Vilnius University

WATER QUALITY PARAMETERS ASSESSMENT IN LITHUANIAN LAKES USING REMOTE SENSING AND MACHINE LEARNING

DALIA GRENDAITĖ
phD student
Supervisor: Edvinas Stonevičius
Institute of Geosciences, Vilnius University,
M. K. Čiurlionio 21/27, LT-03101 Vilnius, Lithuania
Email address: dalia.grendaite@chgf.vu.lt

State monitoring system

357 lakes and ponds

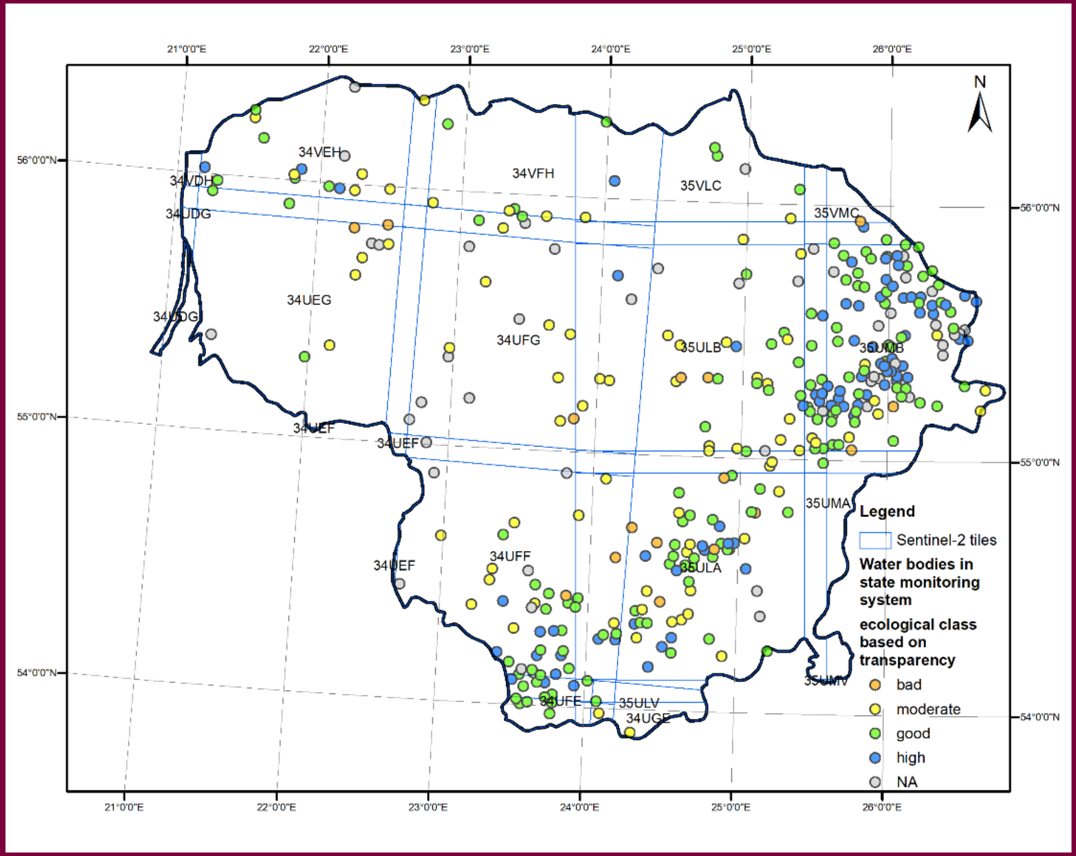
289 are natural, the rest – artificial or heavily modified

During years 2015-2020 308 lakes and ponds were monitored.

60% good/high class (based on transparency)

26% bad/moderate class

14% NA



Vilnius University

Algal blooms

Lake Jieznas

Lake Mastis

Vilnius
University

Remote sensing data

We need well performing **atmospheric correction algorithms** and

good **parameter retrieval algorithms**

to be able to use remote sensing data **more effectively and more reliably**.

