BROCKMANN CONSULT

Cyanobacteria detection with Sentinel 2 MSI

How to use machine learning to detect cyanobacteria risks in inland waters

Jorrit Scholze, Kerstin Stelzer,
Carole Lebreton, Dagmar Müller
Brockmann Consult GmbH

ESA Living Planet Symposium 2022, Bonn, Germany

Importance of Cyanobacteria Detection

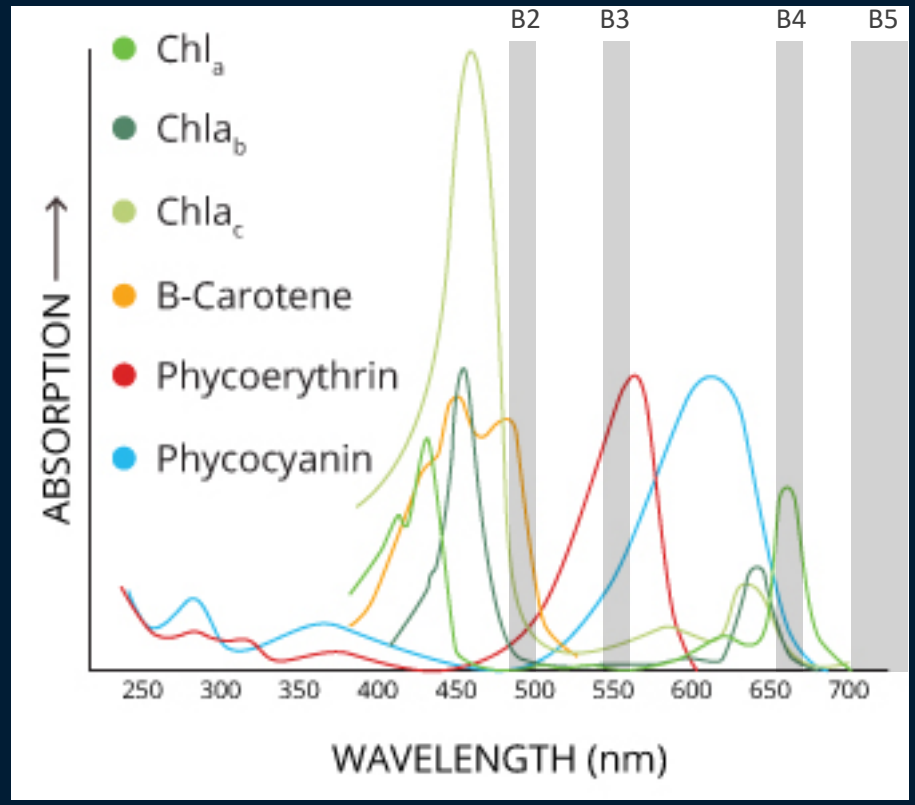
- Potentially harmful cyanobacteria
  - Affect freshwater and coastal ecosystems
  - Health impact
  - Relevant for recreational activities
- Assessing and monitoring water quality
  - Monitoring of lakes for local authorities
  - Potential indicator for climate change

BROCKMANN CONSULT

Earth Observation for Cyano Detection

- Various publications and algorithms address immersed and floating cyano detection[1]
  - Phycocyanin detection, cell counts, abundances
  - Based on water colour sensors with respective bands at absorption features
  - Spatial resolution of medium resolution water colour sensors not sufficient for small inland waters

- S2 MSI proven to be a valuable sensor for monitoring smaller inland waters bodies
  - cyano absorption features not covered with channels



[1] Kutser 2004,2009, Kahru et al. 2007, Simis et al. 2005, Alikas et al 2010, Matthews et al. 2012, Lunetta et al. 2014

