The Earth Observation National Eutrophication Monitoring Program: Phase II

FINAL REPORT

A Report to the
Water Research Commission

by

Mark Matthews and Jeremy Kravitz
CyanoLakes (Pty) Ltd.

WRC Report No. 3081/1/23
ISBN 978-0-6392-0481-9

July 2023

Report obtainable from
Water Research Commission
Lynnwood Bridge Office Park
Bloukrans Building, 2nd Floor
4 Daventry Street
Lynnwood
PRETORIA

hendrickm@wrc.org.za or download from www.wrc.org.za

This is the final report of WRC project no. C2019/2020-00198.

EXECUTIVE SUMMARY

BACKGROUND

Phase II of the Earth Observation National Eutrophication Monitoring Programme (EONEMP II) has developed complimentary mobile technologies to the web application developed in the previous EONEMP project (WRC Project K5/2418). It has seen the development of a highly innovative, novel and commercially viable weather-like mobile application providing information on cyanobacteria blooms and water pollution, with high-definition maps and a forecasting capability based on phenology. Given the risk cyanobacteria bloom pose to public health, the project outcomes are essential to South Africa's sustainable development agenda, and the constitutional obligation to ensure citizens and recreational water users are kept safe from the significant health impacts that can occur as a result of water contamination and pollution. The project will contribute towards building capacity in an area of scarce skills through the awarding of scholarships and internship opportunities to previously disadvantaged students. The mobile application will consist of free and paid versions, and be commercialised according to the IPR Act and WRC policies to generate sustainable revenue and jobs.

AIMS

The following were the aims of the project:

1. To develop a mobile application providing current and forecasted information on cyanobacteria blooms and water pollution in lakes and reservoirs.
2. To implement a method for providing weekly forecasts for cyanobacteria blooms based on remotely sensed phenology to facilitate early warning through the mobile application.
3. To integrate medium- and high resolution maps showing cyanobacteria blooms and high resolution value added layers from Sentinel-3 and Sentinel-2 satellite instruments into the mobile application.
4. To promote uptake of the mobile application with community interest groups, industry and government.
5. To build human capacity in computer science and programming through providing internships and scholarships to previously disadvantaged students.
6. To generate sustainable revenue and employment opportunities through a commercial mobile app solution.

METHODOLOGY

The mobile app was developed using the latest mobile technology (React Native) and released on the Apple App and Google Play Stores. The mobile app relies on the data processor and application programming interfaces (API) of the CyanoLakes Web App. High resolution imagery was integrated into the App using the Sentinel-hub cloud satellite imagery API. The App uses a freemium pricing model, with a free basic app with no advertising or login, with monthly and annual subscription to access premium features with payments processed on the Apple or Google stores. A pricing strategy was determined by comparing with available weather and health apps.

New high-resolution satellite imagery products were developed for Sentinel-2 MSI imagery and implemented using the Sentinel-Hub cloud imagery API. New algorithms were developed for land / water separation, a bloom index, and chlorophyll-a from cyanobacteria blooms. A novel algorithm for chlorophyll-a was trained using a large synthetic dataset enabling estimation between concentrations of 1 and 500 μg/L for cyanobacteria-dominant water.

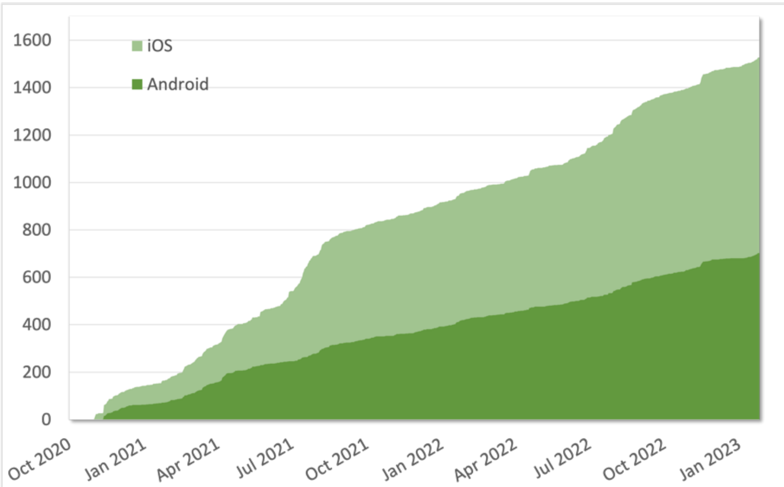
The forecasting model was developed based on simple univariate forecasting methods and logical decomposition with timeseries of spatially aggregated estimates of chlorophyll-a from cyanobacteria and algae from the Ocean and Land from a small test set of 15 waterbodies. Various univariate baseline models were

tested to develop a robust model. The moving average seasonal error adjusted (MASEA) model was developed using a combination of a moving average and seasonal naïve approaches adjusted for seasonal anomalies. Forecasts were implemented for 1-week, 2-week and 4-week time horizons, deemed the most useful for management purposes.

RESULTS AND DISCUSSION

The CyanoLakes mobile App ("Your Weather App for Lakes") is a novel mobile application that provides summary weather-like information on cyanobacteria health risk levels and eutrophication in lakes around the world, based on satellite remote sensing data. The app features a home screen, list screen for managing waterbodies, the ability to add and remove lakes interactively, and premium features including high resolution satellite image products, notifications sent in near real-time for various risk thresholds that can be set by the user, and10-day ahead forecasts. It also includes a FAQ page, as well as terms of service and a privacy policy as required by the Apple and Google Stores. The monthly premium subscription is priced at US$4.99/m ($9.99 for developed countries), with a discounted annual subscription priced at 4 months cost. The app is available free on iOS and Android devices.

The mobile app was promoted at several conferences including the Water Research Foundation (WRF) LIFT Webinar Series (May 2020), the North American Lakes Management Society (NALMS) 12th National Monitoring Conference (April 2021), World Water Tech North America (October 2021), GEOWeek 2021 (November 2021), a dedicated Launch Event hosted by the WRC (August 2021), the Sydney Water Innovation Festival (2022), and the NALMS conference (November 2022). The App was also promoted on the CyanoLakes social media channels (Twitter, Facebook), via the CyanoLakes mailing list (350+ subscribers), and on Google advertising. The App was also made available to a group of pilot users in the utility sector from Rand Water, the City of Cape Town and Umgeni Water. As of end January 2023 the App has 825 installations on Apple, and 702 on Google (total installs of 1527). Most users are South Africa (628 installs, 41%), Sweden (192 installs, 13%), USA (239 installs, 16%), Canada (124 installs, 8%) and the Australia (49 installs, 3%).



Internships during the project provided scholarships and training to two 3rd year computer science and engineering students, providing them with references for the workplace. During the project, there were more than 100 premium subscriptions (monthly or annually) across the Apple App Store and Google Play Store. This is far below the subscription number needed for break even costs. It would need 250 monthly paying subscribers to meet break even costs, however the current ratio of free to pay users is only 1.8%, meaning that more than 13 000 users are required to meet annual operating costs. It is unclear whether this number of users is achievable. This means that in the short term, the app will be subsidised by CyanoLakes from other revenue streams to continue to sustain the app into the future.

High resolution products are available for a premium subscription through the mobile App using the Sentinel-hub cloud API. Based on the synthetic dataset, a novel algorithm for chlorophyll-a was developed, tailored for cyanobacteria blooms, using the normalised difference chlorophyll-a index that had a mean average percentage error of 42% over a range of chlorophyll-a from 1 to 500 μg/L. This produces realistic chlorophyll-a scenes that compared favourably with algorithms used for the Ocean and Land Colour Instrument medium resolution data. A novel bloom index taking advantage of a switching between the 705 nm peak and 740 nm peak was also provided - a novel index for bloom and floating vegetation identification. Products available through the mobile App include chlorophyll-a, bloom index, enhanced vegetation, true colour, and the normalised difference vegetation index.

The simple univariate MASEA model could forecast high risk cyanobacteria blooms with an accuracy of 80% for a 1-week time horizon, which reduce to 71% and 69% for 2-week and 4-week time horizons. This was roughly equivalent to results from more complex models integrating meteorological and other data sources. The 1-week horizon forecast ability for cyanobacteria risk level and trophic state had a 74% and 75% accuracy, respectively. For the 4-week forecast horizon, the forecasting ability drops off to 66% for CRL and 71% for TS, respectively. Median whole-lake chlorophyll-a was forecast with an estimated accuracy of 42% over a 1-week time horizon, which may meet the minimum requirements for applications such as ecological observation, but it is unclear whether this would meet the requirements for other applications such as water treatment or issuing of recreational advisories. We found that the forecast performance was weakly influenced by the size of the waterbody, lending weight to the use of spatially aggregated data even for large lakes. There was a significant correlation between the length of the timeseries and the forecast error, with longer timeseries leading to reduced errors owing to improved characterisation of the seasonal signal.

An open API for web browsers was developed that provides free access to statistical information via the internet via API queries using authentication tokens. This enables users to access and integrate the information into downstream applications, much like how weather information is accessed today.

CONCLUSIONS

The CyanoLakes mobile App is a significant technological innovation as the only app available for iOS and Android with daily updates for cyanobacteria risk levels from lakes worldwide, with integrated high-resolution imagery from Sentinel-2 for paid subscribers. The mobile App has more than 1500 installs across iOS and Android devices, with the most users in South Africa, followed by US, Sweden, Canada, and Europe. The novel forecasting method allows high risk cyanobacteria blooms to be forecast with 80% accuracy for a 1-week time horizon, with additional forecasts at 2-weekly and 4-weekly time horizons for trophic status and chlorophyll-a, with forecasts available via the App. The project has contributed towards capacity building by providing valuable internship experience for two students with revenue and uptake expected to grow significantly.

RECOMMENDATIONS

The mobile app should be promoted through paid advertising, social media, and conferences to increase user uptake, especially in the US, Europe, and Australia to increase the number of users, and especially paying users to ensure financial sustainability. The app should be maintained and improved after the project. Future development of the App should include features for premium users with advanced functionality such as charts, historical time series, weather, and advanced forecasting models. Alternative revenue streams should be explored to increase financial viability. The forecast model should be refined, further developed, and applied to 1000 or more lakes worldwide to increase robustness. Additional remote sensing data streams from additional sensors should be considered for integration in the mid to long term (3+ years).

ACKNOWLEDGEMENTS

The project team wishes to thank the following people for their contributions to the project.

Reference Group	Affiliation
Prof Abel Ramoelo	University of Pretoria
Dr Annelie Swanepoel	Rand Water
Prof Mohamed A.M. Abd Elbasit	Sol Plaatje University
Dr Michael Silberbauer	Private Consultant
Ms Vadhini Moodley	University of KwaZulu-Natal
Dr Ngetar Njoya	University of KwaZulu-Natal
Dr Sabelo Madonsela	CSIR
Ms Tovhowani Nyamande	Department of Water Sanitation
Dr Vusi Ntema	Umgeni Water
Others	
Dr Eunice Ubomba-Jaswa	WRC (Research Manager)
Ms Penny Jaca	WRC (Project Administrator)
Ms Sandra Meyer	CyanoLakes
Mr Jeremy Kravitz	University of Cape Town & CyanoLakes

This page was intentionally left blank

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS & ABBREVIATIONS

API	Application Programming Interface
App	The CyanoLakes Mobile App
Chl-a	Chlorophyll-a
CRL	Cyanobacteria Risk Level
EONEMP	Earth Observation National Eutrophication Monitoring Programme
GEO	Group on Earth Observations
MASEA	Moving average seasonal error adjusted
MSI	Multi Spectral Imager
NDCI	Normalised difference chlorophyll index
NDVI	Normalised difference vegetation index
NPL	Nutrient Pollution Level
OLCI	Ocean and Land Colour Instrument

GLOSSARY

App. A software application accessed through a webpage or a mobile device.

Application programming interface. In this report, means a webpage through which a software program accesses data produced by another software program.

Chlorophyll-a. The pigment found in all green plants and algae

Cyanobacteria. A toxin-producing bacterium present in water forming green blooms that are harmful to human and animal health.

Eutrophication. The process whereby waterbodies become enriched with nutrients and thereby display elevated concentrations of plants, algae and bacteria (including cyanobacteria).

Phenology. The study of cyclic and seasonal natural phenomena, especially in relation to climate and plant and animal life.

Remote sensing. Using a sensor (usually in space) to measure the light spectrum or colour of a target (in this case, water).

CHAPTER 1: BACKGROUND

1.1 INTRODUCTION

This is the second progress report for the Earth Observation National Eutrophication Monitoring Programme Phase II. This project will focus on the development of complimentary mobile technologies to the web application developed in the previous EONEMP project (K5/2418). It will see the development of a highly innovative, novel and commercially viable weather-like mobile application providing information on cyanobacteria blooms and water pollution, with high-definition maps and a forecasting capability based on phenology. Given the risk cyanobacteria bloom pose to public health, the project outcomes are essential to South Africa's sustainable development agenda, and the constitutional obligation to ensure citizens and recreational water users are kept safe from the significant health impacts that can occur as a result of water contamination and pollution. The project will contribute towards building capacity in an area of scarce skills through the awarding of scholarships and internship opportunities to previously disadvantaged students. The mobile application will consist of free and paid versions, and be commercialised according to the IPR Act and WRC policies to generate sustainable revenue and jobs.

1.2 PROJECT AIMS

The aims of this project are as follows:

1. To develop a mobile application providing current and forecasted information on cyanobacteria blooms and water pollution in lakes and reservoirs
2. To implement a method for providing weekly forecasts for cyanobacteria blooms based on remotely sensed phenology to facilitate early warning through the mobile application
3. To integrate medium and high resolution maps showing cyanobacteria blooms and high resolution value added layers from Sentinel-3 and Sentinel-2 satellite instruments into the mobile application
4. To promote uptake of the mobile application with community interest groups, industry and government
5. To build human capacity in computer science and programming through providing internships and scholarships to previously disadvantaged students (Appendix A)
6. To generate sustainable revenue and employment opportunities through a commercial mobile app solution

1.3 SCOPE AND LIMITATIONS

The project will involve the development of a novel mobile application for monitoring the risk to human health from cyanobacteria blooms. The desired functionality for the application includes high resolution imagery as well as a forecasting capability which must be developed in order to fulfil the scope of the application. The project does not however involve the development of a forecasting model that takes account of all variables (e.g. weather, landscape, reservoir size, etc.) but rather will test the idea of employing an approach based on remote sensing phenology (seasonal variation) data that can be used and applied globally. The algorithms developed will not be validated by field work, since this was the topic of previous research projects (e.g. EONEMP).

CHAPTER 2: MOBILE APPLICATION

2.1 INTRODUCTION

2.1.1 Overview

The CyanoLakes mobile application ("Your Weather App for Lakes") (the App) is a novel mobile application aiming to provide weather-like information on cyanobacteria and algal blooms for lakes around the globe from satellite imagery. The App mimics a weather-like application – to display potentially complex information in an intuitive, easy to understand, minimalistic way.