Uncertainty of atmospheric correction algorithm selection for chlorophyll α retrieval

Vilnius University

Acolite and Acolite Rayleigh

Sen2Cor

iCOR

POLYMER

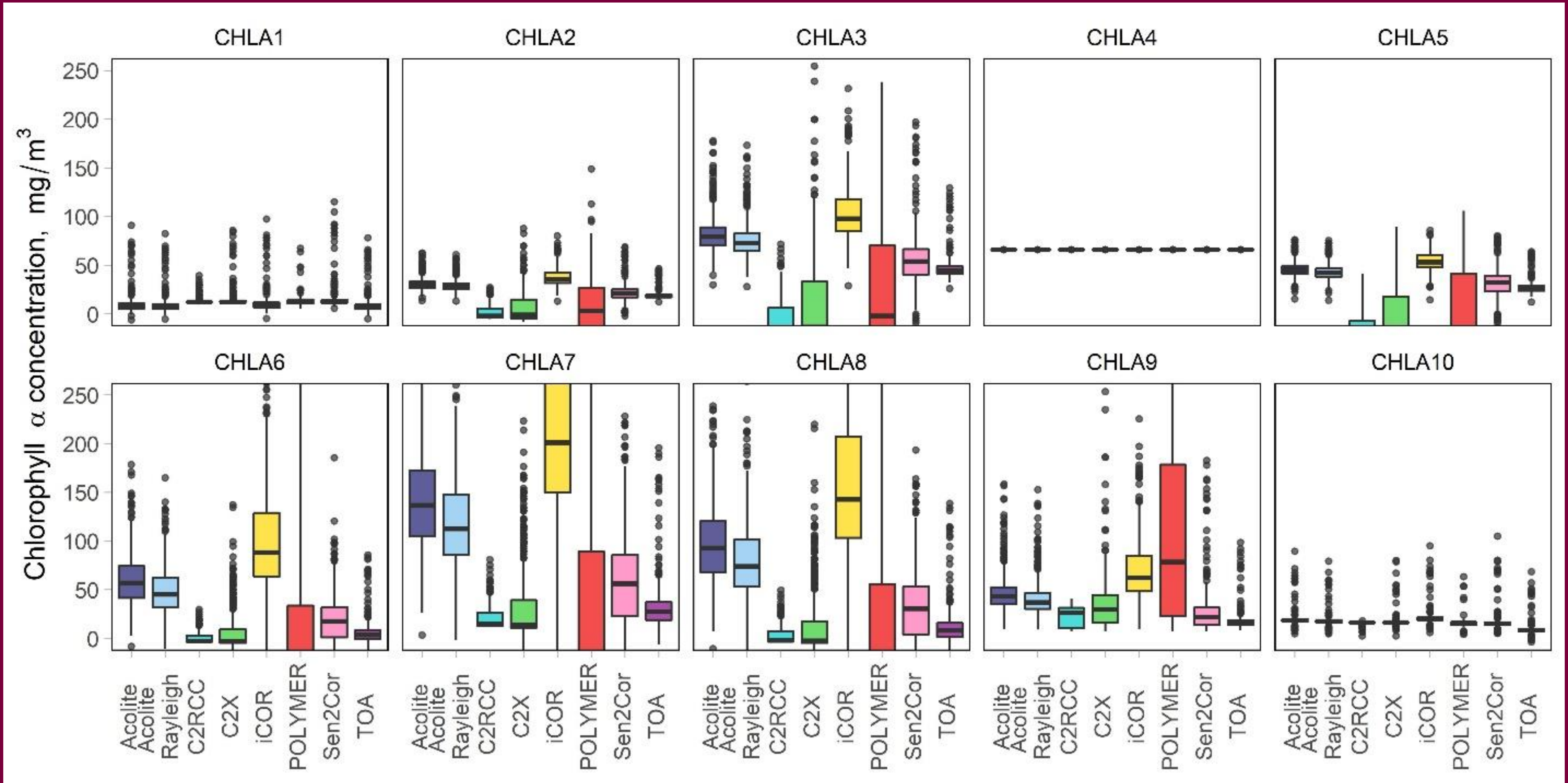
C2RCC C2X

Chlorophyll α algorithm code	Reference of empirical equation	Empirical equation
CHLA1	Toming et al., 2016	$-2231 * \left(R705 - \frac{R665 + R740}{2}\right) + 12.7$
CHLA2	Moses et al., 2009	$61.32 * \frac{R705}{R665} - 37.94$
CHLA3	Watanabe et al., 2019	$185.34 * \frac{R705}{R665} - 125.9$
CHLA4	Watanabe et al., 2019	$0.000001 * \frac{R705}{R665} \Big/ 705 - 665 + 66.038$
CHLA5	Soomets et al., 2020	$\frac{R665}{R705} * (-105.3) + 140.6$
CHLA6	Moses et al., 2009	$-232.29 * (R665^{-1} - R705^{-1}) * R740) + 23.174$
CHLA7	Watanabe et al., 2019	$474.69 * ((R665^{-1} - R705^{-1}) * R740) + 67.502$
CHLA8	Soomets et al., 2020	$-368.5 * \left(\frac{R740}{R705} - \frac{R740}{R665}\right) + 39.1$
CHLA9	Watanabe et al., 2019	$\left(\frac{R705 - R665}{R705 + R665}\right)^2 * (-1093.2) + 283.47 * \left(\frac{R705 - R665}{R705 + R665}\right) + 25.947$
CHLA10	Grendaitė, 2018	$2054(R705 - 1.05 * R665) + 17$

