose yra labai maži kiekiai optiškai aktyvių medžiagų (skaidri klasė) bei, kai vandens telkinyje stebimas silpnas (tarpinė klasė) arba stipresnis vandens žydėjimas (Chla-klasė). Visgi biofizinės klasės nustatymas leidžia tik apytikriai įvertinti, kad vandens telkinys žydi, tačiau žydėjimo stiprumas įvertinamas apytiksliai, dėl to tikslesniam žydėjimo intensyvumo įvertinimui kuriami modeliai chla koncentracijos nustatymui. Šiame skyriuje aptariamas modelių sudarymas stebėjimų klasėms (5.1 skyrius), taip pat jais gauti rezultatai lyginami su modeliu, sukurtu remiantis dirbtiniais neuroniniais tinklais stebėjimų neskirstant į klases bei atsitiktinio miško modeliu, sukurtu duomenų neskirstant į klases (5.2 skyrius), chla išskirčių dėl debesų ir debesų šešėlių pašalinimas (5.3 skyrius), nustatyti potencialiai probleminiai vandens telkiniai pagal chla koncentracijos vidutines ir momentines vertes (5.4 skyrius) bei nagrinėjama chla koncentracijos kaita laike (5.5 skyrius).

### 5.1. Chlorofilo $\alpha$ koncentracijos modeliai, paremti atsitiktinio miško regresijos algoritmu

Siekiant nustatyti chla koncentraciją buvo sudaromi atsitiktinio miško regresijos algoritmu paremti modeliai: skaidriai ežerų klasei (modelis toliau tekste bus vadinamas RFR1s) bei tarpinei ir Chla-klasei kartu (toliau tekste vadinamas RFR1tc). Buvo atlikta vizuali duomenų patikra – peržiūrėtos natūralių spalvų RGB kompozicijos ir duomenys, kurie buvo paveikti debesų ar jų šešėlių buvo pašalinti ir nepateko į modelių apmokymo duomenų aibes.

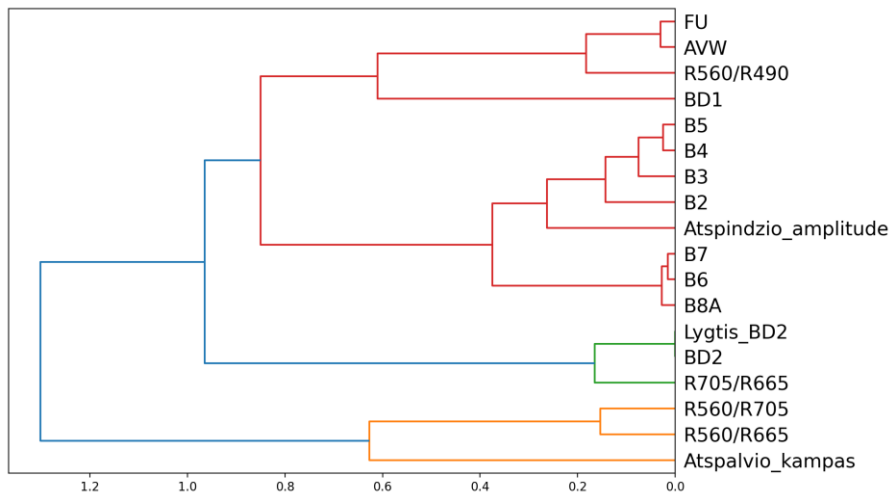
Modeliai sudaryti pasiremiant 749 *in situ* ir palydovinių duomenų poromis ir dalinami į dvi grupes: stebėjimai, kurie pagal biofizinės klasės modelį buvo priskirti skaidriai klasei bei stebėjimai, kurie priskirti tarpinei arba Chla-klasei. RFR1s sudaryti buvo naudojamos 450, o RFR1tc – 299 duomenų poros. Vidutinė chla koncentracija RFR1s duomenų aibėje buvo  $7,0 \pm 6,0 \text{ mg/m}^3$ , o RFR1tc duomenų aibėje  $27,5 \pm 26,8 \text{ mg/m}^3$ .

Pirmiausia buvo sukurti modeliai pagal visus požymius (žr. 2.10 lentelę) atmetus mėnesio požymį, kuris nepagerino biofizinės klasės modelio rezultatų. RFR1s kryžminio validavimo metu, naudojant dalinimą į tris dalis pagal grupes (kur grupė – unikalus ežero matavimo vietos kodas), gautas  $R^2 = 0,31 \pm 0,07$ ,  $RMSE = 4,9 \pm 0,6 \text{ mg/m}^3$ , o RFR1tc –  $R^2 = 0,54 \pm 0,12$ ,  $RMSE = 17,7 \pm 1,2 \text{ mg/m}^3$ . Kryžminis validavimas leidžia įvertinti modelio jautrumą duomenų dalinimui ir modelio generalizavimo galimybes. Šiuo atveju naudojamas dalinimas pagal grupes tam, kad modelis neišmoktų konkretaus ežero spektro, tad visi vieno ežero stebėjimai buvo atskirti ir pateko arba į mokymo, arba į testavimo aibę.

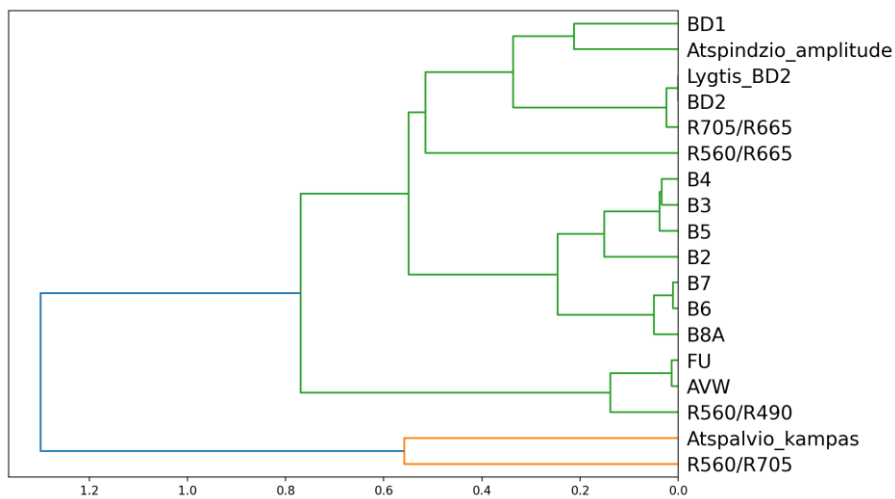
Siekiant pagerinti modelio interpretaciją ir tikslumą bei turėti paprastesnį modelį buvo vykdoma požymių atranka. Buvo atrinkti mažiausiai koreliuojantys ir didžiausią svarbą turintys požymiai pagal tris kriterijus: 1. požymių svarbumą (*angl. – feature importance*), gautą iš sudaryto pradinio atsitiktinio miško modelio, 2. požymių klasterizacijos dendrogramą (RFR1s modelio patekta 5.1 pav., RFR1tc modelio – 5.2 pav.), kuri gaunama apskaičiuojant atstumus tarp požymių koreliacijos, ir 3. remiantis dispersijos padidėjimo faktoriumi (*angl. – variance inflation factor*), kuris parodo, kiek padidėja nustatytų regresijos koeficientų dispersija dėl požymių multikolinearumo, tai yra, nustatomi tarpusavyje koreliuojantys požymiai. Taigi atrinkti požymiai, kurie pasižymėjo didžiausia santykinė svarba, mažu VIF ir pagal dendrogramą pasižymėjo didžiausiu atstumu pagal koreliaciją. RFR1s modeliui parinkti keturi požymiai (skliausteliuose pateikiamas santykinis svarbumas sudarant modelį): BD1 (0,47), atspindžio amplitudė (0,19), R560/R705 (0,19) ir atspalvio kampas (0,16). Atrinkus požymius VIF reikšmė kiekvienam iš jų buvo  $< 10$ . Sumažinus požymių skaičių modelio tikslumas naudojant kryžminį validavimą nepakito. RFR1tc modeliui palikti šeši požymiai: BD1 (0,52), R705/R665 (0,22), R560/R490 (0,07), B8A (0,07), R560/R705 (0,07), R560/R665 (0,06).

Pagal šešis požymius sudarytas RFR1tc modelis išlaikė panašų tikslumą ( $R^2 = 0,5 \pm 0,2$ ,  $RMSE = 18,2 \pm 2,2 \text{ mg/m}^3$ ), tačiau abiejų modelių atveju ryšys tarp išmatuotos ir modeliu nustatytos chla koncentracijos naudojant kryžminį validavimą parodė, kad modelis pernelyg gerai išmoksta mokymo aibės pavyzdžius, todėl mokymo metu gaunamos itin geros metrikos, o atskiri

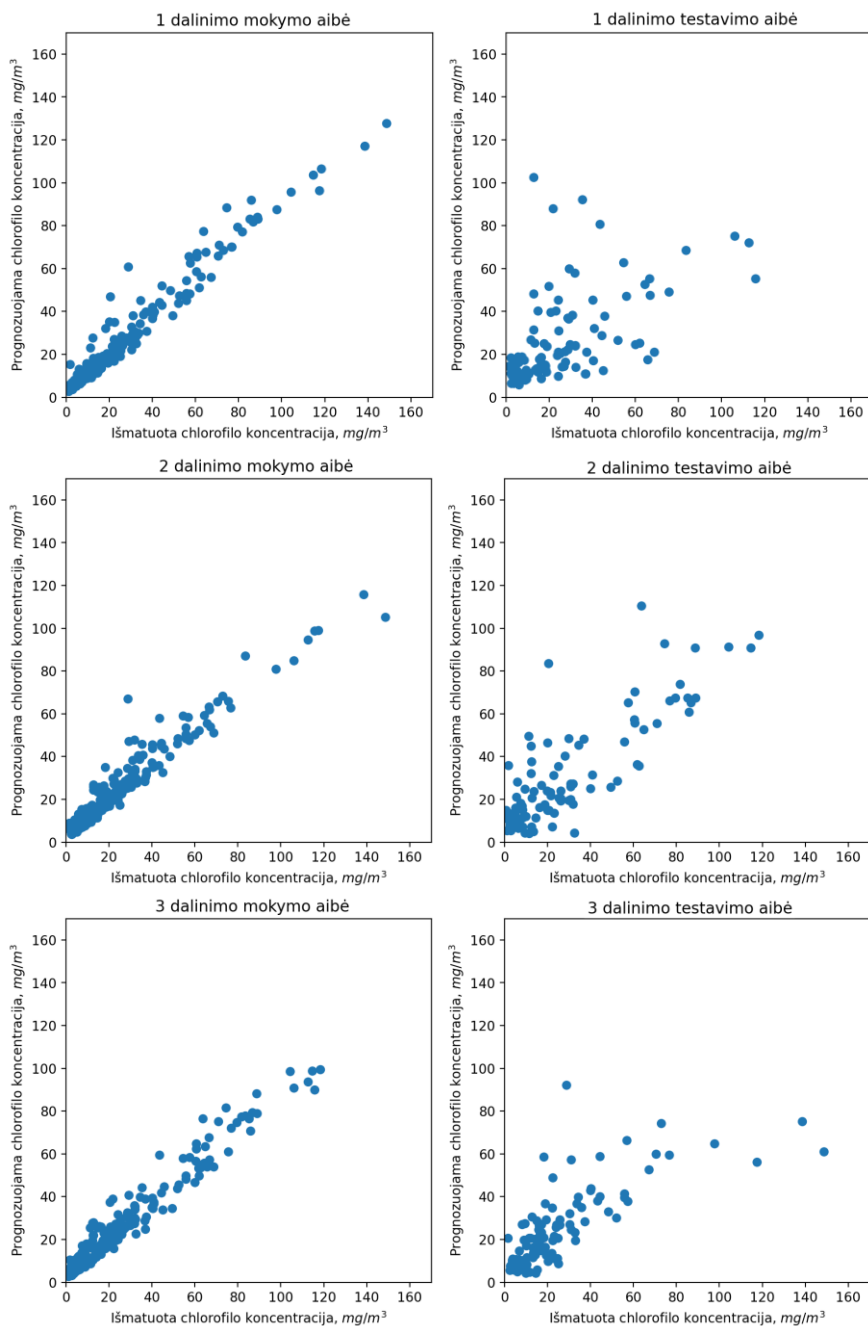
sprendimų medžiai šakojasi tol, kol kiekviename lape lieka po vieną pavyzdį. Testavimo metu modelio rezultatai ganėtinai varijuoja ir ryšys tarp išmatuotos ir modeliu nustatytos koncentracijos nėra glaudus (5.3 pav.).



**5.1 pav.** Požymių koreliacijos dendrograma pagal skaidrios klasės ežerų stebėjimus (pagal 451 duomenų porą).



**5.2 pav.** Požymių koreliacijos dendrograma pagal tarpinės ir Chla-klasės duomenų aibę (299 duomenų poras).



**5.3 pav.** Išmatuotos ir nustatytos RFR1tc modeliu chlorofilo  $\alpha$  koncentracijos sklaidos grafikas mokymo (kairėje) ir testavimo (dešinėje) aibėms kiekvieno iš trijų dalinimų metu naudojant kryžminio validavimo metodą.

Siekiant sumažinti modelio dydį ir sudėtingumą buvo nuspręsta optimizuoti modelio hiperparametrus (medžių skaičių, didžiausią gylį, maksimalų naudojamų požymių skaičių, mažiausią duomenų skaičių medžio lape, mažiausią duomenų kiekį, kai medis šakojasi), naudojant *GridSearchCV()*, o tai pagerina modelio generalizavimo galimybes. Kad būtų galima pilnai įvertinti su hiperparametrais apmokyto modelio veikimą ant mokymui nenaudotų duomenų, duomenų aibė abiejų modelių atveju buvo padalinta į tris dalis – mokymo, testavimo ir validavimo. Modelio hiperparametrų ieškota pasitelkiant mokymo aibę (323 stebėjimai RFR1s, 206 stebėjimai RFR1tc). Iš kelių hiperparametrų reikšmių buvo atrinkta tokia parametru kombinacija, kuri labiausiai sumažina kvadratinę paklaidą (numatytą RFR modelio tikslumo kriterijų).

Nustačius hiperparametrus (5.1 lentelė), su kuriais modeliai pasižymėjo geriausiomis generalizavimo savybėmis, modelis buvo apmokytas pasitelkus mokymo ir testavimo aibes (414 stebėjimai RFR1s, 271 stebėjimas RFR1tc), o jo tikslumas patikrintas naudojant validavimo aibę (RFR1s – 36 stebėjimai, RFR1tc – 28 stebėjimai). Validavimo metu nustatytas geresnis modelio tikslumas (RFR1s modelio  $R^2 = 0,54$ ,  $RMSE = 4,1 \text{ mg/m}^3$ , RFR1tc modelio  $R^2 = 0,71$ ,  $RMSE = 11,5 \text{ mg/m}^3$ ) nei naudojant kryžminio validavimo metodą, kai buvo naudoti visi požymiai ir numatytieji algoritmo hiperparametrai, todėl nuspręsta, kad hiperparametrai buvo parinkti tinkamai ir dėl to sudarytas galutinis modelis, apmokytas naudojant visą modeliui skirtą duomenų aibę (RFR1s modeliui – 450, RFR1tc – 299 stebėjimus).

Siekiant geriau suprasti sudarytų modelių sandarą ir rezultatų gavimą, modelių paaiškinimui naudoti dalinės priklausomybės (*angl. – partial dependence*), sprendimų (*angl. – decision plot*) ir apibendrinimo (*angl. – summary plot*) grafikai, sudaryti *shap* pakete esančiomis *PartialDependenceDisplay()*, *decision\_plot()* ir *summary\_plot()* funkcijomis. Dalinės priklausomybės grafikai (5.4 pav., 5.5 pav.) rodo, kaip kintant požymio reikšmei kito modeliu nustatomo kintamojo – chla koncentracijos – reikšmė. Pvz., didėjantis svarbiausias priklausomas kintamasis – skirtumas BD1 – lėmė vis didesnę reikšmę chla koncentracijai. Ši priklausomybė paremta būdingomis optinėmis savybėmis, t. y., esant didesnėms chla koncentracijoms raudonosios ribos 705 nm bangų ruože atspindys paprastai padidėja, o raudonos spalvos 665 nm bangų ruože stebima sumažėjusi atspindžio reikšmė dėl chla absorbcijos (5.4 pav. ir 5.5 pav.). Dalinis reikšmingumas RFR1s modelio atveju išauga ties  $BD1 = 0,001$ , o RFR1tc modelio atveju ties  $BD1 = 0,011$  reikšme. BD1 apibūdina atspindžio skirtumą tarp 705 nm ir atspindžio vidurkio 665 nm bei 742 nm spektrinių juostų. BD1 svarba didėjo didėjant skirtumo reikšmei. RFR1tc modelio atveju R705/R665

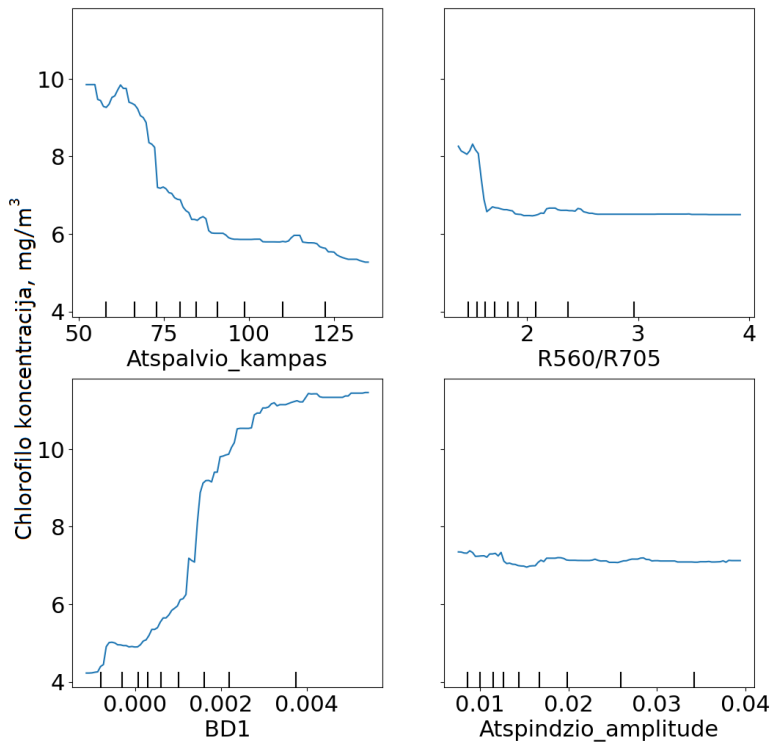
požymio svarba padidėja ties R705/R665 reikšme 1,35 (5.5 pav.), kas parodo 705 nm atspindžio piką. R560/R705 ir B8A kintamieji veikė chla koncentraciją gana panašiai – didesnėms chla koncentracijoms turėjo mažiau reikšmės nei mažesnėms. RFR1s modelio atveju mažesnės atspalvio kampo reikšmės (40–80°) nei didesnės (90–130°) turėjo didesnę reikšmę chla koncentracijai (5.4 pav.).

**5.1 lentelė.** Hiperparametrų reikšmės, naudotos paieškoje bei nustatytos reikšmės RFR1s ir RFR1tc modeliams naudojant mokymo aibę.

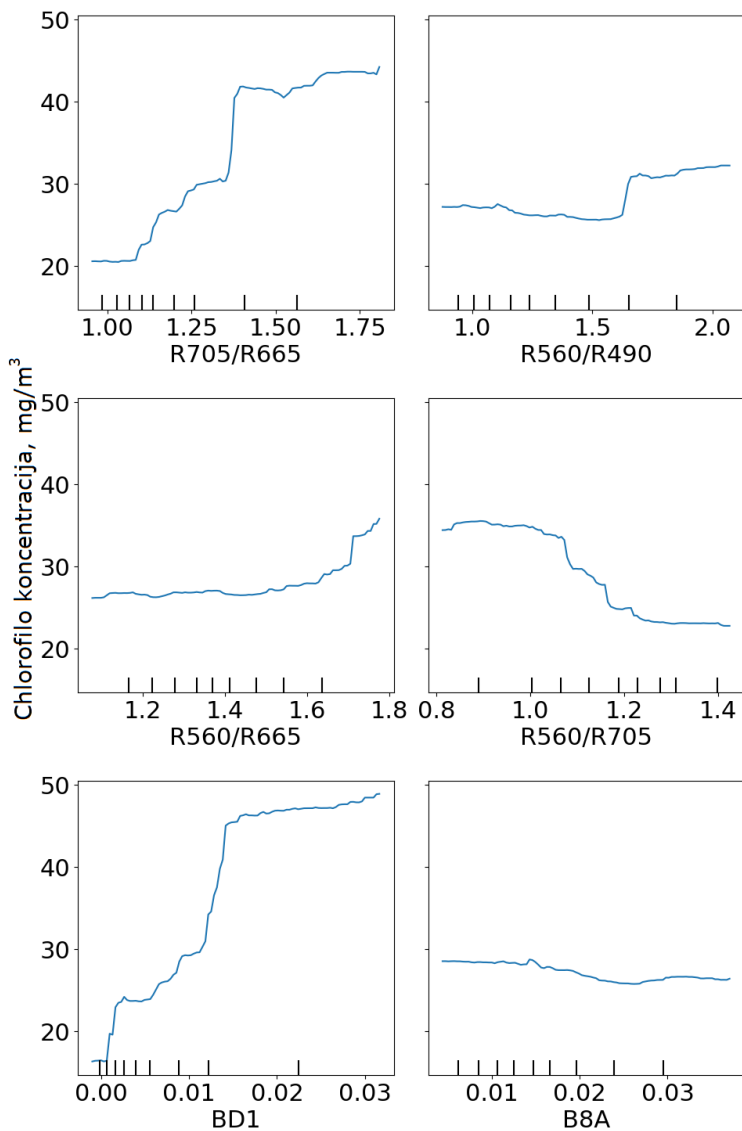
Hiperparametro pavadinimas (pavadinimas kode)	Ieškota tarp šių reikšmių	Nustatytos tinkamiausios hiperparametrų reikšmės RFR1s modeliui	Nustatytos tinkamiausios hiperparametrų reikšmės RFR1tc modeliui
Medžių skaičius ( <i>n_estimators</i> )	50, 75, 100, 150, 200, 300, 400	100	300
Didžiausias gylis ( <i>max_depth</i> )	5, 10, 15, 20	5	10
Maksimalus naudojamų požymių skaičius ( <i>max_features</i> )	„sqrt“, „log2“	šaknis iš požymių skaičiaus, sqrt(4)=2 (nurodoma „sqrt“)	šaknis iš požymių skaičiaus, sqrt(6)=2,4 (nurodoma „sqrt“)
Mažiausias duomenų skaičius medžio lape ( <i>min_samples_leaf</i> )	3, 5, 10, 15	5	3
Mažiausias duomenų kiekis, kai medis šakojasi ( <i>min_samples_split</i> )	5, 10, 20, 30	20	10
Atsitiktinė būseną ( <i>random_state</i> )	—	985	985

Tikėtina modelio reikšmė RFR1tc atveju yra 27,6 mg/m<sup>3</sup>, tad nagrinėjant konkretų atvejį galima sužinoti kaip modelis priima sprendimą – nustatomą chla reikšmę. Pvz., vienu atveju tikėtina modelio reikšmė 27,6 mg/m<sup>3</sup> buvo koreguojama dėl B8A atimant 0,1, dėl R560/R490 atimant 0,7, dėl R560/R665 atimant 2,8, dėl R560/R705 pridėdant 4,2, dėl R705/R665 požymio atimant

8,1 ir dėl BD1 atimant 9,1, tad galutinis modelio rezultatas – 11,0 mg/m<sup>3</sup> (5.6 a pav.). Kitu atveju nustatoma reikšmė buvo gauta koreguojant 27,6 mg/m<sup>3</sup> – dėl B8A atimant 0,7, dėl R560/R490 pridedant 5,3, dėl R560/R705 pridedant 4,2, dėl R560/R665 pridedant 7,1, dėl R705/R665 pridedant 17,0 mg/m<sup>3</sup> ir galiausiai dėl BD1 pridedant 26,6 mg/m<sup>3</sup>, tad galutinis rezultatas išėjo 90,7 mg/m<sup>3</sup> (5.6 b pav.). Dažniausiai didelės BD1 reikšmės stipriai padidindavo chla reikšmę, o mažos BD1 – sumažindavo. Kiek panašiai buvo ir su R705/R665 požymiu, tik jo įtaka buvo mažesnė. Požymio R560/R705 mažos reikšmės didino, o didelės mažino chla koncentraciją. Kiti požymiai dažniausiai nežymiai paveikdavo chla reikšmę – 0–5 mg/m<sup>3</sup> (5.7 pav.). Šie paaiškinamieji grafikai leido geriau suprasti kaip modeliu gautos konkrečios chla koncentracijos.



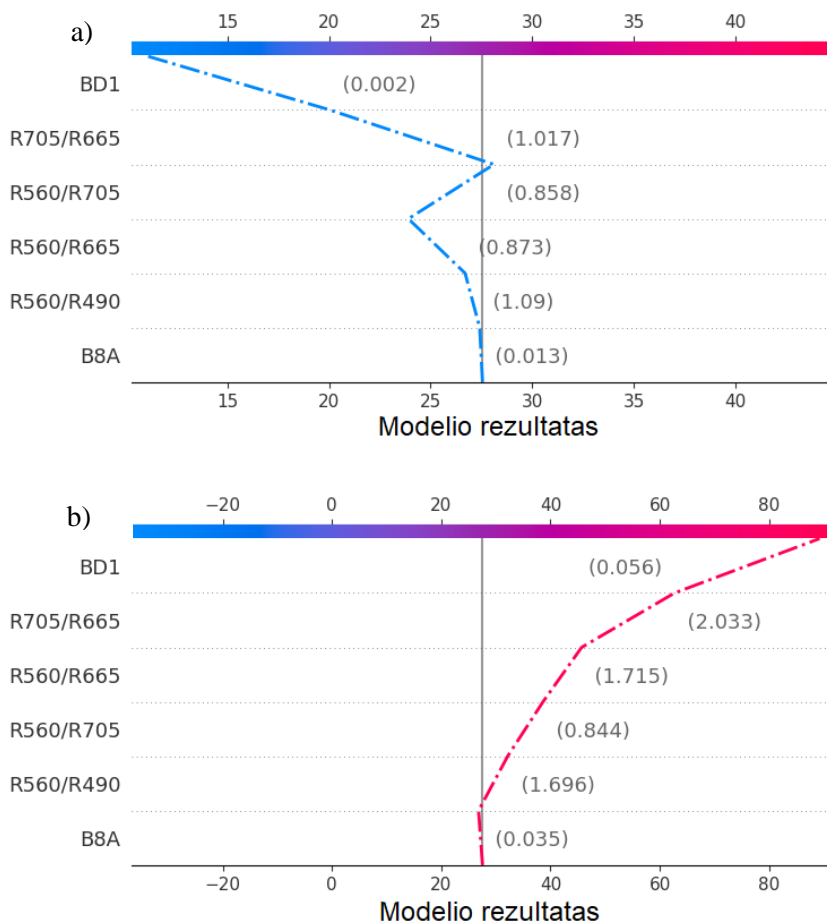
**5.4 pav.** Dalinės priklausomybės grafikai tarp požymių ir chlorofilo  $\alpha$  koncentracijos RFR1s



**5.5 pav.** Dalinės priklausomybės grafikai tarp požymių ir chlorofilo  $\alpha$  koncentracijos RFR1tc modelio atveju.

Pastebėtina, kad RFR1s modelis geriausiai nustatė reikšmes 5–10 mg/m<sup>3</sup> intervale (išmatuotas vidurkis 7,0±1,0 mg/m<sup>3</sup>, o modeliu nustatytas – 7,5±3,1 mg/m<sup>3</sup>), o mažesnes koncentracijas visais atvejais pateikdavo apie 2 mg/m<sup>3</sup> didesnes. Aukštesnes chla koncentracijas (> 20 mg/m<sup>3</sup>) RFR1s modelis nustatė prastai (5.8 pav.), tačiau tai sudarė mažiau nei 1 % stebėjimų šioje duomenų aibėje.

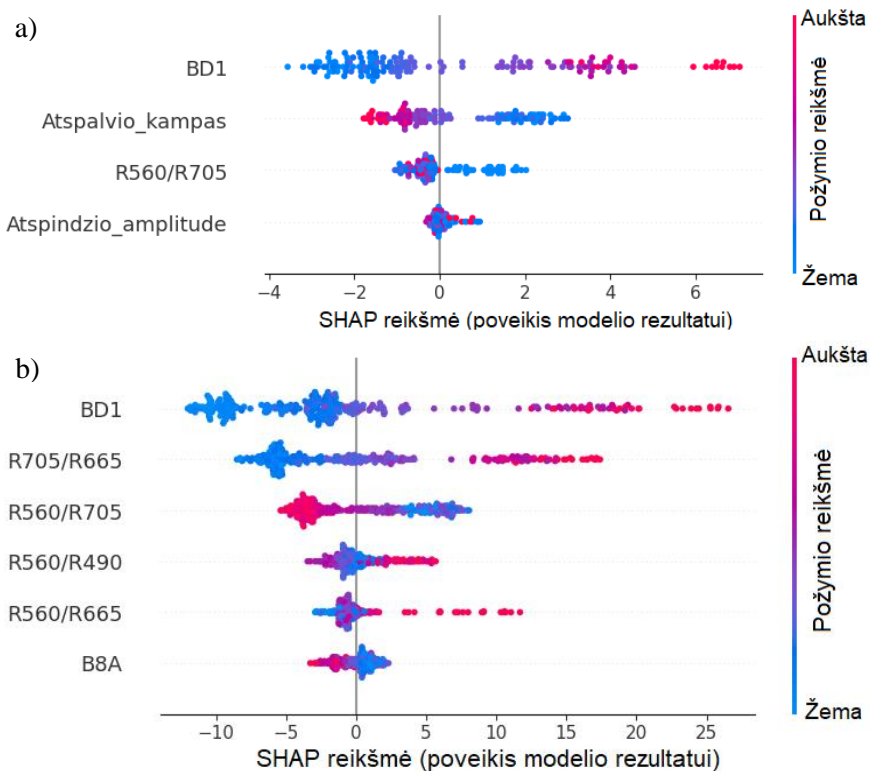




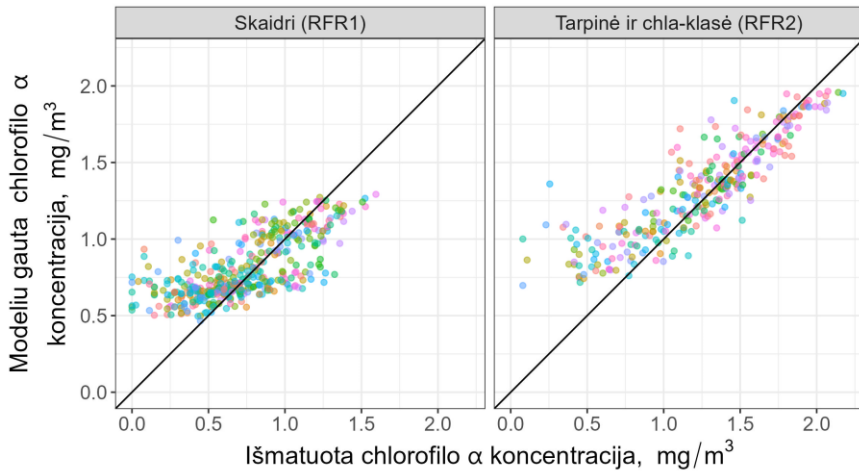
**5.6 pav.** Požymių suminis poveikis galutiniam RFR1tc modelio rezultatui. Pradinė reikšmė (apačioje) yra tikėtina chla reikšmė, o galutinė (viršuje) – modelio rezultatas, a) rodo pavyzdį, kai modeliu nustatyta chlorofilo  $\alpha$  koncentracija Vaitiekūnų tvenkinyje 2017-09-27 buvo 11,0 mg/m<sup>3</sup>, o b) kai modeliu nustatyta chlorofilo  $\alpha$  reikšmė buvo 90,7 mg/m<sup>3</sup> Rimiečio ežere 2016-07-26.