The App displays information through four main views:

1. home – intuitive statistical information for each waterbody
2. list – a summary displaying all of the user's waterbodies
3. add a lake – functionality for adding waterbodies
4. satellite imagery – a premium features providing high resolution satellite imagery

The basic App consists of a free version with the home, list and add a lake features. A premium version provides satellite imagery and notifications for a monthly or yearly subscription. The App is the first of its kind to provide information related to cyanobacteria blooms for the worlds lakes for satellite imagery.

2.1.2 Technology choice

The App was developed using React Native developed by Facebook (https://facebook.github.io/react-native/). React Native is mature technology with many libraries available on the Internet. This enabled the App to be built for both the iOS app store and Android devices. The App code base is hosted in a private repository on Github. A record of the Github development log is shown in Appendix D.

2.1.3 Functionality and Logic

The app has the following basic functionality:

- Each user must accept the Disclaimer and Terms of Use
- Page providing instructions how to use the application
- Display the most recent cyanobacteria risk level and pollution level, from low to very high, as well as a number of other statistics derived from satellite imagery for the users waterbodies
- Add a lake either from the existing database, or a new waterbody by dropping a pin on a map
- Add and remove waterbodies as needed from the list view
- Subscribe or login

The following additional functionality is available to paid users:

- Interactive high resolution satellite imagery products (maps)
- Notifications sent to the users device for updates or various risk levels

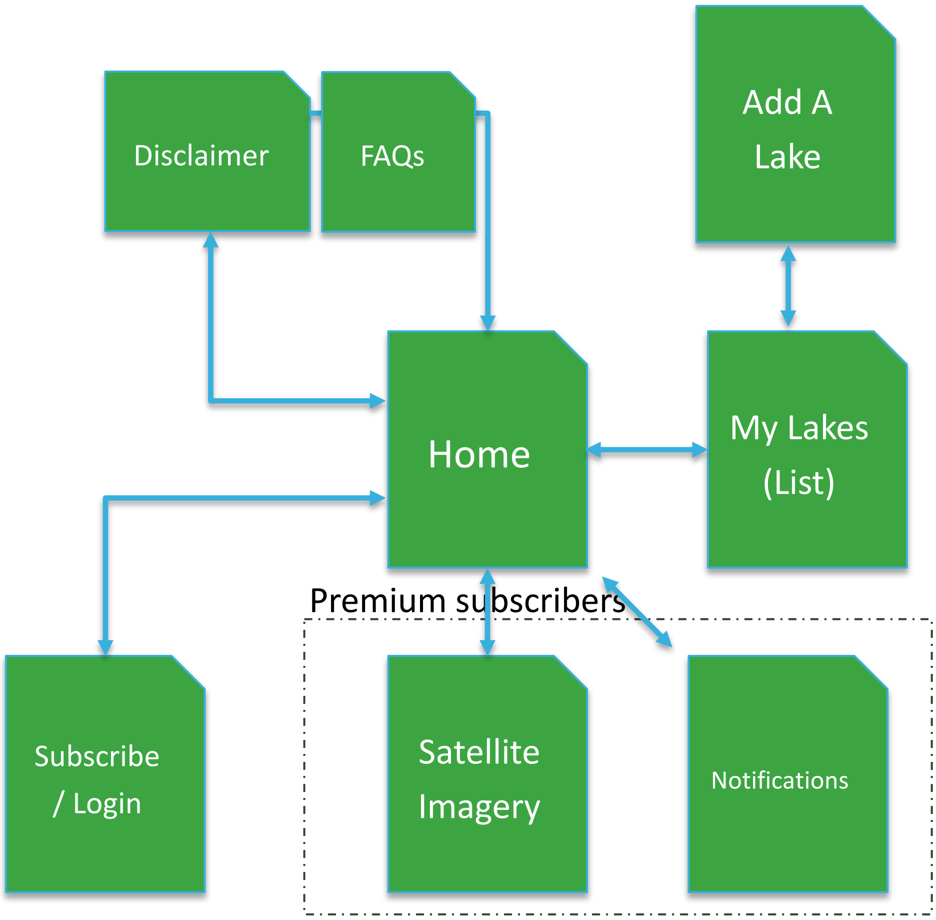


Figure 2.1 Application Logic

The App workflow logic for the application is shown in Figure 2.1.

1. User accepts the Disclaimer and Terms of Use (see Appendix B).
2. User lands on home page
    2.1. User clicks on "Please add a lake" on home screen
        2.1.1. User adds lakes on the Add a Lake Page
        2.1.2. User returns to Home Page
    2.2. User clicks on "Help me get started"
        2.2.1. User reads instructions on FAQs page
3. User views a list of their lakes (app remembers if returning – stores list of lakes already added)
    3.1. User selects or adds dams by the Add a Lake page
    3.2. User adds or removes lakes on the list page
4. User clicks on maps or settings page
    4.1. User subscribes or logs in with their organizational account
        4.1.1. User views the latest satellite imagery on maps view
            4.1.1.1. User views various satellite products using the dropdown menu
        4.1.2. User views settings page
            4.1.2.1. User specifies which notifications they want to receive using button selections

2.1.4 Architecture

A complex architecture is required to operate the App (Figure 2.2). The CyanoLakes Web Application provides various restful Application Programming Interfaces (APIS) that provide the information to the Mobile App through post and get requests. The APIs were built to provide the information to the mobile App. The Django Web App provides the functionality and database for storing the information served by the Mobile App. In addition a separate application for processing medium resolution satellite imagery is required to provide the statistical information from satellite (note: imagery at Medium resolution is not implemented in the App). The Sentinel-hub cloud imagery serves the high resolution imagery though a web mapping service. The Mobile App relies upon the ongoing maintenance of the web Application.

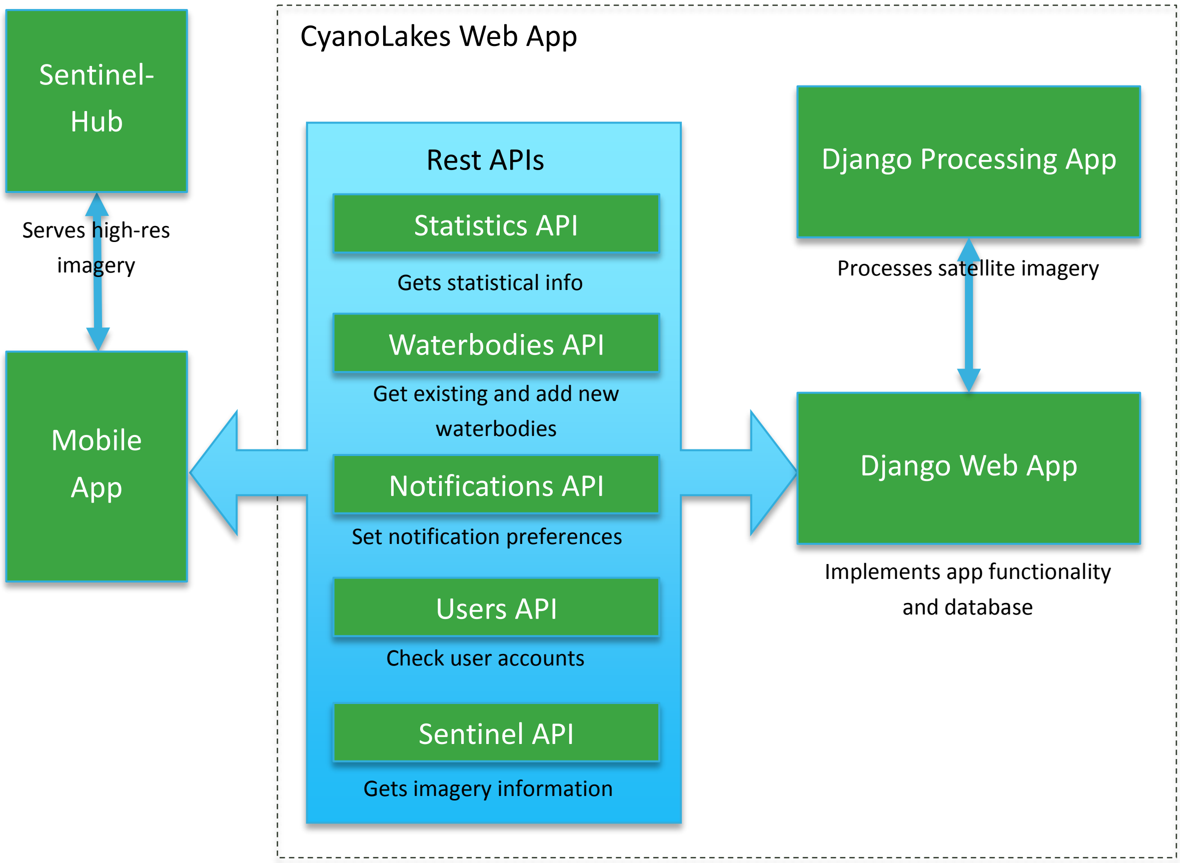


Figure 2.2 Supporting architecture for the mobile app

2.2 FEATURES

2.2.1 Disclaimer and Start Screen

The App is listed in the Apple and Google Stores under the name CyanoLakes. After downloading, the user must accept the Disclaimer and Terms of Use (see Appendix B) before opening the App. The User is then directed to the Start Page in order to Add a Lake.

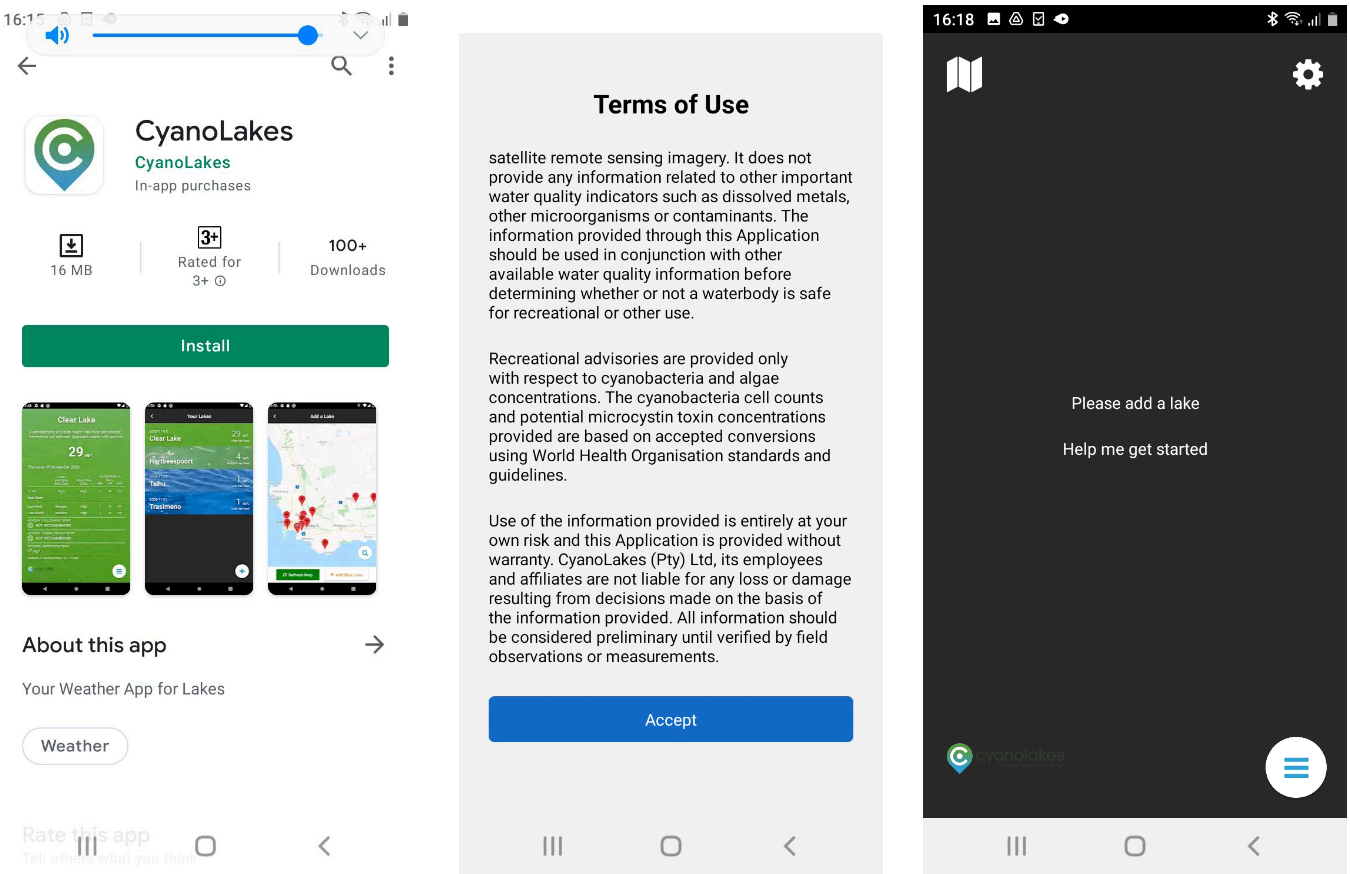


Figure 2.3 Store listing, terms of use and start page (from left).

2.2.2 Home Screen

2.2.2.1 Dynamic background and text

The App employs a dynamic background and text description to allow an intuitive interpretation of the information (Figure 2.4 and Figure 2.5). A dynamic background graphic is used to indicate the current cyanobacteria risk level (CRL) and nutrient pollution level (NPL) (Table 2.1.). In order to display the background graphic, an image library representing three images of each possible variation in water conditions was assembled. These images were sourced from free online pictures or alternative pictures taken on field work excursions. The text description was also determined based on the CRL and NPL (Table 2.1.), as well as on the persistence and coverage of cyanobacteria blooms. The text provides a recreational advisory according the CRL and NPL. The text is visible in Table 2.1.

Figure 2.4 The home page of the CyanoLakes Application showing an example from Taihu Lake.