Grendaitė & Stonevičius, 2021

Chlorophyll α concentration retrieval

Vilnius University

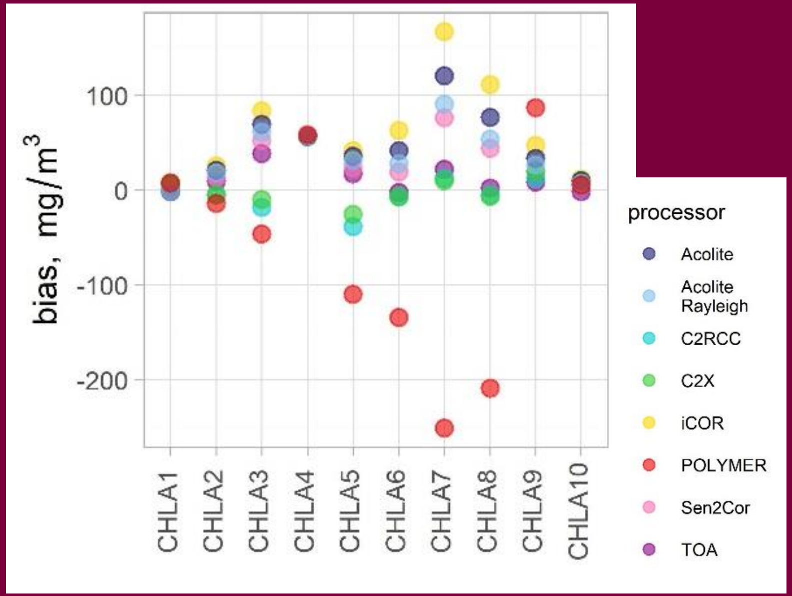


Grendaitė & Stonevičius, 2021

Chlorophyll α concentration retrieval – matchup analysis

30 matchup points

0±1 day difference between the in situ and the satellite acquisition

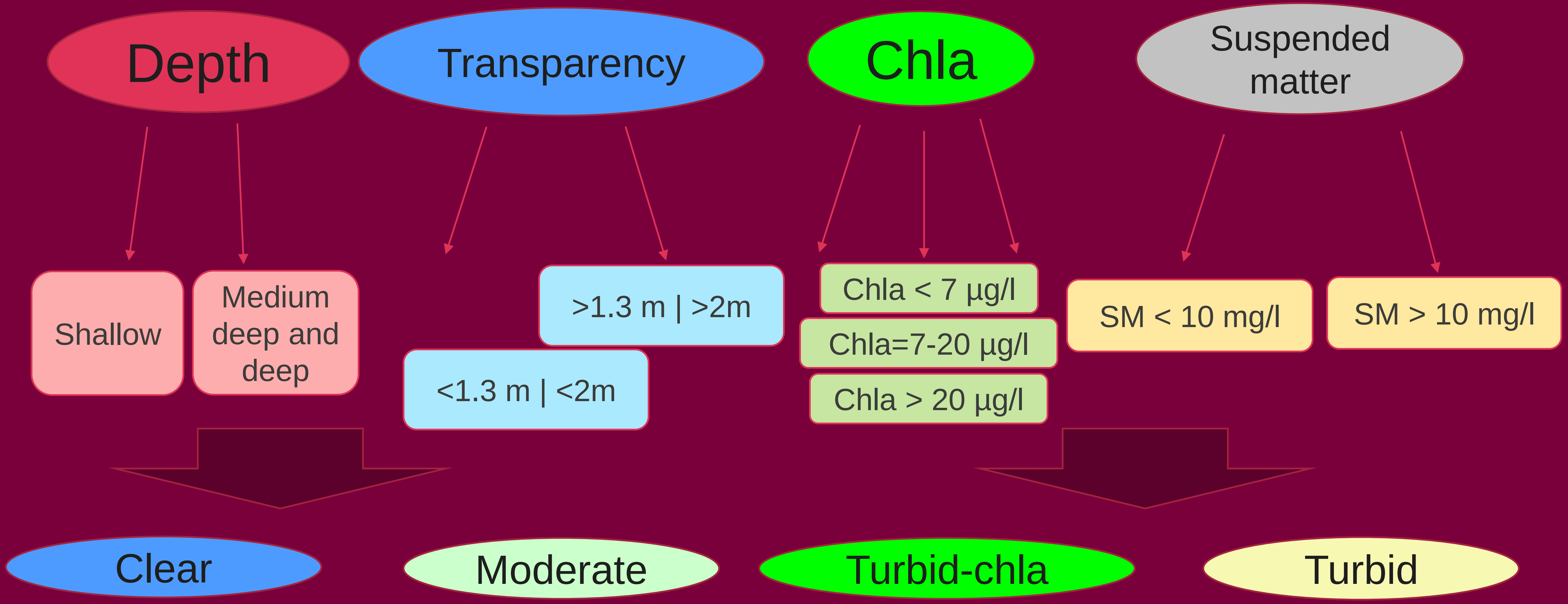


Grendaitė & Stonevičius, 2021

Inland water classification

Vilnius
University

Classification based on in situ measurements



Lake classification

Vilnius University

Two class problem

Class	Transparency, m	Chlorophyll α concentration, mg m^{-3}	Suspended matter concentration, g m^{-3}	Number of lakes	Number of cases
Clear	SD >= 1.3 m for shallow lakes SD >= 2 m for medium deep and deep lakes	Chla < 7.2	SM<10	119	336