RFR1tc modelis geriausiai chla nustatė 20–30 ir 40–50 mg/m<sup>3</sup> reikšmių intervaluose, kurie yra artimi vidutinei modeliu nustatomai koncentracijai (5.8 pav.). Modeliu nustatyta chla koncentracija buvo 5–92 mg/m<sup>3</sup> reikšmių intervale, dėl to didesnės nei 92 mg/m<sup>3</sup> reikšmės buvo įvertinamos su nemaža paklaida. Pastebėtina, kad santykinai žemos koncentracijos (< 10 mg/m<sup>3</sup>) dauguma atvejų buvo pervertinamos. Vidutinė nustatyta koncentracija šiame reikšmių intervale buvo 10,1 mg/m<sup>3</sup>, o vidutinė išmatuota – 5,3 mg/m<sup>3</sup>. Aprašyti modelio rezultatų netikslumai buvo iš dalies nulemti nedidelio ir ne visada reprezentatyvaus naudoto duomenų masyvo, kuriame chla reikšmės

kito santykinai plačiame reikšmių intervale (nuo 0,2 iki 148 mg/m<sup>3</sup>), o tokios pačios išmatuotos chl<sub>a</sub> reikšmės kai kuriais atvejais buvo susietos su skirtingais palydoviniais spektrais, kas apsunkino modelio apmokymą. Visgi panaudoti šeši požymiai ir atsitiktinio miško algoritmas leidžia nustatyti tikslesnę chl<sub>a</sub> koncentraciją nei metodai naudojantys vieną požymį, pvz., tiesinė regresija, kaip kad CHLA1, CHLA10 empirinių lygčių atveju, kadangi atsitiktinio miško algoritmas padeda nustatyti netiesines priklausomybes. Ateityje surinkus daugiau reprezentatyvių duomenų būtų galima modelį mokyti ir validuoti iš naujo ir modelio rezultatai turėtų gerėti.



**5.7 pav.** Požymių poveikis galutiniam a) RFR1s ir b) RFR1tc modelio rezultatui. Požymiai, pagal nustatytą poveikį modelio rezultatui, didino arba mažino modelio rezultatą – chlorofilo  $\alpha$  koncentraciją – nuo tikėtinos reikšmės.



**5.8 pav.** Išmatuotos ir RFR1s bei RFR1tc nustatytos momentinės chlorofilo  $\alpha$  koncentracijos ežeruose pasiskirstymas. Spalvos rodo vieno ežero ir vienos datos kombinaciją, skalės logaritmuotos.

## 5.2. Atsitiktinio miško regresijos ir dirbtiniais neuroniniais tinklais paremtų modelių rezultatų palyginimas

RFR1s ir RFR1tc modelių rezultatai buvo palyginti su RFR2 modeliu, sukurtu naudojant tą pačią pilną, kaip ir RFR1s ir RFR1tc sudėjus kartu duomenų aibę, neskirstant duomenų pagal biofizines klases ir taip pat su sudėtingesnio mašininio mokymosi modelio rezultatais – dirbtiniais neuroniniais tinklais paremto modelio (ANN). ANN buvo apmokytas pasiremiant beveik dvigubai didesne mokymo duomenų aibe, kuri apėmė Lietuvos, Latvijos ir Estijos ežerus. Modelio sudarymui naudotos 909 *in situ* ir palydovinių duomenų poros, o modelis testuotas naudojant 437 duomenų poras. ANN modelio kūrimas buvo pradėtas nuo bazinio regresinio modelio sukūrimo naudojant *python pycaret* paketą, kadangi buvo siekiama sukurti geresnį nei bazinį parentą ANN modelį. Modelio sudarymui naudoti šie požymiai: R490, R560, R665, R705, R740, R783, R842, R865, R705/R665, R560/R490, R560/R665, R560/R705, BD1, Lygtis\_BD2, AVW (4.4 lentelė). Sudarytas galutinis ANN modelis buvo paremtas liekanų giliuoju neuroniniu tinklu (*angl. – Residual deep neural network with Multi-Head Attention*), šio modelio tikslumas testavimo aibėje buvo  $RMSE = 13,5 \text{ mg/m}^3$ , o  $R^2 = 0,55$ . Modelis pasižymėjo geresnėmis generalizavimo savybėmis nei bazinis sprendimų medžiais paremtas regresijos modelis (*angl. – Extra tree regressor*), kurio generalizavimo galimybės buvo prastos mokymo aibėje, nes buvo pasiekta  $0,0 \text{ mg/m}^3$  paklaida.

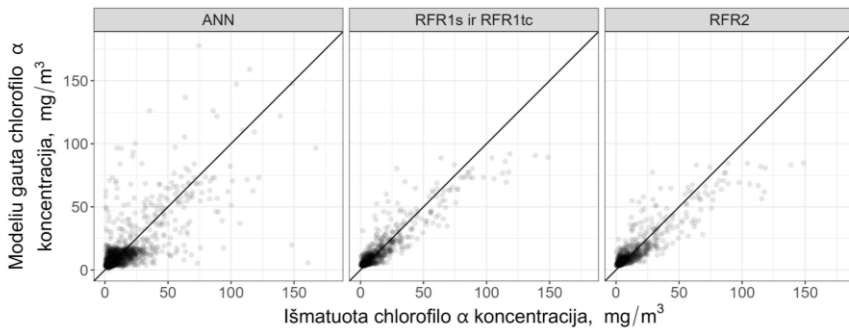
Atsitiktinio miško modelis be skirstymo į klases veikė prasčiau nei du RFR1s ir RFR1tc modeliai biofizinėms klasėms (5.2 lentelė). Skirstymas pagal biofizinės klases beveik nepagerino chla rezultatų skaidrioje klasėje,  $R^2$  ir RMSE pokytis nežymus. Tačiau kitoms klasėms skirstymas pagal biofizinės klases turėjo teigiamos įtakos, kadangi chla modelis klasėms parodė geresnius rezultatus tarpinei ir Chla-klasei. Tarpinė ir Chla-klasė yra svarbesnės, nes jos siejamos su vandens žydėjimu.

**5.2 lentelė.** Atsitiktinio miško modelių metrikų palyginimas.

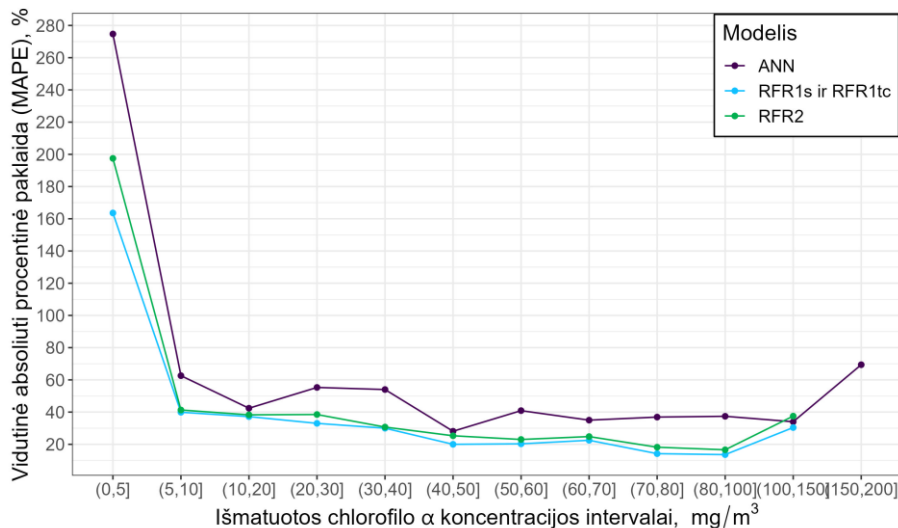
Klasė	N	$R^2$ (naudojant RFR1s ir RFR1tc)	$R^2$ (naudojant RFR2, be skirstymo į klases)	RMSE (naudojant RFR1s ir RFR1tc)	RMSE (naudojant RFR2, be skirstymo į klases)
Skaidri	450	0,56	0,55	4,1	4,2
Tarpinė	109	0,46	0,19	6,4	7,9
Chla-klasė	190	0,76	0,68	14,7	16,9

RFR2 ir ANN modeliai buvo sudaryti neskirstant stebėjimų į klases, o naudojant visą duomenų aibę, kurioje buvo skirtingomis optinėmis savybėmis pasižyminčių ežerų. Mokymo aibės dydis ir tai, kad RFR1s ir RFR1tc modeliai buvo sudaryti atskiroms ežerų klasėms, kurios klasės viduje pasižymėjo panašiomis optinėmis savybėmis, buvo pagrindiniai metodiniai skirtumai tarp RFR2 ir ANN modelių sudarymo.

RFR1s ir RFR1tc modeliai geriausiai nustatė chla koncentracijas 10–50  $\text{mg}/\text{m}^3$  reikšmių intervale, tačiau mokymo aibėje buvusius atvejus, kur chla koncentracija  $> 100 \text{ mg}/\text{m}^3$ , modelis nustatė prasčiau ir nė viena modeliu nustatyta chla koncentracija nesiekė  $100 \text{ mg}/\text{m}^3$  (5.9 pav.). ANN modelio rezultatų sklaida didesnė nei RFR1s, RFR1tc ir RFR2 ir modelis pateikė chla koncentracijų didesnių už  $100 \text{ mg}/\text{m}^3$ , tačiau dauguma atvejų, chla koncentracijos buvo nustatytos netiksliai. Nors modeliai buvo apmokyti naudojant iš dalies skirtingas duomenų bazines, bet konkrečių chla reikšmių intervalų vidurkiai buvo beveik tokie pat, tad duomenys palyginami. Mažiausia vidutinė absoliuti procentinė paklaida RFR1s ir RFR1tc modelių buvo nustatyta 70–100  $\text{mg}/\text{m}^3$  reikšmių intervale (14 %), 40–60  $\text{mg}/\text{m}^3$  intervale – 20 %, o kitur siekė nuo 23 (60–70  $\text{mg}/\text{m}^3$ ) iki 164 % (0–5  $\text{mg}/\text{m}^3$ ) (5.10 pav.).



**5.9 pav.** Išmatuotos ir modeliu apskaičiuotos (ANN – dirbtiniais neuroniniais tinklais paremtas modelis, RFR1s ir RFR1tc – biofizinėms klasėms sukurti atsitiktinio miško algoritmu paremti modeliai ir RFR2 – atsitiktinio miško algoritmu paremtas modelis be klasių skirstymo) chlorofilo α koncentracijos ryšio grafikas.



**5.10 pav.** RFR1s ir RFR1tc, RFR2 ir ANN modeliais gautos chlorofilo α koncentracijos vidutinė absoliuti procentinė paklaida (MAPE), %, pagal *in situ* koncentracijos intervalus.

ANN modelio MAPE visuose chla koncentracijos intervaluose buvo didesnė nei RFR1s ir RFR1tc modelių ir siekė nuo 28 % (40–50 mg/m<sup>3</sup>) iki 275 % (0–5 mg/m<sup>3</sup>). Pastebėtina, kad RFR1s ir RFR1tc modelių žemiausia nustatyta reikšmė buvo 2,9 mg/m<sup>3</sup> (visame duomenų rinkinyje – 2,7 mg/m<sup>3</sup>), o ANN – 1,4 mg/m<sup>3</sup>, tad RFR1s ir RFR1tc modelių ribos gan aiškiai apibrėžtos – modeliais nenustatomos mažesnės koncentracijos nei 2,7 mg/m<sup>3</sup> ir ties ~90 mg/m<sup>3</sup> modelis įsisotina ir didesnių reikšmių nepateikia.

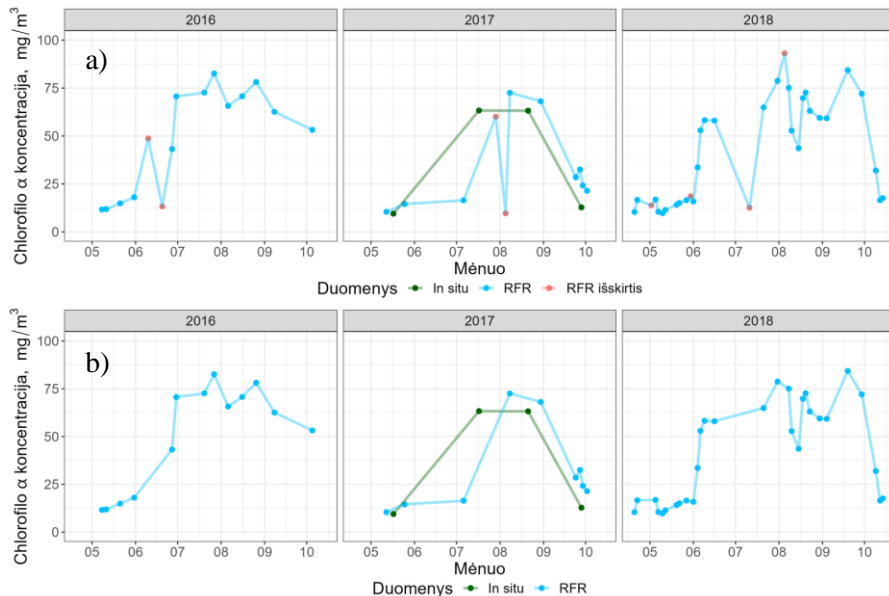
ANN modelis buvo apmokytas panaudojant chl<sub>a</sub> koncentracijas 0–150 mg/m<sup>3</sup> reikšmių intervale iš 477 Baltijos šalių ežerų, o RFR1s ir RFR1tc modelių mokymui panaudoti tokio pat intervalo duomenys, tačiau iš 267 Lietuvos ežerų. ANN mokymo aibėje galėjo būti daugiau ežerų, pasižyminčių kompleksiškomis optinėmis savybėmis, o tai galėjo lemti daugiau išskirčių mokymo aibėje, kai ežero spektras nebuvo reprezentatyvus išmatuotai chl<sub>a</sub> koncentracijai. Kadangi tokių ežerų ir duomenų nėra daug ir kurti jiems atskirą modelį būtų sudėtinga dėl mažo duomenų kiekio, dėl iš mokymo aibės juos geriau pašalinti ir jiems naudoti paprastus algoritmus, pvz., tiesinę regresiją. RFR1s ir RFR1tc modeliai buvo sukurti duomenis dalinant į dvi grupes, atskirtas pagal nustatytą biofizinę klasę. Taigi šiuo atveju ežerų skirstymas pagal klases padėjo sukurti tikslesnius modelius nei naudojant RFR2 modelį visai duomenų aibei ir ANN, todėl tolesnei analizei buvo pasirinkta naudoti RFR1s ir RFR1tc modelius ir jie toliau tekste trumpiau vadinami RFR.

### 5.3. Chlorofilo $\alpha$ koncentracijos išskirčių pašalinimas

RFR modeliai buvo pritaikyti 37450 Lietuvos ežerų spektrams 2015–2021 m. nuo balandžio 15 iki spalio 15 d. Analizuojant laiko eilutės grafikus buvo pastebėti chl<sub>a</sub> koncentracijos svyravimai trumpame laikotarpyje (3–7 dienų tarpe). Patikrinus keletą atvejų, kai koncentracija iš ~60 mg/m<sup>3</sup> nukrenta iki ~20 mg/m<sup>3</sup>, pastebėta, kad tai įvyko dėl ežerą dengiančio debesies šešėlio, kuris itin sumažina chl<sub>a</sub> signalą palydoviniuose duomenyse. Taip pat pasitaikė atveju, kai ežere vyravusi žema koncentracija vieno stebėjimo metu pakyla 15–20 mg/m<sup>3</sup>. Tokie atvejai siejami su plonais debesimis ar debesies kraštu, kurių nuotraukos klasifikacija nepašalino.

Siekiant pašalinti tokius stebėjimus buvo pritaikytas išskirčių pašalinimo metodas. Buvo paskaičiuotas slenkantis vidurkis (3 stebėjimų), slenkantis standartinis nuokrypis ir z-įvertis, kuris gautas iš konkretaus stebėjimo chl<sub>a</sub> koncentracijos atimant atitinkamą slenkantį vidurkį ir šį skirtumą padalinant iš slenkančio vidurkio. Atvejai, kai šios standartizuotos vertės z modulis buvo didesnis už bandymų metu parinktą slenkstinę vertę 1,1, buvo pažymėti kaip išskirtys ir nufiltruoti prieš vykdant tolimesnę analizę. Slenkstinė vertė buvo parinkta Simno ežero pavyzdžiu. Ją parinkus labiausiai išsiskiriantys stebėjimai buvo pažymėti kaip išskirtys. Pavyzdžiui, Simno ežero atveju beveik kiekvienais metais pasitaikė po atvejį, kai chl<sub>a</sub> koncentracija nukrito staigiai. Tokie staigūs pokyčiai yra mažai tikėtini ir apsunkina duomenų interpretaciją. Tokie atvejai daro įtaką skaičiuojamiems parametrams, pvz., vidurkiui. Pritaikius minėtą metodą šie, debesies šešėlio paveikti duomenys

buvo nufiltruoti (pvz., 2016 m. birželio viduryje, 2017 m. rugpjūčio pradžioje ir 2018 m. liepos pirmoje pusėje, 5.11 pav.).



**5.11 pav.** Simno ežero a) pradiniai duomenys prieš išskirčių pašalinimą ir pažymėtos išskirtys, b) laiko eilutės grafikas po išskirčių pašalinimo.

Šiuo metodu iš ežerų spektrų buvo pašalinta nuo 3 iki 38 stebėjimų (mediana = 20), o tai sudarė nuo 7 iki 31 % (mediana 19 %) stebėjimų ežeruose.

#### 5.4. Potencialiai probleminių vandens telkinių išskyrimas

Surinkti duomenys buvo panaudoti potencialiai probleminių ežerų išskyrimui pagal hidrobiologinį kriterijų – chl<sub>a</sub> koncentracijos vidutinės ir momentines reikšmes. Šio tyrimo metu buvo sudaryta *in situ* matavimų ir palydovais užfiksuotų bei atsitiktinio miško modeliais gautų chl<sub>a</sub> koncentracijos duomenų bazė. Pasiremiant minėtais kriterijais bei naujesniais duomenimis buvo atnaujintas ežerų skirstymas. Šiame skyrelyje pateikiama informacija naudojant *in situ* ir iš palydovinių duomenų įvertinta informacija.

Pabrėžtina, kad 43 ežeruose ir tvenkiniuose *in situ* stebėjimų šiuo laikotarpiu nebuvo užfiksuota, todėl modelių pritaikymas pasiremiant palydoviniais duomenimis reikšmingai papildė žinias apie juos. Iš šių ežerų potencialiai probleminiams priskirti: 31 pagal vidurkio ir momentinės vertės kriterijų, 6 pagal vidurkį, o likę 6 priskirtini neprobleminiams.

Pagal *in situ* matavimus buvo išskirta daugiau neprobleminių ežerų nei naudojant RFR, tačiau skirtumai galėjo atsirasti dėl to, kad *in situ* buvo po kelis matavimus per metus ar net 2015–2021 m. laikotarpį, o iš palydovinių duomenų gautos bent kelios dešimtys stebėjimų kiekviename ežere per tyrimo laikotarpį (mediana 88 stebėjimai). Taigi palydoviniais duomenimis užfiksuojami trumpalaikiai chla pokyčiai, kurie atitinkamai veikia vidutinę bei momentinę chla reikšmę.

Tiek pagal *in situ*, tiek pagal palydovinius duomenis 52 ežerai (15 % nuo visų) buvo priskirti neprobleminiams, o 172 ežerai (48 % nuo visų) priskirti potencialiai probleminiams tiek pagal *in situ*, tiek pagal RFR vidutines ir didžiausias chla koncentracijas (5.3 lentelė). Likę 133 (37 %) buvo priskirti potencialiai probleminiams pagal vieną ar kitą kriterijų, arba tik pagal *in situ*, arba tik pagal RFR duomenis.

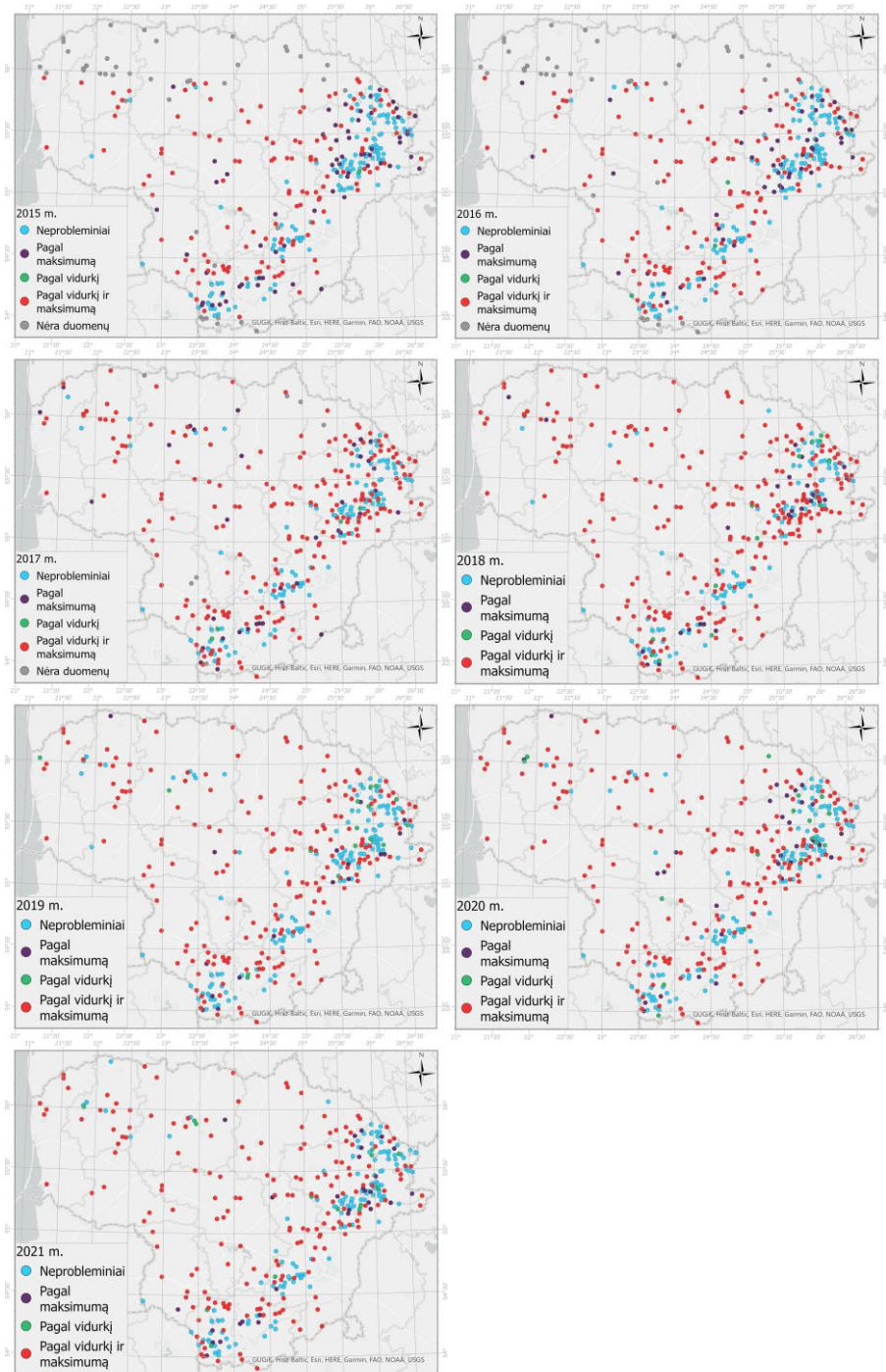
**5.3 lentelė.** Potencialiai probleminių ežerų skaičius, nustatytas pagal chlorofilo  $\alpha$  slenkstines vidurkio ir maksimumo vertes.

Ežerų grupė	Pagal <i>in situ</i>	Pagal RFR	Pagal abu
Neprobleminiai (pagal abu kriterijus)	124	66	52 (6–RFR)
Nėra duomenų	43	–	–
Probleminiai tik pagal chla vidurkį	4	1	0
Probleminiai tik pagal momentinę (maksimalią) chla vertę	4	47	9 (6–RFR)
Pagal vidurkį ir maksimalią vertę	158	227	172 (31–RFR)

Biofizinė klasė taip pat preliminariai gali indikuoti apie telkinio problemišumą. Visi ežerai, kuriuose vyravo tarpinė arba Chla-klasė buvo priskirti potencialiai probleminiams. Tarp ežerų, kuriuose vyravo skaidri klasė, nustatyta didesnė variacija – visi 52 neprobleminiai ežerai buvo priskirti skaidriai biofizinei klasei, o likę priskirti potencialiai probleminiams, iš kurių 77 potencialiai probleminiams šie ežerai priskirti beveik visus tyrimo metus. Taigi biofizinė klasė nėra pakankamas indikatorius apie problemišumą, chla modeliai suteikia galimybę įvertinti tiksliau.

Skirtingais metais nustatytas potencialiai probleminių ežerų skaičius kito. Daugiausiai potencialiai probleminių nustatyta 2018 m. (72 %), kiek mažiau 2017 m. (68 %), 2019–2021 m. 62–66 % ežerai (5.12 pav.). 2015 ir 2016 m. nustatyta po 54 %, tačiau dalyje ežerų (11 %) tais metais duomenų nebuvo surinkta.





**5.12 pav.** Pagal vidutinę ir maksimalią chlorofilo  $\alpha$  koncentraciją per metus nustatyti potencialiai probleminiai ežerai 2015–2021 m.

Ežerai, priskirti potencialiai probleminiams per tyrimo laikotarpį (2015–2021 m.), kai kuriais atvejais tik vienais metais pasiekė maksimalią ar/ir vidutinę chla slenkstinę vertę, pagal kurią ežeras priskiriamas potencialiai probleminių grupei. Tarp potencialiai probleminių buvo 29 ežerai, kuriuose momentinė vertė, viršijanti slenkstinę chla vertę nustatytą RFR, buvo užfiksuota tik 1 kartą, 7 iš šių ežerų buvo priskirti potencialiai probleminiams ir pagal kitą kriterijų, o 22 tik pagal vieną momentinę reikšmę. Tarp šių ežerų yra gilių (3 tipo pagal morfometrines savybes) – Alaušas, Klykių ež., Skaistis, Šakarvai ir Antalieptės HE tvenkinys ir negilių, pvz., Germantas, Labanoras, kurie yra saugomose teritorijose. Šiems ežerams reiktų gilesnės analizės ir vertinimo ar pastaruoju metu jų būklė išties pablogėjo.

Ežerų sąrašas su informacija apie priskyrimą neprobleminiams ar potencialiai probleminiams pateiktas 1 priede.

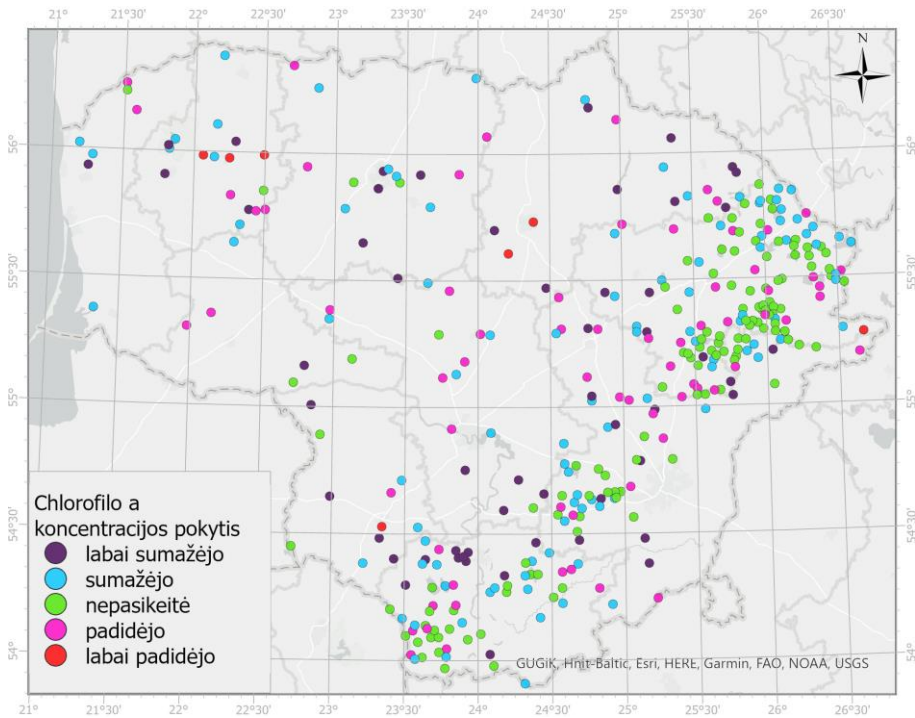
### 5.5. Chlorofilo $\alpha$ kaita ežeruose

Pasitelkus palydovinius duomenis ir sukurtus modelius tiriamuose ežeruose buvo nustatyta chla koncentracija, jos sezoninė ir tarpmetinė kaita 2015–2021 m. laikotarpiu.