Methodology

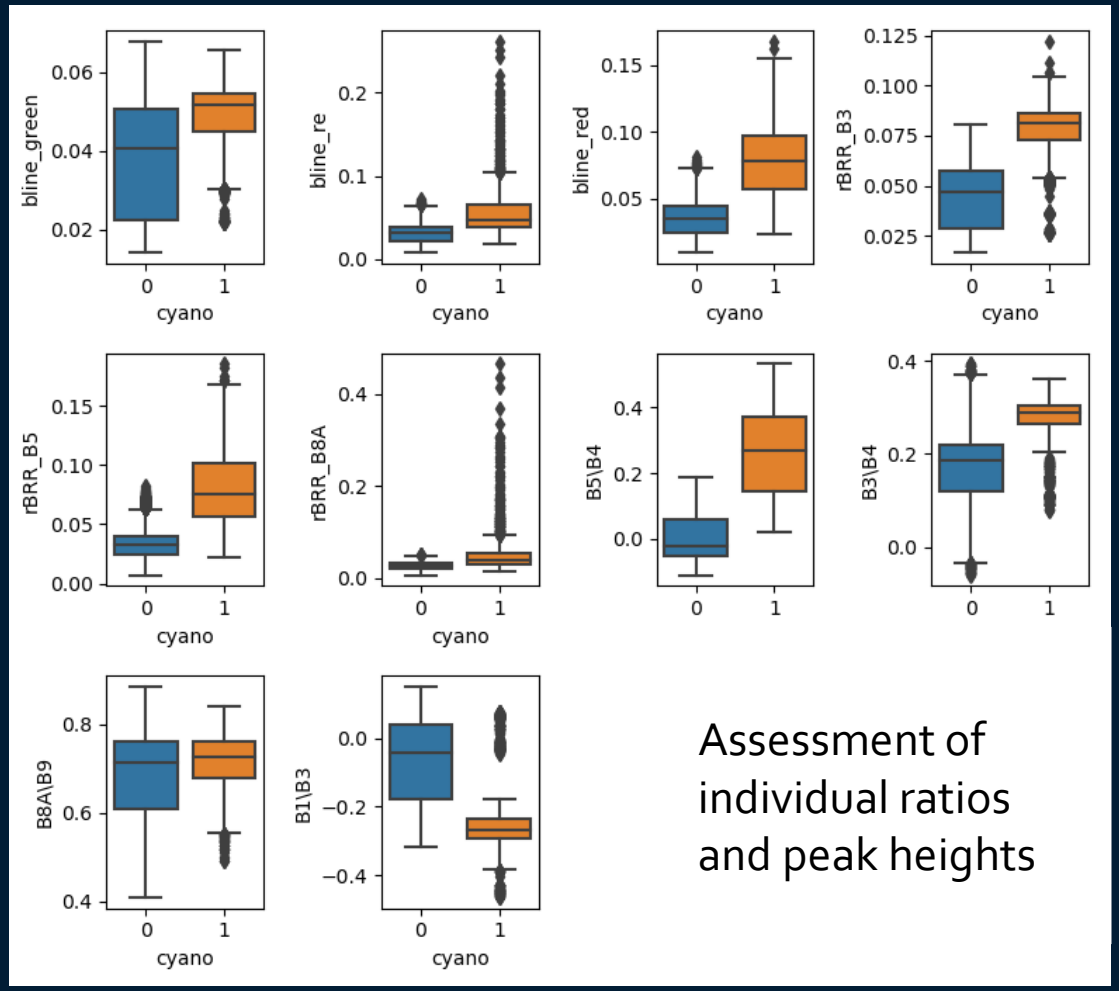
- Maximum Peak Height[1] (MPH) based for MERIS and OLCI
  - good results based on in-situ comparison
- Atmosphere
  - Rayleigh correction → BRR
- In-Water
  - S2 spectral band at 620nm is missing → MPH not applicable
  - Our approach:

Using a Random Forest Model to detect distinct spectral features of cyano blooms with S2.

[1] Matthews, M. W., & Odermatt, D. (2015).