Other	SD < 1.3 m for shallow lakes SD < 2 m for medium deep and deep lakes	any	any	149	375

Three class problem

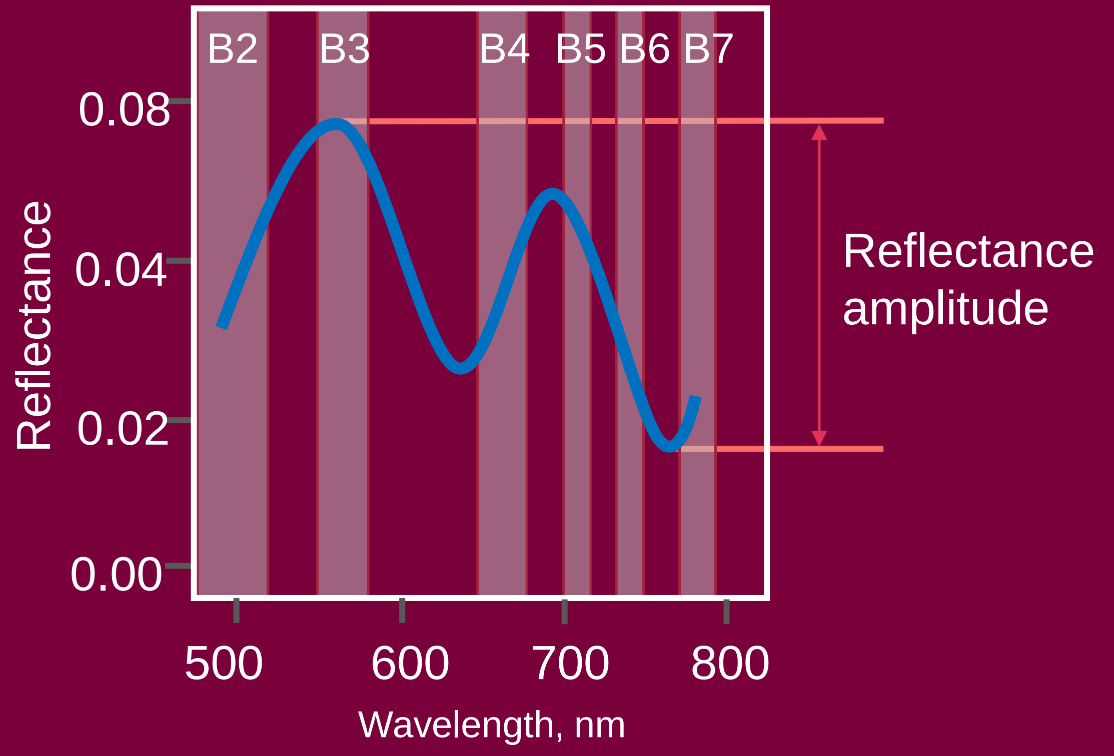
<table>
<tr><th>Class</th><th>Transparency, m</th><th>Chlorophyll α concentration, mg m^{-3}</th><th>Suspended matter concentration,</th><th>Number of lakes</th><th>Number of cases</th></tr>
<tr><td>Moderate</td><td>SD >= 1.3 m for shallow lakes
SD >= 2 m for medium deep and deep lakes</td><td>Chla= [7.2, 20]</td><td>SM<10</td><td>93</td><td>168</td></tr>
<tr><td>Turbid-chla</td><td rowspan="2">SD < 1.3 m for shallow lakes
SD < 2 m for medium deep and deep lakes</td><td>Chla>20</td><td rowspan="2">any</td><td>55</td><td>115</td></tr>
<tr><td>Turbid</td><td>Chla<20</td><td>51</td><td>92</td></tr>
</table>