Vidutinė chla koncentracija per tiriamą laikotarpį padidėjo 76 ežeruose, iš kurių 7-iuose padidėjo labai ( $> 10 \text{ mg/m}^3$ ), pvz., Masčio ežere ir Marijampolės II tvenkinyje (5.13 pav.). 126 ežeruose per tyrimo laikotarpį būklės pokyčio pagal chla nebuvo nustatyta – vidutinė koncentracija nepasikeitė. Tačiau tarp šių ežerų nustatyti ne tik žema vidutine koncentracija pasižymintys, tačiau ir tokie ežerai, kur visais metais laikėsi aukštos koncentracijos (pvz., Latežeris, Totorviečių tvenkinys). Vidutinė chla koncentracija sumažėjo 155 ežeruose, iš kurių 54-iuose sumažėjo labai. Didžioji dalis šių ežerų priskirti potencialiai probleminiams beveik visą tyrimo laikotarpį (5.13 pav.). Šiuose ežeruose stebimas būklės pagerėjimas, tačiau tik dviejuose iš jų chla koncentracija sumažėjo tiek, kad ežeras paskutiniais tyrimo metais buvo priskirtas neprobleminiams.

Didžiausios užfiksuotos chla koncentracijos tarp ežerų varijavo nuo  $3,8 \text{ mg/m}^3$  iki  $92,7 \text{ mg/m}^3$ , vidutinė –  $23,2 \pm 19,2$  (5.14 pav.). Daugiausiai aukštų ( $80\text{--}100 \text{ mg/m}^3$ ) koncentracijų nustatyta 2018 m. – 12-oje ežerų – dėl palankių vandens žydėjimui meteorologinių sąlygų bei didelio ir dažno tais metais užfiksuoto palydovinių stebėjimų kiekio. Tyrimo periodo pradžioje ir pabaigoje nustatytas panašus kiekis šių aukštų koncentracijų (3–4 ežeruose), tad apie maksimalios koncentracijos didėjimą ir mažėjimą ežeruose galima spręsti tik nagrinėjant ežerus individualiai. 23-iuose ežeruose nustatytas statistškai reikšmingas (p-reikšmė 0,05) maksimalios chla koncentracijos

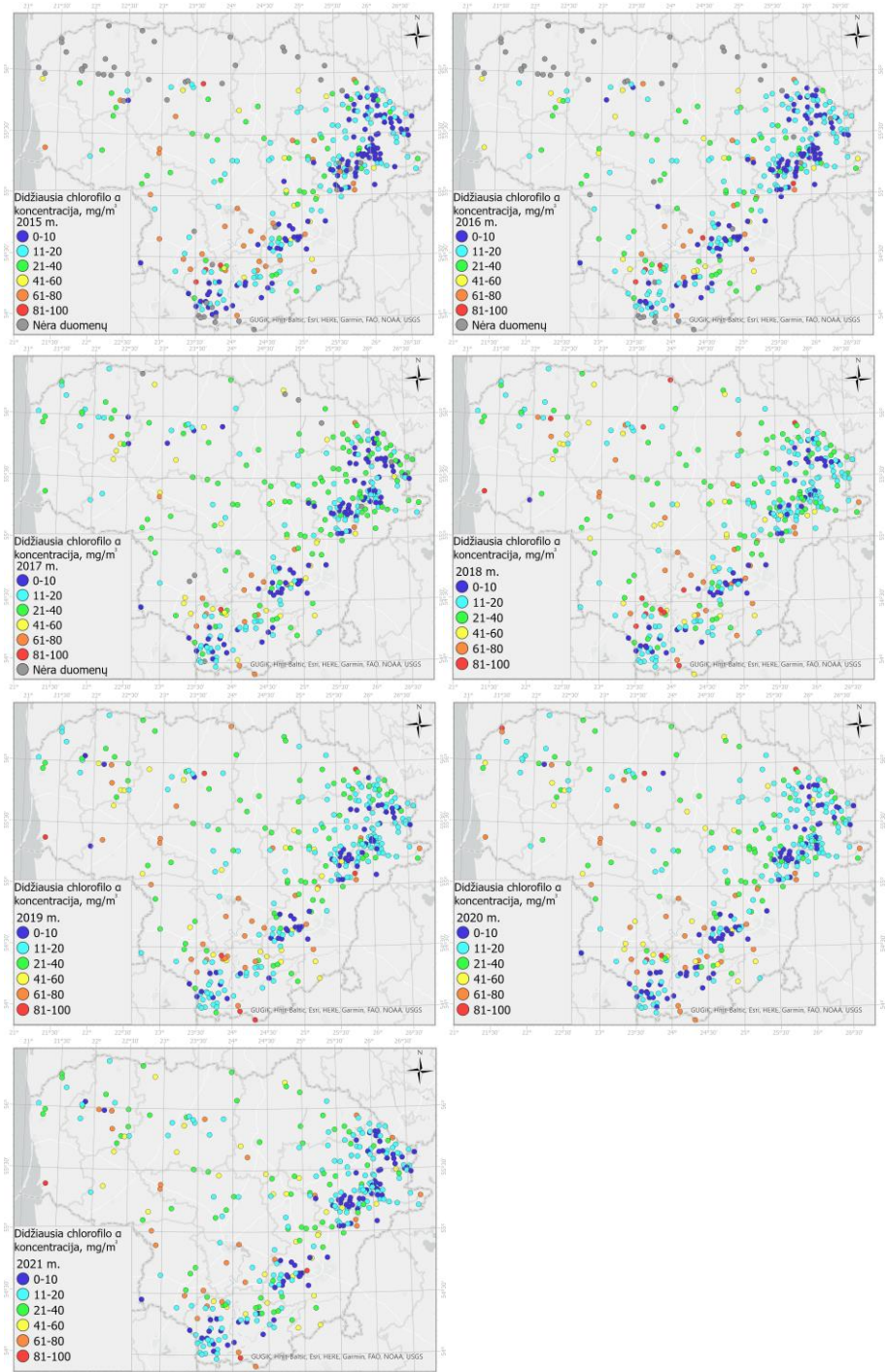
kitimas (mažėjantis ar didėjantis) per tyrimo laikotarpį. Biržulio, Svirkų ir Laukeso ežeruose, Stepanionių ir Totorviečių tvenkiniuose nustatytas  $> 5 \text{ mg/m}^3$  chla didėjimas per metus, o Beržoro, Kalvių ir Stervo ežeruose  $> 5 \text{ mg/m}^3$  mažėjimas per metus.



**5.13 pav.** Metinio chlorofilo  $\alpha$  koncentracijos vidurkio 2015–2021 m. pokytis tirtuose ežeruose. Didesnis sumažėjimas ar padidėjimas rodo pokytį  $> 10 \text{ mg/m}^3$ , mažesnis –  $1\text{--}10 \text{ mg/m}^3$ .

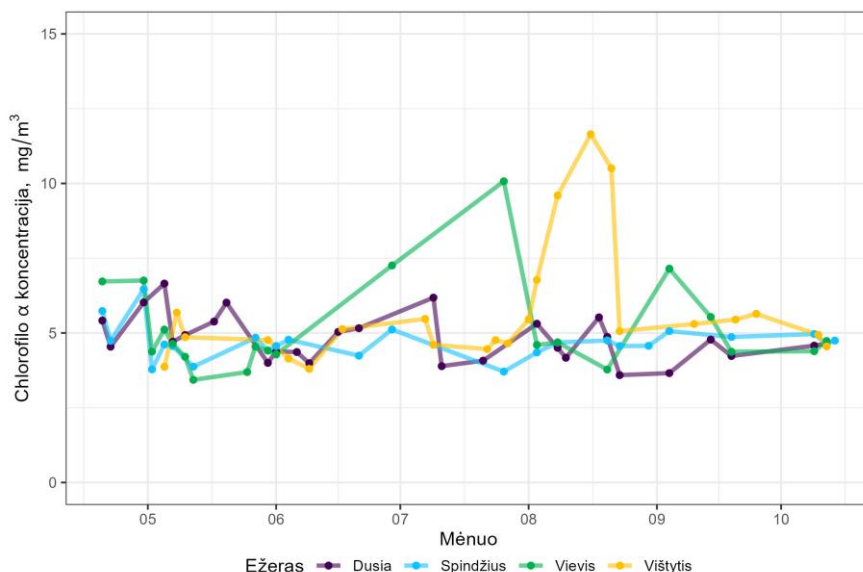
Potencialiai probleminiuose ežeruose vyko didelė chla koncentracijos kaita per šiltąjį sezoną. Pusėje šių ežerų (86) daugiau nei 54 % užfiksuotų palydovinių stebėjimų chla koncentracija buvo aukštesnė nei slenkstinė vertė, pagal kurią ežeras priskiriamas potencialiai probleminiams, o tai siejama su vandens žydėjimais.

Intensyvius žydėjimus galima išskirti tada, kai chla koncentracija viršija  $50 \text{ mg/m}^3$ . Beveik ketvirtyje (82) ežerų chla koncentracija buvo pasiekusi  $> 50 \text{ mg/m}^3$  pagal palydovinius duomenis. Šie ežerai pasižymi kasmetiniais dumblių žydėjimais ir dažniausiai maksimali chla koncentracija pasiekama rugpjūčio arba rugsėjo mėnesį. Šiuose ežeruose stebima gerokai intensyvesnė sezoninė chla kaita (medianinis standartinis nuokrypis  $13,7 \text{ mg/m}^3$ ). Didžioji dalis šių ežerų yra seklūs (1 tipo pagal morfometrines savybes).



**5.14 pav.** Didžiausia chlorofilo  $\alpha$  koncentracija 2015–2021 m. tirtuose ežeruose, įvertinta iš Sentinel-2 duomenų pasitelkus RFR modelius.

Ežeruose, kurie buvo priskirti neprobleminiams, chla koncentracija kito nedaug – didžiausias standartinis nuokrypis  $2,5 \text{ mg/m}^3$ , o variacijos koeficientas 39 %. Daugumoje ežerų sezoninė kaita nežymi, pvz., Dusios ir Spindžiaus ežeruose 2018 m. Kai kuriuose buvo nustatyti trumpalaikiai iki kelių savaičių trunkantys chla koncentracijos padidėjimai, pvz., Vievio ežere birželio–liepos mėnesį užfiksuotas  $5 \text{ mg/m}^3$  koncentracijos padidėjimas, o Vištyčio ežere rugpjūčio mėnesio pirmoje pusėje nustatytas 22 dienas trukęs chla koncentracijos padidėjimas (5.15 pav.). Šie chla padidėjimai siejami su nestipriais vandens žydėjimais, kurie neturi ilgalaikio poveikio ežero ekosistemai ir ekologiinei būklei.

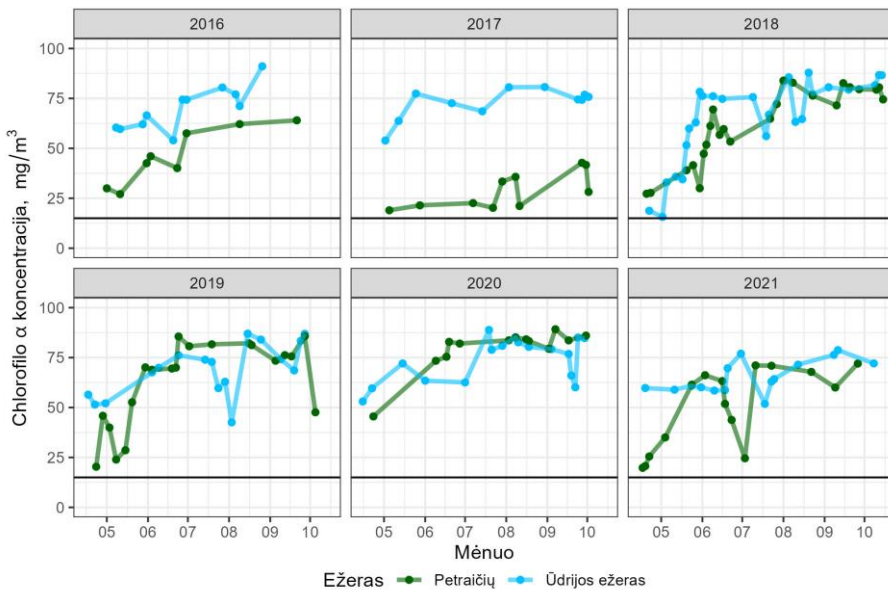


**5.15 pav.** Chlorofilo  $\alpha$  koncentracijos kaita 2018 m. kai kuriuose neprobleminiuose ežeruose (Dusioje, Spindžiuje, Vievio ežere ir Vištytyje).

Tarp potencialiai probleminių ežerų su aukšta maksimalia koncentracija ( $> 50 \text{ mg/m}^3$ ) didžiausia vidutine chla koncentracija (5 % procentilio) pasižymėjo ežerai (Ūdrijos, Obelių, Amalvo, Pravalos ežerai ir Petraičių tvenkinys), kur sezoninė chla kaita buvo kiek mažiau išreikšta (standartinis nuokrypis  $14,5\text{--}22,0 \text{ mg/m}^3$ ) nei kai kuriuose kituose ežeruose su taip pat palyginti aukšta chla koncentracija (standartinis nuokrypis iki  $28,7 \text{ mg/m}^3$ ). Šiuose ežeruose pavasario mėnesiais buvo fiksuojama santykinai aukšta chla koncentracija ( $30\text{--}80 \text{ mg/m}^3$ ), kuri vasaros mėnesiais padidėdavo iki  $90 \text{ mg/m}^3$  pagal palydovinius duomenis, o pagal *in situ* iki  $222 \text{ mg/m}^3$  (Obelių ežeras). Aukšta chla koncentracija išsilaikydavo ir ežerams vėstant rudens

metu – chla koncentracija ir spalio pirmoje pusėje dažnai siekdavo  $70 \text{ mg/m}^3$ , o Ūdrijos ežere ir Petraičių tvenkinyje  $83\text{--}87 \text{ mg/m}^3$  (5.16 pav.). Visgi turint omenyje, kad modelio vidutinė paklaida aukštų chla koncentracijų intervale buvo iki 30 %, tikrosios reikšmės galėjo būti didesnės nei gautos RFR modeliu iš palydovinių duomenų.

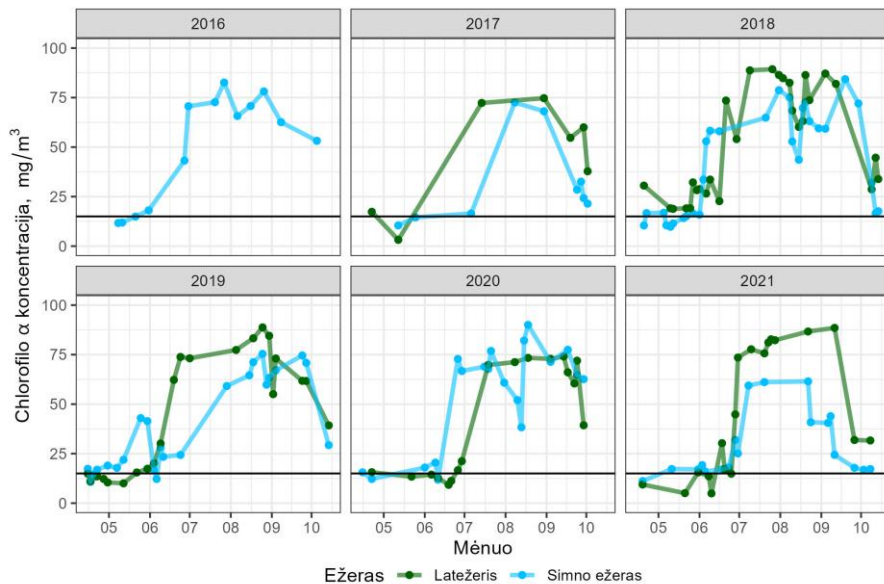
Kituose ežeruose su aukšta maksimalia chla koncentracija sezoninė chla kaita, užfiksuota palydovais, buvo išraiškingesnė. Ši kaita gerai atsispindėjo Latežeryje ir Simno ežere, kur balandžio–gegužės mėnesiais chla koncentracija atitinkamai buvo  $16,0$  ir  $16,6 \text{ mg/m}^3$ , vasarą pakildavo iki vidutiniškai  $56,1$  ir  $54,3 \text{ mg/m}^3$ , o rudenį išlikdavo dar gana aukšta – vidutiniškai  $57,1$  ir  $45,5 \text{ mg/m}^3$ . Atskirais metais eksponentinis dumblių augimas, apibūdinantis vandens žydėjimą, prasidėdavo Simne birželio pirmoje pusėje (2018 m.) arba truputį vėliau – birželio pabaigoje – liepos pirmoje pusėje (pvz., 2019 ir 2021 m.) (5.17 pav.). Eksponentinis augimas užfiksuojamas, kai tarp 2–3 stebėjimų nustatomas didelis chla koncentracijos didėjimo gradientas ( $> 5 \text{ mg/m}^3$  per dieną). Ši vandens žydėjimo fazė vyksta, kai yra palankios aplinkos sąlygos ir pakankamai maisto medžiagų.



**5.16 pav.** Metinė chlorofilo  $\alpha$  kaita pagal palydovinius duomenis Ūdrijos ežere ir Petraičių tvenkinyje, kuriuose nustatyta aukščiausia vidutinė chlorofilo  $\alpha$  koncentracija per 2015–2021 m. laikotarpį. Horizontali linija parodo chlorofilo  $\alpha$  koncentracijos momentinę slenkstinę vertę sekliuose (1 tipo pagal morfometrines savybes) vandens telkiniuose ( $15 \text{ mg/m}^3$ ).

Latežeryje eksponentinis augimas prasidėdavo birželio pirmoje pusėje (2019 m.), birželio viduryje (2018 m.) arba birželio gale (2020 ir 2021 m.). Greičiausio augimo metu chla padidėjimas siekė 9–14 mg/m<sup>3</sup> per dieną. Tačiau buvo užfiksuota atveju, kai augimas buvo dar staigesnis nei Latežeryje ir Simno ežere ir per dieną chla padidėjimas siekė iki 22,4 mg/m<sup>3</sup> (Didžiulis (Vilniaus raj., LTL134). Tokį augimą pavyko užfiksuoti kai kuriuose ežeruose, pasižyminčiuose aukšta vidutine koncentracija (medianinis vidurkis 36,9 mg/m<sup>3</sup>), kai buvo giedri orai ir buvo gaunami kokybiški palydoviniai duomenys kas 2–3 dienas. Iš viso 41 ežere buvo fiksuojamas augimas > 9 mg/m<sup>3</sup> per dieną, o dar 69-iuose – 5–9 mg/m<sup>3</sup> augimas per dieną.

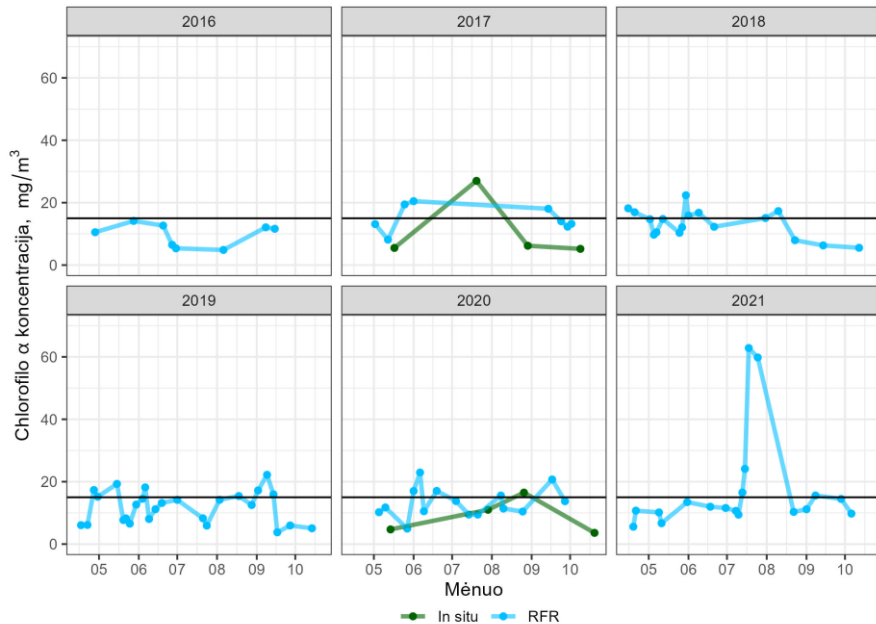
Pagal *in situ* duomenis didžiausias užfiksuotas augimo greitis buvo 6,9 mg/m<sup>3</sup> per dieną Kalvių ežere, dėl to staigų augimą *in situ* stebėjimais galima užfiksuoti, tik jei jie daromi kasdien ar kas kelias dienas, tačiau valstybinio monitoringo programa neatlieka tokio dažnumo matavimų.



**5.17 pav.** Metinė chlorofilo α kaita pagal palydovinius duomenis Latežeryje ir Simno ežere, kuriuose nustatyta viena iš aukščiausių vidutinė chlorofilo α koncentracija per 2015–2021 m. laikotarpį. Horizontali linija parodo chlorofilo α koncentracijos momentinę slenkstinę vertę sekliuose (1 tipo) vandens telkiniuose (15 mg/m<sup>3</sup>).

Tačiau greitas chla didėjimas užfiksuotas ne tik itin probleminiuose ežeruose, bet ir kai kuriuose, pasižyminčiuose palyginti žema vidutine koncentracija, pvz., Laukeso ežere, kurio vidutinė chla koncentracija

13,3 mg/m<sup>3</sup>, o didžiausia užfiksuota palydoviniais duomenimis – 62,8 mg/m<sup>3</sup> 2021 m. liepos mėn. (5.18 pav.), kai ežere staiga sužydėjo dumbliai ir chla augimas siekė 12,9 mg/m<sup>3</sup> per dieną, o tai parodė eksponentinį dumblių augimą. Žydėjimo pabaiga nebuvo užfiksuota, kadangi po stebėjimo su aukšta koncentracija, kitas stebėjimas dėl didelio debesuotumo sekė po 28 dienų, kada jau buvo užfiksuota koncentracija kaip prieš staigų chla didėjimą. Galima apytikriai įvertinti, kad žydėjimas truko nuo apytikriai dviejų savaičių iki mėnesio.

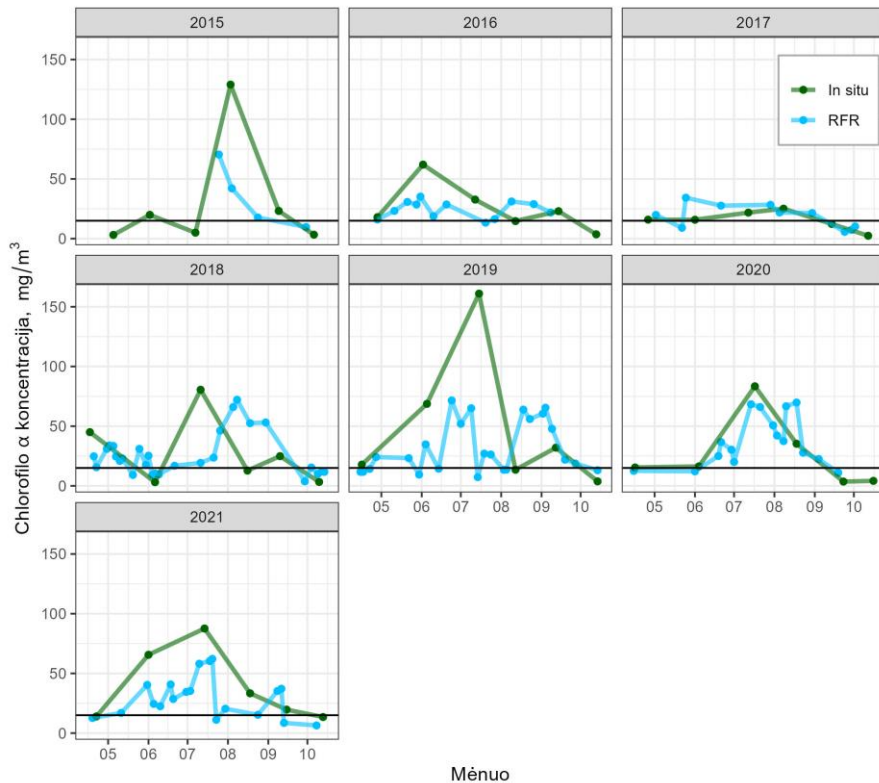


**5.18 pav.** Metinė chlorofilo  $\alpha$  kaita pagal palydovinius duomenis Laukeso ežere 2016–2021 m. Horizontali linija parodo chlorofilo  $\alpha$  koncentracijos momentinę slenkstinę vertę sekliuose (1 tipo pagal morfometrines savybes) vandens telkiniuose (15 mg/m<sup>3</sup>).

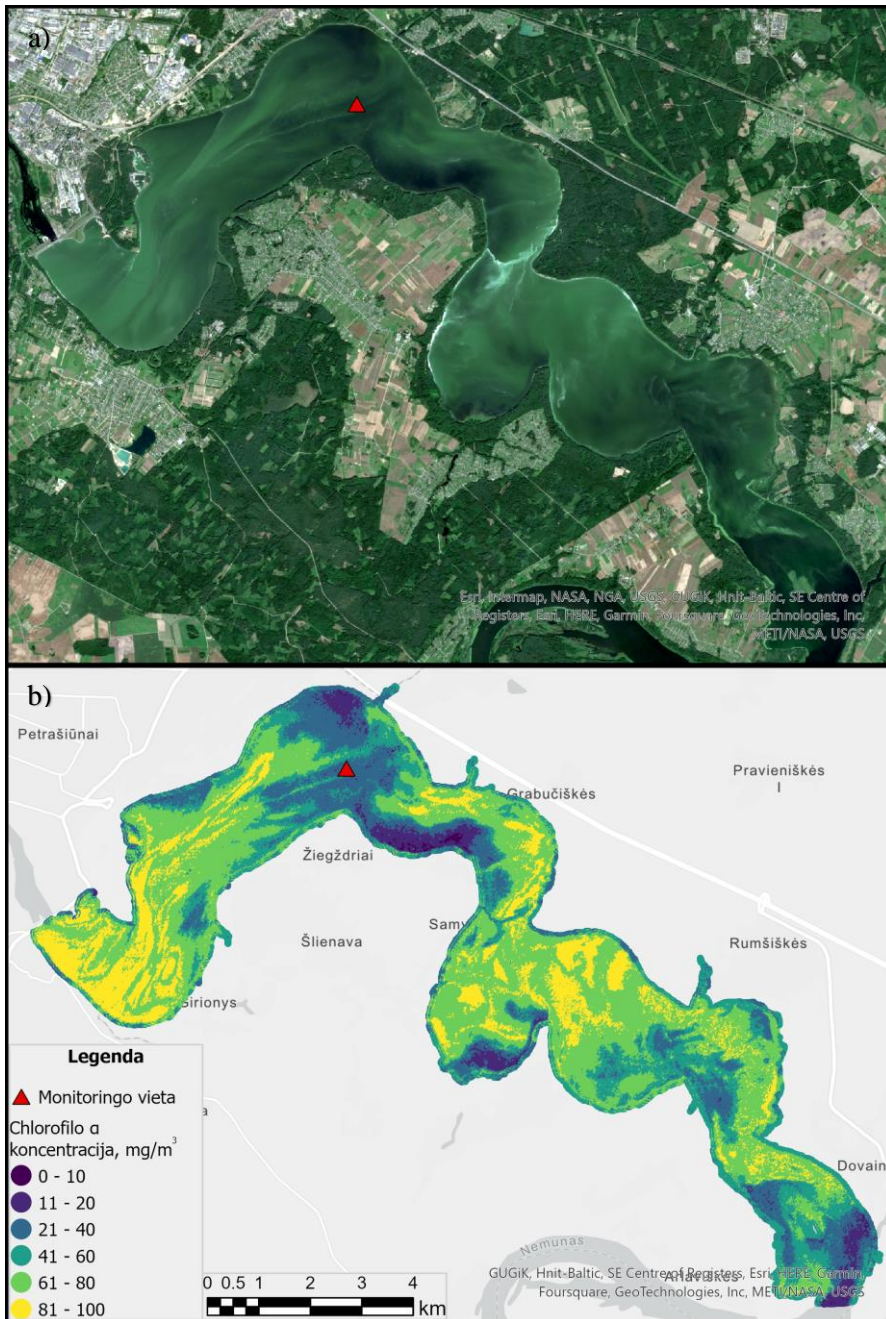
Palydoviniais duomenimis galima užfiksuoti chla koncentraciją ne tik dažnais stebėjimais laike, tačiau ir erdvėje – visame vandens telkinyje. Tai didelis palydovinių duomenų privalumas, ypač dideliems nehomogeniškiems vandens telkiniams. Erdvinė chla analizė gali parodyti valstybinio monitoringo matavimo vietos reprezentatyvumą visam vandens telkiniui. Pavyzdžiui, Kauno marios yra didžiausias vandens telkinys (plotas 63,5 km<sup>2</sup>) valstybinio ežerų ir tvenkinių monitoringo sistemoje, kuriam dažnai buvo priskirta biofizinė Chla-klasė, o pagal chla koncentraciją marios priskiriamos potencialiai probleminiams vandens telkiniams. Kauno mariose, kaip ir



kituose vandens telkiniuose, nepaisant vandens telkinio ploto, matavimai atliekami viename taške. *In situ* ir RFR gauta šio taško chla laiko eilutė 2017 ir 2020 m. labai panaši, tačiau kitais metais užfiksuotos koncentracijos labai skiriasi, o kaita truputį pasislinkusi laike (pvz., 2018 m., 5.19 pav.). Šio dirbtinio vandens telkinio dydis ir forma bei ekspozicija vyraujančioms vėjo kryptims lemia didelį erdvinį chla koncentracijos netolygumą (5.20 pav.) Aukščiausios koncentracijos dažnai užfiksuojamos vakarinėje ir centrinėje marių dalyje, o monitoringo vietoje koncentracija dažnu atveju būna truputį žemesnė, todėl nereprezentuoja vandens žydėjimų šiame vandens telkinyje ir atspindi vandens telkinio būklę tik iš dalies.



**5.19 pav.** Metinė chlorofilo  $\alpha$  kaita pagal palydovinius duomenis Kauno mariose 2015–2021 m. Horizontali linija parodo chlorofilo  $\alpha$  koncentracijos momentinę slenkstinę vertę vidutiniškai giliuose (2 tipo) vandens telkiniuose ( $15 \text{ mg/m}^3$ ).



5.20 pav. Kauno marių a) Sentinel-2 natūralių spalvų RGB kompozicija ir b) RFR modeliais gautos chlorofilo  $\alpha$  koncentracijos pasiskirstymas 2020 rugpjūčio 7 d.