Selection of Model Inputs

- manually selected dataset covering various water types:
  - Cyano blooms
  - high chlorophyll biomass blooms, springblooms
  - clear water cases
- Cyanobacteria blooms defined from in-situ data:
  - Biovolume of Cyano > 50% of total biovolume and a minimum CHL-a concentration of 10 µg/L
- Dataset taken in lakes in Germany and the US
- Analysis of indices and peak heights to distinguish cyanos from green algae blooms

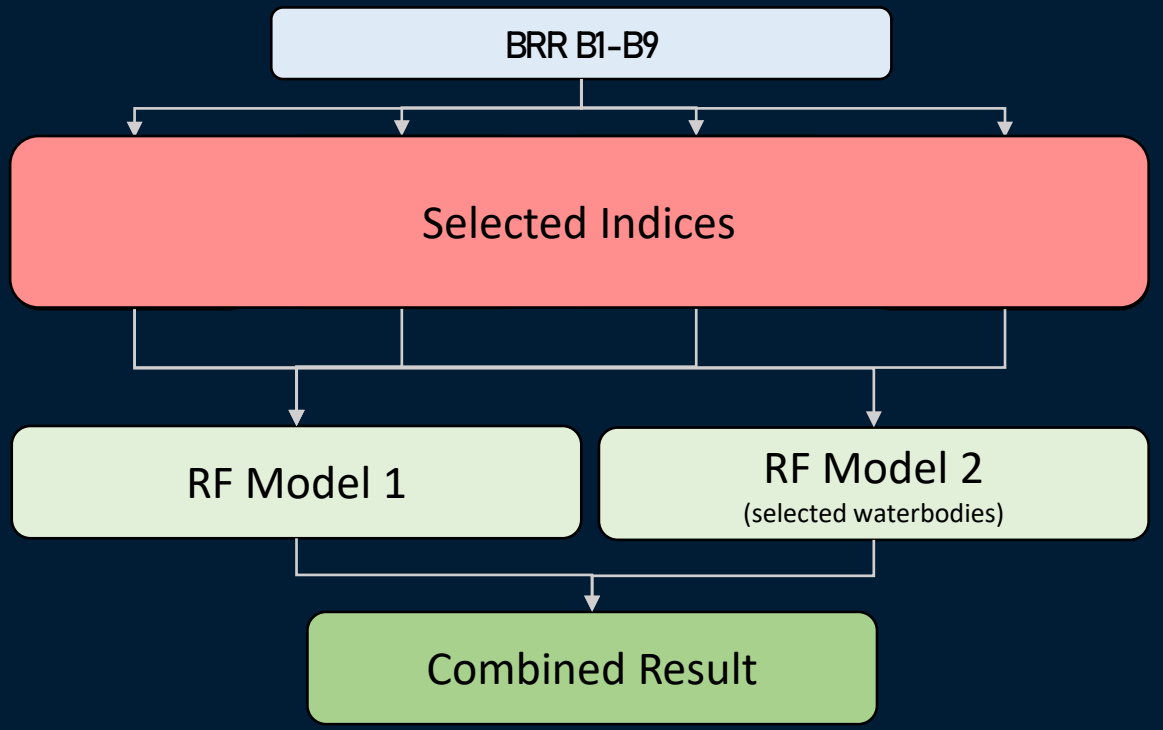


Assessment of individual ratios and peak heights

BROCKMANN CONSULT

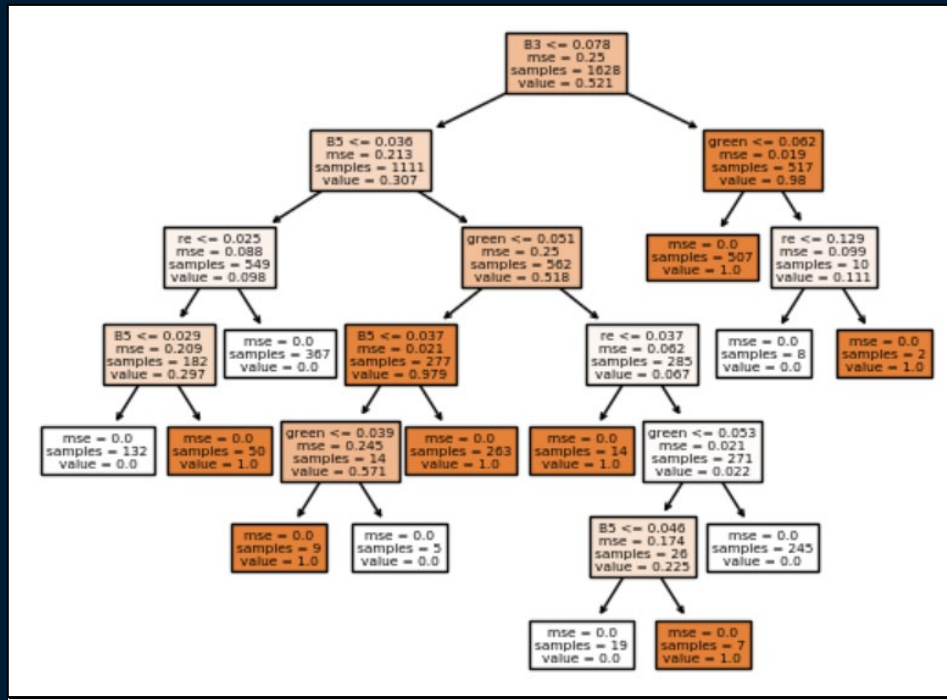
Model Configurations

- Selection of 4 indices based on BRR
- Identification of two Models
  - Model 1: covering various waterbodies
  - Model 2: covering cyano abundances in darker brown lakes
  - Combination of both model results



Random Forest

- Supervised machine learning algorithm
- Multiple decision trees, with each tree representing a class prediction
- Wisdom-of-crowds concept
  - ➢ class with the most votes becomes the model's prediction
- Easy and fast model training
- Full dataset split
  - 70% training data
  - 30% test data



Number of trees	Maximum depth
400	13

Testing of Model

- Test dataset showed very high accuracy
  - OAA > 95%
- Based on manually collected water pixels
  - Biased approach
- Unbiased validation necessary
  - Extensive in-situ dataset from German authorities

Model 1 — RF CB

True CB	Cyano	No Cyano	Sum	U A
Cyano	507	8	515	98.45
No Cyano	2	513	515	99.61
Sum	509	521		
P A	99.61	98.46		OAA: 99.03

Model 2 — RF CB

True CB	Cyano	No Cyano	Sum	U A
Cyano	36	4	40	90.00
No Cyano	2	191	193	98.96
Sum	38	195		
P A	94.74	97.95		OAA: 97.42