A class retrieval from remote sensing data

Vilnius University

Features derived from remotely sensed spectra:

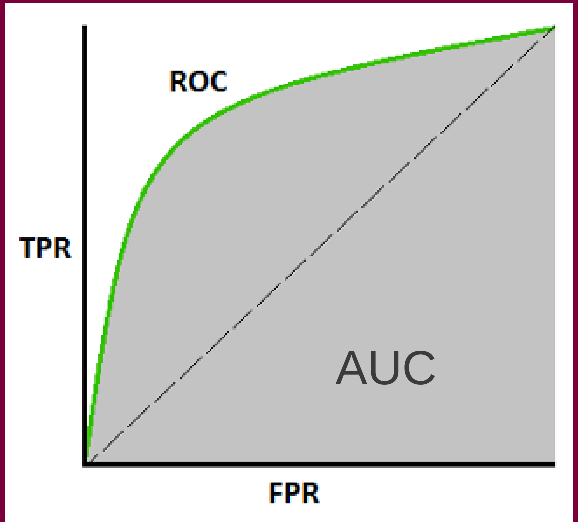
- reflectance amplitude,
- band ratios R705/R665, R560/R665, R560/R705,
- apparent visible wavelength (Vandermeulen et al., 2020),
- hue angle (van der Woerd et al., 2018) calculated based on visible 490-665 nm and red-edge 705 nm bands ,
- colour based on Forel-Ule colour scale as derived from hue angle (van der Woerd et al., 2018),
- month



Machine learning algorithms

Vilnius University

- Logistic regression (LR)
- Support vector machine (SVM)
- Ensemble methods:
  - AdaBoost (Ada)
  - XGBoost (XGB)
  - Random forest (RF)



Model performance metrics:

		Predicted label	
		0	1
True label	0	True negatives (TN)	False positives (FP)
	1	False negatives (FN)	True positives (TP)