## REKOMENDACIJOS

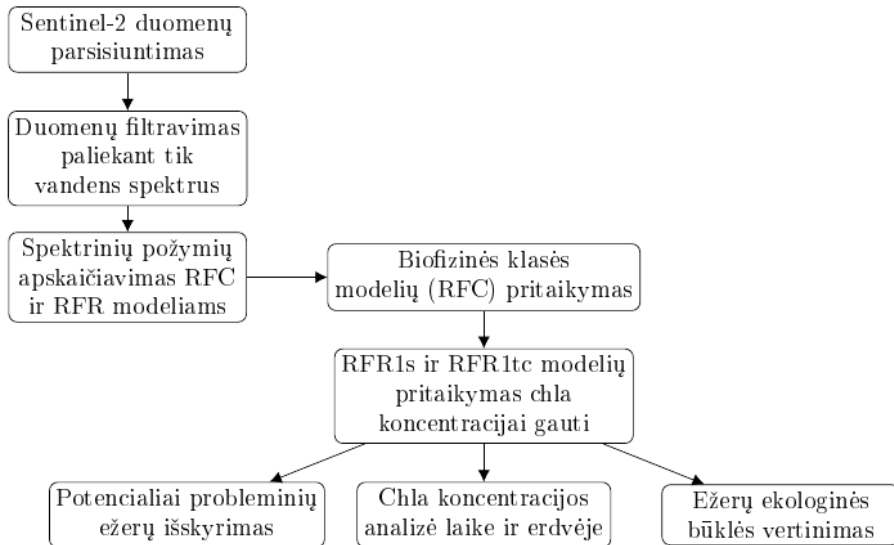
Dėl antropogeninės veiklos ir klimato kaitos vandens telkiniams kyla grėsmių, kurias valdyti galima tik turint pakankamai informacijos apie vandenų būklę. Tradiciniai vandens kokybės matavimai atliekami *in situ* nėra pakankamai dažni, kad leistų stebėti vandens žydėjimų dinamiką. Taip pat apsiribojama matavimu viename taške, todėl trūksta informacijos apie erdvinį vandens parametrų pasiskirstymą. Šiuos trūkumus galima spręsti naudojant palydovinius duomenis, dėl to pritaikius tyrime sukurtus algoritmus galima efektyviai gauti daug vandens telkinių apimančius ir dažnus duomenis.

Tyrimo rezultatai parodė, kad nepaisant didelio nustatyto chl<sub>a</sub> koncentracijos neapibrėžtumo, naudojant skirtingus AK produktus, Sentinel-2 MSI duomenys gali būti naudojami chl<sub>a</sub> koncentracijos gavimui, kai yra taikomi mažiau jautrūs atmosferos korekcijai chl<sub>a</sub> algoritmai/spektriniai požymiai kartu su geresniu atmosferos korekcijos produktu. Taip pat atmosferai nekoreguotų (TOA) duomenų naudojimas galėtų būti tinkamas sprendimas tais atvejais, kai *in situ* spektriniai duomenys, padedantys pasirinkti tinkamesnį AK produktą, yra neprieinami.

Įvertinus atmosferos korekcijos neapibrėžtumą chl<sub>a</sub> koncentracijai gauti buvo nustatyta, kad nuotrauka paremti AK produktai veikia panašiai ir tik ilgesnėse bangose (> 740 nm) pateikia skirtingus atspindžius. Kadangi, kai vandenyje nėra daug priemaišų, infraraudonosios spinduliuotės spektrinėse juostose dėl vandens sugėrimo atspindys yra nežymus, todėl tolesniam duomenų apdorojimui pasirinktas Sen2Cor produktas, pateikiantis mažus atspindžius šiuose bangos ilgiuose. Sentinel-2 MSI duomenis su pašalintu atmosferos poveikiu (L2A) ir duomenų filtravimui reikalinga nuotraukos klasifikacija galima parsisiųsti iš Copernicus Open Access Hub arba naudotis palydovinių duomenų peržiūros ir analizės platformomis, pvz., Google Earth Engine, turinčiomis palydovinių duomenų rinkinius (6.1 pav).

Dėl ežerų ir tvenkinių vandens savybių kompleksškumo ir atmosferos korekcijos neapibrėžtumo vertinimo metu gautų rezultatų buvo nuspręsta tikslesniam chl<sub>a</sub> modelių kūrimui sukurti biofizinę ežerų klasifikaciją. Biofizinės klasifikacijos modeliams kurti buvo panaudoti mašininio mokymosi algoritmai, iš kurių šiam uždaviniui spręsti labiausiai tiko atsitiktinio miško algoritmas. Remiantis tyrime sukurta biofizine klasifikacija atskiriami mažai optiškai aktyvių medžiagų turintys ežerai (skaidri klasė) nuo ežerų su optiškai aktyviomis medžiagomis, kuriems nustatyta tarpinė, Chl<sub>a</sub>-klasė ir drumsta klasė. Pastarosios klasės indikuoja apie potencialų vandens telkinio problemiškumą pagal chl<sub>a</sub> koncentraciją, kadangi visi ežerai, kuriems buvo nustatyta tarpinė ar Chl<sub>a</sub>-klasė kaip vyraujanti per tyrimo laikotarpį buvo

priskirti potencialiai probleminiams. Siekiant gauti biofizinę ežero klasę palydoviniam stebėjimui, paskaičiuojami spektriniai požymiai: BD1, R560/R705 ir AVW (2 klasėms atskirti – skaidriai ir OAS-klasei) ir R705/R665, AVW bei atspindžio amplitudė (3 klasėms atskirti – tarpinei, Chla ir drumstai klasei, 4.4 lentelė) ir pritaikomi RFC modeliai (6.1 pav.).



**6.1 pav.** Darbo eiga chlorofilo  $\alpha$  koncentracijos nustatymui pagal tyrime sukurtą metodiką.

Biofizinė klasifikacija panaudota chla modelių kūrimui naudojant atsitiktinio miško algoritmą. Biofizinė klasifikacija padidino chla algoritmų tikslumą naudojant Sentinel-2 MSI duomenis, ypač tarpinės ir Chla-klasės ežeruose. Skaidrios klasės (RFR1s) ežerams chla nustatymui reikalingi spektriniai požymiai: BD1, atspindžio amplitudė, R560/R705 ir atspalvio kampas, o tarpinės ir Chla-klasės (RFR1tc) ežerams – BD1, R705/R665, R560/R490, B8A, R560/R705, R560/R665.

Sukurti ir validuoti chla modeliai buvo pritaikyti 2015–2021 m. 357 ežerų spektrams, gautiems iš Sentinel-2 MSI duomenų. Tokiu būdu buvo gauti duomenys apie 43 per tyrimo laikotarpį *in situ* nestebėtus ežerus bei surinkta žymiai daugiau dideliu dažniu pasižyminčių duomenų apie visus 357 ežerus. Šiais duomenimis buvo užfiksuoti trumpalaikiai chla koncentracijos padidėjimai (pvz., Laukeso ežere), kurie parodė vandens žydėjimus. Taip pat didesnio dažnio duomenys leido užfiksuoti daugiau ežerų (86) nei *in situ* (53 ežerai), kuriuose buvo nustatyta  $> 50 \text{ mg/m}^3$  chla koncentracija, kadangi esant giedresniems orams palydoviniai stebėjimai buvo gaunami kas 2–3 dienas.

Turint tokius dažnus duomenis kai kuriuose ežeruose buvo stebima eksponentinė – staigaus sužydėjimo – fazė (41-ame ežere chla augimas per dieną  $9 \text{ mg/m}^3$ , o dar 69-ame  $5\text{--}9 \text{ mg/m}^3$ ).

Surinkta chla duomenų bazė gali būti panaudota ir chla didėjimo/mažėjimo per laiką nustatymui bei potencialiai probleminių ežerų išskyrimui, kadangi neapsiribojama keliais stebėjimais per metus ar net visą tyrimo laikotarpį, o viename vandens telkinyje turima keliasdešimt stebėjimų (mediana 88 per tyrimo laikotarpį). Pagal vidutines ir maksimalias chla koncentracijas buvo išskirti potencialiai probleminiai ežerai – nustatyti 52 neprobleminiai ir 305 potencialiai probleminiai ežerai. Šie surinkti duomenys galėtų būti naudingi vandenų būklės vertinimo ekspertams, sprendimų priėmėjams bei visuomenei, kadangi tik gera būkle pasižymintis vanduo – saugus naudoti.

Sukurtų modelių tolesnis panaudojimas galėtų padėti optimizuoti monitoringo programas (pvz., parenkant stebėjimų laiką pagal palydovinius duomenis nustatytą vandens žydėjimo dinamiką, parenkant matavimo vietą pagal erdvinį chla pasiskirstymą) ir prisidėti prie geresnio vandens telkinių būklės nustatymo tiek ilgalaikiu laikotarpiu vertinant ekologinę ežerų būklę ir jos kaitos tendencijas, tiek vertinant trumpalaikius chla svyravimus laike ir erdvėje, kadangi nauji duomenys gali būti gaunami beveik realiu metu. Taip pat šie duomenys galėtų padėti ieškant prastos/prastėjančios būklės priežasčių ir pagrįsti atitinkamų vandens būklei gerinti sprendimų priėmimą.

Sukurti modeliai gali būti naudojami Lietuvos ežerams ir kitiems vidutinių platumų ežerams, pasižymintiems panašiomis savybėmis, t. y., pagrindinė medžiaga juose – chlorofilas  $\alpha$ . Aprašyta biofizinė klasifikacija, naudojama kaip pirmas žingsnis chla modeliams kurti, ir suteikia pirminę informaciją apie ežero būklę. Ši klasifikacija gali būti praplečiama išskiriant kitas 4.1 lentelėje paminėtas klases, kai turima daugiau jas apibūdinančių duomenų. Tokiu atveju į biofizinės klasifikacijos RFC modelį reiktų įtraukti šiuos papildomus duomenis ir RFC modelį atnaujinti. Modelių atnaujinimui būtų galima panaudoti naujas pasaulines duomenų bazines, pvz., GLORIA (Lehmann ir kt., 2023). Atnaujinus biofizinės klasės modelį, būtų reikalinga sukurti tai klasei naują chla modelį. Tokiu būdu sukurta metodika galėtų būti plėtojama toliau ir pritaikoma įvairesniems vandens telkiniams.

## IŠVADOS

1. Spektrinėse juostose 665–740 nm, kurios yra svarbios chlorofilo  $\alpha$  (chl $\alpha$ ) nustatymui iš palydovinių duomenų, 7 skirtingos architektūros atmosferos korekcijos algoritmais gauti skirtingi atspindžiai. Skirtingi spektrai lėmė didelę chl $\alpha$  koncentracijos variaciją naudojant 7 atmosferos korekcijos produktus ir chl $\alpha$  algoritmus. Neapibrėžtumas naudojant TOA duomenis ir chl $\alpha$  algoritmus buvo mažesnis nei naudojant atmosferos korekcijos produktus.

2. Biofizinei ežerų klasifikacijai pagal tikslumo metrikas buvo parinktas atsitiktinio miško klasifikacijos algoritmas. Sukurti klasifikacijos modeliai leidžia atskirti keturias klases (skaidri, tarpinė, Chl $\alpha$ -klasė ir drumsta). Per tyrimo laikotarpį dažniausiai ežerų stebėjimams priskirta skaidri klasė (58,4 %), pasižyminti mažais optiškai aktyvių medžiagų kiekiais. Antra pagal dažnumą – Chl $\alpha$ -klasė (21,6 %), kuri siejama su vandens žydėjimais.

3. Skaidriai (RFR1s) ir tarpinei su Chl $\alpha$ -klase (RFR1tc) sukurti chl $\alpha$  modeliai, paremti atsitiktinio miško regresijos algoritmu, pasižymėjo didesniu tikslumu nei visus duomenis naudojantis RFR2 modelis ir dirbtiniais neuroniniais tinklais paremtas modelis, sukurti stebėjimų neskirstant į klases.

4. Remiantis chl $\alpha$  koncentracija, gauta RFR1s ir RFR1tc modeliais, buvo užfiksuoti trumpalaikiai (kelių dienų, kelių savaitių) bei staigūs chl $\alpha$  padidėjimai, kurie parodė skirtingo intensyvumo vandens žydėjimus. Modeliais taip pat galima užfiksuoti erdvinį chl $\alpha$  koncentracijos pasiskirstymą ežeruose ir tvenkiniuose, kas gali būti panaudojama reprezentatyvios monitoringo vietos pasirinkimui.

5. Per 2015–2021 m. laikotarpį nustatyta, kad 126 ežeruose būklė pagal vidutinę chl $\alpha$  koncentraciją nepakito, 155 ežeruose nustatytas būklės pagerėjimas – vidutinė chl $\alpha$  koncentracija sumažėjo, o 76 ežeruose stebėtas būklės pablogėjimas dėl padidėjusios vidutinės chl $\alpha$  koncentracijos.

6. Pritaikius RFR1s ir RFR1tc modelius palydoviniams duomenims nustatyta chl $\alpha$  koncentracija 2015–2021 m. visiems 357 valstybinio monitoringo ežerams ir tvenkiniams papildant *in situ* stebėjimų sekas dažniais duomenimis bei surenkant duomenis apie 43 *in situ* nestebėtus ežerus. Pagal nustatytas vidutines ir maksimalias chl $\alpha$  koncentracijas bei ežero tipą buvo išskirti 52 neprobleminiai ir 305 potencialiai probleminiai ežerai. Nustatyta, kad vyraujanti tarpinė ir Chl $\alpha$  biofizinė klasė visais atvejais buvo siejama su potencialiai problemineis ežerais, tačiau skaidrios klasės ežerų problemškumo įvertinimui biofizinė klasė nėra pakankamas indikatorius ir geriau remtis chl $\alpha$  rezultatais.

## LITERATŪROS SĄRAŠAS

- Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., & Networks, P. (2019). *Optuna: A Next - generation Hyperparameter Optimization Framework*. 1–10.
- Al-Kharusi, E. S., Tenenbaum, D. E., Abdi, A. M., Kutser, T., Karlsson, J., Bergström, A.-K., & Berggren, M. (2020). Large-Scale Retrieval of Coloured Dissolved Organic Matter in Northern Lakes Using Sentinel-2 Data. *Remote Sensing*, *12*(1), 157. <https://doi.org/10.3390/rs12010157>
- Ammenberg, P., Flink, P., Lindell, T., Pierson, D., & Strömbeck, N. (2002). Bio-optical modelling combined with remote sensing to assess water quality. *International Journal of Remote Sensing*, *23*(8), 1621–1638. <https://doi.org/10.1080/01431160110071860>
- Ansper, A., & Alikas, K. (2019). Retrieval of chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes. *Remote Sensing*, *11*(64). <https://doi.org/10.3390/rs11010064>
- Aplinkos Apsaugos Agentūra. (2018). *Žemės ūkis ir Lietuvos vandenys. Žemės ūkio veilos pproveikis Lietuvos upių būklei ir taršos apkrovoms į Baltijos jūrą*.
- Balevičienė, J., Balevičius, A., Bukelskis, E., Ciūnys, A., Valiuškevičius, G., Kalytytė, D., Ūselytė, R., & Šalčiūnienė, K. (2009). *Restauruotinių Lietuvos ežerų nustatymas ir preliminarus restauravimo priemonių parinkimas šiems ežerams, siekiant pagerinti jų būklę. I dalis. Bendrieji reikalavimai vandens telkinių monitoringui, patvirtinti 2003-12-31 Lietuvos Respublikos aplinkos ministro įsakymu Nr. 726*. (2003).
- Blix, K., & Eltoft, T. (2018). Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval. *Remote Sensing*, *10*(5). <https://doi.org/10.3390/rs10050775>
- Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–31. <https://doi.org/https://doi.org/10.1023/A:1010933404324>
- Bresciani, M., Adamo, M., De Carolis, G., Matta, E., Pasquariello, G., Vaičiūtė, D., & Giardino, C. (2014). Monitoring blooms and surface accumulation of cyanobacteria in the Curonian Lagoon by combining MERIS and ASAR data. *Remote Sensing of Environment*, *146*, 124–135. <https://doi.org/10.1016/j.rse.2013.07.040>
- Bricaud, A., Morel, A., & Prieur, L. (1981). Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, *26*(1), 45–53. <https://doi.org/10.4319/lo.1981.26.1.0043>
- Brockmann, C., Doerffer, R., Peters, M., Stelzer, K., Embacher, S., Ruescas, A., & (1). (2016). Evolution of the C2RCC Neural Network for Sentinel 2 and 3 for the Retrieval of Ocean Colour Products in Normal and Extreme Optically Complex Waters. *Proc. 'Living Planet Symposium 2016'*, *5*, 1–6.
- Caballero, I., Steinmetz, F., & Navarro, G. (2018). Evaluation of the first year

- of operational Sentinel-2A data for retrieval of suspended solids in medium- to high-turbidity waters. *Remote Sensing*, 10(7). <https://doi.org/10.3390/rs10070982>
- Carlson, R. E. (1977). A trophic state index for lakes. *Limnology and Oceanography*, 22(2), 361–369. <https://doi.org/10.4319/lo.1977.22.2.0361>
- Chávez, P. S. J. (1996). Image-based atmospheric corrections - revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62(9), 1025–1036. <https://doi.org/0099-1112/96/6209-1025>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chollet, F., & others. (2015). *Keras*.
- Coffer, M. M., Schaeffer, B. A., Salls, W. B., Urquhart, E., Loftin, K. A., Stumpf, R. P., Werdell, P. J., & Darling, J. A. (2021). Satellite remote sensing to assess cyanobacterial bloom frequency across the United States at multiple spatial scales. *Ecological Indicators*, 128(May), 107822. <https://doi.org/10.1016/j.ecolind.2021.107822>
- da Silva, E. F. F., Novo, E. M. L. de M., Lobo, F. de L., Barbosa, C. C. F., Noernberg, M. A., Rotta, L. H. da S., Cairo, C. T., Maciel, D. A., & Flores Júnior, R. (2020). Optical water types found in Brazilian waters. *Limnology, Kirk 2011*. <https://doi.org/10.1007/s10201-020-00633-z>
- Dall’Olmo, G., & Gitelson, A. A. (2005). Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Applied Optics*, 44(3), 412–422. <https://doi.org/10.1364/AO.45.003577>
- Dall’Olmo, G., Gitelson, A. A., & Rundquist, D. C. (2003). Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. *Geophysical Research Letters*, 30(18), 8–11. <https://doi.org/10.1029/2003GL018065>
- De Keukelaere, L., Sterckx, S., Adriaensens, S., Knaeps, E., Reusen, I., Giardino, C., Bresciani, M., Hunter, P., Neil, C., Van der Zande, D., & Vaiciute, D. (2018). Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: validation for coastal and inland waters. *European Journal of Remote Sensing*, 51(1), 525–542. <https://doi.org/10.1080/22797254.2018.1457937>
- Ding, X., Gong, F., Zhu, Q., Li, J., Wang, X., & Bai, R. (2022). Using geostationary satellite ocean color data and superpixel to map the diurnal dynamics of water transparency in the eastern China seas. *Ecological Indicators*, 142(July), 109219. <https://doi.org/10.1016/j.ecolind.2022.109219>
- Doerffer, R. (2015). *C2RCC - Algorithm Theoretical Bases Document (ATBD) for L2 processing of MERIS data of case 2 waters, 4th reprocessing*. 2, 55.



- Donis, D., Mantzouki, E., McGinnis, D. F., Vachon, D., Gallego, I., Grossart, H. P., de Senerpont Domis, L. N., Teurlincx, S., Seelen, L., Lüring, M., Verstijnen, Y., Maliaka, V., Fonvielle, J., Visser, P. M., Verspagen, J., van Herk, M., Antoniou, M. G., Tsiarta, N., McCarthy, V., ... Ibelings, B. W. (2021). Stratification strength and light climate explain variation in chlorophyll a at the continental scale in a European multilake survey in a heatwave summer. *Limnology and Oceanography*, 66(12), 4314–4333. <https://doi.org/10.1002/lno.11963>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- Environmental Protection Agency. (2018). *Agriculture and Lithuanian waters. Impact of agricultural activities to the status of Lithuanian rivers and pollution loads to the Baltic Sea (in Lithuanian)*. Vilnius.
- ESA. (2021). *Gearing up for third Sentinel-2 satellite*. [https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-2/Gearing\\_up\\_for\\_third\\_Sentinel-2\\_satellite](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Gearing_up_for_third_Sentinel-2_satellite)
- Eskandari, E., Mohammadzadeh, H., Nassery, H., Vadiati, M., Zadeh, A. M., & Kisi, O. (2022). Delineation of isotopic and hydrochemical evolution of karstic aquifers with different cluster-based (HCA, KM, FCM and GKM) methods. *Journal of Hydrology*, 609(February), 127706. <https://doi.org/10.1016/j.jhydrol.2022.127706>
- European Parliament and The Council. (2014). *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*. 2000. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000L0060>
- Europos Parlamentas ir Taryba. (2014). *Vandens pagrindų direktyva*. 2000. <https://eur-lex.europa.eu/legal-content/LT/ALL/?uri=celex:32000L0060>
- Fan, Y., Li, W., Gatebe, C. K., Jamet, C., Zibordi, G., Schroeder, T., & Stamnes, K. (2017). Atmospheric correction over coastal waters using multilayer neural networks. *Remote Sensing of Environment*, 199, 218–240. <https://doi.org/10.1016/j.rse.2017.07.016>
- Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. *ICML*, 96, 148–156.
- Gailiušis, B., Jablonskis, J., & Kovalenkoviėnė, M. (2001). *Lietuvos upės. Hidrografija ir nuotėkis* (Numeris July). Lietuvos energetikos institutas.
- Gilerson, A. A., Gitelson, A. A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., & Ahmed, S. A. (2010). Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. *Optics Express*, 18(23), 24109.

<https://doi.org/10.1364/oe.18.024109>

- Gitelson, A. A., Gurlin, D., Moses, W. J., & Barrow, T. (2009). A bio-optical algorithm for the remote estimation of the chlorophyll-a concentration in case 2 waters. *Environmental Research Letters*, 4(4), 2–7. <https://doi.org/10.1088/1748-9326/4/4/045003>
- Gómez, D., Salvador, P., Sanz, J., & Casanova, J. L. (2021). A new approach to monitor water quality in the Menor sea (Spain) using satellite data and machine learning methods. *Environmental Pollution*, 286(May). <https://doi.org/10.1016/j.envpol.2021.117489>
- Gordon, H. R., & Clark, D. K. (1980). Atmospheric effects in the remote sensing of phytoplankton pigments. *Boundary-Layer Meteorology*, 18(3), 299–313. <https://doi.org/10.1007/BF00122026>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.06.031>
- Grendaitė, D. (2018). *Chlorophyll-A Concentration Assessment in Lithuanian Lakes Using Remote Sensing*. Vilnius University.
- Grendaitė, D., & Stonevičius, E. (2021). Uncertainty of atmospheric correction algorithms for chlorophyll  $\alpha$  concentration retrieval in lakes from Sentinel-2 data. *Geocarto International*, 0(0), 1–25. <https://doi.org/10.1080/10106049.2021.1958014>
- Grendaitė, D., Stonevičius, E., Karosienė, J., Savadova, K., & Kasperovičienė, J. (2018). Chlorophyll-a concentration retrieval in eutrophic lakes in Lithuania from Sentinel-2 data. *Geologija. Geografija*, 4(1), 15–28. <https://doi.org/10.6001/geol-geogr.v4i1.3720>
- Guanter, L. (2006). *New algorithms for atmospheric correction and retrieval of biophysical parameters in Earth Observation . Application to ENVISAT / MERIS data*. 285.
- Gurlin, D., Gitelson, A. A., & Moses, W. J. (2011). Remote estimation of chl-a concentration in turbid productive waters - Return to a simple two-band NIR-red model? *Remote Sensing of Environment*, 115(12), 3479–3490. <https://doi.org/10.1016/j.rse.2011.08.011>
- Hafeez, S., Wong, M., Ho, H., Nazeer, M., Nichol, J., Abbas, S., Tang, D., Lee, K., & Pun, L. (2019). Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II Waters: A Case Study of Hong Kong. *Remote Sensing*, 11(6), 617. <https://doi.org/10.3390/rs11060617>
- Hieronymi, M., Krasemann, H., Müller, D., Brockmann, C., Ruescas, A., Stelzer, K., Nechad, B., Ruddick, K., Simis, S., Tilstone, G., Steinmetz, F., & Regner, P. (2016). Ocean colour remote sensing of extreme case-2 waters. *European Space Agency, (Special Publication) ESA SP, SP-740*, 9–13.
- Hieronymi, M., Müller, D., & Doerffer, R. (2017). The OLCI neural network swarm (ONNS): A bio-geo-optical algorithm for open ocean and coastal

- waters. *Frontiers in Marine Science*, 4(140), 1–18. <https://doi.org/10.3389/fmars.2017.00140>
- Hovis, W. A., Clark, D. K., Anderson, F., Austin, R. W., Wilson, W. H., Baker, E. T., Ball, D., Gordon, H. R., Mueller, J. L., El-Sayed, S. Z., Sturm, B., Wrigley, R. C., & Yentsch, C. S. (1980). Nimbus-7 Coastal Zone Color Scanner: System Description and Initial Imagery. *Science*, 210(4465), 60–63. <https://doi.org/http://www.jstor.org/stable/1684603>
- Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, 16(8), 471–483. <https://doi.org/10.1038/s41579-018-0040-1>
- Hussein, N. M., & Assaf, M. N. (2020). Multispectral Remote Sensing Utilization for Monitoring Chlorophyll-a Levels in Inland Water Bodies in Jordan. *Scientific World Journal*, 2020. <https://doi.org/10.1155/2020/5060969>
- Ioannou, I., Gilerson, A., Gross, B., Moshary, F., & Ahmed, S. (2013). Deriving ocean color products using neural networks. *Remote Sensing of Environment*, 134, 78–91. <https://doi.org/10.1016/j.rse.2013.02.015>
- IOCCG. (2019). Uncertainties in Ocean Colour Remote Sensing. F. Melin (Sud.), *Journal of Chemical Information and Modeling* (T. 53, Numeris 9). International Ocean Colour Coordinating Group. <https://doi.org/10.1017/CBO9781107415324.004>
- Jaagus, J., Briede, A., Rimkus, E., & Remm, K. (2014). Variability and trends in daily minimum and maximum temperatures and in the diurnal temperature range in Lithuania, Latvia and Estonia in 1951–2010. *Theoretical and Applied Climatology*, 118(1–2), 57–68. <https://doi.org/10.1007/s00704-013-1041-7>
- Kashulin, N., Kashulina, T., & Bekkelund, A. (2021). Long-term eutrophication and dynamics of bloom-forming microbial communities during summer HAB in large arctic lake. *Environments - MDPI*, 8(8). <https://doi.org/10.3390/environments8080082>
- Kilkus, K., & Stonevičius, E. (2012). *Lietuvos vandenų geografija*. Vilniaus universitetas.
- Kim, Y. H., Im, J., Ha, H. K., Choi, J. K., & Ha, S. (2014). Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GIScience and Remote Sensing*, 51(2), 158–174. <https://doi.org/10.1080/15481603.2014.900983>
- Koponen, S., Jenni, A., Pulliainen, J., Kallio, K., Pyhälähti, T., Lindfors, A., Rasmus, K., & Hallikainen, M. (2007). A case study of airborne and satellite remote sensing of a spring bloom event in the Gulf of Finland. *Continental Shelf Research*, 27(4), 228–244. <https://doi.org/10.1016/j.csr.2006.10.006>
- Kratzer, S., Kyriliuk, D., Edman, M., Philipson, P., & Lyon, S. (2019). Synergy of Satellite, In Situ and Modelled Data for Addressing the Scarcity of Water Quality Information for Eutrophication Assessment

- and Monitoring of Swedish Coastal Waters. *Remote Sensing*, 11(17), 2051. <https://doi.org/10.3390/rs11172051>
- Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., & Casal, G. (2016). Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. *Remote Sensing*, 8(497), 1–15. <https://doi.org/10.3390/rs8060497>
- LAND 69-2005 „Vandens kokybė. Biocheminių parametru matavimas. Spektrometrinis chlorofilo-a koncentracijos nustatymas“, patvirtintas 2005-12- 28 Lietuvos Respublikos aplinkos ministro įsakymu Nr. D1-648. (2005).
- Lehmann, M. K., Gurlin, D., Pahlevan, N., Alikas, K., Anstee, J., Balasubramanian, S. V., Barbosa, C. C. F., Binding, C., Bracher, A., Bresciani, M., Burtner, A., Cao, Z., Dekker, A. G., Di Vittorio, C., Drayson, N., Errera, R. M., Fernandez, V., Ficek, D., Fichot, C. G., ... Yue, L. (2023). GLORIA - A globally representative hyperspectral in situ dataset for optical sensing of water quality. *Scientific Data*, 10(1), 1–14. <https://doi.org/10.1038/s41597-023-01973-y>
- Ligi, M., Kutser, T., Kallio, K., Attila, J., Koponen, S., Paavel, B., Soomets, T., & Reinart, A. (2017). Testing the performance of empirical remote sensing algorithms in the Baltic Sea waters with modelled and in situ reflectance data. *Oceanologia*, 59(1), 57–68. <https://doi.org/10.1016/j.oceano.2016.08.002>
- Loizzo, R. ., Guarini, R., Longo, F., Scopa, T., Formaro, R., Facchinetti, C., & Varacalli, G. (2018). PRISMA : The Italian Hyperspectral Mission. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 175–178.
- LR Aplinkos ministerija. (2018). *Paviršinių vandens telkinių tipų aprašas*.
- LR Aplinkos Ministerija. (2019). *Paviršinių vandens telkinių būklės nustatymo metodika*.
- Maier, P. M., & Keller, S. (2018). Estimating Chlorophyll a Concentrations of Several Inland Waters with Hyperspectral Data and Machine Learning Models. *arXiv Computer Science*, 1–6. <https://doi.org/arXiv:1904.02052v1>
- Maier, P. M., & Keller, S. (2019). *Application of Different Simulated Spectral Data and Machine Learning to Estimate the Chlorophyll A Concentration of Several Inland Waters*. <https://doi.org/10.48550/arXiv.1905.12563>
- Matthews, M. W. (2017). Chapter 6 – Bio-optical Modeling of Phytoplankton Chlorophyll-a. *Bio-optical Modeling and Remote Sensing of Inland Waters* (p. 157–188). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-804644-9.00006-9>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817.