2.2.2.2 Statistical variables

The following statistical information is pulled from the CyanoLakes Web Application API:

- The current nutrient pollution level / trophic status (low, moderate, high, very high)
- The current cyanobacteria risk level (low, moderate, high, very high)
- The current chlorophyll-a concentration and cell count for cyanobacteria
- The information for last week and last month, calculated as averages
- A recreational advisory for full and partial contact water sports
- Other variables shown in Table 2.2

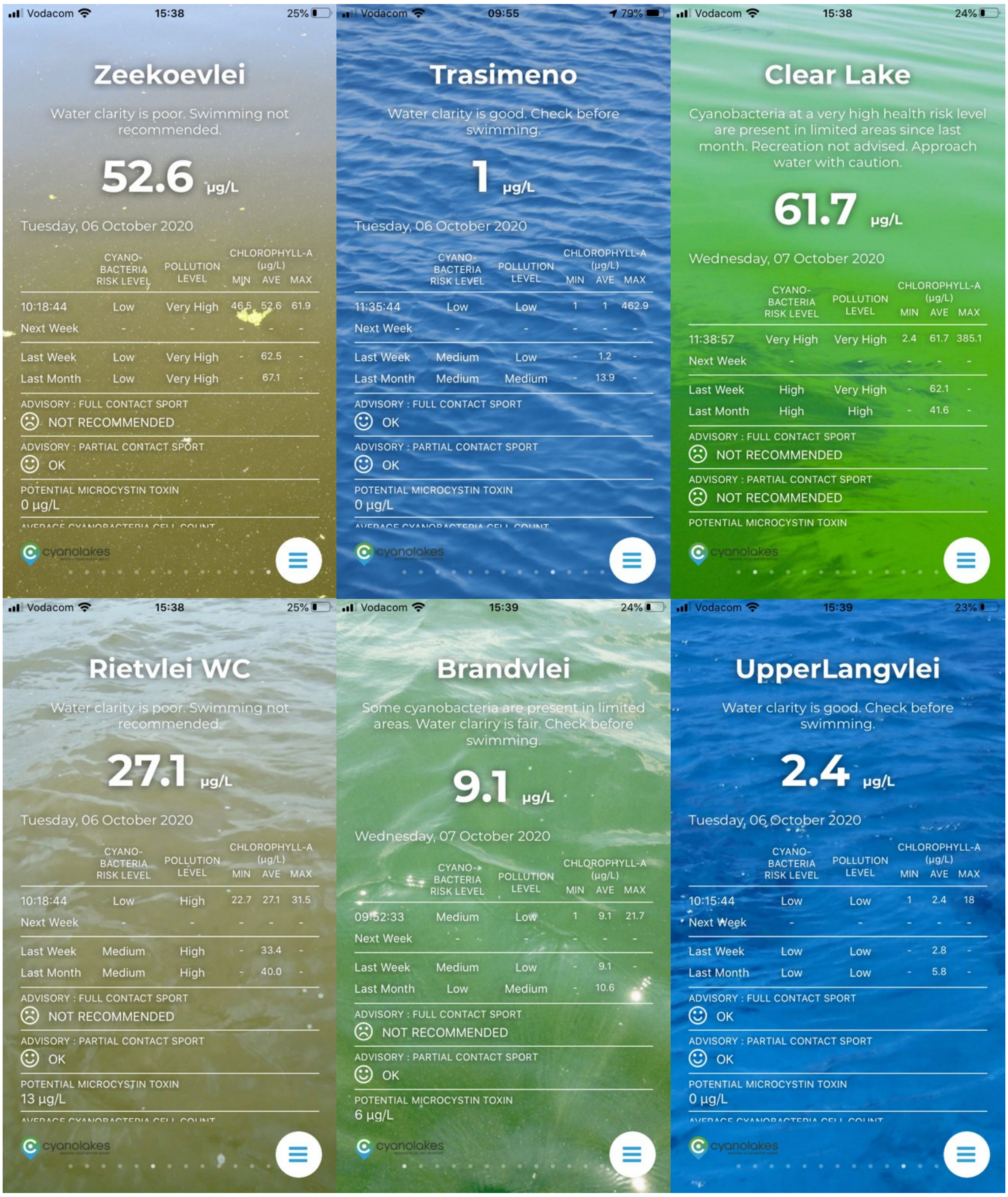


Figure 2.5 Home page showing variable background and text depending upon the CRL and NPR levels.

Table 2.1 Key to text and image mapping on home page

CRL	NPL	Descriptive text	Image prefix
Very high	Any	Cyanobacteria at a very high health risk level are present * &. Recreation not advised. Approach water with caution.	vhigh
High	Any	Cyanobacteria at a high health risk level are present * &. Recreation not advised. Approach water with caution.	high
Medium	High or Very High	Some cyanobacteria are present *. Water clarity is poor. Swimming not recommended. &	medium_high
Medium	Medium or Low	Some cyanobacteria are present *. Water clarity is fair. Check before swimming. &	medium_low
Low	High or Very High	Water clarity is poor. Swimming not recommended. &	low_high
Low	Medium	Water clarity is fair. Check before swimming. &	low_medium
Low	Low	Water clarity is good. Check before swimming. &	low_low

Percentage coverage of high or very high risk cyanobacteria blooms (denoted by *)

Value	Text
0-5%	in limited areas
5-50%	in some areas
50%+	throughout

Persistence of cyanobacteria blooms (denoted by &)

	CRL	Low or Medium	High or very high
last month	high or very high	Use caution: cyanobacteria at high risk level detected during the last month.	since last month
last week	high or very high	Use caution: cyanobacteria at high risk level detected within the last week.	since last week

Table 2.2 Variables displayed on the home page

Name	units	database variable
Cyanobacteria risk level		cyanobacteria_risk_level
Pollution level		trophic_status
Chlorophyll-a low	μg/L	chla_min
Chlorophyll-a average	μg/L	chla_med
Chlorophyll-a high	μg/L	chla_max
Advisory: partial contact sports		partial_contact
Advisory: full contact sports		full_contact
Potential microcystin toxin	μg/L	microcystin_toxin
Cyanobacteria cell count	cells/mL	cyanobacteria_cell_count
Cyanobacteria coverage	%	area_cyano_percentage
Cyanobacteria scum coverage	%	area_scum_percentage
Average cyanobacteria chl-a	μg/L	chla_cyano
Average algae chl-a	μg/L	chla_noncyano
Vegetation coverage	%	area_veg_percentage
St. Dev. chl-a	μg/L	chla_std

2.2.3 My Lakes List Screen

The list page has the following features (Figure 2.6):

- A dynamic graphic showing an image representing the water conditions to indicate the current status for each water body in the list (colour image of water)
- The current chlorophyll-a value
- The current cyanobacteria risk level (low, moderate, high, very high) for each water body
- A plus button to add water bodies to your list

The list is sortable, and users can delete a waterbody by swiping.

Figure 2.6 List page showing waterbodies selected by the user

2.2.4 Add a Lake Screen

The add a lake page allows the user to select from the existing database of lakes or add a new waterbody that is not in the database (Figure 2.7).

2.2.4.1 Adding an existing lake

The lakes are represented by pins shown on the world map. When selecting a pin, the user has the option to add the lake to their account. The user can also search for a lake in the database using the search function.

2.2.4.2 Adding a new lake

The user can also add a new lake by dropping a pin on the map. The location of the pin is retrieved and a request sent to the Open Street Map Overpass API (https://wiki.openstreetmap.org/wiki/Overpass_API) to retrieve a water related feature. If successful the geometry of the waterbody is added to the Web Application, and the lake processed. If the criterion (shown in Table 2.3) are not met, an error message is raised. Waterbodies that are too small or too large are filtered out (based on medium resolution minimum size of 600 x 600 m).

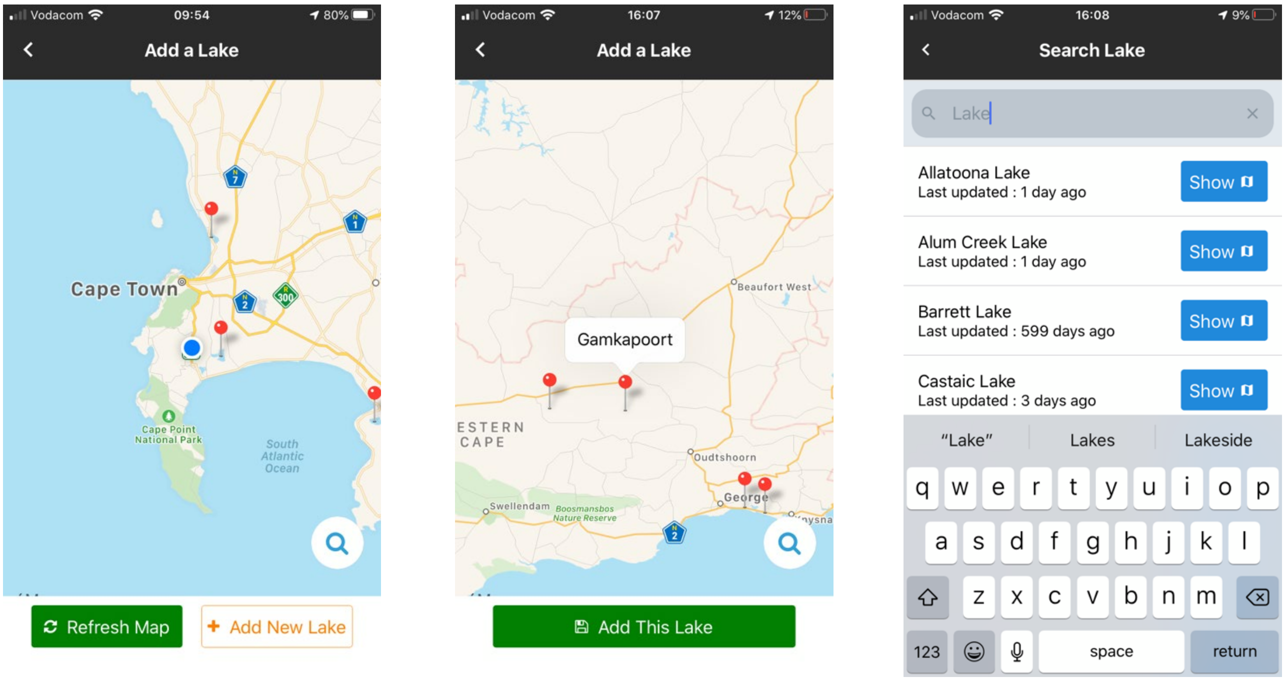


Figure 2.7 Add a lake page

Table 2.3 Overpass API criterion and error messages

Criterion	Value	Error Message
Minimum size threshold	0.81 km^2	This lake is too small, please select another one.
Maximum size threshold	10 000 km^2	This lake is too large, please select another one.
Minimum width threshold	600 m	This lake is too narrow, please select another one
Small waterbody warning	2 km^2	WARNING: The waterbody you have chosen is small. Data frequency may be reduced.
Waterbody intersects existing one		This waterbody already exists!
No waterbody		Your pin is not over a waterbody!

2.2.5 Subscription Offering Screen

The subscription offering view (Figure 2.8) enables the user to upgrade to a premium subscription in order to unlock premium features. The page has the following functionality:

- Provide a description of premium features highlighting the advantages of subscribing
- Monthly subscription option (button)
- Annual subscription option (button)
- Link to frequently asked questions
- Login functionality for organisational account users
- Functionality to restore purchases made previously
- Links to the terms of use and privacy policy

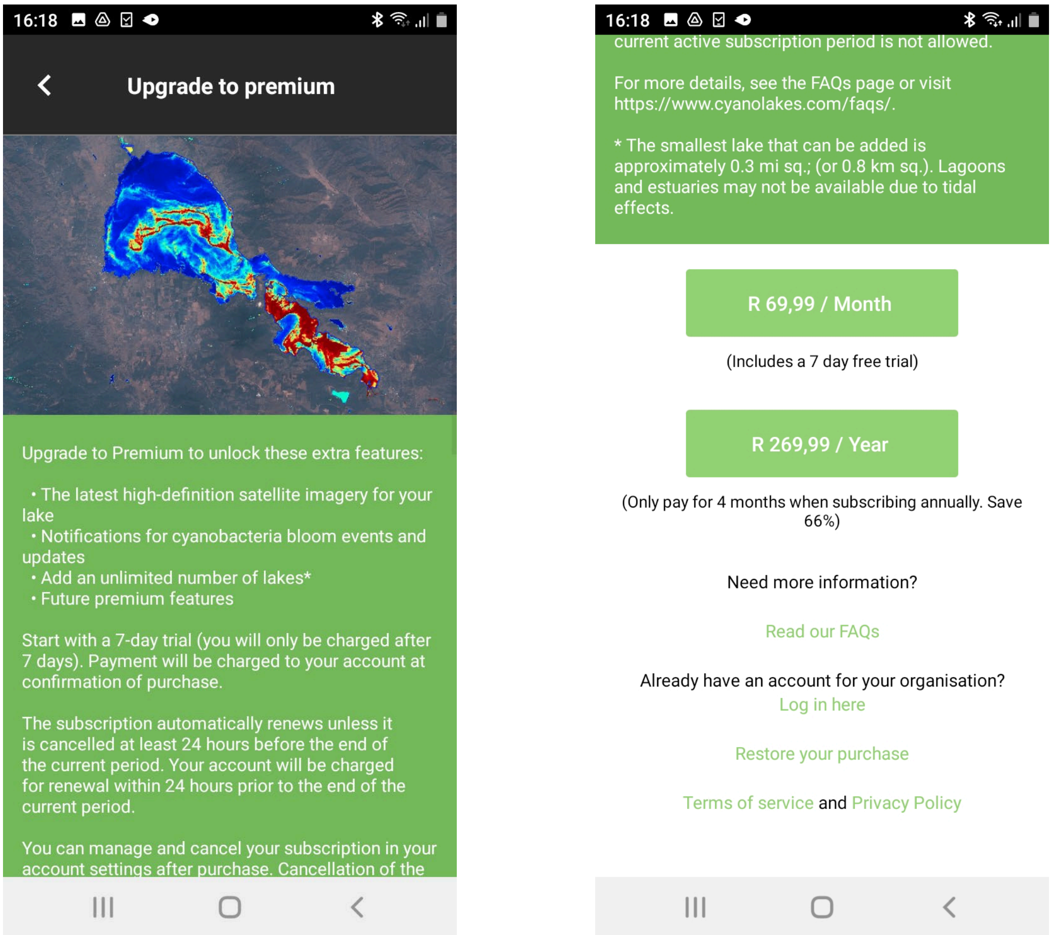


Figure 2.8 Subscription offering view (upgrade view)

2.2.6 Satellite Imagery Screen

The satellite image view displays the custom value-added high resolution satellite imagery from Sentinel-2 from the Sentinel-hub web mapping service:

- True colour – what the human eye would perceive (default)
- Chlorophyll-a – including a scale ranging from 1 to 1000 μg/L
- Bloom index – index for algae and cyanobacteria blooms
- Normalised difference vegetation index – NDVI – for visualising vegetation
- Enhanced vegetation – for visualising vegetation

The maps are visualised by selecting the layer from a dropdown menu at the bottom of the page.