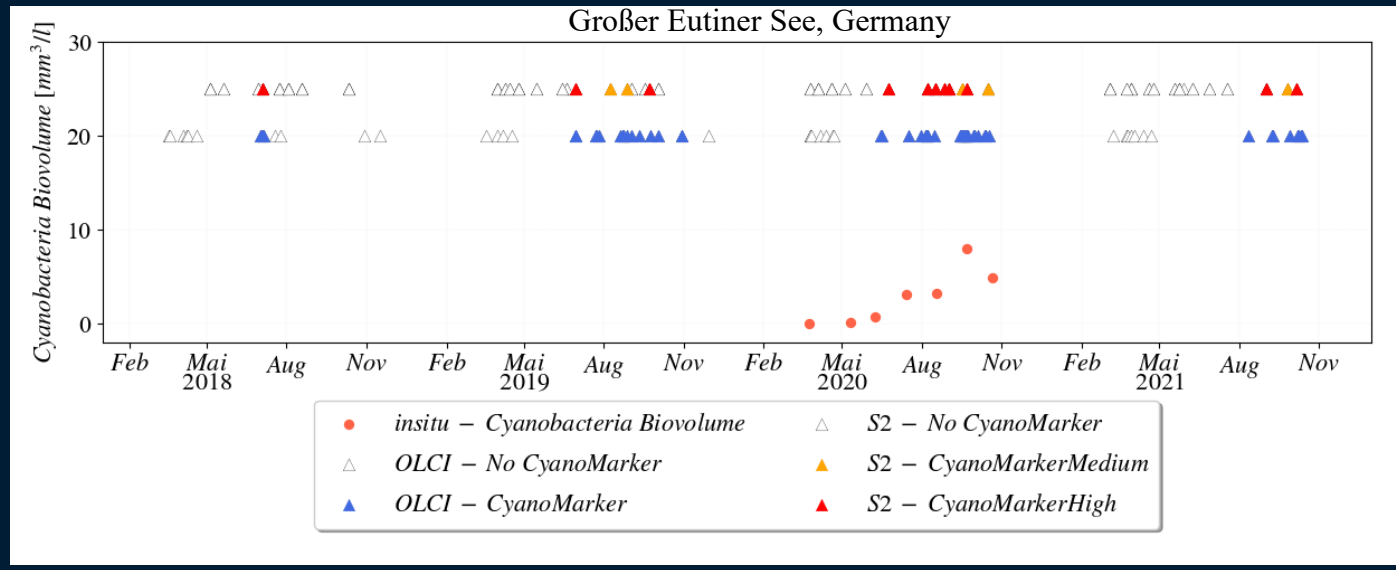
BROCKMANN CONSULT

Validation of Model

- Validation of model based on:
  - In-situ measurements
    - Cyano Biovolume > 50% of total Biovolume
    - CHL-a > 10µg/L
    - Same day
  - OLCI MPH CyanoMarker
    - Time series of German lakes
    - Check periodic cyano abundance
- High confidence in cyano risk assessment with random forest model

Germany — RF CB

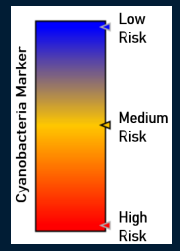
In-situ CB \ Class	Cyano	No Cyano	Sum	U A
Cyano	15	2	17	88.24
No Cyano	2	30	32	93.75
Sum	17	32	49	
P A	88.24	93.75		OAA: 91.84