- <https://doi.org/10.1080/01431161.2018.1433343>
- McClain, C. R. (2009). A decade of satellite ocean color observations. *Annual review of marine science*, 1, 19–42. <https://doi.org/10.1146/annurev.marine.010908.163650>
- Minomura, M., Kuze, H., & Takeuchi, N. (2001). Adjacency effect in the atmospheric correction of satellite remote sensing data: Evaluation of the influence of aerosol extinction profiles. *Optical Review*, 8(2), 133–141. <https://doi.org/10.1007/s10043-001-0133-2>
- Mishra, D. R., & Mishra, S. (2010). Plume and bloom: Effect of the mississippi river diversion on the water quality of lake pontchartrain. *Geocarto International*, 25(7), 555–568. <https://doi.org/10.1080/10106041003763394>
- Mishra, S., & Mishra, D. R. (2012). Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of Environment*, 117, 394–406. <https://doi.org/10.1016/j.rse.2011.10.016>
- Mittenzwey, K. -H, Ullrich, S., Gitelson, A. A., & Kondratiev, K. Y. (1992). Determination of chlorophyll a of inland waters on the basis of spectral reflectance. *Limnology and Oceanography*, 37(1), 147–149. <https://doi.org/10.4319/lo.1992.37.1.0147>
- Modabberi, A., Noori, R., Madani, K., Ehsani, A. H., Danandeh Mehr, A., Hooshyaripor, F., & Kløve, B. (2020). Caspian Sea is eutrophying: The alarming message of satellite data. *Environmental Research Letters*, 15(12). <https://doi.org/10.1088/1748-9326/abc6d3>
- Morel, A. (2001). Bio-Optical Models. *Encyclopedia of Ocean Sciences: Second Edition*, 385–394. <https://doi.org/10.1016/B978-012374473-9.00407-0>
- Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color. *Limnology and Oceanography*, 22(4), 709–722. <https://doi.org/10.4319/lo.1977.22.4.0709>
- Moses, W. J., Gitelson, A. A., Berdnikov, S., & Povazhnyy, V. (2009). Satellite Estimation of Chlorophyll- a Concentration Using the Red and NIR Bands of MERIS — The Azov Sea Case Study. *IEEE Geoscience and Remote Sensing Letters*, 6(4), 845–849. <https://doi.org/10.1109/LGRS.2009.2026657>
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Mueller-Wilm, U., Devignot, O., & Pessiot, L. (2019). *S2 MPC Sen2Cor Configuration and User Manual*. 2, 1–59.
- Ndebele-Murisa, M. R., Musil, C. F., & Raitt, L. (2010). A review of phytoplankton dynamics in tropical African lakes. *South African Journal of Science*, 106(1–2), 13–18. <https://doi.org/10.4102/sajs.v106i1/2.64>
- Neil, C., Spyrakos, E., Hunter, P. D., & Tyler, A. N. (2019). A global approach for chlorophyll-a retrieval across optically complex inland waters based

- on optical water types. *Remote Sensing of Environment*, 229(4), 159–178. <https://doi.org/10.1016/j.rse.2019.04.027>
- O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., & McClain, C. R. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, 103(C11), 24937–24953. <https://doi.org/10.1029/98JC02160>
- O'Shea, R. E., Pahlevan, N., Smith, B., Bresciani, M., Egerton, T., Giardino, C., Li, L., Moore, T., Ruiz-Verdu, A., Ruberg, S., Simis, S. G. H., Stumpf, R., & Vaičiūtė, D. (2021). Advancing cyanobacteria biomass estimation from hyperspectral observations: Demonstrations with HICO and PRISMA imagery. *Remote Sensing of Environment*, 266. <https://doi.org/10.1016/j.rse.2021.112693>
- Ogashawara, I. (2020). *Determination of Phycocyanin from Space — A Bibliometric Analysis. I*, 1–16.
- Pahlevan, N., Mangin, A., Balasubramanian, S. V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikawa, J., Kratzer, S., Lehmann, M. K., Ligi, M., ... Warren, M. (2021). ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sensing of Environment*, 258(February). <https://doi.org/10.1016/j.rse.2021.112366>
- Papathanasopoulou, E., Simis, S., Alikas, K., Ansper, A., Anttila, S., Jenni, A., Barillé, A.-L., Barillé, L., Brando, V., Bresciani, M., Bučas, M., Gernez, P., Giardino, C., Harin, N., Hommersom, A., Kangro, K., Kauppila, P., Koponen, S., Laanen, M., ... Zoffoli, M. L. (2019). Satellite-assisted monitoring of water quality to support the implementation of the Water Framework Directive. *EOMORES white paper, November*, 28. <https://doi.org/10.5281/zenodo.3463051>
- Paulino, R. S., Martins, V. S., Novo, E. M. L. M., Barbosa, C. C. F., de Carvalho, L. A. S., & Begliomini, F. N. (2022). Assessment of Adjacency Correction over Inland Waters Using Sentinel-2 MSI Images. *Remote Sensing*, 14(8). <https://doi.org/10.3390/rs14081829>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pereira-Sandoval, M., Ruescas, A., Urrego, P., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J., & Moreno, J. (2019). Evaluation of atmospheric correction algorithms over Spanish inland waters for Sentinel-2 multi spectral imagery data. *Remote Sensing*, 11(12), 1–23. <https://doi.org/10.3390/rs11121469>
- Philipson, P., Kratzer, S., Ben Mustapha, S., Strömbeck, N., & Stelzer, K. (2016). Satellite-based water quality monitoring in Lake Vänern, Sweden. *International Journal of Remote Sensing*, 37(16), 3938–3960.

- <https://doi.org/10.1080/01431161.2016.1204480>
- Plymouth Marine Laboratory. (2012). *Ocean Colour Climate Change Initiative (OC\_CCI) - Phase One*. <https://doi.org/10.1007/s11055-008-9074-6>
- Rao, P. K., Holmes, S. J., Anderson, R. K., Winston, J. S., & Lehr, P. E. (1990). A History of Civilian Weather Satellites. P. K. Rao, S. J. Holmes, R. K. Anderson, J. S. Winston, & P. E. Lehr (Sud.), *Weather Satellites: Systems, Data, and Environmental Applications* (p. 7–19). American Meteorological Society. [https://doi.org/10.1007/978-1-944970-16-1\\_2](https://doi.org/10.1007/978-1-944970-16-1_2)
- Reinart, A., Herlevi, A., Arst, H., & Sipelgas, L. (2003). Preliminary optical classification of lakes and coastal waters in Estonia and south Finland. *Journal of Sea Research*, 49(4), 357–366. [https://doi.org/10.1016/S1385-1101\(03\)00019-4](https://doi.org/10.1016/S1385-1101(03)00019-4)
- Rezaei, K., & Vadiati, M. (2020). A comparative study of artificial intelligence models for predicting monthly river suspended sediment load. *Journal of Water and Land Development*, 45, 107–118. <https://doi.org/10.24425/jwld.2020.133052>
- RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. <http://www.rstudio.com/>
- Saberioon, M., Brom, J., Nedbal, V., Souček, P., & Císař, P. (2020). Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters. *Ecological Indicators*, 113(February), 106236. <https://doi.org/10.1016/j.ecolind.2020.106236>
- Savadova-Ratkus, K., Mazur-Marzec, H., Karosienė, J., Sivonen, K., Suurnäkki, S., Kasperovičienė, J., Paškauskas, R., & Koreivienė, J. (2022). Cyanobacteria and Their Metabolites in Mono- and Polidominant Shallow Eutrophic Temperate Lakes. *International Journal of Environmental Research and Public Health*, 19(22). <https://doi.org/10.3390/ijerph192215341>
- Savadova, K., Mazur-Marzec, H., Karosienė, J., Kasperovičienė, J., Vitonytė, I., Toruńska-Sitarz, A., & Koreivienė, J. (2018). Effect of increased temperature on native and alien nuisance cyanobacteria from temperate lakes: An experimental approach. *Toxins*, 10(11), 1–15. <https://doi.org/10.3390/toxins10110445>
- Schalles, J. F., Gitelson, A. a, Yacobi, Y. Z., & Kroenke, A. E. (1998). Estimation of Chlorophyll a from Time Series Measurements of High Spectral Resolution Reflectance in an Eutrophic Lake. *Water*, 34(2), 383–390.
- Schindler, D. E., & Scheuerell, M. D. (2002). Habitat coupling in lake ecosystems. *Oikos*, 98(2), 177–189. <https://doi.org/10.1034/j.1600-0706.2002.980201.x>
- Schott, J. R., Gerace, A., Woodcock, C. E., Wang, S., Zhu, Z., Wynne, R. H., & Blinn, C. E. (2016). The impact of improved signal-to-noise ratios on algorithm performance: Case studies for Landsat class instruments.

- Remote Sensing of Environment*, 185, 37–45.  
<https://doi.org/10.1016/j.rse.2016.04.015>
- Song, W., Dolan, J. M., Cline, D., & Xiong, G. (2015). Learning-based algal bloom event recognition for oceanographic decision support system using remote sensing data. *Remote Sensing*, 7(10), 13564–13585.  
<https://doi.org/10.3390/rs71013564>
- Soomets, T., Uudeberg, K., Jakovels, D., Brauns, A., Zagars, M., & Kutser, T. (2020). Validation and comparison of water quality products in Baltic lakes using sentinel-2 MSI and Sentinel-3 OLCI data. *Sensors*, 20(742), 1–22. <https://doi.org/10.3390/s20030742>
- Sòria-Perpinyà, X., Vicente, E., Urrego, P., Pereira-Sandoval, M., Tenjo, C., Ruíz-Verdú, A., Delegido, J., Soria, J. M., Peña, R., & Moreno, J. (2021). Validation of water quality monitoring algorithms for sentinel-2 and sentinel-3 in mediterranean inland waters with in situ reflectance data. *Water (Switzerland)*, 13(5). <https://doi.org/10.3390/w13050686>
- Spyrakos, E., O'Donnell, R., Hunter, P. D., Miller, C., Scott, M., Simis, S. G. H., Neil, C., Barbosa, C. C. F., Binding, C. E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A. A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M. W., ... Tyler, A. N. (2018). Optical types of inland and coastal waters. *Limnology and Oceanography*, 63(2), 846–870. <https://doi.org/10.1002/lno.10674>
- Steinmetz, F., Deschamps, P.-Y., & Ramon, D. (2011). Atmospheric correction in presence of sun glint: application to MERIS. *Optics Express*, 19(10), 9783. <https://doi.org/10.1364/oe.19.009783>
- Steinmetz, F., & Ramon, D. (2018). *Sentinel-2 MSI and Sentinel-3 OLCI consistent ocean colour products using POLYMER*. 13. <https://doi.org/10.1117/12.2500232>
- Sterckx, S., Knaeps, S., Kratzer, S., & Ruddick, K. (2015). SIMilarity Environment Correction (SIMEC) applied to MERIS data over inland and coastal waters. *Remote Sensing of Environment*, 157, 96–110. <https://doi.org/10.1016/j.rse.2014.06.017>
- Sun, D., Li, Y., & Wang, Q. (2009). A unified model for remotely estimating chlorophyll a in Lake Taihu, China, based on SVM and in Situ hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8), 2957–2965. <https://doi.org/10.1109/TGRS.2009.2014688>
- Tarrio, K., Tang, X., Masek, J. G., Claverie, M., Ju, J., Qiu, S., Zhu, Z., & Woodcock, C. E. (2020). Comparison of cloud detection algorithms for Sentinel-2 imagery. *Science of Remote Sensing*, 2(September), 100010. <https://doi.org/10.1016/j.srs.2020.100010>
- The Ministry of Environment. (2018). *Description of types of surface waters (In Lithuanian)*. Lithuania.
- Tiškus, E., Bučas, M., Vaičiūtė, D., Gintauskas, J., & Babrauskienė, I. (2023). An Evaluation of Sun-Glint Correction Methods for UAV-Derived Secchi Depth Estimations in Inland Water Bodies. *Drones*, 7(9), 546.



- <https://doi.org/10.3390/drones7090546>
- Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First experiences in mapping lakewater quality parameters with sentinel-2 MSI imagery. *Remote Sensing*, 8(8). <https://doi.org/10.3390/rs8080640>
- Toming, K., Kutser, T., Uiboupin, R., Arikas, A., Vahter, K., & Paavel, B. (2017). Mapping water quality parameters with Sentinel-3 Ocean and Land Colour Instrument imagery in the Baltic Sea. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9101070>
- Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., Dillon, P., Finlay, K., Fortino, K., Knoll, L. B., Kortelainen, P. L., Kutser, T., Larsen, S., Laurion, I., Leech, D. M., McCallister, S. L., McKnight, D. M., Melack, J. M., Overholt, E., ... Weyhenmeyer, G. A. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography*, 54(6part2), 2298–2314. [https://doi.org/10.4319/lo.2009.54.6\\_part\\_2.2298](https://doi.org/10.4319/lo.2009.54.6_part_2.2298)
- Vaičiūtė, D., Bresciani, M., & Bučas, M. (2012). Validation of MERIS bio-optical products with in situ data in the turbid Lithuanian Baltic Sea coastal waters. *Journal of Applied Remote Sensing*, 6(1), 063568–1. <https://doi.org/10.1117/1.jrs.6.063568>
- Vaičiūtė, D., Bučas, M., Bresciani, M., Dabulevičienė, T., Gintauskas, J., Mėžinė, J., Tiškus, E., Umgiesser, G., Morkūnas, J., De Santi, F., & Bartoli, M. (2021). Hot moments and hotspots of cyanobacteria hyperblooms in the Curonian Lagoon (SE Baltic Sea) revealed via remote sensing-based retrospective analysis. *Science of the Total Environment*, 769. <https://doi.org/10.1016/j.scitotenv.2021.145053>
- van der Woerd, H. J., & Wernand, M. R. (2018). Hue-angle product for low to medium spatial resolution optical satellite sensors. *Remote Sensing*, 10(2), 1–18. <https://doi.org/10.3390/rs10020180>
- Vandermeulen, R. A., Mannino, A., Craig, S. E., & Werdell, P. J. (2020). 150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean. *Remote Sensing of Environment*, 247(December 2019), 111900. <https://doi.org/10.1016/j.rse.2020.111900>
- Vanhellemont, Q. (2019). Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sensing of Environment*, 225(November 2018), 175–192. <https://doi.org/10.1016/j.rse.2019.03.010>
- Vanhellemont, Q., & Ruddick, K. (2018). Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. *Remote Sensing of Environment*, 216(July), 586–597. <https://doi.org/10.1016/j.rse.2018.07.015>
- VITO. (2019). *iCOR Landsat-8 and Sentinel-2 plugin for SNAP toolbox, Software User Manual. version 2.*, 1–21.
- Wang, M. (2007). Remote sensing of the ocean contributions from ultraviolet

- to near-infrared using the shortwave infrared bands: simulations. *Applied Optics*, 46(9), 1535–1547. <https://doi.org/10.1364/AO.46.001535>
- Wang, S.-C. (2003). Artificial Neural Network. *Interdisciplinary Computing in Java Programming* (p. 81–100). Springer US. [https://doi.org/10.1007/978-1-4615-0377-4\\_5](https://doi.org/10.1007/978-1-4615-0377-4_5)
- Warren, M. A., Simis, S. G. H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., & Ansper, A. (2019). Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sensing of Environment*, 225(3), 267–289. <https://doi.org/10.1016/j.rse.2019.03.018>
- Watanabe, F., Alcântara, E., Bernardo, N., de Andrade, C., Gomes, A. C., do Carmo, A., Rodrigues, T., & Rotta, L. H. (2019). Mapping the chlorophyll-a horizontal gradient in a cascading reservoirs system using MSI Sentinel-2A images. *Advances in Space Research*, 64(3), 581–590. <https://doi.org/10.1016/j.asr.2019.04.035>
- Woolway, R. I., Dokulil, M. T., Marszelewski, W., Schmid, M., Bouffard, D., & Merchant, C. J. (2017). Warming of Central European lakes and their response to the 1980s climate regime shift. *Climatic Change*, 142(3–4), 505–520. <https://doi.org/10.1007/s10584-017-1966-4>
- Woolway, R. I., Sharma, S., Weyhenmeyer, G. A., Debolskiy, A., Golub, M., Mercado-Bettín, D., Perroud, M., Stepanenko, V., Tan, Z., Grant, L., Ladwig, R., Mesman, J., Moore, T. N., Shatwell, T., Vanderkelen, I., Austin, J. A., DeGasperi, C. L., Dokulil, M., La Fuente, S., ... Jennings, E. (2021). Phenological shifts in lake stratification under climate change. *Nature Communications*, 12(1), 1–11. <https://doi.org/10.1038/s41467-021-22657-4>
- Yacobi, Y. Z., Gitelson, A., & Mayo, M. (1995). Remote sensing of chlorophyll in Lake Kinneret using highspectral-resolution radiometer and Landsat TM: spectral features of reflectance and algorithm development. *Journal of Plankton Research*, 17(11), 2155–2173.
- Žematijos nacionalinis parkas. (2019). *No Title*. Platelių ežero “žydėjimą“ galima pamatyti iš palydovinių nuotraukų. <https://zemaitijosnp.lt/plateliu-ezero-zydejima-galima-pamatyti-palydoviniu-nuotrauku/>
- Zhai, P.-W., Hu, Y., Winker, D. M., Franz, B. A., Werdell, J., & Boss, E. (2017). Vector radiative transfer model for coupled atmosphere and ocean systems including inelastic sources in ocean waters. *Optics Express*, 25(8), A223. <https://doi.org/10.1364/oe.25.00a223>

## CURRICULUM VITAE

Disertacijos autorė Dalia Grendaitė gimė 1993 m. liepos 15 d. Vilniuje. 2012 m. baigė Vilniaus Žirmūnų gimnaziją ir įstojo į Vilniaus universiteto Gamtos mokslų fakulteto meteorologijos ir hidrologijos studijų programą, kurią sėkmingai baigė 2016 m. cum laude diplomu, apsigynusi bakalauro darbą tema „Oro tarša ir jos poveikis sveikatai Oslo miesto mikroaplinkose“. Dalia toliau tęsė studijas Hidrologijos ir klimatologijos katedroje hidrometeorologijos magistrantūroje. Erasmus studijų 2017 m. Lundo universitete ir erasmus praktikos Nanseno aplinkos ir nuotolinių tyrimų centre metu, autorė susidomėjo nuotoliniais stebėjimais ir magistro darbą paruošė tema „Chlorofilo-a koncentracijos nustatymas distanciniais metodais Lietuvos ežeruose“. Pagal magistro darbą buvo paruoštas pirmasis autorės straipsnis „Chlorophyll-a concentration retrieval in eutrophic lakes in Lithuania from Sentinel-2 data“. 2018 m. po magistro studijų baigimo magna cum laude diplomu autorė įstojo į fizinės geografijos krypties doktorantūrą Vilniaus universiteto Chemijos ir geomokslų fakulteto Hidrologijos ir klimatologijos katedroje. Nuo antrų doktorantūros metų Dalia pradėjo vesti ežerotyros pratybas, o šiuo metu dėsto ežerotyros, Lietuvos vandenu geografijos kursus, veda hidrologijos seminarus ir GIS pagrindų pratybas.

## PRIEDAI

**1 priedas.** Ežerų ir tvenkinių sąrašas su *in situ* ir iš palydovinių duomenų bei atsitiktinio miško modeliu gautomis chlorofilo  $\alpha$  vidutinėmis ir maksimalios koncentracijomis. **Melsva spalva** pažymėta ten, kur pagal slenkstines vertes chlorofilo  $\alpha$  koncentracija nebuvo viršyta ir ežeras pagal tą kriterijų priskirtas neprobleminiams ežerams. **Žalsva spalva** pažymėta ten, kur ežere nustatyta vidutinė ar momentinė chlorofilo  $\alpha$  koncentracija viršijo slenkstinę vertę (vidutinė  $> 8 \text{ mg/m}^3$  1 ir 2 tipo ežerams,  $> 6 \text{ mg/m}^3$  3 tipo ežerams, maksimali  $> 15 \text{ mg/m}^3$  1 ir 2 tipo ežerams,  $12 \text{ mg/m}^3$  3 tipo ežerams). **Pilkai** pažymėta, kur nebuvo duomenų. Ežerų tipai pagal morfometrines savybes apibūdinti 2.1 lentelėje.

Nr.	Pavadinimas	MV kodas	T i p a s	N / n s i t u	N RFR	Chla vid. <i>In situ</i>	Chla maks. <i>In situ</i>	Chla vid. RFR	Chla maks. RFR
1	"Ekranas" gamyklos tvenkinys	LTL406	1		80			17,5	38,3
2	Aisetas	LTL336	3	8	88	2,7	7,1	7,6	15,8
3	Akmėna	LTL460	3	8	83	4,1	5,3	4,9	11,9
4	Akmenių ežeras	LTL516	1	8	100	15	28,4	16,1	38,2
5	Alaušai	LTL452	2	4	85	7	9,4	8,3	26,7
6	Alaušas	LTL231	3	8	92	2,5	3,2	4,6	14,5
7	Alksnas(Alksnait is)	LTL228	1		98			7	13,7
8	Alksnas(Ignalino s)	LTL331	2	4	95	9,4	14,5	6,2	12,8
9	Almajas	LTL322	1	4	85	5	6,8	6,6	15,3
10	Alnis	LTL65	2	4 1	91	2,9	6,5	4,3	9,3
11	Alovės ežeras	LTL192	1	8	97	37,2	70,6	31,2	77,1
12	Alsakys	LTL366	1	4	95	32,5	50,2	29,2	67,4
13	Alsėdzių ežeras	LTL258	1	8	63	17,6	37,3	29,3	68,5
14	Alys	LTL361	1		81			10,8	37,4
15	Amalvas	LTL244	1		102			57,1	80,2
16	Ančia	LTL195	2	8	112	8,5	24,2	6,6	17,3
17	Angirių tvenkinys	LTL410	2	4	47	6,8	10,9	11,2	28,1

18	Antakmenių ežeras	LTL115	2	8	85	16,9	37	9,4	25
19	Antalieptės HE tvenkinys	LTL433	3	4	91	3,1	3,9	5,4	13,4
20	Antanavo HE tvenkinys	LTL243	1	4	37	20,2	54,2	28,4	63,9
21	Apvardai	LTL225	1		110			12	34,9
22	Arimaičių ežeras	LTL266	1		98			14,7	28,4
23	Arinas	LTL230	2	4	76	3,6	5,9	6,5	15,3
24	Asalnai	LTL324	3	4	92	3,1	5,3	5,8	17,9
25	Asalnykštis	LTL329	2	4	89	5,4	10,1	6	21,2
26	Asavas	LTL310	1	$\frac{3}{6}$	88	5,2	14,1	6,3	14,8
27	Asveja(Dubingių ežeras)	LTL449	3	4	89	2,4	3	5,2	9,8
28	Atesys	LTL248	2	8	89	6,5	14,1	9,1	20
29	Aukštadvario HE tvenkinys(ež.Bal uosis ež.Sienis)	LTL110	1	4	96	4,7	5,2	5,1	12,9
30	Auslas	LTL303	1		74			7,7	17,2
31	Avilys	LTL299	1	4	99	5,4	6,9	6,6	15
32	Aviris	LTL181	3	4	102	3,9	5,4	5,3	10,2
33	Ažvintis	LTL318	2	8	88	5,7	11,4	6,1	14,6
34	Babrų ežeras	LTL235	2	8	100	6,8	15	7,8	16,4
35	Bagdononių HE	LTL537	1	4	100	12,8	25,6	6	12,1
36	Balandis	LTL546	2	8	96	5,5	12,7	5,5	12,8
37	Balskų tvenkinys	LTL419	2	8	95	23	117,5	16,3	78,1
38	Balsys(Žalieji ežerai)	LTL509	3	8	79	6,8	13,7	4,8	11,4
39	Baltas	LTL338	2	8	71	3,9	5,9	4,5	8,3
40	Baltausių tvenkinys	LTL404	1	8	59	30,1	57,2	25,9	80,4
41	Baltieji Lakajai	LTL157	3	4	96	6,7	9,5	5,3	11,9
42	Baltis(Molėtų)	LTL342	3	4	83	5,5	10,4	4,3	10,5
43	Baltosios Ančios HE tvenkinys	LTL189	2	8	84	13,6	24	10,2	21,8
44	Baltys(Ign.)	LTL323	2		103			4,7	10,1
45	Baluošai	LTL358	3	8	93	3,6	5	4,8	10,1
46	Baluošas	LTL211	3	4	91	2,5	5,2	6,4	16,3
47	Bartkuškio tvenkinys	LTL128	1	8	63	17	66,9	23,2	38,5
48	Bebrusai	LTL345	2	4	90	6,1	7,6	8,1	17

49	Beičių tvenkinys	LTL524	1	4	23	16,9	57,4	24,4	38,6
50	Beržoras	LTL469	1	2	66	5,4	7	10,9	34,6
51	Bijotė	LTL259	1		87			6,6	17,1
52	Biržulis	LTL208	1		68			32,3	50,9
53	Bubių tvenkinys	LTL503	1	3	99	12,9	26,3	10,2	18,6
54	Bublių tvenkinys	LTL414	2	4	93	13,9	23,9	12,8	46,3
55	Čičirys	LTL222	3	<sup>3</sup> / <sub>5</sub>	88	3	6,6	5,3	13,3
56	Didysai Siaurys	LTL217	2	4	80	4,7	7,1	8,5	19,4
57	Didžiulis(Alytaus )	LTL197	3	8	101	5,3	11,6	6,4	13
58	Didžiulis(Dusmena)	LTL274	1	8	101	35,8	67,3	47,2	77,2
59	Didžiulis(Vilniaus raj.)	LTL134	2	4	105	27,6	44,5	36,9	85,3
60	Drabužis	LTL116	2	7	90	8,7	23,8	5,5	12,8
61	Draudenių ežeras	LTL205	1	8	96	6,1	18,9	7	17
62	Dringis	LTL229	2	4	87	3,3	4,4	6,4	14,5
63	Drūkšiai	LTL52	3	4	105	3,5	4,6	6,4	20,4
64	Dūburaitis	LTL528	2	8	83	11,4	26,8	7,6	21,3
65	Dūburys	LTL300	2	4	89	3,5	6,7	5,8	12,5
66	Dūkštas	LTL313	1	4	95	3,4	4,4	5,4	13,6
67	Dūriai	LTL159	2	8	86	10	14,6	13,1	21,3
68	Dusia	LTL43	3	4	104	4,2	4,5	4,8	9,4
69	Dusynas	LTL289	1	8	86	8,8	20	11,4	23,3
70	Dvariukų tvenkinys	LTL403	1	8	44	6,6	12,4	16,7	39,3
71	Dviragis	LTL148	1	4	101	34,6	48	18,2	47,2
72	Dysnai	LTL223	1	4	50	11,9	16,9	17,2	44,9
73	Dysnykštis	LTL156	1		51			14,1	27,2
74	Eišiškių HE tvenkinys	LTL426	1	8	75	28,9	40,7	22,4	70,8
75	Elektrėnų marios	LTL125	2	8	85	5,5	18,5	5,4	13
76	Erzvėtas	LTL387	2	7	88	7,8	11,4	9,7	20,1
77	Gačionių ežeras	LTL294	2	4	86	11,9	14,7	9,5	25,1
78	Gailintas	LTL251	3	4	88	4,2	6,2	6,7	12,4
79	Galadusys	LTL187	2	4	90	3,7	4,7	5,2	19,9
80	Galstas	LTL390	3		100			5,2	11,9
81	Galunai	LTL224	2	4	96	7	10,7	6,3	16,6

82	Galuonis	LTL341	2	8	91	5,7	10	9,9	24,3
83	Galvė	LTL168	3	4	91	3,3	4,5	4,7	8,7
84	Gauštvinis	LTL518	1	4	100	43,2	62,6	40	70,3
85	Gaveikių ežeras	LTL525	1	4	92	18,1	26	11,8	33,8
86	Gaviekas	LTL534	2	8	75	8,7	22,3	11,4	21,5
87	Gavys	LTL443	3	4	99	8,4	11,7	6	15,3
88	Gėlių ežeras	LTL283	1	8	90	39,7	63,3	25,1	71,1
89	Gelvanės ežeras	LTL520	1	8	84	27,3	37,2	20,5	45,6
90	Germantas	LTL256	1	1 0	64	3,1	5,9	6,4	19,8
91	Giedavardys	LTL392	1	8	96	6,6	10,4	7,9	15,1
92	Giluitis	LTL500	2		111			8	16,5
93	Ginkūnų tvenkinys	LTL272	1	8	67	8,4	23,2	8,1	15,7
94	Girdžių tvenkinys	LTL422	1		72			18,9	59,1
95	Girutiškis	LTL521	1		81			14,6	37,5
96	Glėbas	LTL377	1	4	91	3,7	6,1	5,7	15,2
97	Gludas	LTL501	1	1	94	3,4	3,4	8,3	19,3
98	Glūkas	LTL234	3	8	95	13	63,3	5,5	11,8
99	Gondingos HE tvenkinys	LTL401	2	1 2	86	10,9	35,6	8,4	21,7
100	Grabuostas	LTL360	2	4	92	20,7	28,9	15	52,5
101	Grūda	LTL396	1	1 2	82	54,3	88,2	40,9	82,6
102	Grūtas	LTL514	1	1	102	10,9	10,9	42,9	73,8
103	Gudelių ežeras	LTL264	2		107			7,2	25,4
104	Gudelių ežeras(Alytaus)	LTL253	1	7	88	14,6	65,7	29,1	78,6
105	Guostus	LTL454	3	4	92	2,5	3,3	5,6	13
106	Ilgai	LTL368	3	4	87	3	4,3	4,5	7,7
107	Ilgajis	LTL279	1	8	98	41,4	54,5	25,4	68,2
108	Ilgas	LTL359	2	8	83	15,3	39,4	11,3	33,3
109	Ilgės	LTL135	1	8	105	17,2	31	17,8	77,3
110	Ilgiai	LTL397	2	8	93	9,7	16,6	9,5	24,7
111	Ilgis(Alytaus)	LTL374	1	1 2	90	21,6	52,7	20	69,3
112	Ilgis(Lazdijų)	LTL394	2	4	101	8,2	17,7	6,2	13
113	Ilgis(Molėtų)	LTL356	1	4	83	14,6	32,7	8,6	17,4
114	Ilgis(Plungės)	LTL204	1	3 6	73	5,7	25,8	7,4	27,4