-

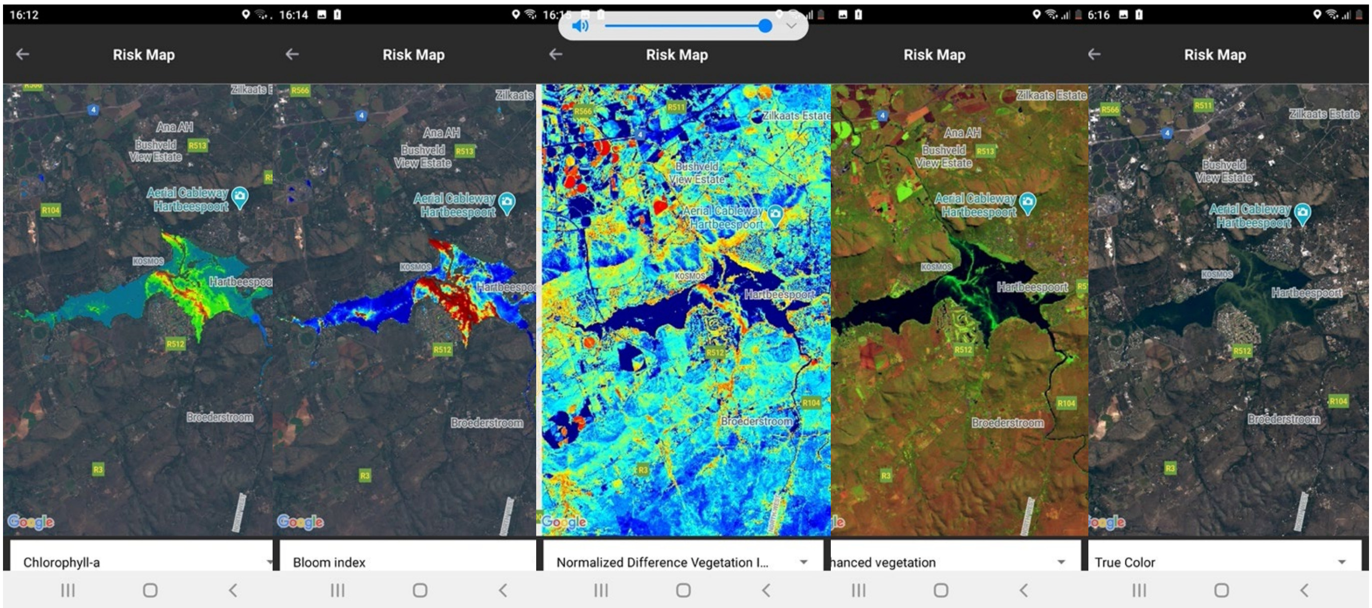


Figure 2.9 High resolution image layers displayed in the app

2.2.7 Settings Screen

The settings view allows the user to manage the notifications they receive through the App (Figure 2.9). The following notification options are available:

- High or very high risk events – the user only receives alerts for high or very high risk events
- Medium risk events – the user receives updates for medium risk events for early warning
- All updates – the user receives all updates

The settings view also has a link to FAQs page and to the terms of use.

2.2.8 Frequently Asked Questions Screen

The frequently asked questions content (Figure 2.10) in the App is provided in Appendix C.

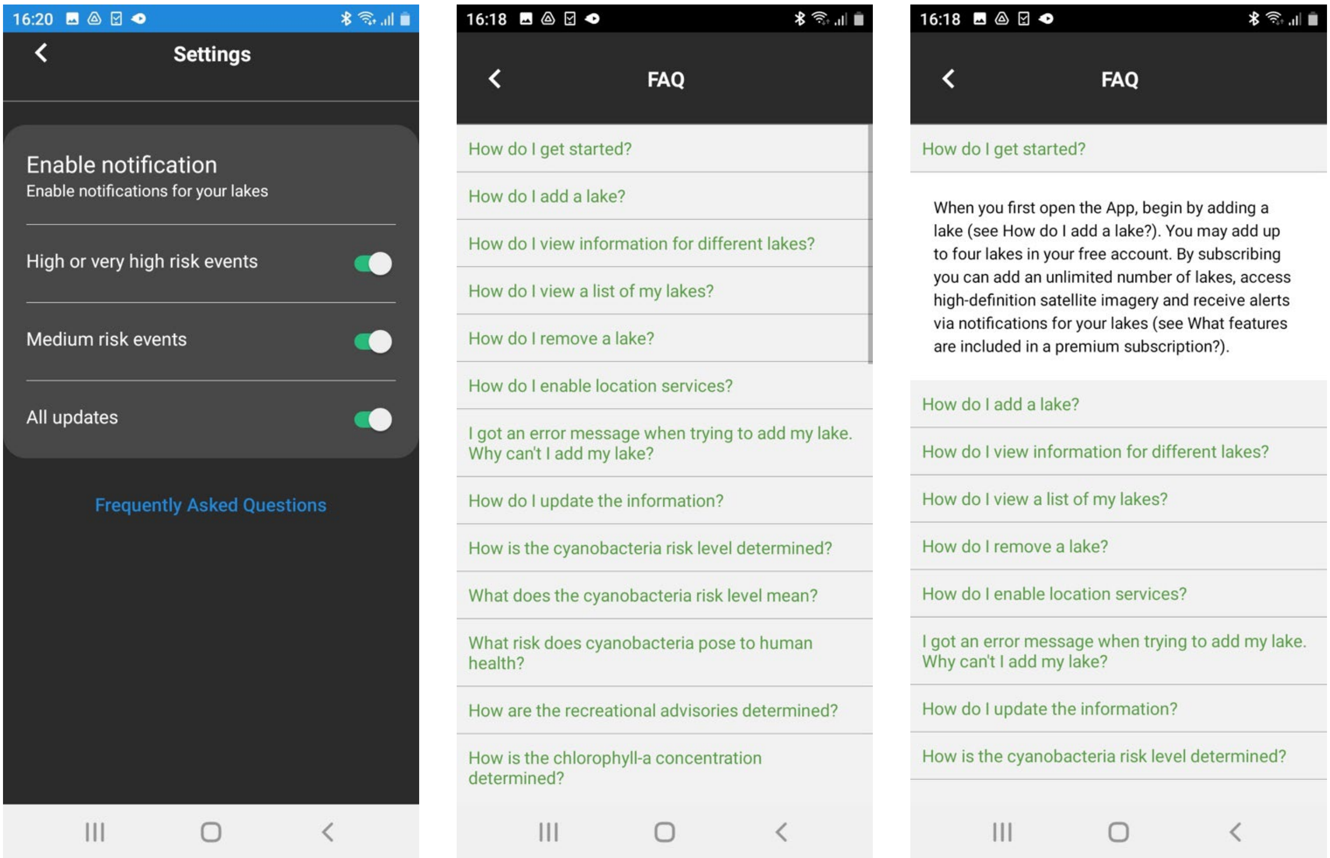


Figure 2.10 Settings and FAQs views

2.3 FORECASTING API

2.3.1 Description

The forecasting API implements the forecasting model for operational usage. The details behind the forecasting methodology are provided elsewhere. The forecasting method has been adapted to provide an interpolated estimate for the near term from a longer 4-week time horizon forecast. This may be referred to as a "background" and "foreground" forecasts as illustrated in Figure 2.11.

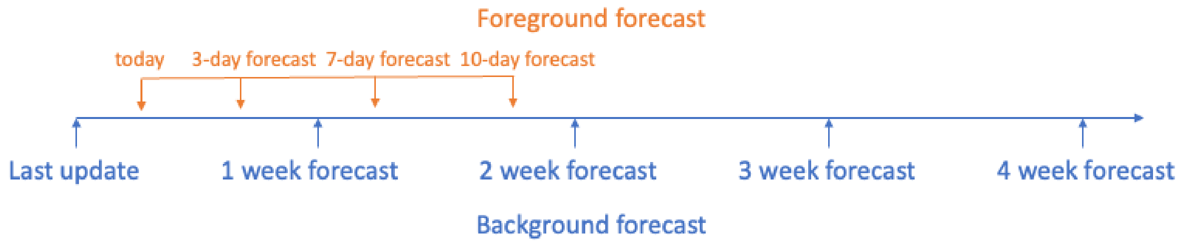


Figure 2.11 Schematic representation of forecasting approach

The background forecast is produced on a weekly timescale using the weekly annual averages (weekly averages for all previous years) plus the weekly averages of the last two observations according to the Moving Average Seasonal Error Adjusted (MASEA) model. However, the weekly timescale is too course for near-term forecasts digested by a user. Thus, a foreground forecast is produced that interpolates the background forecasts from the weekly timescale onto a finer temporal resolution:

- Today – the forecast for the current day
- 3-day – the forecast 3 days from the current day
- 7-day – the forecast one week ahead of the current day
- 10-day – the forecast 10 days from the current day

The foreground forecast updates daily and depends on the values of the background forecast, which may update only when more data are acquired. The foreground forecast is calculated using a simple linear interpolation method:

$$x_t = x_{to} + t * \frac{x_{to} - x_{t1}}{7}$$

where x is the forecast variable, t is the number of days to forecast, x_{to} is the most recent update or forecast value, and x_{t1} is the one week ahead forecast value.

The forecast variables are *chla_med*, the median chlorophyll-a average, and *chla_cyano*, the mean chlorophyll-a value for cyanobacteria.

The forecasted values are converted to a text representation using the logic presented in Table 2.4.

2.3.2 Querying the API

The API is queried as follows:

```
root/api/forecast/{id}
```

Fields:

root – the api root (e.g. https://online.cyanolakes.com)
id – the unique dam identifier

Example response:


```
{
 last_update: "2022-10-26",
 d3: "2022-10-29",
 d7: "2022-11-02",
 d10: "2022-11-05",
 days_since_update: 0,
 text: "High cyanobacteria risk level expected around Saturday October 29",
 chla_cyano_now: "74.7",
 chla_med_now: "71.9",
 cyanobacteria_risk_level_now: "High",
 trophic_status_now: "Hypertrophic",
 microcystin_now: "29.9",
 cyanobacteria_cell_count_now: 149400,
 full_contact_now: ":(",
 partial_contact_now: ":(",
 chla_cyano_3day_forecast: "74.1",
 chla_med_3day_forecast: "71.4",
 cyanobacteria_risk_level_3day_forecast: "High",
 trophic_status_3day_forecast: "Hypertrophic",
 microcystin_3day_forecast: "29.6",
 cyanobacteria_cell_count_3day_forecast: 148200,
 full_contact_3day_forecast: ":(",
 partial_contact_3day_forecast: ":(",
 chla_cyano_7day_forecast: "73.3",
 chla_med_7day_forecast: "70.6",
 cyanobacteria_risk_level_7day_forecast: "High",
 trophic_status_7day_forecast: "Hypertrophic",
 microcystin_7day_forecast: "29.3",
 cyanobacteria_cell_count_7day_forecast: 146600,
 full_contact_7day_forecast: ":(",
```

```
  partial_contact_7day_forecast: ":(",
  chla_cyano_10day_forecast: "72.2",
  chla_med_10day_forecast: "69.5",
  cyanobacteria_risk_level_10day_forecast: "High",
  trophic_status_10day_forecast: "Hypertrophic",
  microcystin_10day_forecast: "28.9",
  cyanobacteria_cell_count_10day_forecast: 144300,
  full_contact_10day_forecast: ":(",
  partial_contact_10day_forecast: ":("
}
```


Variables:

- *last update* – date of most recent update
- *d3, d7, d10* – date for 3-day, 7-day and 10-day horizon
- *days since update* – number of days since update
- *text* – a text description of the forecast

Each of the following variables are available on a now, 3, 7 and 10-day time horizons:

- *chla_cyano* – chlorophyll-a from cyanobacteria
- *chla_med* – median chlorophyll-a value
- *cyanobacteria_risk_level*
- *trophic_status*
- *microcystin* – toxin potential
- *cyanobacteria_cell_count* – estimated cell count
- *full_contact* – full contact advisory
- *partial_contact* – partial contact advisory

2.3.3 Forecast text

The forecast text has been implemented using the following logic (Table 2.4):

- Cyanobacteria forecasts take precedence over algal bloom forecasts
- Higher risk or trophic status levels take precedence over lower ones
- The highest risk level or trophic status for the 3-, 7- and 10-day forecasts takes precedence

In the cast of a cyanobacteria bloom at medium or higher risk level detected in any of the forecast horizons, the forecast will be:

"[risk level] cyanobacteria risk level expected around [date]"

Where “risk level” and “date” are the expected risk level and the date on which the first highest event is detected.

In cases where no cyanobacteria blooms are detected the forecast will be:

"[level] algal concentrations expected around [date]"

where "level" is the algal bloom concentration and "date" is the date on which the first highest event is detected.

Table 2.4 Forecast text description

Forecast Text		
CRL	**Trophic Status (TS)**	**Text**
Very High, High, Medium	N/A	{CRL} cyanobacteria risk level expected around {D}.
Low	Very High, High, Medium	{TS} algal levels expected around {D}.
Low	Low	Low cyanobacteria and algal concentrations expected.

{D} is the date in format "Monday October 12"

2.3.4 Back-processing design

The forecasts rely on historical data to provide seasonal changes in the chlorophyll-a concentration or cyanobacteria presence for the waterbody. This means that historical imagery needs to be processed before reliable seasonal based forecasts can be produced. The following methodology has been implemented to enable back processing of historical data when a user adds a dam to their premium subscription:

- Forecasts are only produced for lakes that belong to Premium accounts
- Premium users have been limited to 10 lakes to avoid needing to back process an unlimited number of lakes
- Dams added in premium accounts are marked as "premium" in the database, using a new API
- An automated script checks for historical data for premium dams, and if less than 2-years of historical data are found, an automated download and process script is triggered to produce a timeseries of historical data for a 2-year period in the background
- A waiting period of 1-week is implemented before data are back processed. This prevents needless processing for lakes that may be haphazardly added or removed by a premium user (abuse)
- The 2-year minimum historical data will be used to characterize the seasonal signal for forecasts
- Before historical data are available, a moving average approach is implemented for forecasts

Back processing is an operational procedure that is performed on an *ad hoc* basis once the new app has been released to the public via an upgrade.

2.3.5 Home Page Redesign for Forecasts

The new home page has several re-designed elements:

- Summary text – a new summary text description
- Floating boxes containing current information
- Forecast and update text – text description of the forecast
- 10 day forecast table – all information forecast with a 10-day time horizon
- A new sideways scroll table with additional information and statistics for current (most recent measurement), last week and last month summary statistics

The improved summary text logic is shown in Table 2.5. The summary text logic has been updated as follows:

- Algal bloom concentration is now stated explicitly as “High algal levels detected”, etc.
- Coverage notations (“throughout”, “in some areas”, “in limited area”) is now also applied to algal blooms, not just cyanobacteria blooms
- Persistence notations are now applied to algal blooms as well as cyanobacteria blooms (“since last month”, etc.)
- Recreational advisories are less prescriptive, with partial contact being determined as “exercise caution” and full contact as “avoid contact with water”. This is in preference to the “avoid swimming” which is not relevant in many waterbodies.
- The reference to water clarity has also been removed, as the water may be turbid even when algae are at a low concentration due to inorganic matter.
- The coverage statistic text has been corrected for a bug, and the recreational advisory has been updated to not make any reference to suitability for swimming. Rather, it now refers to “Exercise caution” or “Avoid contact with water” for partial and full contact advisories, respectively.
- The summary text description has been moved from the mobile app to the summary API to enable automated update and future changes without the need to release a new version of the App.

The last update text has also been added to the summary API call and corrected for time zone information. The update text will be in the format: “Updated x days ago.”

The format of the home page has been redesigned to make the information easier to read with the use of floating opaque round cornered boxed containing the statistical information.

The forecast text and information will only appear if the subscriber has a premium subscription. The forecast text information is shown in Table 2.5 and is used in the format “Cyanobacteria risk level expected around Monday October 14th”.

The 10-day forecast table Provides information on the following time horizons:

- Today’s value (i.e. the current day)
- 3-day forecast (from current day)
- 7-day forecast (1 week from current day)
- 10-day forecast (from current day)

The forecast variables include:

- Cyanobacteria risk level
- Trophic status
- Advisories
- Cyanobacteria cell count and microcystin production potential
- Chlorophyll-a from cyanobacteria (which is the forecast variable)

Dam Name

Summary text

6 ug/L

L:4 ug/L H: 9 ug/L

Cyanobacteria Risk Level	Trophic Status
Low	Low

Full Contact Advisory	Partial Contact Advisory
😁	😕

Forecast and Update text.