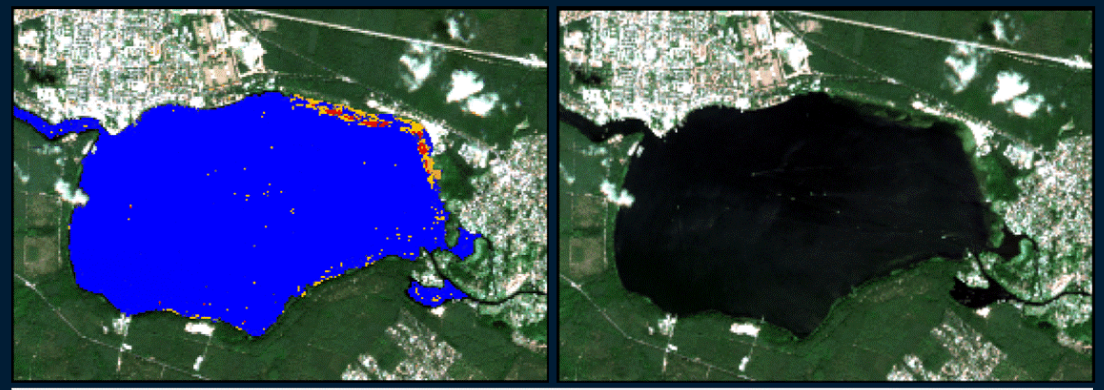


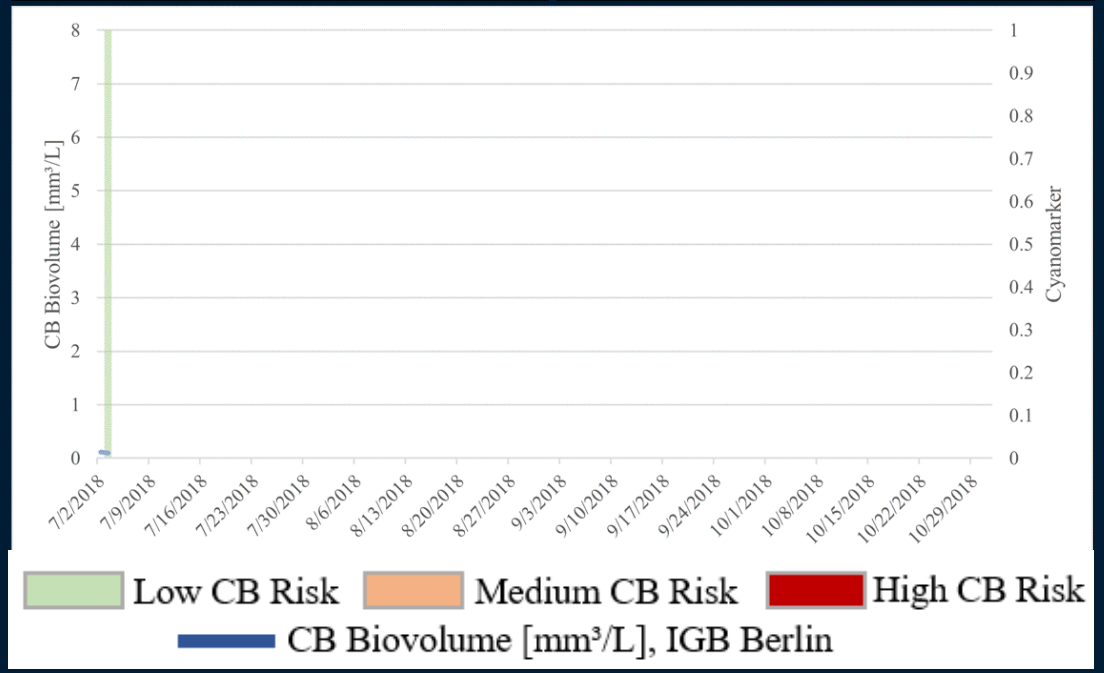
BROCKMANN CONSULT

Case Study – Müggelsee

- Müggelsee (Berlin, GER)
  - Frequent cyano abundance
  - Monthly in-situ measurements
- S2 MSI cyano risk identifies cyano bloom in time and space
  - Spatial information cyano bloom crucial for public bathing places
  - Static in-situ measurements station not covering full spread of cyano bloom







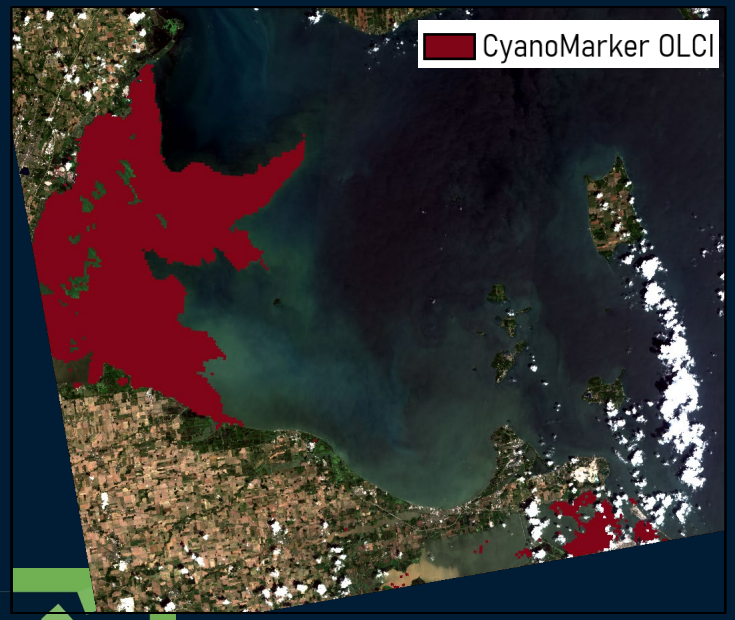
BROCKMANN CONSULT

Validation with the CyanoTRACKER

- Using public monitoring programs to validate model
  - CyanoTRACKER (University of Georgia)