115	Ilgis(Varėnos)	LTL273	2	8	104	8,6	33,9	5,8	15,8
116	Ilgis(Zarasų)	LTL306	2	4	97	6,7	10,1	9,3	35,9
117	Ilgutis(Vilniaus)	LTL380	1	4	95	2,5	3,8	5,7	10
118	Ilgys(Daugailių sen.)	LTL316	2	4	74	2,7	4,1	9	22,3
119	Ilgys(Zarasų raj.)	LTL305	2		87			7,1	22,6
120	Ilmėdas	LTL351	2	8	89	6,1	15,4	4,8	7,7
121	Imbradas	LTL441	1	4	96	7,4	15,1	11,7	24
122	Indrajai	LTL232	2	8	89	4,4	7,3	7,3	12,7
123	Išnarai	LTL344	1	4	92	5,6	6,4	11,9	21,7
124	Janušonių tvenkinys	LTL412	1	4	62	26,7	55,7	18,6	58,6
125	Jiezno ežeras	LTL463	1	1 4	108	55,3	89,3	39,7	81,4
126	Juodas Kauknoris	LTL174	2	8	77	15	35,1	11,5	18,3
127	Juodeikių tvenkinys	LTL268	2		64			8,6	16
128	Juodieji Lakajai	LTL347	3	4	89	5,4	10,4	5,3	10
129	Juodis(Panevėžio)	LTL541	1	4	65	29	70,9	20,7	58,4
130	Juodkiškių tvenkinys	LTL161	2	4	81	6,2	9,5	10,9	23,3
131	Jurbarkų tvenkinys	LTL423	1	4	92	25,9	40,5	20	58,2
132	Jurgonių tvenkinys	LTL428	1	8	88	21,2	34,2	13,2	45,9
133	Kadrėnų tvenkinys	LTL162	1	8	79	14,3	31,4	14,9	40
134	Kairių ežeras	LTL265	1	8	97	13,1	23,4	8,7	17,6
135	Kalvių ežeras	LTL141	1	4	95	81,3	251	27,4	70,8
136	Kampuolis	LTL526	2	8	76	19,7	31,7	13,6	46,1
137	Kančioginas	LTL386	2	1 2	87	7,8	15,8	9,5	22,8
138	Karališkių ežeras	LTL282	2	4	88	25,3	39,3	14,1	36,5
139	Karvys(Vilniaus)	LTL363	2	4	86	8,2	13,3	12,2	34,5
140	Kauno marios	LTL71	2	4 2	113	30,7	161	29	72,1
141	Kavalys	LTL465	1	8	104	26,9	55,9	16,7	62,2
142	Kavarsko tvenkinys	LTL416	1	4	46	5,8	17,6	20,8	42,2
143	Kaviškis	LTL395	2	8	81	12	21,3	10,7	19,1
144	Kemėšys	LTL214	1	4	67	18,6	24,8	26,2	66,6
145	Kernavas	LTL507	1		83			35,6	60,8



146	Kernų tvenkinys	LTL261	1	8	67	19,9	32,1	16,4	67,8
147	Kertuojai	LTL227	1	4	88	7,9	10,4	9,8	19,9
148	Kiaunas	LTL334	2		81			8,4	20,5
149	Kiementas	LTL450	1	4	91	23,4	52,6	12,6	41,6
150	Kilučių ežeras	LTL438	1	3	54	14,2	17,5	21,3	44,2
151	Kirneilis	LTL349	2	4	88	6,1	9,5	6	11,1
152	Kivylių tvenkinys	LTL270	1	8	45	18,2	33,5	24,9	36,9
153	Klykių ežeras	LTL539	3	4	86	3,1	8	4,6	16,5
154	Kretuonas	LTL219	1	4	87	8,5	16,8	7,9	17,7
155	Kretuonykštis	LTL339	1		58			13,7	26,4
156	Krivėnų tvenkinys	LTL413	2	4	60	12,7	17,8	15,1	44
157	Krokialaukio tvenkinys	LTL188	1	8	38	19	38,7	27,4	59,8
158	Krokų Lanka	LTL277	1		94			28,9	87,7
159	Krūminių tvenkinys	LTL184	1	1 2	74	26,5	41,6	17,3	35,1
160	Kumpuolis(Anta zavės sen.)	LTL530	1	4	89	25,2	74,6	6,3	16,4
161	Kumpuolis(Deg učių sen.)	LTL301	2	8	85	8,1	22,1	17	40,2
162	Kupiškio tvenkinys	LTL417	2	3	99	24,2	30,9	19,9	63,1
163	Kurėnų ežeras	LTL281	1	8	101	29,5	55	37	61,8
164	Labanoras	LTL337	1	8	86	6,1	10,4	6,6	15,2
165	Labūnavos tvenkinys	LTL411	1	4	92	12	28,7	12,8	67,3
166	Laičių I tvenkinys	LTL544	1	4	59	34,6	93,8	8,9	17,8
167	Lakajas	LTL353	1	4	81	3,8	5	7,9	22,8
168	Lamėstas	LTL444	1	4	82	17,7	24,7	14,4	35,3
169	Lampėdis	LTL510	1		94			7,3	20,7
170	Latežeris	LTL513	1	8	93	65	124,3	46,8	89,3
171	Laukesas	LTL527	1	8	96	10	27	13,3	62,8
172	Lavysas	LTL276	1	4	85	3,3	5,1	4,9	10,1
173	Lazdininkų tvenkinys	LTL263	1	8	68	4,2	7,3	9,4	18,1
174	Lazdinių ežeras	LTL389	2	8	77	9	22,5	15,1	30,7
175	Lėnas	LTL278	1	4	99	13,9	23,1	17,4	33,4
176	Liberiškio tvenkinys	LTL405	1	3	65	12	27,5	18,6	55,1
177	Liedis	LTL335	1	4	86	14,6	21,8	13,2	33,6

178	Lielukas	LTL455	1	8	97	23,8	49,3	20,2	62,2
179	Ligajai	LTL315	2	4	77	3,4	4,4	8,3	16,4
180	Linkmenas	LTL330	1	8	89	5,8	9,6	6	13,8
181	Liškiavis	LTL194	2	4	99	7,2	17,2	6,2	12,5
182	Luka	LTL468	2	4	87	8,9	11,8	5,2	10,4
183	Luknas	LTL290	1	4	80	11,2	11,7	12,5	33,9
184	Luksnēņu ežeras	LTL466	1	8	93	20,5	45,5	13,9	69,8
185	Lukštas	LTL302	1	4	87	8,4	13,7	14,5	33,5
186	Lūkstas	LTL48	1	8	92	17,6	32,1	18,9	48,1
187	Luodis	LTL312	2	10	93	3	5,5	5,8	16,4
188	Luokesai	LTL158	3	4	83	3,3	5,2	5	20,8
189	Lūšiai	LTL327	3	10	92	2,9	5,4	6	15,4
190	Makys	LTL284	2	4	81	8,4	21,1	5,9	21,5
191	Malkēstas	LTL350	2	4	80	5,4	11,6	8,5	22,2
192	Mantviliškio tvenkinys	LTL409	1	4	34	32,2	43,7	24,1	56,3
193	Margis	LTL367	2	4	87	5,2	7,6	6,9	16,4
194	Marijampolės II tvenkinys	LTL198	1	4	50	22,4	37,3	21,3	40,8
195	Mastis	LTL435	1	8	72	32,7	86,8	40,4	80,2
196	Mažasai Siaurys	LTL352	2	4	88	5,6	8,3	6,1	19,5
197	Metelys	LTL236	2	8	102	4,2	9,6	5,5	23,7
198	Monis	LTL370	3	4	74	9,4	18,4	4,7	9,8
199	Mosėdžio I tvenkinys	LTL262	1	8	59	10,6	23,7	9,1	17,3
200	Motiejūnų HE tvenkinys	LTL131	1	4	76	7,5	10,9	18,9	37,4
201	Mūšėjus	LTL288	2	4	70	21,9	27,2	18,8	38,3
202	Musia	LTL508	1	8	82	46	73,4	27,6	65
203	Našys	LTL440	1	4	80	5	6,1	11,9	39,9
204	Nečiūnų ežeras	LTL372	3	4	92	8,6	9,2	10,1	26,9
205	Nedzingis	LTL238	1	4	88	12,2	18,7	11,8	25
206	Nemeikščių tvenkinys	LTL432	2	4	92	8,6	12,8	12,7	22,9
207	Nestrėvantys	LTL535	1	8	92	12,9	22,7	11,5	26,2
208	Netečius	LTL464	1	8	87	22	52,2	27,4	41,1
209	Neveiglas	LTL457	2	8	85	26,9	48,2	19,3	67,3
210	Nevėža	LTL285	1	4	95	10,3	19,9	16,3	71,2

211	Niedulis	LTL456	1	8	88	37,1	125,5	30,1	72,6
212	Niedus	LTL240	2	8	71	13,1	29,8	8,4	14,8
213	Nikajis	LTL355	1	8	81	9,2	22,5	9,1	20,4
214	Notigalė	LTL147	1	4	49	32	87	13,8	27,8
215	Obelija	LTL247	1	8	111	7,1	9,7	6,5	13,9
216	Obelių ežeras	LTL439	1	8	95	88,7	222	48,3	86,7
217	Orija	LTL202	1	4	57	21,8	37,2	11,3	33,2
218	Pabezninkų ežeras	LTL375	1	8	90	27,4	65,7	18	34,4
219	Padvarių tvenkinys	LTL257	1	1/2	76	20	45,9	18,7	51,5
220	Padysnio HE tvenkinys	LTL164	1		44			12,8	22,6
221	Paežerių ežeras	LTL246	2	4	97	7	7,9	15,6	66
222	Paežerių ežeras(Vilk.)	LTL461	1	1/2	66	26,6	52	19,9	47,4
223	Pajiesio tvenkinys	LTL425	1	4	99	23,4	34,6	31,1	76,5
224	Pakalas	LTL333	2		73			4,8	9,8
225	Pakasas	LTL321	2	4	90	8,6	13	6,4	14,5
226	Papilio tvenkinys	LTL163	1		43			16,9	39,3
227	Papis	LTL459	1		88			46,1	74,8
228	Paršežeris	LTL436	1	8	102	33,1	89,3	26,5	58,6
229	Parsvėtas	LTL532	1		62			5,2	13,7
230	Paštys	LTL293	1	8	90	9,1	23,5	11,7	31
231	Paupio tvenkinys	LTL427	1		109			46,7	73,9
232	Peršokšnai	LTL340	3	8	84	3,1	7,6	6,8	15,3
233	Petraičių	LTL545	1	8	94	71,4	172,4	58,6	89,2
234	Petrošiškio ežeras	LTL291	1		97			10,5	26,9
235	Pienionių tvenkinys	LTL430	1	8	85	37	91,5	26,5	69,5
236	Pikeliškių ežeras(ež.Žalesas)	LTL364	2	8	83	17,2	28,6	17,6	36,1
237	Pilvės-Vabalkšnės tvenkinys	LTL543	1	4	24	26,1	54	32	42,3
238	Platelių ežeras	LTL34	3	4/1	79	3,2	9,1	5,6	18,4
239	Plinkšių ežeras	LTL434	2		69			11,8	25,1
240	Pluvija	LTL512	1	8	97	37,8	81,9	36,3	85,6

241	Prapuntas	LTL391	2	4	100	6,4	15,8	6,9	21,2
242	Pravalas(Vidugiris)	LTL354	1	8	84	50,6	148,7	55,7	92,7
243	Praviršulis	LTL502	1		82			20	37,7
244	Prūtas	LTL506	1	4	97	4	5,9	5,5	13,1
245	Rašai	LTL462	1	4	88	7,5	10,3	10,9	26,8
246	Rašia	LTL218	2	4	85	5,7	12,5	5,1	10,1
247	Rėkyvos ežeras	LTL64	1	8	123	64,3	148,1	32,3	60,5
248	Riešė	LTL165	1	1/2	93	34,9	59,9	25,4	65,4
249	Rimietis	LTL254	1	8	113	93,2	164	47,3	90,8
250	Rubikių ežeras	LTL10	2	4	98	12	20,2	8,9	21,7
251	Rūžas	LTL399	1	4	100	4	5,6	7,8	16,1
252	Sablausių tvenkinys	LTL271	1	8	70	19,2	50,3	21,2	59,3
253	Sągardas	LTL317	2	8	80	6,7	15,7	5,8	14,6
254	Sagavas	LTL249	2	8	104	7,8	14,2	9,2	18,1
255	Šakarvai	LTL328	3	8	86	4,6	10,9	5,7	12,1
256	Šakių ežeras	LTL531	1	8	72	10	24,7	10,6	23,7
257	Salotas	LTL504	1	1	91	6,2	6,2	41,6	71,2
258	Samavas	LTL309	2	4	89	5,6	13,5	6,4	15,8
259	Samis	LTL109	1	4	98	5,9	15,2	19,4	61,9
260	Sarių ežeras	LTL381	2	4	53	3,3	6,6	7,6	35,7
261	Sartai	LTL169	2	8	91	10,7	27,4	9,3	24,4
262	Sausvingis	LTL378	2	4	81	7,9	10,6	7,4	17,2
263	Savistas	LTL250	2	1/2	95	15,7	33,7	10,3	23,5
264	Seirijis	LTL193	2	4	98	3,3	4,6	4,9	11,8
265	Šiometis	LTL155	2	4	84	4,7	7,3	4,7	11,2
266	Siesartis	LTL348	3	4	101	6,1	9	5,8	14,8
267	Siesikų ež.	LTL280	1	8	93	18	30,8	13,3	25,4
268	Simno ežeras	LTL196	1	8	117	48	112,7	42,7	90,4
269	Širvėnos ežeras	LTL437	1	4	66	8,1	13,8	18,9	47,4
270	Širvintų HE tvenkinys	LTL540	1	4	56	23,1	37,9	22,9	34,6
271	Širvio ežeras	LTL451	1	2/1	92	40,7	97,6	33,8	73,3
272	Šiurpys	LTL311	2	3/6	102	2,3	5,9	4,9	9,4
273	Skaistė	LTL385	2	4	42	10	12,4	8,5	25,4
274	Skaistis	LTL136	3	4	97	2,5	4,5	4,8	13,9

275	Skuodo tvenkinys	LTL260	1	4	66	12,2	19	15,4	80,7
276	Šlavantas	LTL177	2	4	119	2,5	3,2	5,3	10,2
277	Šlavantėlis	LTL515	2		109			5,3	12,2
278	Smalvas	LTL220	2	4	48	2,2	3,8	5,2	11,7
279	Smalvykštis	LTL308	1	4	85	6,7	15,1	8,4	22,5
280	Snaignas	LTL252	2	8	94	3,2	4,6	4,8	9,6
281	Spenglas	LTL173	1	7	96	51,2	81,7	47	76,3
282	Spėra	LTL114	1	8	79	51,6	76,7	39,6	68,4
283	Spindžius	LTL106	2	$\frac{1}{3}$	112	3,8	7,4	4,9	8,6
284	Stavarygalos ežeras	LTL362	1	4	86	21,6	26,5	20,9	44,8
285	Stebuliškių tvenkinys	LTL542	1	4	53	38,6	55,7	30,8	63,2
286	Stepanionių tvenkinys	LTL407	1		81			19,7	67,4
287	Stervas	LTL203	1		84			31,5	73,9
288	Stirniai	LTL343	3	$\frac{4}{2}$	84	2,9	9	5,2	10,6
289	Stirtos	LTL517	2	8	53	6,5	9,5	7,7	15,3
290	Sujainių tvenkinys	LTL420	1	4	91	56,9	115,6	40,4	72,5
291	Suosa	LTL382	1	4	76	13,6	18,1	16	30
292	Suvieko ežeras	LTL304	1	4	98	6,4	8,1	6,6	15,1
293	Suvingis	LTL376	2	8	89	6,8	14,8	6,4	12,4
294	Svėdasas	LTL170	2		97			11	28
295	Švenčius	LTL144	1	4	93	62,9	100,4	45,1	82,4
296	Šventas(Švenčio nėlių sen.)	LTL446	1	8	88	33,8	58,7	23,4	53,4
297	Šventas(Švenčio nių sen.)	LTL522	1	8	45	16,6	48,5	13,5	20,4
298	Šventas(Zarasų)	LTL39	2	$\frac{4}{2}$	87	2,3	6,2	4,2	7,4
299	Šventosios tvenkinys	LTL536	1	4	88	6,9	11,1	10,1	14,7
300	Svirkų ežeras	LTL388	1	8	88	59,1	95,9	40,8	77,5
301	Talkša	LTL267	1	8	75	17,8	24	17,7	61
302	Tauragnas	LTL442	3	4	90	1,8	2,7	6	19,8
303	Tausalas	LTL206	1	8	75	12,5	35,5	14,5	29,4
304	Totoriškių ežeras	LTL467	2	4	84	3,9	4,5	6,1	14,7
305	Totorviečių tvenkinys	LTL424	1	4	65	20,9	22,8	12,4	67,3

306	Tūbasių I tvenkinys	LTL207	1	1 2	77	15,2	36,7	18,2	46,3
307	Ubiškės tvenkinys	LTL400	1	8	49	14,2	21,2	17,3	35
308	Ūdrijos ežeras	LTL511	1	8	101	109	167,7	69	91,1
309	Ūkojas	LTL221	3	4	83	5,5	10,4	6,5	15,3
310	Uolys	LTL533	2	4	78	3,6	6,1	5,7	16,4
311	Urkis	LTL346	1		92			34,9	77,7
312	Ūsiai	LTL326	2	4	79	3,9	5,7	5,9	14
313	Utenas	LTL319	2	4	83	3,5	5,1	7,9	19,1
314	Utenos tvenkinys	LTL431	2	4	84	6,5	12,2	9	31,1
315	Utenykštis	LTL320	1	3	96	17,8	32,6	9,3	21,6
316	Vabalių ežeras	LTL379	2	1 2	91	7,5	13,9	7,5	16,4
317	Vaisietis	LTL371	2	4	94	6,8	10,9	6,5	22,5
318	Vaisinis	LTL307	1		96			15,2	39,7
319	Vaitiekūnų tvenkinys	LTL408	1	8	75	10,9	49,4	13,7	29,5
320	Vajuonis	LTL325	2	4	95	5,3	5,9	5,7	11,4
321	Varėnos m. I tvenkinys	LTL538	1	4	87	15,7	27,5	9,7	20
322	Vasaknas	LTL297	1	8	92	8,3	15,7	7	16
323	Vastapas	LTL286	1	4	90	15,9	25,4	13	36,7
324	Vazajis	LTL292	1	4	86	13,3	22,9	10,3	21,5
325	Veisiejis	LTL237	3	8	114	17	27,2	10,5	22,2
326	Vencavas	LTL226	3	4	103	1,8	2,6	4,7	11,9
327	Veprys	LTL298	1	4	100	15,4	22,6	9,4	22,1
328	Verniejus	LTL117	3	4	93	6	7,9	5,6	11,8
329	Vernijis	LTL393	2	4	105	9,2	14,5	9,1	18,8
330	Vidinkstas	LTL287	1	8	91	13	29,6	8,8	20,7
331	Viešnių ežeras	LTL505	1	1	93	5,7	5,7	13,1	43,6
332	Viešintas	LTL383	1		101			12,1	20,6
333	Vievis	LTL137	3	4	88	5,4	6,8	5,1	10,1
334	Vilkas	LTL180	1	4	99	9,4	10,7	10,3	21,7
335	Vilkinys	LTL239	2	8	109	6,4	9,6	7,8	17
336	Vilkokšnis	LTL373	2	8	96	13,3	44,9	7,4	15,2
337	Vilnoja	LTL365	3	4	83	3,1	4	5,2	10,1
338	Virintai	LTL447	3	4	90	11	18,7	8	15,4
339	Visaginas	LTL314	1	4	43	12,3	21,8	12,8	22,3

340	Vištytis	LTL68	3	4	98	3,2	4	5,2	11,6
341	Volungiškių tvenkinys	LTL421	1		89			17,5	39,7
342	Voverių tvenkinys	LTL519	1		31			40,7	68,3
343	Zaduojys	LTL295	2	4	92	9,6	19,8	7,3	24,6
344	Žaltytis	LTL245	1	4	109	3,5	5,6	30	74,5
345	Zalvas	LTL523	2	4	88	3,3	4,4	4,3	6,2
346	Zapsys	LTL241	2	8	118	11,3	18,6	10,5	17,8
347	Zarasas	LTL171	3	4	91	7,8	9	9,7	23,9
348	Žaslių ežeras	LTL453	1	4	91	10,3	14	11,1	19,5
349	Žeimenys	LTL445	2	4	77	7,2	8,8	5,7	28,7
350	Želvos	LTL448	2	8	91	17,7	39,3	11,4	21,3
351	Žiegas	LTL296	2	4	88	4,9	7,9	5,5	16
352	Žiezdras	LTL332	3		80			7,9	17,8
353	Žiezdrelis	LTL529	2		80			8,1	24,7
354	Žilmas	LTL398	2	4	88	9,1	26	6,2	16,9
355	Žirnajai	LTL154	2	4	93	4,4	6,9	7,2	16,4
356	Žuvintas	LTL29	1	4	91	6,9	8,7	18	38,3
357	Žvernas	LTL357	1	4	86	6,9	9,5	8,7	15,2

## SUMMARY

### 1. INTRODUCTION

Water from lakes and ponds is used for recreation, industry, drinking water. Lakes and ponds are an important part of the global carbon cycle and therefore a component of the climate system (Tranvik et al., 2009). Lakes and ponds are habitats for most living organisms (Schindler & Scheuerell, 2002). Water quality is important for any water use, and due to human activities, when water bodies are altered or when pollutants, nutrients, the main of which are nitrogen and phosphorus, enter water bodies, eutrophication of water bodies accelerates, water quality deteriorates, and water organisms have to adapt to changed conditions. Water ecosystems experience increasing challenges due to climate change, water scarcity and quality issues. Climate change cause more suitable conditions for algal and cyanobacterial blooms. Algal blooms can disrupt water use due to their negative effects to water quality. Algal blooms are an under-monitored phenomenon that can have a significant impact on both ecosystems and humans.

There are about 6000 lakes and ponds in Lithuania, which constitute 1.4% of the country's area (Kilkus & Stonevičius, 2012). Water bodies are constantly monitored to ensure the safe use of water and the good status of water systems in accordance with the Water Framework Directive of the European Union (European Parliament and The Council, 2014). Information collected by traditional methods allows accurate assessment of the water quality parameters of water bodies and bathing sites but requires a lot of time and resources. However, in Lithuania, about 40% of water bodies in the category of lakes do not meet the criteria of good status, and due to high diffuse pollution, the entire territory of Lithuania is declared a sensitive area according to the Nitrates Directive (91/676/EEC) (Environmental Protection Agency, 2018).

Optical Earth observation satellites overcome some of the shortcomings of traditional methods. With satellites, it is possible to monitor water bodies much more often, and the information received covers the entire water body and all remote, hard-to-reach water bodies. With satellites we can observe optically active substances, such as, chlorophyll  $\alpha$  (chl $a$ ), suspended matter (SM) and coloured dissolved organic matter (CDOM). The inclusion of satellite data would improve the reliability of the ecological status assessment due to the greater representation of the diversity of water bodies and the detection of changes in water quality in them (Papathanasopoulou et al., 2019). With the use of satellite data, the phenomenon of algal blooms could



be monitored better – for determination of its onset, assessment of intensity, spatial distribution and to inform people about the current situation. However, the use of satellite data for assessing the state of waters and transmitting information about them in accordance with the European Union Water Framework Directive is still limited, since satellite data as a data source is not included in the monitoring policy system and there is a lack of a unified standardized methodology for data use (Papathanasopoulou et al., 2019).

Determining water parameters from satellite data is complicated due to water properties themselves and the mixture of optically active substances. Pure water absorbs most of the radiation, so the signal captured by the satellite is weak (McClain, 2009). Also, the signal is strongly influenced by the atmosphere. Removal of atmospheric effects is one of the main challenges of remote sensing, hindering the qualitative determination of biogeochemical parameters in water (Pahlevan et al., 2021). Thus, when using optical satellite data, it is important to be aware of the uncertainties arising from the inherent properties of satellite data and the choice of processing methodology. Due to the water properties and atmospheric effects, the choice of algorithm can be very significant for the results obtained. In addition, to obtain biophysical indicators, the algorithms developed in other territories may not be suitable for the studied water bodies due to different characteristics of the basin and a water body, thus it is usually necessary to adapt the algorithms to the studied water bodies.

The Sentinel-2 satellite A of the Copernicus program has been in orbit since 2015, and satellite B since 2017. Due to Sentinel-2 sensor suitability for monitoring water bodies, the continuity of the mission envisaged in the program (ESA, 2021) and the characteristics of regional water bodies, there is a need to create algorithms for the estimation of water quality parameters based on the data of these sensors. Recently, machine learning algorithms have been applied to process large amounts of satellite data and extract useful information from them. In this dissertation, models based on a machine learning random forest algorithm were developed to determine the concentration of chlorophyll  $\alpha$ .

**The object of the work** is a methodology for determining lake water quality parameters, based on satellite data.

**The aim of the work** is the development of a methodology for determining the concentration of chlorophyll  $\alpha$  from satellite data.

**The objectives:**

1. To evaluate the uncertainty of atmospheric correction algorithms for the determination of chlorophyll  $\alpha$  concentration in Lithuanian lakes.

2. Create a classification of optical types of lakes based on satellite data and apply it to the studied lakes.
3. Develop and validate chlorophyll  $\alpha$  concentration assessment algorithms adapted to different biophysical classes of lakes.
4. Apply the developed algorithms to the studied lakes and determine potentially problematic lakes based on the concentration of chlorophyll  $\alpha$ .
5. Analyse chlorophyll  $\alpha$  concentration changes in lakes.

#### **Defending statements:**

1. Determination of chlorophyll  $\alpha$  concentration from Sentinel-2 MSI data is sensitive to selection of atmospheric correction product and chlorophyll  $\alpha$  algorithm.
2. The model of biophysical lake classification increases the accuracy of chlorophyll  $\alpha$  concentration models.
3. Using satellite data and chlorophyll  $\alpha$  concentration models based on random forest regression algorithm, it is possible to determine the spatial distribution of chlorophyll  $\alpha$  concentration in lakes and its short-term changes in lakes associated with algal blooms.
4. The developed algorithms allow identification of potentially problematic lakes using average and maximum values of chlorophyll  $\alpha$  concentration.

#### **Scientific novelty**

This dissertation aims to fill the lack of observations of water bodies in Lithuania based on satellite data. The presented results contribute to the greater use of data by proposing models of water parameters describing the ecological state of the lake. Scientific value of the work:

1. The uncertainty of atmospheric correction algorithms for obtaining chlorophyll  $\alpha$  concentration from Sentinel-2 MSI data was evaluated.
2. A model of the biophysical class of a lake was created based on the derived spectral parameters and the random forest (RFC) algorithm.
3. Models based on random forest regression algorithm were developed for the biophysical classes of lakes to obtain the concentration of chlorophyll  $\alpha$ .
4. Potentially problematic water bodies were identified according to the hydrobiological criterion – the concentration of chlorophyll  $\alpha$ . Short-term changes in the concentration of chlorophyll  $\alpha$  associated with algal blooms were determined.

The developed methodology for evaluating the concentration of chlorophyll  $\alpha$ , based on machine learning algorithms, provides an opportunity to efficiently collect information about a large number of lakes without going

to them. This work contributes to lake water quality research using remote sensing methods in mid-latitude climate conditions.

### **Relevance and applicability of the work**

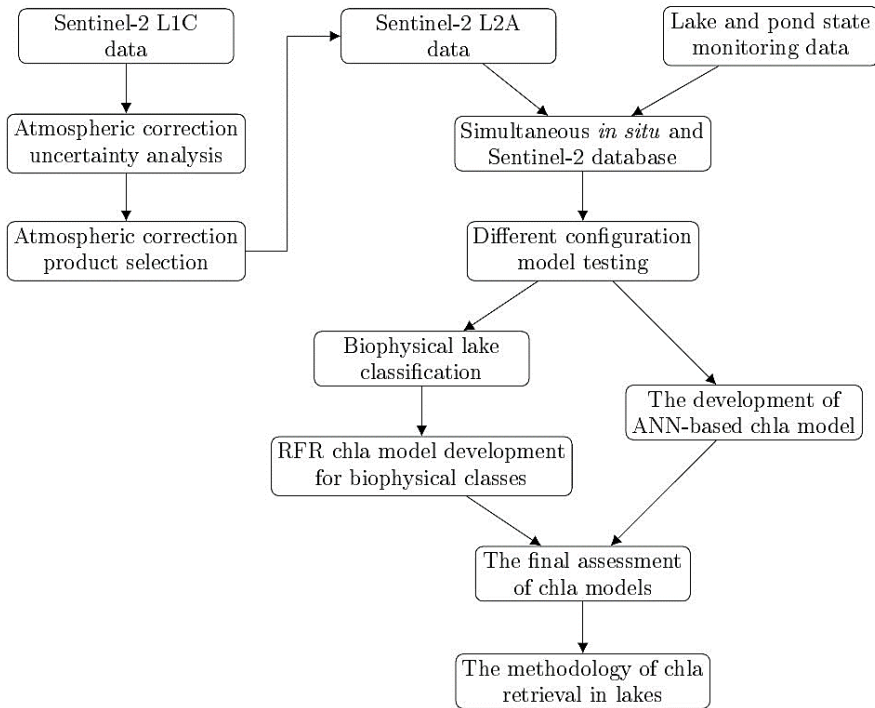
With traditional methods, observations are carried out relatively rarely and do not allow the assessment of short-term changes of water quality in lakes. The amount of information received by satellites is increasing every day, and machine learning methods are increasingly being used to process them. The machine learning methods proposed in this paper can process large amounts of data and quickly obtain information on water parameters covering the entire water body and covering many water bodies.

Applying the models developed in this work since the year of 2015, time series can be supplemented with more frequent observations. More information is collected about the duration of algal blooms, and trends of their change can be determined. The created models can be used in the work of experts to assess the status of a lake in near-real time and also in long-term assessments of status of water bodies. Satellite observations can help optimize the monitoring program – select the time, frequency of monitoring, or monitoring site location of considering the spatial distribution of chlorophyll  $\alpha$  concentration.

## **2. DATA AND METHODS**

To estimate the concentration of chl $a$  in lakes, several steps are performed to evaluate the uncertainty of data, develop models, verify their performance and carry out temporal and spatial analysis of chl $a$  concentration (Figure 2.1).

This study examined 357 water bodies located in Lithuania and included in the state monitoring program. These lakes and ponds belong to the Baltic Sea ecoregion, and their absolute height does not reach 200 m above sea level. According to Koppen's climate classification, Lithuania is in the zone of humid continental climate with warm summers (Dfb). Diffuse pollution from agricultural fields is a major source of water pollution (Environmental Protection Agency, 2018). The lakes range in size from 0.35 to 46.8 km<sup>2</sup> (mean 2.24 km<sup>2</sup>, standard deviation 4.1 km<sup>2</sup>). Lakes and ponds are divided into three types according to the average and maximum depth (The Ministry of Environment, 2018). There are 190 shallow lakes (medium depth < 3 m or medium depth > 3 m and maximum depth < 11 m), 121 medium deep (medium depth > 3 m, maximum depth 11–30 m), and 46 deep (maximum depth > 30 m) lakes in the monitoring system.