<----- table scrolls sideways off screen ----->

10 day forecast	Chlorophyll-a	Cyanobacterial Risk Level	Trophic Status	Full Contact Advisory	Partial Contact Advisory	Cyanobacterial cell count	Microcystin production potential	Cyanobacterial chl-a
Today	6	Low	Low	😁	😕	15 800	3	6
Thurs	5	Medium	Medium	😁	😕	15 800	3	6
Next Mon	4	Low	Low	😁	😕	15 800	3	6
Next Thurs	4	Low	Low	😁	😕	15 800	3	6
2nd Sept	6	Low	Low	😁	😕	15 800	3	6
Last week	5	Medium	Medium	😁	😕	15 800	3	6
Last month	4	Low	Low	😁	😕	15 800	3	6

Cyanobacterial cell count	Microcystin production potential
15 800 ug/L	3 ug/L

Cyanobacterial cover	Vegetation Cover
5%	1%

Cyanobacterial scum cover	Latest Update
0%	3 September 2020

Cyanobacterial chl-a	Algae chl-a
6 ug/L	10 ug/L

* 1ug/L Chlorophyll-a ~ 0.4 ug/L Microcystin ~ 2000 cells/mL

Figure 2.12 Design of App Home Page with Forecasts

The sideways scroll table allows more information to be provided and avoids cluttering the home page with information that may not be desired by the user.

If the user is subscribed to premium, the values on the home page will be today's values, with the coverage estimates being the last update values.

Table 2.5 Summary text

Summary text

CRL	Trophic State (TS)	Text
Very high, High or Medium	Any	{CRL} cyanobacteria risk level detected {C} {P} {A}.
Low	Very High, High or Medium	{TS} algal concentrations detected {C} {P} {A}.
Low	Low	Low cyanobacteria and algal levels detected {P}.

Coverage {C}

Value	Text
0-5%	in limited areas
5-50%	in some areas
50%+	throughout

Persistence {P}

	CRL	Current value is Low or Medium	Current value is high or very high
Last Month	Very High or High	Use caution: cyanobacteria at high risk level detected during the last month.	since last month
Last Week	Very High or High	Use caution: cyanobacteria at high risk level detected within the last week.	since last week

Recreational Advisory {A}

Partial contact	:(	Exercise caution.
Full contact	:(	Avoid contact with water.

2.4 COMMERCIALIZATION

2.4.1 Business model

In order to achieve the aim of the project to generate sustainable revenue and create employment opportunities, it is necessary to design a business model for the App.

The design of the business model was guided by the following aims:

- To maximise uptake of the App (paid and non-paid users, professional and novice)
- To provide a frictionless user experience from installation, use and purchase
- To generate an ongoing revenue stream supporting operational costs
- To meet the needs of both novice and professional users

Based on a review of App store documentation, there are several options for generating revenue from mobile applications:

1. Paid app (e.g. Bird Books, etc.) – once off fee charged up-front
2. Free app with in-app purchases (e.g. games) – free to download with content or features available to purchase
3. Free app with a subscription for premium features (e.g. News apps, etc.) – free to download with subscription for in app features or content
4. Free app with ads revenue (games, etc.) – free to download with revenue generated from advertising
5. Combination of the above

The business model chosen was guided by the aims:

Maximise uptake:

1. A free basic application – users can download and try the app for free and decide later if they want premium features

A frictionless experience:

2. No adverts – Advertising is a turn off for users, creating friction – no irritating adverts that deters users
3. No login – Removing the need for login of users (needless friction)
4. No external payment gateways – payments and subscriptions processed easily via the users account (no external payment gateways)

Ongoing revenue:

5. Subscription-based access to premium features

Novice and professionals:

6. Premium features for paid users is optional – login for professional organisational accounts

Given the above considerations, the business model adopted was as follows:

A free basic App, with no advertising or login, offering monthly and annual subscriptions for premium features processed through the Apple or Google Stores.

2.4.2 User types

There are three types of App users.

1. Free users – users that download and use the free version of the App (no login information is required)
2. Paid users – users that upgrade to access premium features for a monthly or annual subscription
3. Organisational users – users that access the app through their organisational login for the CyanoLakes Web App (e.g. water utility customers, municipalities, etc.)

2.4.3 Subscriptions and billing

The application offers value-added features on a subscription basis through the application stores. Billing has been implemented through Apple App and Google Play Stores. The subscription options and pricing is configured in the developer accounts for each store. It is also required to have a terms of service / use and privacy policy (Appendix B).

2.4.4 Paid features

The premium-features are:

- High resolution satellite imagery from the Sentinel-Hub API
- In-app notifications and alerts
- Forecasts

There are three types of notifications:

- High or very high risk events – notify only for high and very high risk events
- Medium risk events – notify only if a medium risk event occurs
- All updates – provide a notification every time there is an update

2.4.5 Pricing strategy

The price of the App was determined through performing a market analysis comparing the cost of similar applications already available on the Apple and Google App stores.

2.4.5.1 Market analysis of similar apps

Table 2.6 shows the pricing of similar types of apps from the App store. The average pricing for subscription-based health apps ranges widely from as low as $3 per month to as high as $15 per month. For weather-type applications, several apps are available for a once-off charges between $2 to $5. One exception is the successful App Carrot Weather, which offers several different premium features that cost as high as $17/m with a low cost of $4.99/m for a standard premium account. Paid yearly, premium access costs just $1.50 per month. Apps like Carrot Weather prove that you can have premium pricing for a weather-like App, provided you have enough features to warrant the price.

Table 2.6 Pricing of similar apps in the Apple App store

Weather apps	Price	Health Apps	Price
Dark Sky	$3.99 (once off)	Peloton Fitness	$12.99/m
Radar Pro	$4.99 (once off)	Fitbit	$9.99/m
Carrot Weather	$4.99/m up to $17/m, $1.5/m for annual	Flo period tracker	$10/m
Weather on the Way	$0.99	Natural cycles	$9.99/m
Weather Up	$1.99	Sleep Cycle	$3.25/m
RadarScope	$9.99	Powernap	$1.99
Skyview	$1.99	Stoic - journaling	$3/m
Skysafari	$2.99	Headspace	$12.99/m
Star rover	$1.99	Calm	$14.99/m

2.4.5.2 Pricing psychology

When deciding on how to price any App, it is important to consider how the user perceives the value of the Application. There are several different pricing strategies:

- Start with a high price, then go lower over time – it is more difficult to increase the price later
- Start at a high price point to attract fewer premium users (create perceived exclusivity)
- Create an incentive for long-term users (significantly reduced annual subscriptions) by offering annual subscription to avoid cancellations and generate more revenue up-front

The following factors could suggest that the App can command a relatively high price point:

- There are no competitors (free EPA CyAN app on Android, not on iOS, US only)
- There is a small, specialized market (boaters, biologists, concerned citizens) willing to pay
- The relative scarcity of the information commands a higher price (high resolution satellite imagery)
- There are no ads in free version – creating a better user experience

2.4.5.3 Case Study – Carrot Weather

Carrot weather is a premium and costly weather app with more than 29 000 reviews and a huge user base. The description on the App store is quoted below:

> *Carrot Weather:*
>
> *CARROT Weather uses auto-renewing subscriptions.*
>
> *Premium provides additional features like weather data sources, notifications, customization, widgets, and Apple Watch complications.*
>
> *Premium Ultra includes all the features of Premium, plus rain, lightning, and storm cell notifications (where available), a weather maps widget, and quick data source switching.*
>
> *Premium Family includes all the features of Premium Ultra, with the added benefit of being shareable with up to five family members via Apple's Family Sharing service.*
>
> *From Carrot Weather*

The pricing tiers adopted by Carrot Weather is show in Table 2.7.

Table 2.7 Carrot Pricing Tiers

Tier	Monthly	Annual	Monthly (annual)	Details
Premium	$4.99 R89.99	$19.99 R349.99	$1.66 R29.17	7 day free trial. Save 66% for annual.
Premium Ultra	$10.99 R174.99	$44 R699.99	$3.66 R58.33	No free trial. Save 66% for annual.
Premium Ultra Family	$17 R259.99	$68 $1049.99	$5.66 R87.50	Family access. No free trial. Save 66% on annual.

2.4.5.4 Break-even analysis

In order to determine a realistic price for the App, we need to determine the break-even point where costs meet revenue. In order to do this we need ascertain the following:

1. What does it cost to run the App per year?
2. How many new users can we get in one year?
3. What percentage of new users become paying users?

The key variables answering the above questions are show in Table 2.8.

The annual operating cost for running the App takes into account operational, maintenance and optional extra development costs. Based on the current uptake of the App since launch (650 users in 10 months), it is estimated that there could be an additional 800 new app users per year (although the growth rate could vary widely). Using a conservative estimate, it is expected that 10% or less of these users will purchase a premium subscription (1 in 10 users). This means we could expect up to new 80 paying users per year. Under the above scenario, there would be approximately 250 paying users after three years (from 2500 users).

Table 2.8 Break even analysis variables

Key variables	
Annual operating cost	R140 000-R225 000
Number of new users per year (est.)	Up to 800
Percentage of paying users (est.)	±1-10%
Number of new paying users per year (est.)	Up to 80
App store commission (fixed)	15%
VAT (inclusive in price) (SA only / varies)	15%

In order to determine the pricing point for the application that meets the required revenue in a 3-year time horizon, we need to take into account the following:

- The monthly subscription price (default price)
- The annual discounted subscription price (discounted price)

- The inclusion of VAT in the sale price (varies, 15% in RSA)
- The service fee commission of the Apple and Google stores (15% or 30%)

The annual revenue target should be met at the discounted annual pricing for 250 users.

Annual revenue (monthly pricing)					
Users / Price	**R 35**	**R 70**	**R 105**	**R 140**	**R 175**
10	R 3 092	R 6 196	R 9 301	R 12 405	R 15 509
50	R 15 460	R 30 981	R 46 503	R 62 025	R 77 547
75	R 23 189	R 46 472	R 69 755	R 93 037	R 116 320
100	R 30 919	R 61 963	R 93 006	R 124 050	R 155 093
200	R 61 839	R 123 926	R 186 013	R 248 099	R 310 186
250	R 77 298	R 154 907	**R 232 516**	R 310 124	R 387 733
300	R 92 758	R 185 888	R 279 019	R 372 149	R 465 280
500	R 154 597	R 309 814	R 465 031	R 620 249	R 775 466

Figure 2.13 Revenue potential for different default pricing (after VAT and fees)

Annual revenue (yearly pricing)					
Users / Price	**R 349**	**R 699**	**R 1 049**	**R 1 189**	**R 1 749**
10	R 2 577	R 5 164	R 7 751	R 8 785	R 12 925
50	R 12 883	R 25 818	R 38 753	R 43 927	R 64 623
75	R 19 325	R 38 727	R 58 129	R 65 890	R 96 934
100	R 25 766	R 51 636	R 77 506	R 87 854	R 129 245
200	R 51 533	R 103 272	**R 155 012**	R 175 707	R 258 491
250	R 64 416	R 129 090	**R 193 765**	**R 219 634**	R 323 113
300	R 77 299	R 154 908	R 232 518	R 263 561	R 387 736
500	R 128 831	R 258 180	R 387 529	R 439 269	R 646 227

Figure 2.14 Revenue potential for discounted pricing (after VAT and fees)

2.4.6 CyanoLakes pricing

After analysis of the market related data, and a break-even analysis above, CyanoLakes Pro will initially adopt the following pricing strategy:

- Offer a 7-day trial of premium version (only charged after 7 days)
- A monthly starting price of $4.99 (or R74.99 per month) for a premium subscription. Additional tiers could be added later on once more features are added.
- Discounted annual subscription offer 60% discount, get 8 months free when paying annually – a significant discount to generate revenue up-front and secure long-term users

Duration	Discount	USD	ZAR
Monthly renewal	0%	$4.99	R74.99
Annual renewal	60%	$19.99	R299.99

2.5 RELEASES AND INSTALLATIONS

2.5.1 Releases

This is a record of all updates and bug fixes implemented for the app.

Table 2.9 App releases and updates

Apple		
Version	Release date	Summary of changes
1.0	2020-11-02	First release
1.1	2020-12-02	Centre the number on the home page. Fin incorrect number displayed on the home page. Fix error when adding a new lake.
2.1 Beta testing	2021-07-15	Pro version beta testing
Google		
1.1.0	2020-11-08	Initial release
1.2.0	2020-12-18	Fix issue that caused some users not to be able to accept terms and conditions.
1.5.0	2021-07-21	Pro version initial release
1.5.0	2021-07-29	Bug fixes and formatting

2.5.2 First release

Uses satellite remote sensing data to provide cyanobacteria health risk levels and advisories for full and partial contact water sports based on World Health Organisation guidelines and standards. Add any lake on earth and easily manage your list of lakes. Daily updates depending on cloud or ice and snow cover.

Subscribe for advanced features (coming in version 2):

Forecasts
High resolution maps
Notifications and alerts
Data: Copernicus Sentinel-3 and Sentinel-2.

© CyanoLakes (2020)

2.5.3 Pro release

Your weather app for lakes! Provides daily updated cyanobacteria health risk levels for any lake on earth* based on the World Health Organization's guidelines using satellite imagery. Make informed decisions and stay clear of potentially toxic cyanobacteria blooms with advisories for full and partial contact water sports. Search for lakes near you or add a new one by dropping a pin on the map. Easily add or remove lakes from your list. Daily updates depend on cloud, ice and/or snow cover.

Features of the free version:

- No ads, ever
- Cyanobacteria health risk levels based on WHO guidelines
- Recreational advisories for full and partial-contact water sports
- Pollution levels (trophic status levels) based on OECD guidelines
- Last weeks and last month's data
- Estimated cell count and microcystin toxin (derived from chlorophyll-a)
- Coverage statistics for cyanobacteria and vegetation
- Chlorophyll-a statistics

* The smallest lake that can be added is 0.3 mi^2 (1000 x 1000 ft) or 0.8 km^2 (900 x 900 m). Availability of lagoons and estuaries is limited due to tidal effects.

Subscribe to unlock premium features:

- The latest high-definition satellite imagery showing the location of cyanobacteria and algal blooms from chlorophyll-a, bloom-index, vegetation and true-colour maps (smallest target is 30 x 30 m)
- Alerts for cyanobacteria bloom events and updates sent to your notifications centre
- Add an unlimited number of lakes to your subscription
- Forecasts (not currently available – to be released in a future version)
- Family sharing

Payment will be charged to your Apple ID account at the confirmation of purchase. The subscription automatically renews unless it is cancelled at least 24 hours before the end of the current period. Your account will be charged for renewal within 24 hours prior to the end of the current period. You can manage and cancel your subscriptions by going to your App Store account settings after purchase. Cancellation of the current active subscription period is not allowed.