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

$$recall = \frac{TN}{TN + FN}$$

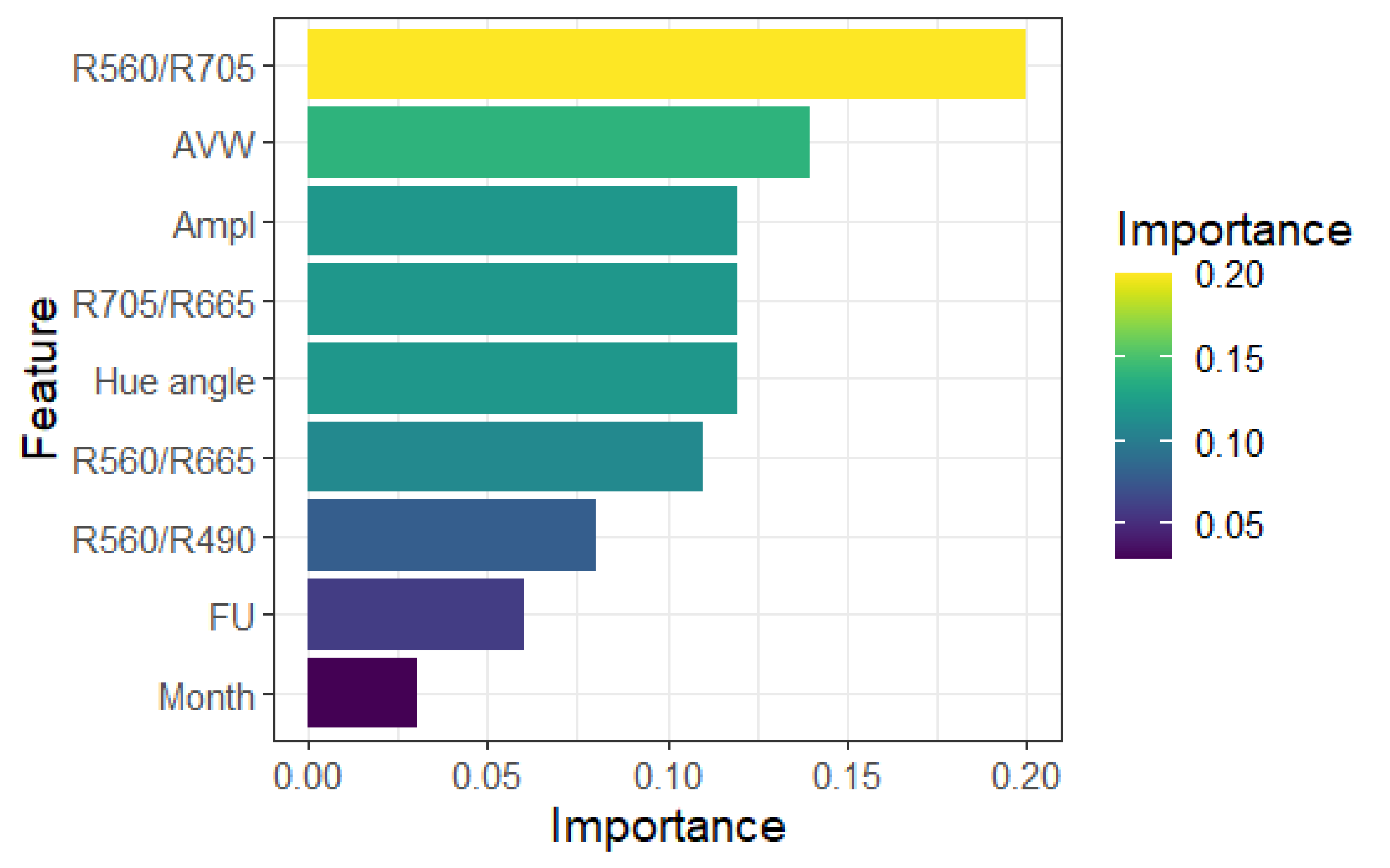
Two class problem

Vilnius University

Classifier	Hyper-parameter optimization	Validation accuracy	Validation AUC	Test accuracy	Test AUC
Logistic regression	No	0.8	0.86	0.79	0.8
	Yes	0.79	0.87	0.76	0.75
Support Vector Machine	No	0.79	0.87	0.76	0.77
	Yes	0.8	0.87	0.78	0.79
Random Forest	No	0.78	0.85	0.74	0.74
	Yes	0.78	0.85	0.78	0.78
AdaBoost	No	0.75	0.83	0.75	0.75
	Yes	0.76	0.83	0.73	0.73
XGBoost	No	0.78	0.85	0.78	0.78
	Yes	0.77	0.83	0.75	0.75

Important features (random forest)

Vilnius University



Feature importance as used by RF model

Vilnius
University

Conclusions

Large uncertainties come from atmospheric correction algorithm and parameter retrieval algorithm selection.

We try to create a data-driven approach to get a class of a lake based on its spectrum from remotely sensed data. This could be used in areas where in situ spectral data is not available.

Important features: R560/R705 ratio, apparant visible wavelength, amplitude.

Lake classes can be used to create water quality parameter retrieval models.

Vilnius University

Thank you for your attention!

CONTACTS

DALIA GRENDAITĖ

phD student

Supervisor: Edvinas Stonevičius

Institute of Geosciences, Vilnius University,
M. K. Čiurlionio 21/27, LT-03101 Vilnius, Lithuania

Email address: dalia.grendaite@chgf.vu.lt