**Figure 2.1.** The scheme representing steps carried out in this study.

Open access medium spatial resolution MultiSpectral Instrument (MSI) optical data of the Sentinel-2 mission satellites of the European Space Agency (ESA) were selected for chla assessment. The MSI data are of sufficient spatial resolution for monitoring relatively small Lithuanian lakes (Grendaitė et al., 2018). The MSI has a relatively high signal-to-noise ratio, so using its data it is possible to capture small changes in water parameters. The Sentinel-2 MSI monitors the Earth in 13 spectral bands with spatial resolution of 10, 20 and 60 m, which cover the part of the electromagnetic spectrum from visible waves (from 440 nm) to short infrared radiation (2190 nm). The radiometric resolution of the sensor is 12 bits. The data contain the dimensionless reflectance value. Values range from 0 to 1, where 0 is no reflection from the surface and 1 is the maximum amount of energy reflected.

Atmospheric correction is a satellite data processing step that is necessary to obtain water leaving reflectance ( $\rho_{\text{row}}$ ), which is used to estimate the concentration of chla. There is no agreement on one AC algorithm which would be the most suitable for inland waters. This study evaluated the uncertainty of spectra obtained with seven atmospheric correction products (Acolite, Acolite Rayleigh, C2RCC, C2X, iCOR, POLYMER, and Sen2Cor) and 10 radiometric chla algorithms (CHLA1-CHLA10) for estimation of chla

concentration. The results allow ranking the utility of atmospheric correction algorithm together with chl<sub>a</sub> algorithms, taking into account the calculated smallest deviations from the measured chl<sub>a</sub> concentration.

The uncertainty of atmospheric correction algorithms was estimated using seven images covering 136 lakes of eastern Lithuania of the years 2018 and 2019 (tile T35UMB). The data were filtered based on the IDEPIX image classification, keeping only water pixels. The assessment of the uncertainty of the atmospheric correction examines the reflections in the 10 m resolution visible and near infrared spectral bands 490–783 nm (B2–B7). The resolution of the satellite data was resampled to 20 m to avoid resampling the 705 nm (B5) spectral band, since this spectral band is used in all chl<sub>a</sub> algorithms.

Chl<sub>a</sub> data from an *in situ* state monitoring program were used for satellite data validation. Usually, in one water body four observations are carried out during the warm season (one in spring, two in summer, and one in autumn). Environmental Protection Agency (EPA) performs about 300 measurements per year, but only a part of them can be used for satellite data validation due to cloudiness and separation in time.

The uncertainty of the atmospheric correction algorithms was highlighted using different radiometric chl<sub>a</sub> algorithms, and after calculating model performance metrics, the best combination of atmospheric correction and chl<sub>a</sub> algorithm was determined. The deviations between the measured chl<sub>a</sub> values and those calculated by the chl<sub>a</sub> algorithm are quite different in different intervals of chl<sub>a</sub> values, which implies that chl<sub>a</sub> algorithms developed for groups of lakes that have similar optical properties (in this case, low or high chl<sub>a</sub> concentration) could work more accurately than using all observations without dividing them into groups. Therefore, considering the fact that the optical properties of lakes are determined by chl<sub>a</sub>, SM and CDOM, it was decided to evaluate whether the results of the chl<sub>a</sub> model improve when models are created for classes of lakes. In this study, a biophysical class model was developed to describe the optical properties of the lake. The developed models for biophysical classes are used as part of methodology for determination of chl<sub>a</sub> concentration.

The preparation of the simultaneous *in situ* and Sentinel-2 database was necessary for training and testing biophysical class models. Simultaneous observations were the ones with a time gap of no more than three days between the *in situ* and satellite observation. The satellite data from a lake were collected in 9 20 m resolution pixels (3 x 3 macropixel), centred at the locations of national monitoring sites. Sen2Cor image classification was used to select only water pixels (water mask value – 6). There were 563 (226

different water bodies) simultaneous observations in 2015–2020 that were used to build biophysical class models.

For the classification of lake observations into biophysical classes, the *in situ* data collected in Lithuanian lakes during the state monitoring and the typology of lakes according to morphometric properties were used (LR Ministry of Environment, 2018). Chla, SM and transparency were used. The final biophysical class models determine one of four biophysical classes: Clear, Moderate, Chla-dominated class, or Turbid. Observations were narrowed down to four classes due to a small number of observations, e.g. a class dominated by SM was not discriminated separately due to the extremely small amount of data. However, collecting more representative data would allow more classes to be distinguished using the workflow of this study.

Machine learning models allow using more spectral information than a single spectral band or a single radiometric index, thus machine learning models were chosen to create the models. Input to the models – in this case, spectral features calculated from satellite spectra (reflectances in spectral bands, differences, ratios, as well as other derived indicators using multiple bands to describe the colour of a water body) and model output in this case is the biophysical class.

In different model application cases one or another algorithm can work better using a specific database, therefore, six different machine learning classification algorithms were tested to find the most suitable algorithm for biophysical classification. The tested algorithms are logistic regression (LR), support vector machine (SVM), random forest (RFC), adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), and artificial neural networks (ANN). The algorithm was chosen based on the calculated model performance metrics (accuracy, precision, recall, F1, area under curve (AUC), and log\_loss) and considering the simplicity of the model.

The biophysical class allowed different observations to be separated, and two models were further developed to determine chla concentrations in lakes based on the available database of simultaneous observations. One model covered mainly low chla concentrations ( $\text{chla} < 7.2 \text{ mg/m}^3$ , Clear biophysical class), the other medium and high ( $\text{chla} < 7.2 \text{ mg/m}^3$ , Moderate and Chla-dominated class). Observations of the Moderate class were combined with the Chla-dominated class due to the small amount of data. A chla model based on machine learning algorithms was also not developed for the fourth class, Turbid, because this class includes lakes dominated by other optically active substances (likely, CDOM), which complicates the calculation of chla from satellite data.

Chla models were constructed using the Random Forest Regression (RFR) algorithm. The performance of the Chla models was evaluated against a model built using the entire dataset. Also, its performance was compared with a more sophisticated ANN-based model. A data set supplemented with lake spectra, images with higher cloud cover (< 60%) and 2021 was used to build the chla models. Data were supplemented using the Google Earth Engine (GEE) data access service (Gorelick et al., 2017) from the S2\_SR collection, where level two data (L2A) are available. With the addition of data from GEE, the dataset covered 2015–2021 and increased to 984 simultaneous observations; however, the visual check (using RGB compositions) showed some cloud and cloud shadow affected cases, thus they were removed and remaining 749 data points were used for development of chla models.

After compiling and evaluating the models based on the machine learning algorithm, the models were applied to the entire available dataset for 2015–2021. The data were used to determine the maximum and average concentrations of lakes. Based on Lake Restoration Report, published in 2009, lakes were classified as potentially problematic when the average chla concentration in shallow and moderately deep lakes was  $> 8 \text{ mg/m}^3$ , and in deep lakes  $> 6 \text{ mg/m}^3$ , as well as when the maximum concentration in shallow and moderately deep lakes exceeded  $15 \text{ mg/m}^3$ , and in deep lakes it was  $> 12 \text{ mg/m}^3$  (Balevičienė et al., 2009). Based on these criteria, potentially problematic lakes were determined. Also, temporal and spatial analysis of chla concentration was performed in the studied 357 lakes and ponds.

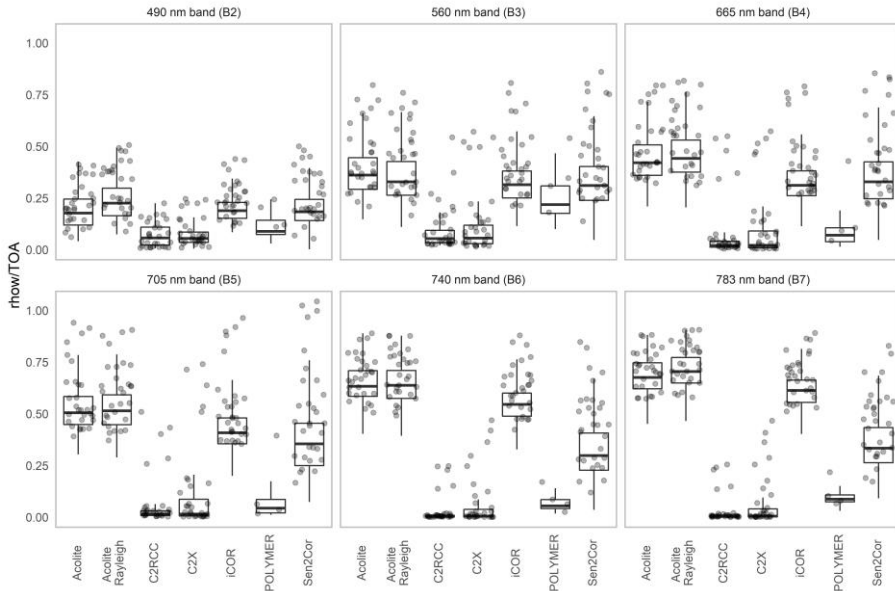
### 3. UNCERTAINTY ASSESSMENT OF ATMOSPHERIC CORRECTION ALGORITHMS FOR THE RETREIVAL OF CHLOROPHYLL A CONCENTRATION FROM SATELLITE DATA

#### 3.1. Effects of atmospheric correction products on estimated water leaving reflectance

Ratios of water reflectance ( $\rho_{\text{row}}$ ) and Top of Atmosphere (TOA) data were calculated for the spectra of 136 lakes that were in the T35UMB Sentinel-2 MSI region. The data variation was caused by the effect of the AC products, the inherent optical properties of lakes and their dynamics during the observation period.

The image-based AC products (Acolite, Acolite Rayleigh, and iCOR) presented the smallest changes in the shorter wavelength bands compared to other AC products (Figure 3.1). The  $\rho_{\text{row}}$ /TOA ratio calculated using these three products showed a similar nature and magnitude of transformation: the

median reflectance ratio for all lakes and dates gradually increased with increasing wavelength. Image-based products reduce reflectance similarly to Sen2Cor in the 490–705 nm bands, but reflectance values are reduced less in the 740 and 783 nm bands. Sen2Cor presents a consistent transformation across all wavelengths.



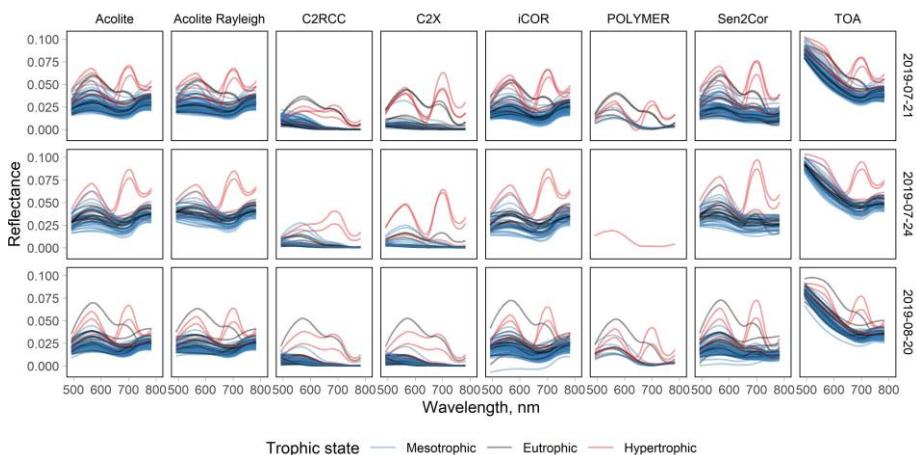
**Figure 3.1.** A boxplot figure represents transformation of the Top of Atmosphere (TOA) reflectance to the aquatic reflectance ( $\rho_{hov}$ ). The lower and upper hinges correspond to the first and third quartiles, the horizontal line shows the median, and the whisker extends from the hinges to the largest and smallest values no further than 1.5 inter-quartile range. The dots represent lakes that were used to match with *in situ* chlorophyll  $\alpha$  data.

When using other AC products, the changes in reflectance were characterised by a different nature and magnitude of the transformation. POLYMER and C2RCC products reduced the reflectance in the 490 nm and 560 nm spectral bands less than in the 665–783 nm bands. The C2RCC product produced extremely low water reflectance values compared to the other AC products used in this study. The median ratio of  $\rho_{hov}$  to TOA calculated by the POLYMER product is significantly higher than that calculated with the C2RCC. The smallest differences in reflectance are found in the 490 nm and 560 nm spectral bands. In the longer wavelength spectral bands, the C2RCC, C2X and POLYMER products reduce the TOA signal much more than the other four products (Figure 3.1).



Reflectances at shorter wavelengths are more corrected for Rayleigh scattering, while all reflectances at all wavelengths are corrected for aerosol absorption. The differences in the magnitude of correction among the image-based products are caused by the algorithms in products that estimate aerosol contribution. Thus, Acolite, Acolite Rayleigh, iCOR, and Sen2Cor exhibited similar transformations in the shorter wavelength spectral bands, with greater differences observed in the longer wavelength spectral bands.

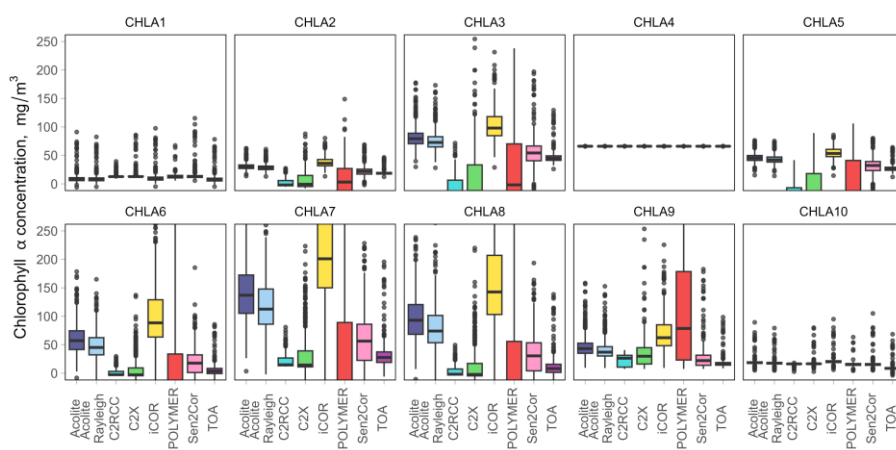
Most of the analysed lakes are mesotrophic based on state monitoring data (100 of 136), and the spectral profiles of all mesotrophic lakes obtained by each AC product for each photo are similar in shape and size (Figure 3.2). In the second half of summer, an increased reflectance is observed in the 705 nm and 560 nm spectral bands in the spectra of several lakes, mainly eutrophic and hypertrophic (Figure 3.2). A clear peak in the 705 nm band is visible in the spectral profiles calculated using all AC products. In most of the examined cases, the shape of the spectral profile in the 665–740 nm wave range depends mostly on the selected AC product. On average, the variation in R705/R740 and R705/R665 is more driven by the selected AC product than by the lake-specific dynamics of optically active substances such as coloured dissolved organic matter, total suspended solids, and chl<sub>a</sub>.



**Figure 3.2.** Top of Atmosphere (TOA) and water reflectance ( $\rho_{\text{row}}$ ) of 136 lakes, processed with even atmospheric correction algorithms on three days in summer when algal blooms potentially occur in lakes with higher productivity in 2019.

### 3.2. The uncertainty of the retrieved chlorophyll $\alpha$ concentration

Chla algorithms showed different skill levels to estimate chla concentration in the 136 analysed lakes. The uncertainties of R705/R665 and R705/R740 were transferred to the uncertainty of chla concentration. The three spectral band CHLA6-CHLA8 algorithms were the most sensitive to the selection of the AC product based on the median of the concentration amplitude. The influence of the 740 nm band to the retrieved chla concentration is lower when the average of the 665 nm and 740 nm spectral bands is used, as in the case of the CHLA1 algorithm. The lowest variation range of chla concentration obtained using different AC products was determined using spectral band difference algorithms (CHLA1 and CHLA10, Figure 3.3).



**Figure 3.3.** Chlorophyll  $\alpha$  concentrations calculated using Top of Atmosphere (TOA) and water reflectance ( $\rho_{\text{row}}$ ) data and 10 chlorophyll  $\alpha$  algorithms. The graph was obtained using data from 136 lakes and seven images. The graph shows the concentration range of 0–250  $\text{mg}/\text{m}^3$ , excluding negative and extreme values.

The concentrations obtained with the CHLA2 and CHLA3 algorithms, where the most commonly used band ratio R705/R665 was included, were sensitive to the AC product used. The CHLA4 algorithm showed no response to using any AC product. The CHLA5 algorithm, which uses the R665/R705 ratio, and the CHLA9 algorithm, which is based on the normalized chlorophyll  $\alpha$  difference, also showed a large variation between the AC products used.

Some algorithms (CHLA3, CHLA5, CHLA6, CHLA7 and CHLA8) yielded quite a number of negative concentration cases: using Sen2Cor and TOA, C2X, POLYMER and C2RCC data. Atmospheric correction algorithms

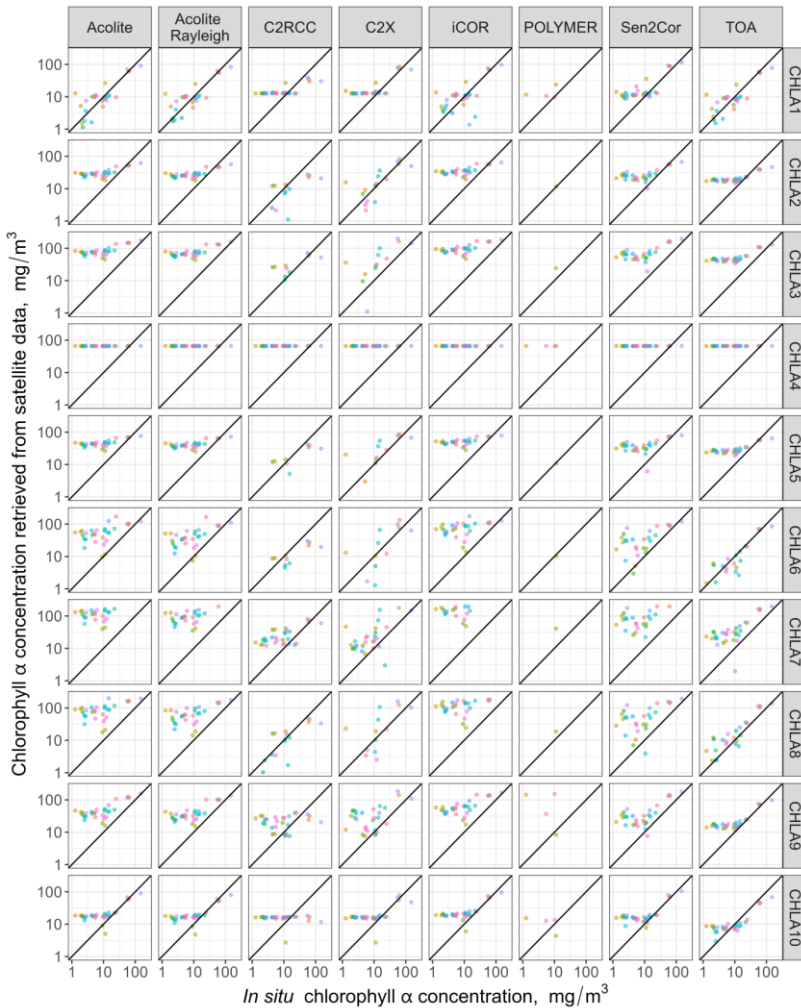
reduced the already small reflectance from lakes, and using different radiometric chlA algorithms increased the uncertainty.

### 3.3. The evaluation of chlorophyll $\alpha$ concentration algorithms

According to the model evaluation metrics, the strength of the relationship between chlA concentration derived from satellite data and *in situ* concentration differed considerably between chlA algorithms and AC products. The coefficient of determination of 17 models was greater than 0.8, but the deviation of these algorithms ranged from  $-0.9 \text{ mg/m}^3$  (CHLA1 with Acolite) to  $38.6 \text{ mg/m}^3$  (CHLA3 with TOA data).

The algorithm of three spectral bands (CHLA1) showed a small deviation (up to  $7.8 \text{ mg/m}^3$ ), and the coefficient of determination of four AC products (Sen2Cor, iCOR, Acolite, Acolite Rayleigh) and TOA was higher than 0.8. Another two-band difference-based algorithm (CHLA10) showed similar results, but the  $R^2$  was slightly higher (0.87–0.88) and the deviation was higher in most cases (up to  $11.9 \text{ mg/m}^3$ ). Only when the algorithm was used with TOA data, the  $R^2$  was higher and the bias slightly smaller than that with AC products ( $-1.6 \text{ mg/m}^3$ ).

Only in some cases did the measured and calculated concentrations based on satellite data agree well (Figure 3.4). In some cases where a high coefficient of determination ( $R^2 > 0.8$ ) was observed (for example CHLA2 and CHLA3 with TOA data, CHLA9 with Sen2Cor and CHLA10 with Acolite, Acolite Rayleigh, iCOR and Sen2Cor), the agreement of chlA concentration obtained from satellite data with *in situ* measurements was poor in the range of low chlA concentrations ( $\text{chlA} < 10 \text{ mg/m}^3$ ). The lowest mean deviation of  $1.7 \text{ mg/m}^3$  for concentrations below  $10 \text{ mg/m}^3$  was determined using CHLA1 and TOA data, and for concentrations above  $10 \text{ mg/m}^3$  the lowest mean deviation of  $0.0 \text{ mg/m}^3$  was determined using the CHLA1 algorithm and Sen2Cor data. CHLA1 and CHLA10 showed the most stable results for the selection of AC products, except for the ANN-based AC products C2RCC and C2X. The results show that relatively larger errors are observed at lower chlA concentrations, so more reliable chlA concentrations could be derived from remote sensing data by applying algorithms to a group of lakes with similar optically active substance content (Spyrakos et al., 2018), as was done by Neil et al. (2019), Soomets et al. (2020), and Pahlevan et al. (2021).



**Figure 3.4.** Graphs of the relationship between measured and algorithmically obtained chlorophyll  $\alpha$  concentrations (scales are logarithmic).

#### 4. BIOPHYSICAL CLASSIFICATION OF LAKES

##### 4.1. Biophysical classification based on *in situ* measurements

EPA measured chla concentration, water transparency, and suspended matter concentration were used to distinguish lake biophysical classes, which then were used to develop chla concentration algorithms. After dividing the observations into groups, 12 classes were distinguished (Table 4.1). Class 1\_clear (46% of observations), characterised by low chla and SM concentrations, and high transparency. The next larger class, 2\_chla\_clear (21% of observations), was characterised by higher chla concentrations, good transparency, and low SM concentrations. Another class, assigned to 9% of

observations – 3\_chla\_turbid is associated with medium, poor or very poor ecological class according to transparency and high chla. The remaining classes were not frequent, accounting for 0 to 7% of the observations. It is also noteworthy that the observations where SM predominates (1\_SM\_turbid, 2\_chla\_SM\_turbid) were very rare, thus these observations were not included in the training-test dataset due to insufficient representative data.

In the first step machine learning models distinguish between measurements with good transparency and low concentrations of optically active substances (Clear class based on 1\_clear, Table 4.1) from others (OAS-class). Separating clear lakes from others allows the separation of observations that are not associated with problematic waters. Therefore, two classes were used in the binary classification – the Clear class and the OAS-class. Classes with only a few measurements were not included in the algorithms. The classes used for the binary classification task were of different sizes, with 46% of observations in the Clear class and 54% of observations in the OAS-class. In the second classification step, the OAS-class was divided into three classes: Moderate, Chla-dominated, and Turbid class.

#### 4.2. Biophysical classes of lakes

The Clear class had an average chla concentration of 3.8 mg/m<sup>3</sup> and an average transparency of 4.5 m (Table 4.2), while the Moderate classes had a higher chla concentration (an average of 11.7 mg/m<sup>3</sup>) and a lower transparency (an average 2.5 m) than the Clear class. The ecological status of the other two classes was generally moderate, poor or bad. The Chla-dominated class was characterised by algal blooms, and the Turbid class water bodies had low transparency despite the relatively low chla concentration. These classes were used as target labels in machine learning algorithms during model building, i.e., the target label is what the model aims to identify.

#### 4.3. Development of biophysical class models

Biophysical class models were built using machine learning classification algorithms. Algorithms calculate and determine target labels, in this case biophysical classes, based on the features provided to them. Various spectral features describing the reflectance spectrum of lake water were used for model training: reflectance amplitude – the difference between the maximum and minimum reflectance at 490–865 nm wavelengths, spectral band ratios R705/R665 (Ammenberget al., 2002), R560/R490 (Hussein & Assaf, 2020), R560/R665 (Toming et al., 2016), and R560/R705 (Sòria-Perpinyà et al., 2021), spectral band differences  $BD1 = R705 - (R665 + R740)/2$  (Toming et al.,

**Table 4.1.** Definition of classes based on water quality parameters. To solve the binary biophysical classification problem, two classes were assigned: Clear class or OAS-class. In the multiple class problem Moderate, Chla-dominated and Turbid classes labels were used. Classes marked with "-" were not included in the training-test dataset due to small amount of data.

Class	Transparency class*	Chlorophyll $\alpha$ class	Suspended Matter Concentration Class	Number of Cases (Match-ups)	Label for Binary Classification	Label for Multi-Class Classification
1_clear	Clear *	Low chla < 7,2 mg/m <sup>3</sup>	< 10 mg/m <sup>3</sup>	260	Clear	—
1_turbid	Turbid			15	OAS-class	Turbid
1_SM_clear	Clear		≥ 10 mg/m <sup>3</sup>	0	—	—
1_SM_turbid	Turbid			1	—	—
2_chla_clear	Clear	Medium chla 7,2–20 mg/m <sup>3</sup>	< 10 mg/m <sup>3</sup>	119	OAS-class	Moderate
2_chla_turbid	Turbid			38	OAS-class	Turbid
2_chla_SM_clear	Clear		≥ 10 mg/m <sup>3</sup>	3	—	—
2_chla_SM_turbid	Turbid			10	OAS-class	Chla-dominated
3_chla_clear	Clear	High chla > 20 mg/m <sup>3</sup>	< 10 mg/m <sup>3</sup>	31	OAS-class	Chla-dominated
3_chla_turbid	Turbid			49	OAS-class	Chla-dominated
3_chla_SM_clear	Clear		≥ 10 mg/m <sup>3</sup>	2	OAS-class	Chla-dominated
3_chla_SM_turbid	Turbid			38	OAS-class	Chla-dominated

\* Clear is defined with water transparency > 1.3 m for shallow lakes and > 2 m for medium deep and deep lakes. While turbid is the opposite—< 1.3 m or 2 m.

**Table 4.2.** Values of water parameters by biophysical class based on 2015–2020 data.

Class	N	Chla range, mg/m <sup>3</sup>	Mean chla, mg/m <sup>3</sup>	Transparency Range, m	Mean transparency, m	SM Range	Mean SM
<b>Clear</b>	260	0.2–7.1	3.8	1.4–11.0	4.5	0.9–7.5	2.5
<b>OAS-class:</b>	303						
<b>Moderate</b>	119	7.3–19.8	11.7	1.3–6.8	2.7	0.9–8.0	3.7
<b>Chla-dominated</b>	131	11.3–148.1	42.4	0.3–5.0	1.1	1.0–56.0	9.3
<b>Turbid</b>	53	0.8–19.4	11.6	0.5–1.9	1.3	1.7–9.6	4.9

2016),  $BD2=R705-1.05*R665$ , and an empirical equation using  $BD2$  (Grendaitė & Stonevičius, 2021), Apparent Visible Wavelength (AVW) (Vandermeulen et al., 2020), hue angle, colour according to the Forel-Ule colour scale (FU) (van der Woerd & Wernand, 2018) and month, which could help distinguish algal blooms that are more common in summer.

During model development, the available data set is divided into a training set (80%) and a testing set (20%) by lake (all observations from one lake fall into the training or testing set so that the model does not learn a lake-specific combination of features). For the binary problem, the training set contained 440 observations and the testing set contained 123 observations. A dataset of 303 observations was used to solve the three-class problem – 245 observations assigned to the training set and 58 observations to the testing set.

Model estimation was performed using stratified three-fold cross-validation, which is suitable for model estimation using small training sets and a testing set. The variation of the metrics during different splits shows the sensitivity of the model to the data split.

#### 4.4. Important spectral features of lakes for biophysical class modelling

The selection of features used to build models is an important step in the development of machine learning models. All features, that are frequently used individually in other studies, were at first used to develop the models in this study; however, some of them are similar (e. g.  $R705/R665$  and  $BD2$ ) and only in individual cases some can provide more information than others.

The features used can be divided into two groups: features that are directly calculated from the reflectance data – either the difference of the spectral bands or the ratio of the bands – and another group – more complex features derived from the reflectance of several spectral bands (AVW, hue angle, and FU). Since AVW, hue angle, and FU were highly correlated ( $R = -0.94, 0.9,$  and  $0.94$ ), only one of them – AVW – was chosen to be used, because AVW had the highest correlation with biophysical class. Furthermore, the feature distribution showed better class separability using AVW, as the dominant colour-characterizing wavelength shifts to longer wavelengths as the class changes from Clear through Moderate and into the Chla-dominated class.

$BD1$  and  $R560/R705$  features were selected from another group of features because they are less correlated than the other features ( $R = -0.23$ ) and correlate well enough with the biophysical class ( $R = -0.38$  and  $R = 0.3$ , respectively).

## 4.5. Results of validation and testing of biophysical lake class models

### 4.5.1. Binary classification

The accuracy of the classification models reached 79–81% using the training data set. The performance of the models was generally quite similar; however, based on the AUC score, the RFC model performed best on the test data (Table 4.3). In addition, RFC provided the lowest value of `log_loss`, indicating good performance of the model. It was decided to use the RFC model because it provided the lowest number of misclassified lakes ( $n = 19$  out of a total of 123), as well as a relatively small number of cases misclassified to OAS-class (FN,  $n = 7$ ). Attempts were made to optimize the hyperparameters of the RFC model, but this did not improve the accuracy of the model, so the original version with default parameters was used. In the RFC classifier BD1 was the most important feature (relative importance = 0.4), while R560/R705 and AVW had equal importance—0.30. The median `chla` concentration of misclassified data points were  $10.3 \text{ mg/m}^3$ , indicating that the majority of data points with high `chla` concentrations were assigned to the correct OAS-class.