For more details, see the FAQs page in the app or visit https://www.cyanolakes.com/faqs/.

Funded by the Water Research Commission.

© CyanoLakes (2015-2021)

2.5.4 App Release 2.0 with Forecasts

The re-designed application was released as an update on iOS and Android. The final design of the new home page and other screens is shown in Figure 2.13 and Figure 2.14.

2.5.4.1 Disclaimer

The disclaimer is now loaded from an API, enabling the application to be modified as needed. The new disclaimer is:

- Easier to read
- Refers users to guidelines from the WHO and their local authorities
- Credits the funding from the WRC

> This Application provides information related to the risk to human health from cyanobacterial and algal blooms in lakes and reservoirs using satellite imagery.
>
> The advisories are based on World Health Organization standards and guidelines for cyanobacterial blooms in recreational waters. These guidelines may differ from those used by your local authority. Always follow advisories issued by your local authority.

This Application does not provide other important water quality information and must be used in conjunction with other available information before determining whether a waterbody is safe for recreational or other use. The information should be considered preliminary until verified by field observations or measurements.

By accessing and using this Application, you agree that you have read, understood, and agree to the Terms of Use which may be updated from time to time.

This Application was developed with funding from the Water Research Commission.

2.5.4.2 New features and bug fixes

The following list of new features have been developed in this round of upgrades:

- Pull down to update feature – user can pull down which shows a loading icon while the app updates itself
- Disclaimer is moved to an API and loaded from there, allowing automatic updates
- The image library for the background images has been completely updated and revised, providing better images for medium concentration blooms
- The summary text is now retrieved from the API, enabling automated updates of the wording used in the summary text. The summary text wording has been updated.
- The FAQ pages has been deleted from the app and moved to the CyanoLakes website, enabling updating of documentation without having to release another app version
- Premium subscribers have been limited to up to 10 lakes, to prevent overloading the system with back processing requests from premium accounts. This does not apply to organizational users who log in using their organization accounts.
- The WRC has been acknowledged in the disclaimer for funding

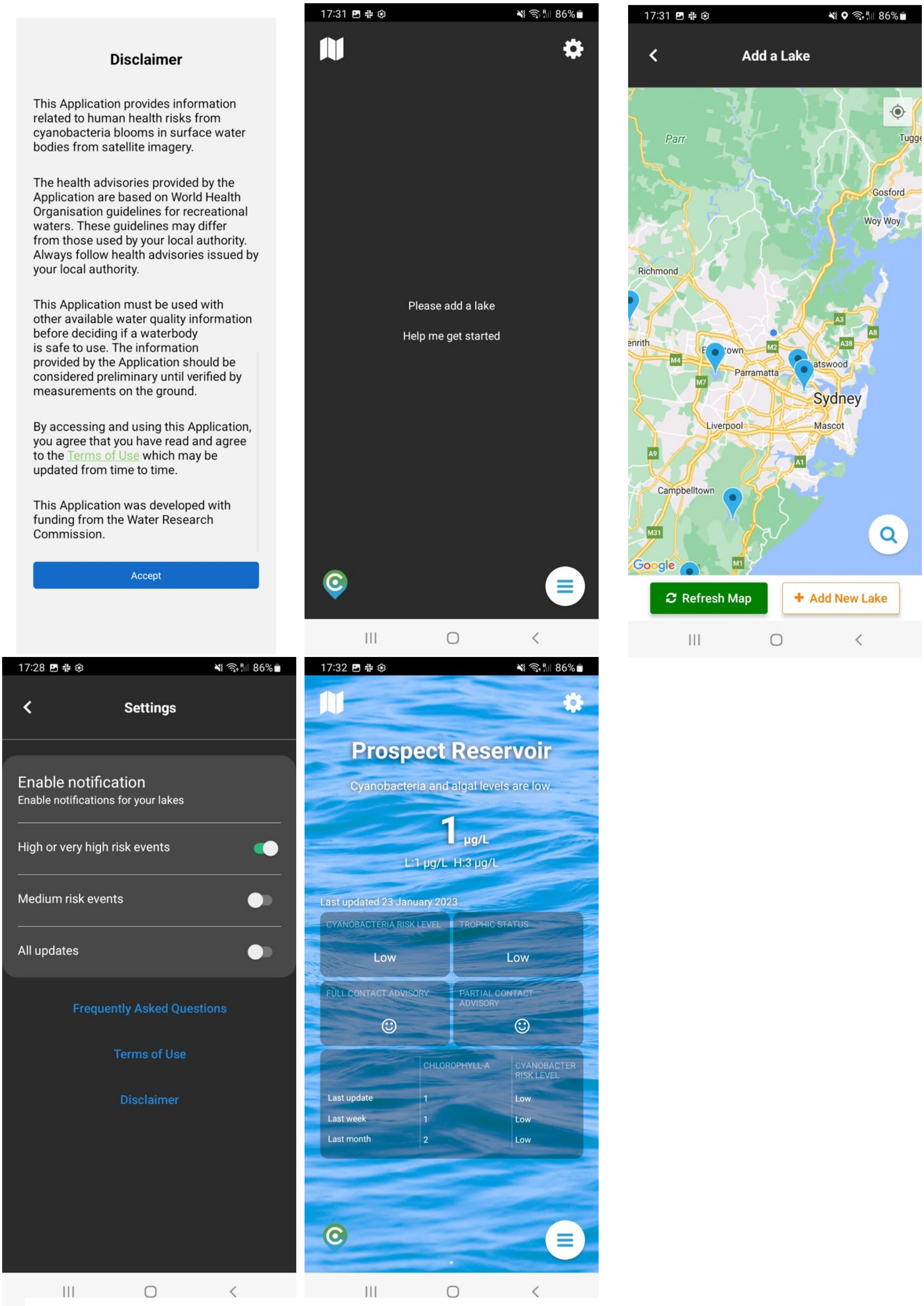


Figure 2.13 Screenshots showing improvements.

Figure 2.14 Redesigned home page showing forecasts.

- The terms and conditions have been updated to account for forecasts
- If a lake name cannot be found, the user can now optionally provide a name for it
- Lakes are now marked as Premium in the database to enable back processing to improve forecast accuracy
- Sentinel layer names are now pulled from an API. This enables adding and removing of satellite image layers for all users without releasing an App upgrade.
- Pins on the add a dam map were changed to blue.
- Added additional variables to the last week and last month variables: microcystin toxin, cyanobacteria cell count and chlorophyll-a for cyanobacteria.
- Use CyanoLakes favicon instead of whole logo at bottom of the screen.
- Migration from the CODA archive to the Eumetsat Data Store (major processing upgrade)
- Creation of system error reports for processing
- Dams are marked as premium for priority for back processing

The following bug fixes have been implemented:

- A bug that caused the API to time out when adding a dam on Android and Apple
- A bug that failed to zoom to a dam in the satellite imagery viewer on Android
- A link to the disclaimer was added to the settings page, along with the terms of use and privacy policy
- A bug that caused dams in certain time zones to update a day later than others
- Errors in processing that were not caught

Refer to the Appendix for the development log that shows the number of commits and lines of code.

2.5.4.3 Store description

The description for the App store has been updated:

> Access cyanobacteria health risk levels based on the World Health Organisation’s guidelines for any lake on earth* from satellite imagery.
> Make informed decisions and stay clear of potentially toxic cyanobacteria blooms with advisories for full and partial contact water sports. Use when on fishing, boating, skiing and doing water sports.
> Search for lakes near you or add a new one by dropping a pin on the map. Easily add or remove lakes from your list.
> Daily updates depend on cloud, ice and/or snow cover.
>
> Subscribe monthly or annually to unlock premium features:
>
> - 10-day forecasts
> - High-definition satellite imagery with maps for chlorophyll-a, blooms, vegetation and more Alerts sent to your notifications centre
> - Additional statistical information including areal coverage statistics
> - Up to 10 lakes
>
> Free version includes:
>
> - Cyanobacteria health risk levels based on WHO guidelines
> - Recreational advisories for full and partial-contact water sports
> - Trophic status based on OECD guidelines
> - Up to 4 lakes
>
> Payment will be charged to your account at the confirmation of purchase. The subscription automatically renews unless it is cancelled at least 24 hours before the end of the current period. Your account will be charged for renewal within 24 hours prior to the end of the current period. You

can manage and cancel your subscriptions by going to your account settings after purchase.
Cancellation of the current active subscription period is not allowed.
For more details, visit https://www.cyanolakes.com/faqs/
Terms of Use: https://www.cyanolakes.com/terms-of-use/
Privacy Policy: https://www.cyanolakes.com/privacy-policy/

* size limits apply, must at least 500 m wide

2.5.4.4 Pricing

The pricing on the store has been updated to reflect current market values. The monthly subscription cost has been increased from $4.99 to $9.99 in developed countries. The pricing remains $4.99 for developing countries (including South Africa). The annual pricing is also changed accordingly for developed countries to $39.99 for 12 months.

2.6 KEY METRICS

At time of writing (27 January 2023):

Google statistics:

- Subscriptions: 33
- New users' acquisition: 702
- Installed audience: 261

Apple statistics:

- 825 first-time downloads (units)
- 78 in-app purchases

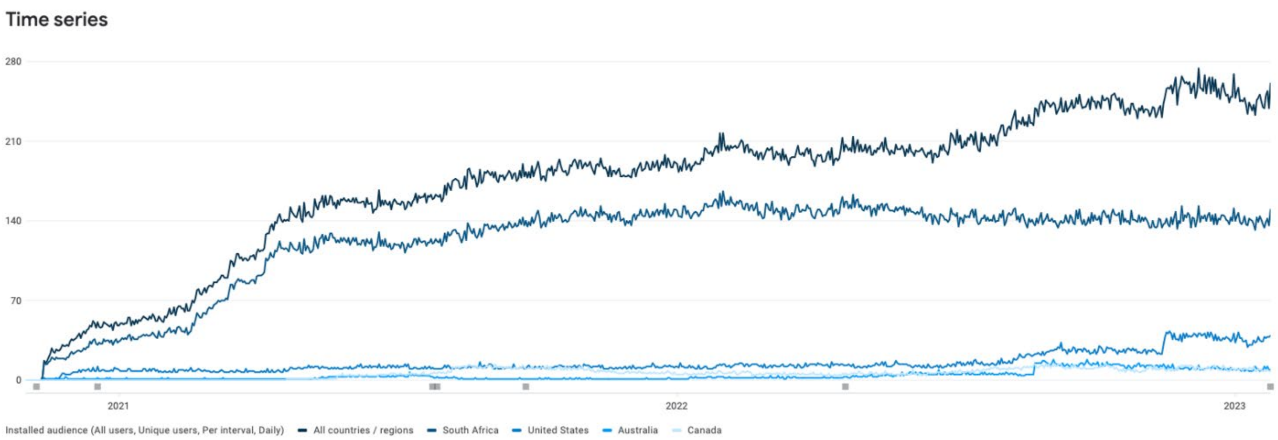


Figure 2.15 Google Play cumulative app installations

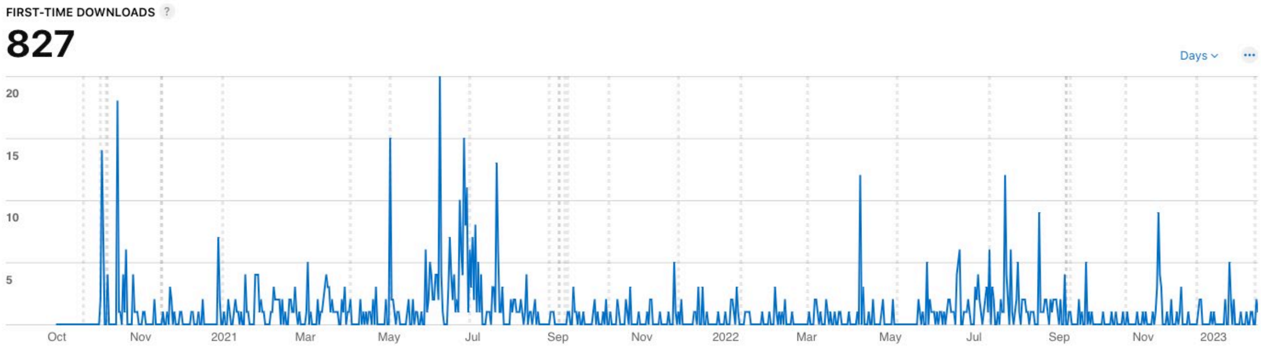


Figure 2.16 Apple App Store daily installations

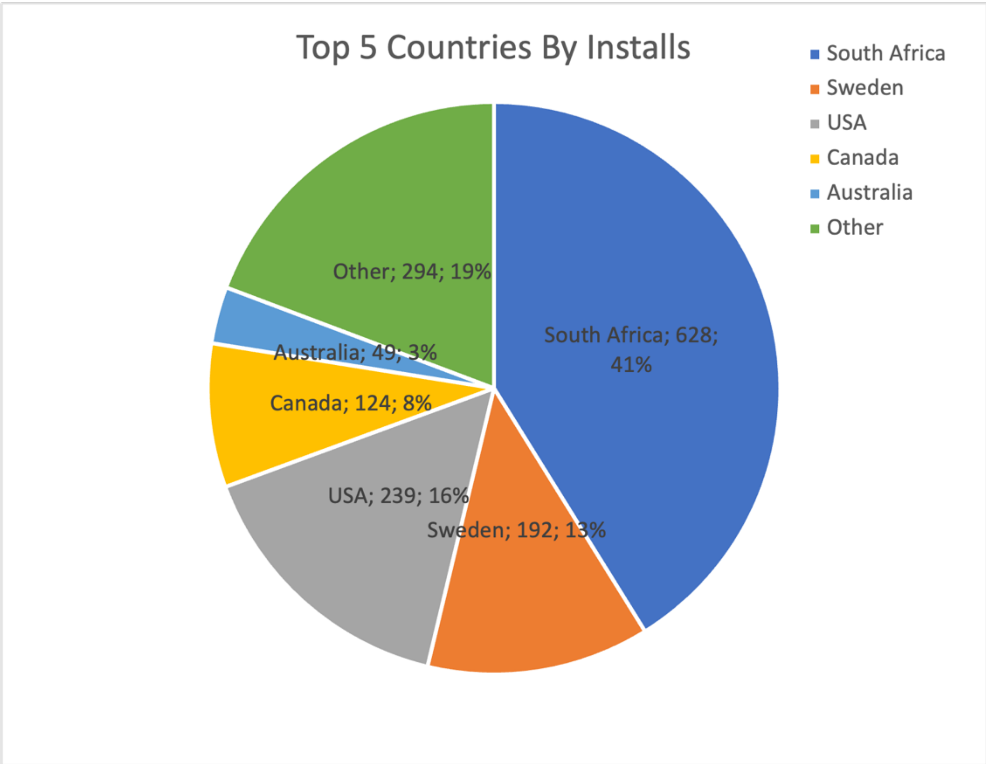


Figure 2.17 Top 5 Countries by Installations

Table 2.10 Sale Units from each platform

Platform	Sales Type	Sales Units
Android	Annual subscription	19
	Monthly subscription	14
iOS	Annual subscription	22
	Monthly subscription	56