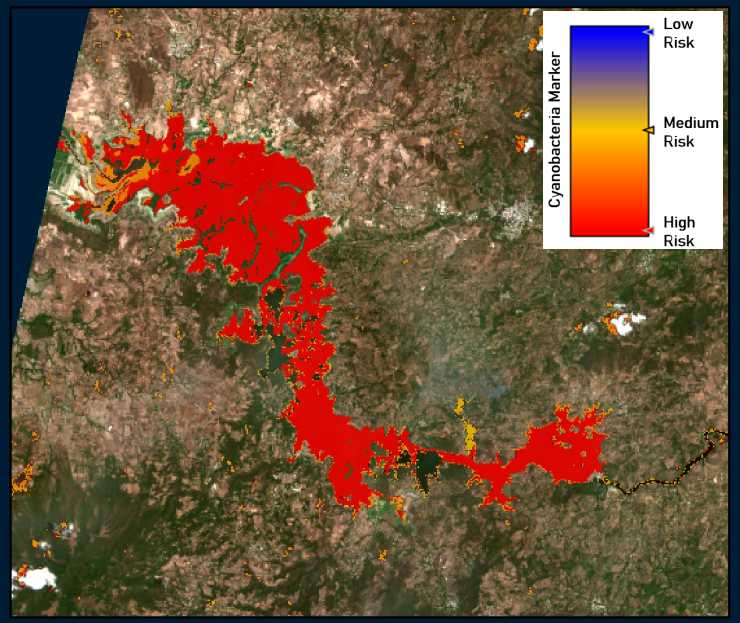
Lake Erie - USA

24.07.2019



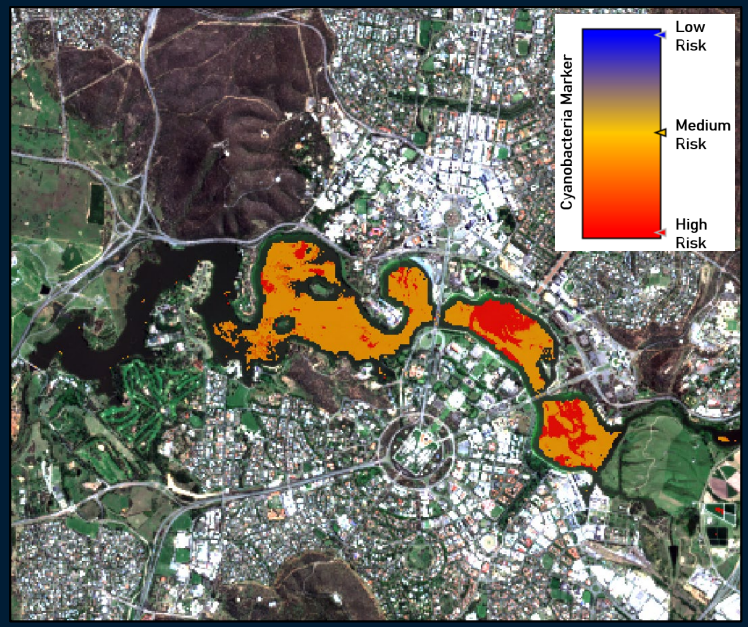
Embalse Cerrón Grande - El Salvador

22.04.2022



Lake Burley Griffin - Australia

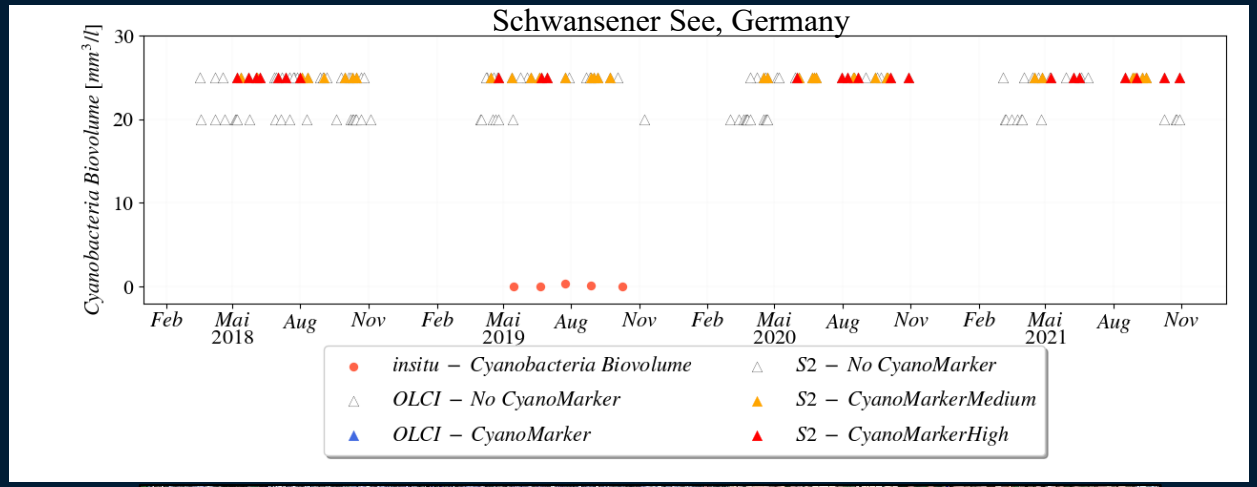
28.02.2021

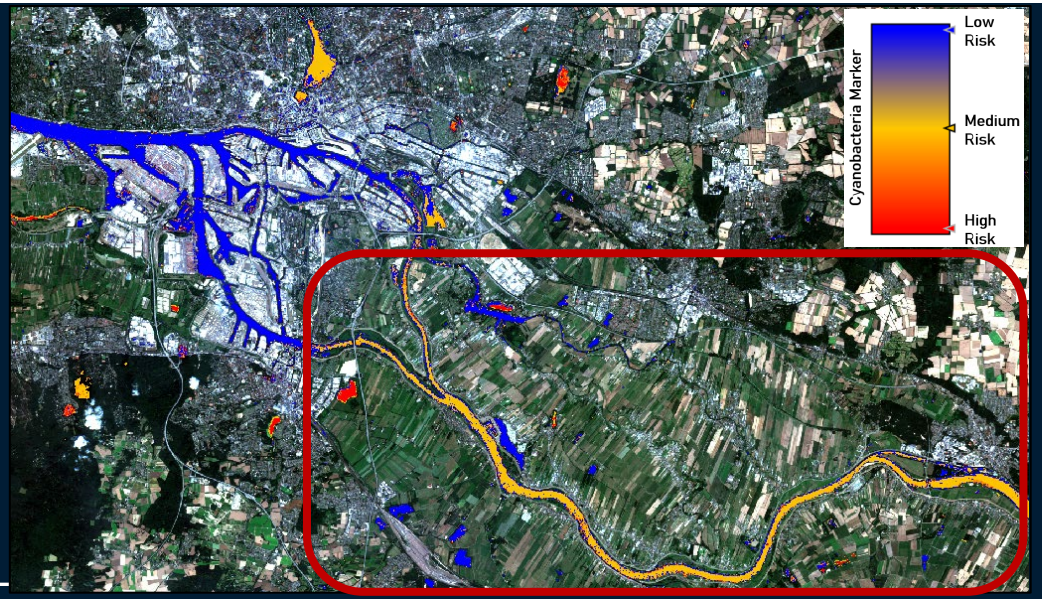


BROCKMANN CONSULT

Limitation encounters

- Limitations for S2 MSI cyanobacteria risk assessment
  - High algae blooms with CHL-a concentrations > 100µg/L
  - Shallow waters/bottom reflection
  - Rivers with high CHL-a concentrations
- Risk assessment
  - No cyano biovolume
  - Low biomass cyano blooms undetected





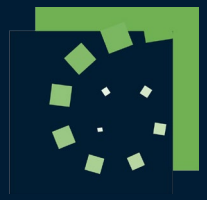
Improvements & Conclusion

Improvements

- Continuation of improving the model
  - Extension of training dataset
  - Extension with a third model
- Limitations for S2 MSI cyanobacteria detection
  - Phycocyanin absorptions wavelength 620nm not covered with S2 MSI
  - Upcoming high-resolution sensors may include bands covering these wavelengths

Conclusion

- Random Forest Model approach based on selected indices and a manually selected training dataset
- High confidence in cyanobacteria risk detection based on S2 for various lakes



Thank you!

Brockmann Consult