**Table 4.3.** Model performance metrics on test data.

Classifier	Test Accuracy	Test Precision	Test Recall	Test F1 Score	Test AUC	Test Log Loss
LR	0.83	0.73	0.87	0.80	0.84	5.9
SVM	0.83	0.72	0.89	0.80	0.84	5.9
RF	0.85	0.77	0.85	0.81	0.85	5.1
Ada	0.81	0.79	0.70	0.74	0.79	6.5
XGB	0.82	0.75	0.81	0.78	0.82	6.2
ANN	0.83	0.76	0.81	0.78	0.83	6.0

### 4.5.2. Multi-class classification

To distinguish between different lakes in the OAS-class, a model based on a random forest algorithm was further built for the OAS-class consisting of measurements from lakes with high amounts of optically active substances, `chla` and SM. Spectral feature selection was also performed. Three features were selected—reflection amplitude, R705/R665 ratio, and AVW. The R705/R665 feature had the highest relative importance—0.38, while the other features had a relative importance of 0.31 each. After fitting the model using the test data, the accuracy between classes ranged from 27% for Turbid class, 70% for `Chla`-dominated, to 81% for Moderate class.

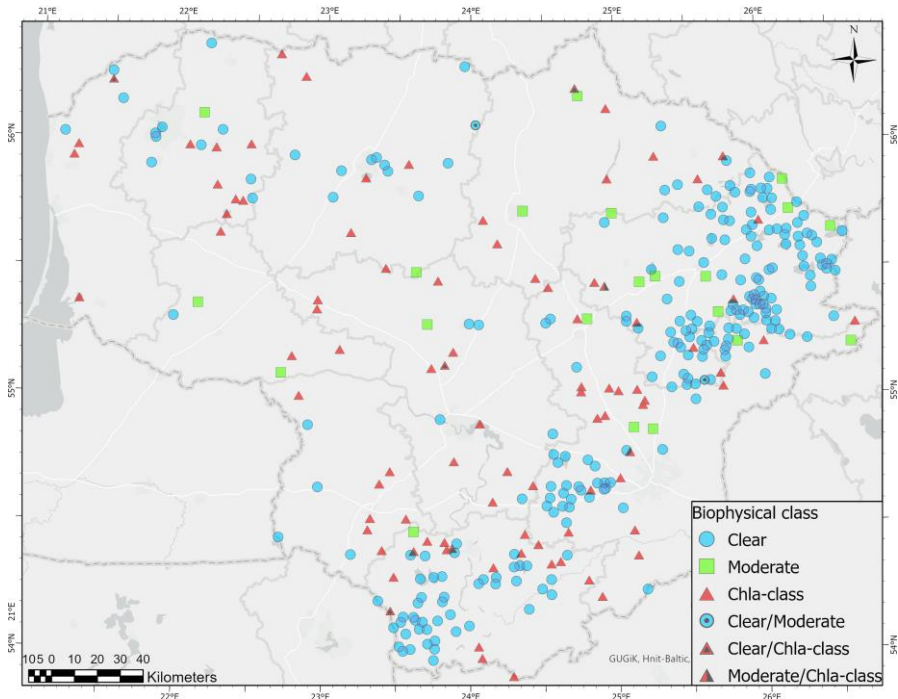


#### 4.6. Biophysical classification of lakes using RFC models

The biophysical class model was constructed to separate different observations and lakes that have more similar optical properties within a class than individual classes. Biophysical classification is further used for more accurate determination of water parameters such as chl<sub>a</sub>.

The two-lake class model was applied to 43 022 lake spectra obtained from 357 lakes for the seven-year period 2015–2021. There were between 36 and 160 observations per lake (mean = 121, standard deviation, sd = 21).

Clear class was detected in 58.4% of lake spectra, a three-class model was applied to the rest of the spectra (17 903) and 18.1% of the observations were assigned to the Moderate class, 21.6% to the Chl<sub>a</sub>-dominated, and 1.9% to the Turbid class. The dominant class throughout the study period 2015–2021 was determined for each lake. The Clear class was found to be dominant in 242 lakes, of which in 161 lakes Clear class dominated throughout 2015–2021 period (Figure 4.1). Other classes were also sometimes observed in these lakes—algal bloom-associated Moderate, Chl<sub>a</sub>-dominated class, and Turbid class were recorded on average in 19.4%, 14.7%, and 6.8% of observations in a lake during the study period. Chl<sub>a</sub>-dominated class was dominant in 87 lakes, from which in 40 it was dominant each year of 2015–2021.



**Figure 4.1.** The prevalent class of lakes throughout the 2015–2021 period.

#### 4.7. Recommendations for assessment of biophysical classes

The created biophysical class model can be applied without adaptation in other regions where waters have similar optical properties. The model was developed based on observations from 226 lakes and focuses on phytoplankton-dominated waters. The class of Turbid lakes (due to other than SM or chl<sub>a</sub> substances) was also included in the classification, but to improve model results of this class, a better description of this class is needed, for example, using the absorption coefficient at 440 nm wavelength, which well describes the amount of CDOM in water (Nima et al., 2019). In other regions, the class definition described in Table 4.1 could be supplemented with this parameter to update the grouping of classes for training machine learning models. More classes could be added to the constructed model, such as a class dominated by SM (e. g., class 1\_SM\_turbid as shown in Table 4.1). In this case, the model should be retrained and the inclusion of new spectral features should be considered.

Biophysical classification of lakes can be a good supporting tool for water status assessment, it can inform about the condition of a lake between *in situ* measurements. Also, biophysical classes can be used to develop algorithms for determining water parameters from satellite data.

### 5. CHLOROPHYLL A CONCENTRATION MODELS AND ANALYSIS OF ITS CHANGE

#### 5.1. Development of chlorophyll $\alpha$ concentration models using random forest regressor

To determine the chl<sub>a</sub> concentration, models based on the random forest regression algorithm were created: for the Clear class (hereinafter RFR1s) and for the Moderate and Chl<sub>a</sub>-dominated classes together (hereinafter called RFR1tc). 450 observations were used to create RFR1s, and 299 data pairs were used for RFR1tc. The average chl<sub>a</sub> concentration in the RFR1s data set was  $7.0 \pm 6.0$  mg/m<sup>3</sup>, and in the RFR1tc data set  $27.5 \pm 26.8$  mg/m<sup>3</sup>.

To improve the interpretation and accuracy of a model and to have a simpler model, feature selection was carried out. The least correlated and most important features were selected. The features that had the highest relative importance, low variance inflation factor, and were less related according to the features dendrogram were selected. Four features were selected for the RFR1s model (relative importance in parentheses): BD1 (0.47), reflectance amplitude (0.19), R560/R705 (0.19), and hue angle (0.16). Reducing the

number of features did not decrease the accuracy of the model using cross-validation. For the RFR1tc model, six attributes were selected: BD1 (0.52), R705/R665 (0.22), R560/R490 (0.07), B8A (0.07), R560/R705 (0.07), R560/R665 (0.06). The hyperparameters were optimized using *GridSearchCV()*, which improves the generalization skills of a model. Model hyperparameters were searched using the training set (323 observations for RFR1s, 206 observations for RFR1tc). The combination of hyperparameters that minimized the squared error (the accuracy criterion of a RFR algorithm) was selected. Then the models were trained using the training and testing sets (414 observations for RFR1s, 271 observations for RFR1tc), and its accuracy was tested using the validation set (RFR1s—36 observations, RFR1tc—28 observations). During the validation, better accuracy of models was determined (RFR1s model  $R^2 = 0.54$ , RMSE = 4.1 mg/m<sup>3</sup>, RFR1tc model  $R^2 = 0.71$ , RMSE = 11.5 mg/m<sup>3</sup>) than when using the cross-validation method, thus it was decided that the hyperparameters were chosen correctly. The final models were trained on the full dataset (450 observations for RFR1s, 299 observations for RFR1tc).

The RFR1s model best determined values in the range of 5–10 mg/m<sup>3</sup> (measured average  $7.0 \pm 1.0$  mg/m<sup>3</sup>, and model determined— $7.5 \pm 3.1$  mg/m<sup>3</sup>), while lower concentrations in all cases were estimated about 2 mg/m<sup>3</sup> higher.

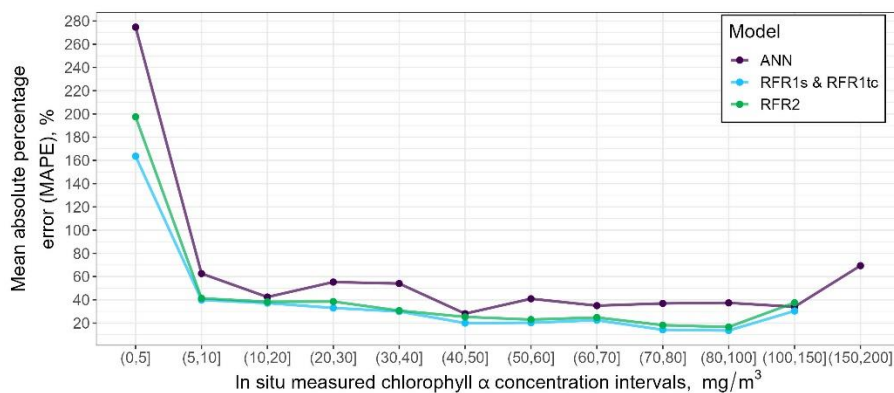
The RFR1tc model best determined chl<sub>a</sub> in ranges—20–30 and 40–50 mg/m<sup>3</sup>, which are similar to the average concentration determined by the model. The chl<sub>a</sub> concentration determined by the RFR1tc was in the range of 5–92 mg/m<sup>3</sup> values, so values higher than 92 mg/m<sup>3</sup> were estimated with a considerable error. Collecting more representative data in the future would allow the model to be trained and revalidated and model performance should improve.

## 5.2. Chlorophyll $\alpha$ models comparison

The results of the RFR1s and RFR1tc models were compared to the RFR2 model that was built using the same complete dataset as RFR1s and RFR1tc combined, without separating the data by biophysical classes. Also, the results of RFR1tc and RFR1tc were compared with a more sophisticated machine learning model, an artificial neural network-based (ANN) model. The ANN was trained on an almost twice as large training dataset, which included lakes in Lithuania, Latvia and Estonia. The final ANN model was based on Residual deep neural network with Multi-Head Attention, the metrics of this model in using the testing set was  $R^2 = 0.55$  and RMSE = 13.5 mg/m<sup>3</sup>

The RFR2 performed worse than the two RFR1s and RFR1tc models developed for biophysical classes. Classification by biophysical classes hardly improved chl<sub>a</sub> results in the Clear class, with little change in R<sup>2</sup> and RMSE. The Chl<sub>a</sub> model for classes showed better results for Moderate and Chl<sub>a</sub>-dominated class (R<sup>2</sup> for Moderate class improved from 0.19 to 0.46 and for Chl<sub>a</sub>-dominated class from 0.68 to 0.76). Moderate and Chl<sub>a</sub>-dominated classes are more important as they are associated with algal blooms.

The lowest mean absolute percentage error (MAPE) for the RFR1s and RFR1tc models was found in the range of 70–100 mg/m<sup>3</sup> (14%), in the range of 40–60 mg/m<sup>3</sup>—20%, and elsewhere it ranged from 23 (60–70 mg/m<sup>3</sup>) to 164 % (0–5 mg/m<sup>3</sup>) (Figure 5.1). The MAPE of the ANN model was higher than that of the RFR1s and RFR1tc models in all ranges of chl<sub>a</sub> concentration and ranged from 28% (40–50 mg/m<sup>3</sup>) to 275% (0–5 mg/m<sup>3</sup>).



**Figure 5.1** The mean absolute percentage error (MAPE) of chlorophyll  $\alpha$  models by *in situ* concentration intervals.

Therefore, the classification of lakes to classes helped to create more accurate chl<sub>a</sub> models than using the random forest RFR2 model using the entire data set, as well as ANN. Thus, for further analysis RFR1s and RFR1tc models were chosen and they will be referred to as RFR in the text below.

### 5.3 Removal of chlorophyll $\alpha$ concentration outliers

RFR models were applied to the spectra of 37 450 Lithuanian lakes in 2015–2021. Analysing the time series graphs, fluctuations in chl<sub>a</sub> concentration were observed in a short period of time (between 3–7 days). It was observed that the sharp decreases in concentration were due to the cloud shadow covering the lake, which greatly reduces the chl<sub>a</sub> signal in satellite data. There were also cases when the prevailing low concentration in the lake

increased by 15–20 mg/m<sup>3</sup> during one observation, such cases were associated with thin clouds or the edge of cloud, which were retained by image classification algorithm.

To eliminate such observations, an outlier removal method based on the standardized z-module value was applied. A threshold value of 1.1 was selected according to several tests. Observations with > 1.1 z-module value were considered as outliers and filtered before further analysis. This method removed between 3 and 38 observations (median 20) from chl<sub>a</sub> observations, representing 7 to 31% (median 19%) of observations in lake.

#### 5.4. Identification of potentially problematic water bodies

The collected data were used to identify potentially problematic lakes based on average and maximum values of chl<sub>a</sub> concentration. It should be noted that in 43 lakes and ponds, *in situ* observations were not carried out during the study period, so the application of models based on satellite data significantly adds to the knowledge about them. 31 of these lakes were determined as potentially problematic by both mean and maximum concentration criteria, six according to the mean, and the remaining 6 were identified as non-problematic (Table 5.2).

The *in situ* measurements identified more non-problematic lakes than the RFR, but the differences could be due to the availability of only several *in situ* measurements per year or even for the entire 2015–2021 period, and at least a few dozen observations were obtained from satellites in each lake during the study period (88 observations on median). Thus, short-term chl<sub>a</sub> changes are recorded by satellite data, which respectively affect the mean and maximum chl<sub>a</sub> values.

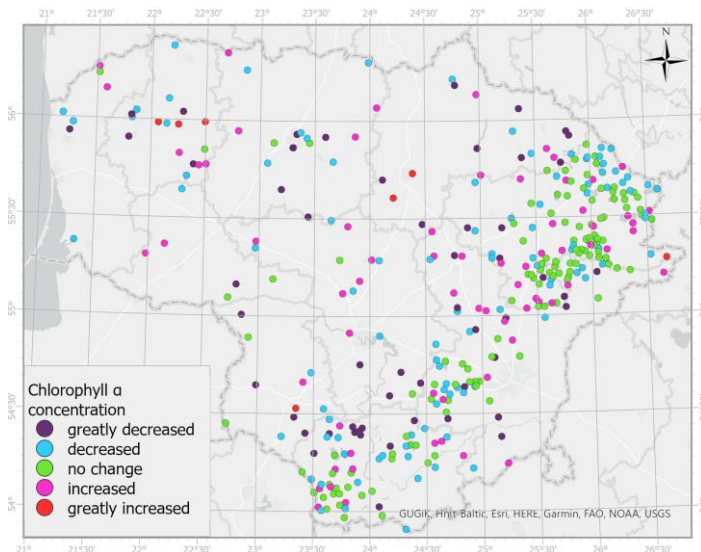
**Table 5.2.** Number of potentially problematic lakes determined by mean and maximum chlorophyll  $\alpha$  threshold values.

Lake group	Using <i>in situ</i>	Using RFR	Using both
Non-problematic (using mean and maximum criteria)	124	66	52 (6—RFR)
No data	43	—	—
Problematic only according to mean chl <sub>a</sub> concentration	4	1	0
Problematic only according to maximum chl <sub>a</sub> concentration	4	47	9 (6—RFR)
Problematic according to both mean and maximum chl <sub>a</sub>	158	227	172 (31—RFR)

Based on both *in situ* and satellite data, 52 lakes (15% of the total) were identified as non-problematic and 172 lakes (48% of the total) were determined as potentially problematic based on both *in situ* and RFR mean and maximum chl<sub>a</sub> concentrations (Table 5.2). The remaining 133 (37%) were classified as potentially problematic by one or the other criteria, either by *in situ* only or by RFR data only.

### 5.5. Chlorophyll $\alpha$ concentration variation in lakes

Using satellite data and developed models the concentration of chl<sub>a</sub> in studied lakes was determined, as well as its seasonal and interannual changes in 2015–2021 period. The average chl<sub>a</sub> concentration increased in 76 lakes during the studied period, in seven of which the increase was significant ( $> 10 \text{ mg/m}^3$ ), e. g. Mastis lake and Marijampolė II pond (Figure 5.2). In 126 lakes during the study period, no change in chl<sub>a</sub> status was detected—the average concentration did not change. The average chl<sub>a</sub> concentration decreased in 155 lakes, of which 54 showed a significant decrease. The majority of these lakes were classified as potentially problematic for almost the entire study period (Figure 5.2). An improvement in condition is observed in these lakes, but only in two of them the concentration of chl<sub>a</sub> decreased to such an extent that the lake was determined as non-problematic in the last year of the study.



**Figure 5.2** The change of chlorophyll  $\alpha$  concentration 2015–2021 in the studied lakes. A larger decrease or increase indicates a change  $> 10 \text{ mg/m}^3$ , a smaller one— $1\text{--}10 \text{ mg/m}^3$ .

In potentially problematic lakes, there was a large seasonal variation in chl<sub>a</sub> concentration. Almost a quarter (82) of lakes experienced chl<sub>a</sub> concentrations > 50 mg/m<sup>3</sup> according to satellite data. These lakes are characterised by annual algal blooms and the maximum concentration of chl<sub>a</sub> is usually reached in August or September.

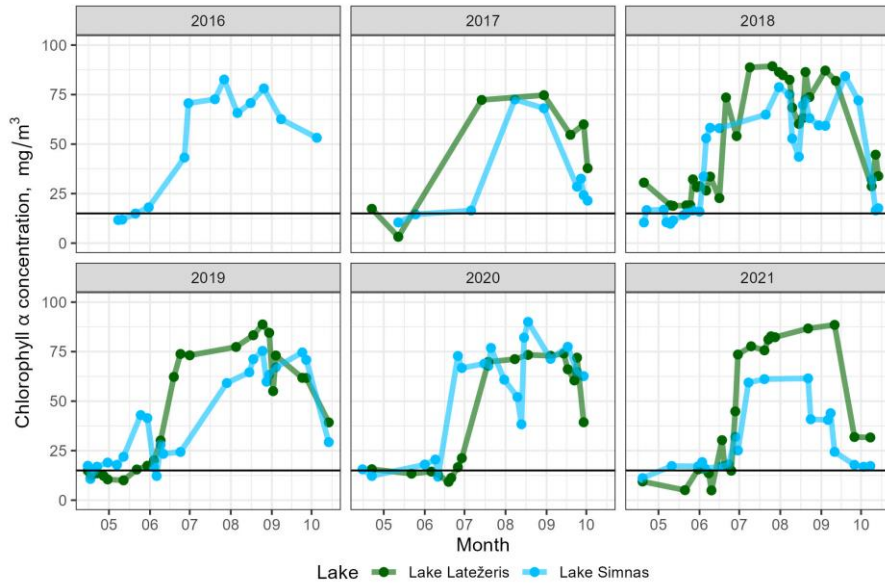
In the lakes that were classified as non-problematic, the concentration of chl<sub>a</sub> did not change much through the season, but in some, short-term increases in the concentration of chl<sub>a</sub> lasting up to several weeks were detected, for example, in Lake Vievis in June–July, and in Lake Vištytis in the first half of August. These increases in chl<sub>a</sub> are associated with light algal blooms, which do not have long-term effects on the ecosystem and ecological status of a lake.

Some potentially problematic lakes (Lake Ūdrija, Lake Obeliai, Lake Amalvas, Lake Pravalas, and Petraičiai pond) with a high maximum concentration (> 50 mg/m<sup>3</sup>) showed less pronounced seasonal chl<sub>a</sub> variation—these lakes were characterised by a relatively high chl<sub>a</sub> concentration from spring to autumn. However, taking into account that the average error of the model in the range of high chl<sub>a</sub> concentrations was up to 30%, the real values could be somewhat higher than those obtained by the RFR model from satellite data.

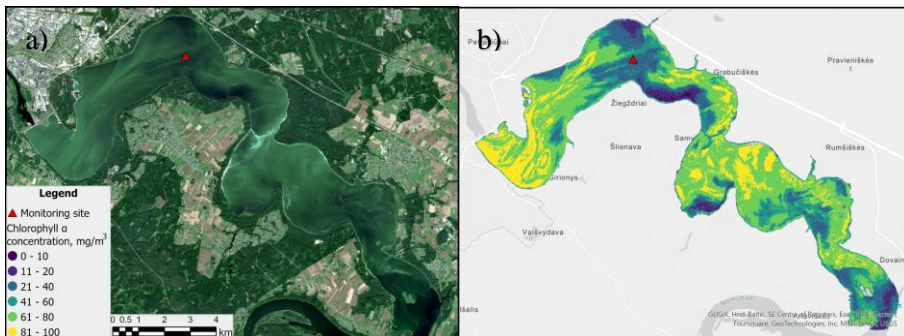
In other lakes with high maximum chl<sub>a</sub> concentration, the seasonal chl<sub>a</sub> variation was more pronounced. The variation was well reflected in Lake Latežeris and Lake Simnas, where the concentration of chl<sub>a</sub> was relatively low in April–May, increased in summer, and remained relatively high in autumn (Figure 5.3). Exponential algal growth has also been recorded in individual years (e. g. Simnas Lake in 2018). Exponential growth is observed when a large gradient of increase in chl<sub>a</sub> concentration is detected between 2–3 observations (> 5 mg/m<sup>3</sup> per day). This algal bloom phase occurs when environmental conditions are favourable, and nutrients are not lacking.

With satellite data, it is possible to record the spatial concentration of chl<sub>a</sub> in the entire water body. This is a big advantage of satellite data, especially for large inhomogeneous bodies of water. Spatial chl<sub>a</sub> analysis can show the representativeness of the state monitoring measurement site for the entire water body. For example, in Kaunas reservoir (area 63.5 km<sup>2</sup>), a large spatial heterogeneity of chl<sub>a</sub> concentration was recorded (Figure 5.4), which is due to the size and shape of the water body and its exposure to the prevailing wind directions. The highest concentrations are often recorded in the western and central part of the lagoon, and the concentration at the monitoring site is often

lower, so it does not represent algal blooms in this water body and only partially reflects the condition of the water body.



**Figure 5.3.** Seasonal chlorophyll  $\alpha$  concentration variation according to satellite data in Lake Latežeris and Lake Simnas, in which one of the highest average chlorophyll  $\alpha$  concentrations were observed during 2015–2021 period. The horizontal line shows the threshold value of chl $a$  concentration in shallow (type 1) water bodies ( $15 \text{ mg/m}^3$ ).



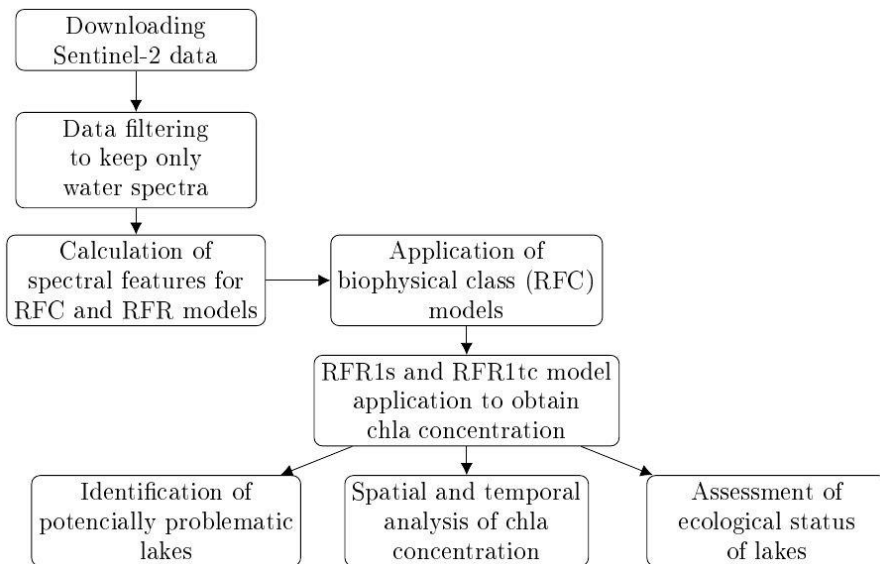
**Figure 5.4.** Kaunas reservoir a) Sentinel-2 natural colour RGB composition and b) distribution of chlorophyll  $\alpha$  concentration obtained by RFR models on August 7, 2020.



## RECOMMENDATIONS

Atmospherically corrected Sentinel-2 MSI (L2A) data and the image classification required for data filtering can be downloaded directly from Copernicus Open Access Hub or using satellite data viewing and analysis platforms such as Google Earth Engine that have satellite image collections (Figure 6.1).

Based on the biophysical classification developed in the study, lakes with low optically active substances (Clear) are separated from lakes with optically active substances, for which the Moderate, Chla-dominated class and Turbid class have been determined. To obtain the biophysical class of a lake, spectral features are calculated: BD1, R560/R705, and AVW (to separate two classes—Clear and OAS-class) and R705/R665, AVW, and reflectance amplitude (to separate three classes—Moderate, Chla-dominated, and Turbid class, Table 4.1) and apply RFC models (Figure 6.1).



**Figure 6.1.** The scheme representing chlorophyll  $\alpha$  concentration determination using the methodology proposed in this study.

Biophysical classification increased the accuracy of chla algorithms using Sentinel-2 MSI data, especially in Moderate and Chla-dominated class lakes. For Clear class (RFR1s) lakes, spectral features are required for chla determination: BD1, reflectance amplitude, R560/R705, and hue angle, while for Moderate and Chla-dominated class (RFR1tc) lakes—BD1, R705/R665, R560/R490, B8A, R560/R705, and R560/R665.

Developed and validated chl<sub>a</sub> models were applied to 357 lakes for 2015–2021 study period. These data helped observe short-term increases in chl<sub>a</sub> concentrations that indicated algal blooms. Also, long-term changes of mean concentration showed increase/decrease in water quality. Potentially problematic lakes were identified using average and maximum chl<sub>a</sub> concentrations, since one water body has several tens of observations (median 88 during the study period). These collected data could be useful for water quality experts, decision makers, and the public, since only water with a good status is safe to use.

The further use of the created models could help optimize monitoring programs (e.g. by choosing the time of observations based on the dynamics of algal bloom determined by satellite data, by choosing the measurement site based on the spatial distribution of chl<sub>a</sub>) and contribute to a better determination of status of water bodies. It could be used both in the long-term assessment of the ecological state of lakes, and in observing short-term chl<sub>a</sub> fluctuations in time and space, since new data can be obtained almost in near-real time.

The developed models can be used for Lithuanian lakes, as well as for other lakes and ponds in middle latitudes with similar characteristics, i.e., the main substance in them is chlorophyll  $\alpha$ . Biophysical classification can be extended by distinguishing additional classes mentioned in Table 4.1 when more data describing them become available. New global databases such as GLORIA (Lehmann et al. 2023) could be used to update the RFC models. Updating the biophysical class model would require creating a new chl<sub>a</sub> model for that class as well. The created methodology could be further developed in this way and applied to more diverse water bodies.

## CONCLUSIONS

1. In the spectral bands 665–740 nm, which are important for chlorophyll  $\alpha$  retrieval from satellite data, different reflectances were obtained with seven different architecture atmospheric correction algorithms. The different spectra resulted in a large variation of chl<sub>a</sub> concentration obtained using seven atmospheric correction products and chl<sub>a</sub> algorithms. Uncertainty using TOA data and chl<sub>a</sub> algorithms was lower than using atmospheric correction products.

2. Random forest classification algorithm was chosen for the biophysical classification of lakes based on accuracy metrics. The developed classification models allow the distinction of four classes (Clear, Moderate, Chl<sub>a</sub>-dominated class and Turbid). During the study period, most of the lake observations were

assigned to the Clear class (58.4%), characterised by low amounts of optically active substances. The second most common was the Chla-dominated class (21.6%), which is associated with algal blooms.

3. The Clear (RFR1s) and Moderate with Chla-dominated class (RFR1tc) models based on the random forest regression algorithm showed higher accuracy than the RFR2 model and the model based on artificial neural networks, created without classifying the observations.

4. Based on the concentration of chlorophyll  $\alpha$  obtained with the RFR1s and RFR1tc models, short-term (several days, several weeks) and sudden increases in chlorophyll  $\alpha$  concentration were observed, which showed algal blooms of different intensity. The models can also capture the spatial distribution of chla concentration in lakes and ponds, which can be used to select a representative monitoring site.

5. During the 2015–2021 period, the status based on the average chla concentration did not change in 126 lakes. There was an improvement in the status in 155 lakes – the average chla concentration decreased, while in 76 lakes a deterioration of the status was observed due to an increase in the average chla concentration.

6. The application of RFR1s and RFR1tc models to satellite data allowed to determine the concentration of chla in 2015–2021 for all 357 state monitoring lakes and ponds, supplementing the *in situ* time series data with frequent data and collecting data on 43 lakes that were not monitored *in situ*. Based on average and maximum chla concentrations and lake type 52 non-problematic and 305 potentially problematic lakes were identified. The predominant Moderate and Chla biophysical classes were in all cases associated with potentially problematic lakes, but for the assessment of the problematic nature of Clear class lakes, the biophysical class is not a sufficient indicator and it is advised to rely on chla results.

## PUBLIKACIJŲ SĄRAŠAS

Pagrindiniai darbo rezultatai buvo publikuojami dviejuose straipsniuose Clarivate Analytics Web of Science duomenų bazėje ir konferencijų medžiagoje:

1. **Grendaitė D.** & Stonevičius E. Uncertainty of Atmospheric Correction Algorithms for Chlorophyll  $\alpha$  Concentration Retrieval in Lakes from Sentinel-2 Data. *Geocarto International* 37.23. 2022: 6867-6891. <https://doi.org/10.1080/10106049.2021.1958014>
2. **Grendaitė D.** & Stonevičius E. Machine Learning Algorithms for Biophysical Classification of Lithuanian Lakes Based on Remote Sensing Data. *Water* 2022, 14, 1732. <https://doi.org/10.3390/w14111732>

## UŽRAŠAMS

# UŽRAŠAMS

# UŽRAŠAMS

Vilniaus universiteto leidykla  
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius  
El. p. [info@leidykla.vu.lt](mailto:info@leidykla.vu.lt), [www.leidykla.vu.lt](http://www.leidykla.vu.lt)  
[bookshop.vu.lt](http://bookshop.vu.lt), [journals.vu.lt](http://journals.vu.lt)  
Tiražas 25 egz.