2.7 OPEN API

2.7.1 Landing page

A redesigned landing page was developed for the open web app (Figure 2.18). The page has several basic elements and is akin to an open weather app:

- Search box (search by lake name)
- Map box showing nearest lakes zoomed to the user's location
- A dynamic, intuitive data display box, showing the data that appears in the mobile app in a similar form to the mobile app
- A table containing some historical information (basic information only)

The open page has the following functions:

- Zoom to users' location
- Load nearest dam when dragging map
- Display summary statistical data in boxes and tables
- Dynamic background image

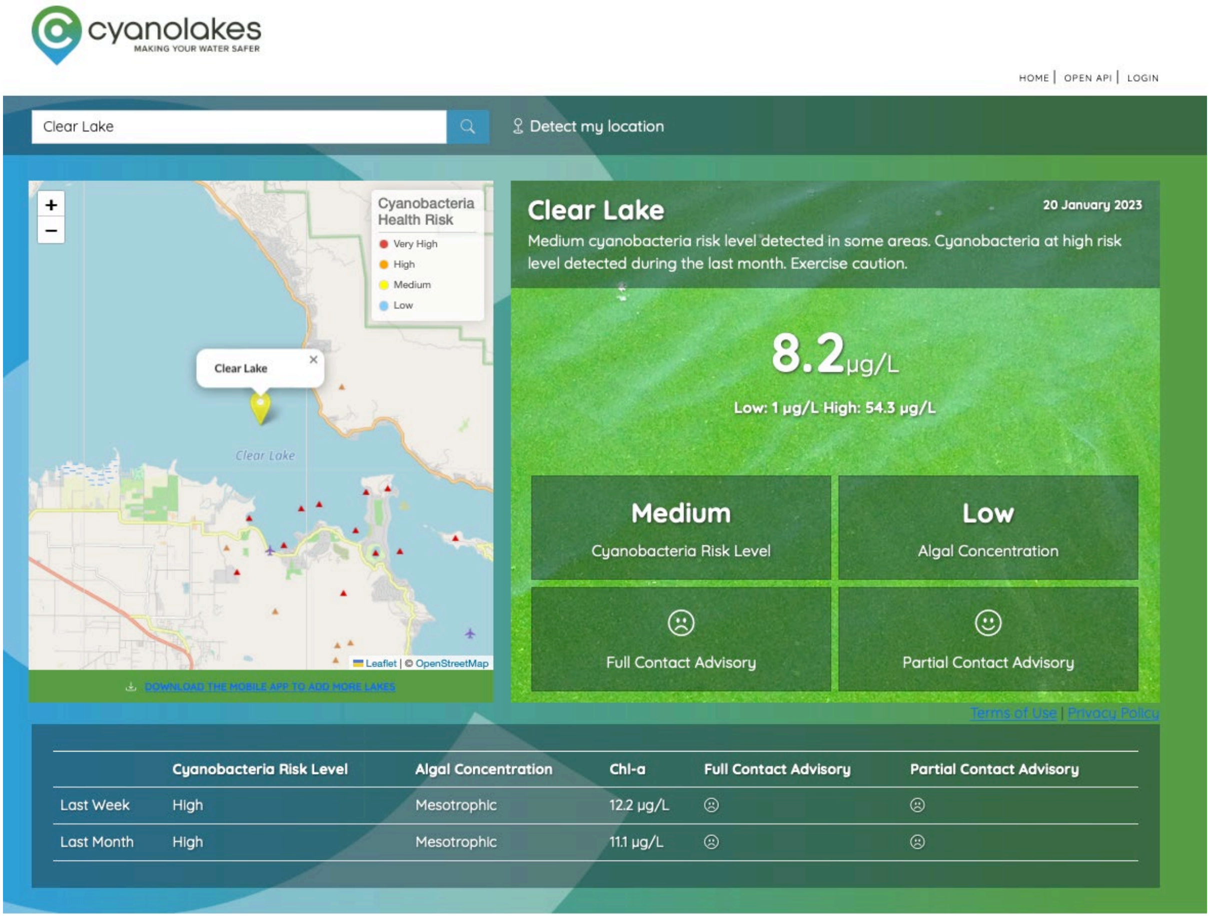


Figure 2.18 Landing page for the open API

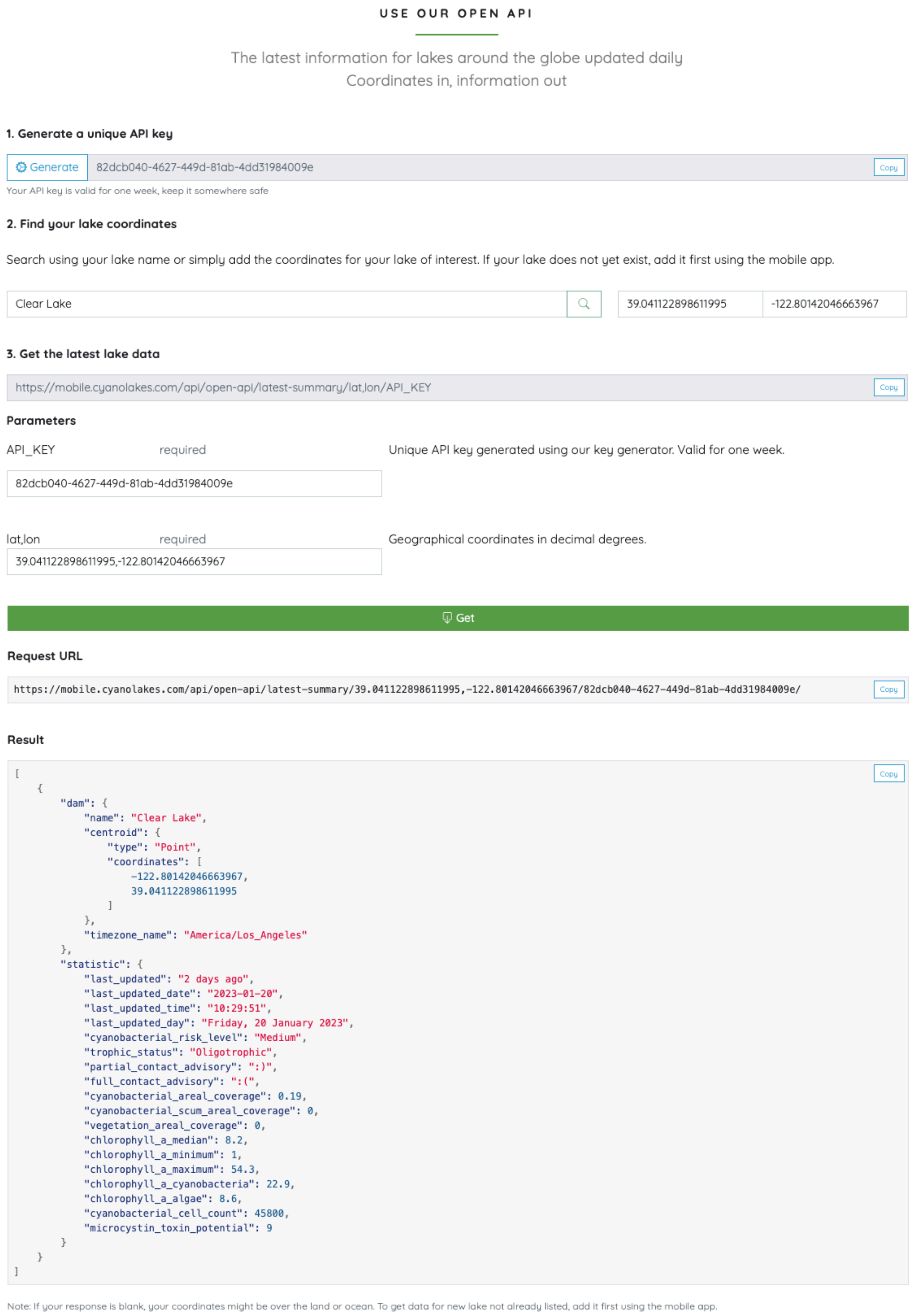


Figure 2.19 Open API page (part 1).

Documentation

cyanobacteria_risk_level	The risk to human health from cyanobacteria.
trophic_status	The trophic status or level of nutrient enrichment.
partial_contact_advisory	Advisory for partial contact recreational use.
full_contact_advisory	Advisory for full contact recreational use.
cyanobacterial_areal_coverage	The area of the lake (in percent) covered by cyanobacteria.
cyanobacterial_scum_areal_coverage	The area of the lake (in percent) covered by cyanobacteria scum.
vegetation_areal_coverage	The area of the lake (in percent) covered by floating vegetation.
chlorophyll_a_median	The median value of chlorophyll-a in µg/L.
chlorophyll_a_minimum	The minimum value of chlorophyll-a in µg/L.
chlorophyll_a_maximum	The maximum value of chlorophyll-a in µg/L.
chlorophyll_a_cyanobacteria	The average value of chlorophyll-a in µg/L for pixels identified as cyanobacteria.
chlorophyll_a_algae	The average value of chlorophyll-a in µg/L for pixels not identified as cyanobacteria.
cyanobacterial_cell_count	The cell count for cyanobacteria in cells/mL calculated from chlorophyll-a using 2000 cells per 1 µg/L chlorophyll-a.
microcystin_toxin_potential	The potential microcystin toxin production calculated from chlorophyll-a using 0.4 µg/L microcystin per 1 µg/L chlorophyll-a.

For a detailed description of the information, please see our frequently asked questions page.

Enjoying our API? Need more data? Please get in touch.

Figure 2.20 Open API page (part 2) showing documentation.

2.7.2 Open API design

A completely new open API page has been designed (Figure 2.19, Figure 2.20). TRhe opne API has the following features:

- A unique key generator for authenticating open API calls. This is used to ensure secure access of the API. The key is valid for a period of 1 week.
- A co-ordinates lookup using the lake name. Users can also enter their co-ordinates manually.
- An example request URL for the API
- An API call that dynamically presents the returned data in JSON format on the page
- Documentation providing a detailed description of the data provided by the API

2.7.3 API description

The open API is accessed using the following URL:

```
/api/open-api/latest-summary/lat,lon/API_KEY
```

where

`lat,lon` – are the coordinates in decimal degrees and
`API_KEY` – the key generated by calling the key from the open API page

The final API call will look as follows for a specific lake and KEY:

https://mobile.cyanolakes.com/api/open-api/latest-summary/39.041122898611995,-122.80142046663967/82dcb040-4627-449d-81ab-4dd31984009e/

The data returned is as follows:


```
[
    {
        "dam": {
            "name": "Clear Lake",
            "centroid": {
                "type": "Point",
                "coordinates": [
                    -122.80142046663967,
                    39.041122898611995
                ]
            },
            "timezone_name": "America/Los_Angeles"
        },
        "statistic": {
            "last_updated": "Today",
            "last_updated_date": "2023-01-23",
            "last_updated_time": "10:13:06",
            "last_updated_day": "Monday, 23 January 2023",
            "cyanobacterial_risk_level": "High",
            "trophic_status": "Oligotrophic",
            "partial_contact_advisory": ":(",
            "full_contact_advisory": ":(",
            "cyanobacterial_areal_coverage": 0.16,
            "cyanobacterial_scum_areal_coverage": 0,
            "vegetation_areal_coverage": 0,
            "chlorophyll_a_median": 8,
            "chlorophyll_a_minimum": 1,
            "chlorophyll_a_maximum": 79.9,
            "chlorophyll_a_cyanobacteria": 20.6,
            "chlorophyll_a_algae": 8.4,
            "cyanobacterial_cell_count": 41200,
            "microcystin_toxin_potential": 8
        }
    }
]
```


The documentation provides an explanation for each of the variables:

- cyanobacteria_risk_level – The risk to human health from cyanobacteria.
- trophic_status – The trophic status or level of nutrient enrichment.
- partial_contact_advisory – Advisory for partial contact recreational use.
- full_contact_advisory Advisory for full contact recreational use.
- cyanobacterial_areal_coverage – The area of the lake (in percent) covered by cyanobacteria.
- cyanobacterial_scum_areal_coverage – The area of the lake (in percent) covered by cyanobacteria scum.
- vegetation_areal_coverage – The area of the lake (in percent) covered by floating vegetation.
- chlorophyll_a_median – The median value of chlorophyll-a in µg/L.
- chlorophyll_a_minimum – The minimum value of chlorophyll-a in µg/L.
- chlorophyll_a_maximum – The maximum value of chlorophyll-a in µg/L.

- chlorophyll_a_cyanobacteria – The average value of chlorophyll-a in µg/L for pixels identified as cyanobacteria.
- chlorophyll_a_algae – The average value of chlorophyll-a in µg/L for pixels not identified as cyanobacteria.
- cyanobacterial_cell_count – The cell count for cyanobacteria in cells/mL calculated from chlorophyll-a using 2000 cells per 1 µg/L chlorophyll-a.
- microcystin_toxin_potential – The potential microcystin toxin production calculated from chlorophyll-a using 0.4 µg/L microcystin per 1 µg/L chlorophyll-a.

CHAPTER 3: HIGH RESOLUTION PRODUCTS

3.1 INTRODUCTION

The Sentinel-2 series of high spatial resolution instruments (A and B) serve as a complimentary data source to Sentinel-3. Unlike Sentinel-3 however, Sentinel-2 was designed for land-based and not water applications. Sentinel-2 has lower observational frequency (roughly every 16 days, or 8 days with 2 sensors), poorer radiometric quality or signal-to-noise ratio, and much fewer spectral bands for algorithm application or detecting the spectral features of cyanobacteria. However, initial studies have highlighted the advantages of the higher spatial resolution, mainly because it enables monitoring of water bodies less than 1 km squared. This would greatly increase the number of water bodies that could be monitored, however, at a lower quality as far as quantitative estimates of chlorophyll-a are concerned.

This project seeks to develop and apply methods for retrieving chlorophyll-a and detecting cyanobacteria using Sentinel-2 data through the provisions of complimentary value-added layers via the mobile application. Available algorithms will be tested and applied with Sentinel-2, or developed.

3.2 METHODOLOGY AND MATERIALS

Sentinel-2 data was derived from the Sentinel-Hub cloud-based storage and processing provider (https://www.sentinel-hub.com). Storing and archiving of Sentinel-2 data is resource intensive and very expensive, due the very high data volumes (Petabyte dimensions). Sentinel-Hub's cloud processing and infrastructure serves up images via a web mapping service. CyanoLakes payment of the subscription for Sentinel-hub amounts to an in-kind contribution of R84 000 per year.

Sentinel-hub has its own JavaScript based scripting language called EvalScript (https://docs.sentinel-hub.com/api/latest/evalscript/v3/). Algorithms are written in EvalScript and deployed on the Sentinel-hub configurations utility.

3.3 ALGORITHMS

3.3.1 Land / Water Masking

Separation of land and water is a challenging task and is essential for water remote sensing applications. Land / water masking tested several available algorithms, as well as novel approaches. We selected the Water Bodies Mapping Script from the Sentinel Hub script repository (https://github.com/sentinel-hub/custom-scripts/tree/master/sentinel-2/water_bodies_mapping-wbm). The script, written by Mohor Gartner, utilises the most recent research into land / water separation of urban waterbodies (Du et al., 2016; Feyisa et al., 2014; Acharya et al., 2018). The script was modified for use for a single scene (Figure 3.1), and is shown in Appendix E.

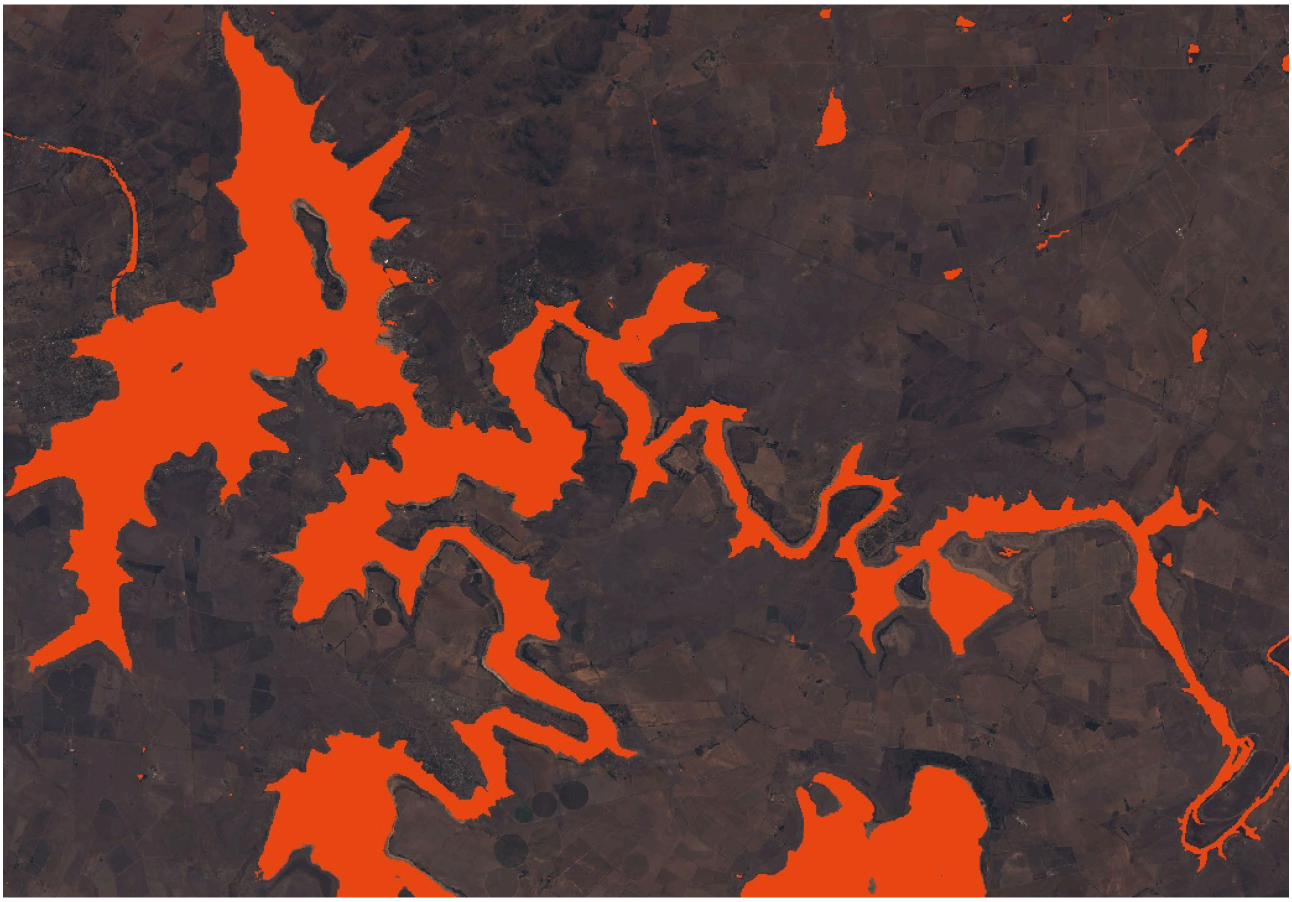


Figure 3.1 Water mask showing land / water separation at the Vaal Dam.

3.3.2 Bloom index

The CyanoLakes bloom index is based on the Maximum Peak Height algorithm approach (Matthews et al., 2012) which is adapted for Sentinel-2. It uses the Maximum Chlorophyll Index (Gower, 2008) and the Floating Algae Index (Hu, 2009) baseline subtraction algorithms:

$$MCI = L705 - L665 - (L740 - L665) \times (705 - 665) / (740 - 665))$$

$$FAI = L740 - L665 - (L842 - L665) \times (740 - 665) / (842 - 665))$$

Where Ln is uncorrected top of atmosphere radiance, and n is the central wavelength of the relative band.

The algorithm switches depending on the magnitude of the index as follows:

Condition	FAI > MCI	MCI > FAI
Index used	FAI	MCI

The script applies the waterbody mask to discriminate between land and water (Figure 3.2) and is shown in Appendix E.

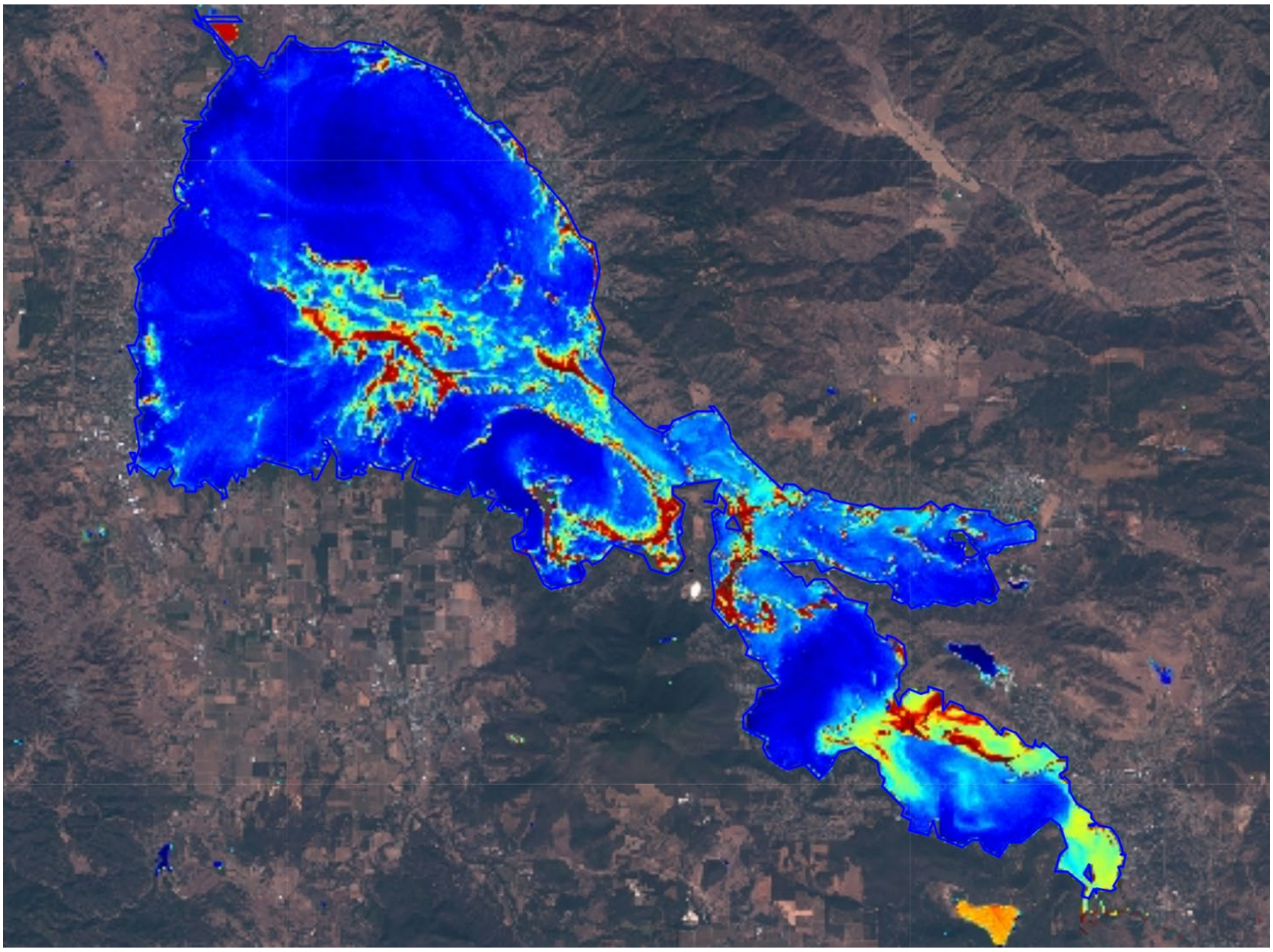


Figure 3.2 CyanoLakes Bloom Index for Clear Lake, California.

3.3.3 Chlorophyll-a

A chl-a retrieval model was calibrated using a novel, high quality synthetic dataset of coincident S2-MSI remote sensing reflectance (Rrs) and pigment concentrations (Figure 3.3). The parameterization of the dataset was informed by the LIMNADES dataset (Lake Bio-optical Measurements and Matchup Data for Remote Sensing: http://www.limnades.org) and compiled using the Hydrolight radiative transfer software (version 5.2, Sequoia Scientific, USA; Mobley, 1994). The unique dataset accounts for the optical complexity of mixed cyanobacteria phytoplankton assemblages, and immense optical variability typically found in global inland water bodies.

In order to produce a fast and robust chl-a estimate from S2, the normalized difference chlorophyll index (NDCI) was used to establish an empirically derived chl-a model (Mishra and Mishra, 2012). The NDCI model utilizes MSI spectral bands located at 665 nm and 705 nm, which are maximally sensitive to chl-a absorption, and backscattering induced reflectance, respectively. NDCI is calculated by taking the spectral difference of these two bands, and normalizing by their sums, as follows:

$$NDCI = [Rrs(705) – Rrs(665)] / [Rrs(705) + Rrs(665)] \qquad (1)$$

The index was trained on synthetic data with the following restrictions to limit uncertainty due to extreme cases:

1. Concentration of non-algal particles (Cnap) < 10 mg/m^3
2. Absorption due to coloured dissolved organic matter (CDOM) < 3 m^{-1}
3. Chl-a concentrations less than 500 mg/m^3
4. Only cyanobacteria *M. aeruginosa* used in training.

This resulting in 9836 samples, which were separated into a training dataset of 7868 samples (80%) and a dataset to test the predictive capability of the model of 1968 samples (20%). An exponential model resulted in the best fit on the training data (eq. 2) based on Pearson correlation coefficient (R^2=0.52), with predictive statistics shown in Figure 3.4.

$$\text{Chl-a} = 17.441e^{(4.7038*\text{NDCI})} \quad (2)$$

Table 3.1 Chlorophyll-a algorithm performance

Log Bias	MAPE (%)	RMSE (mg/m³)	Log RMSE	Relative RMSE (%)
0.0023	42.3	84.2	0.99	95.8

The script is shown in Appendix E.

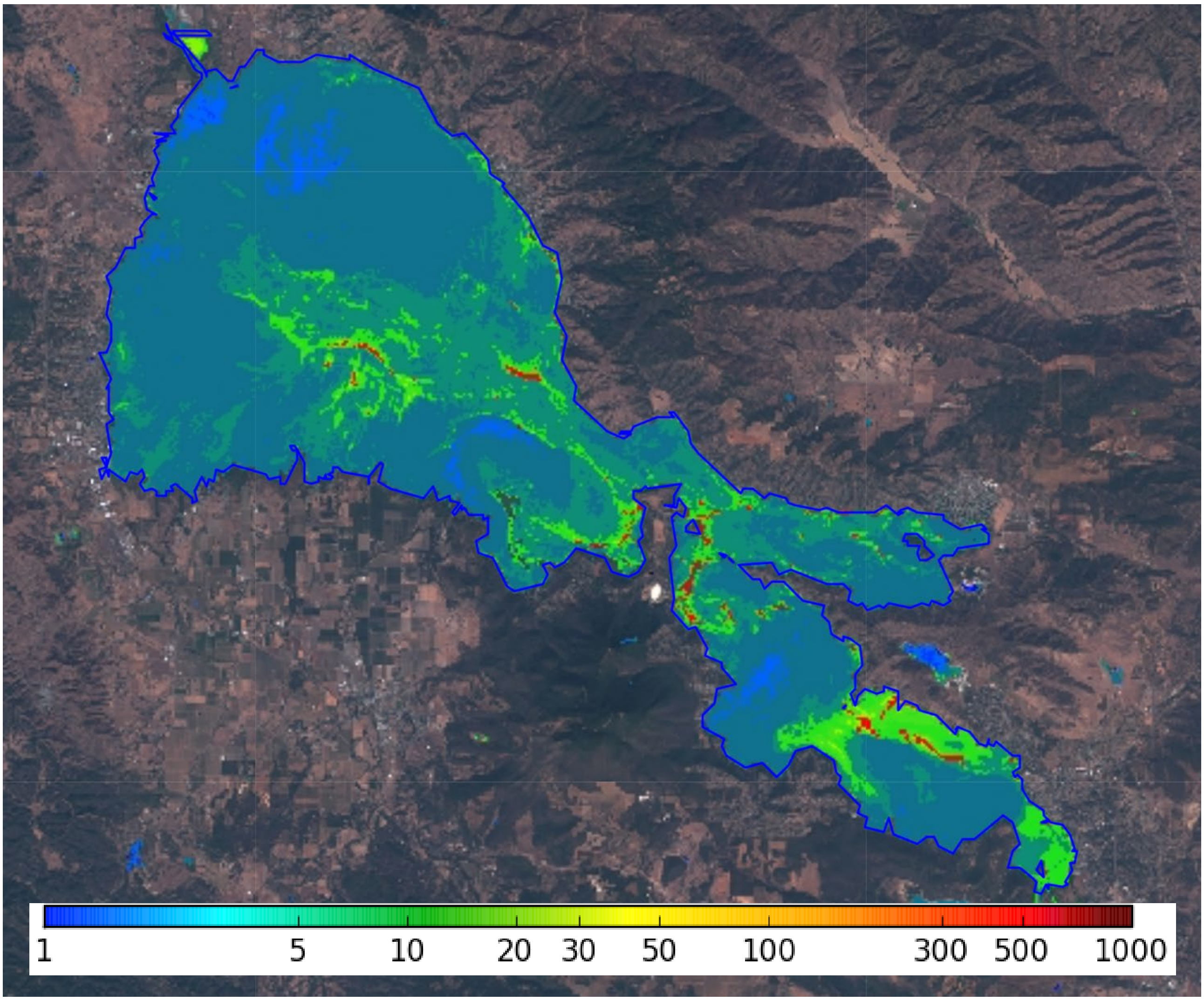


Figure 3.3 CyanoLakes Chlorophyll-a for Clear Lake, California (units is µg/L).

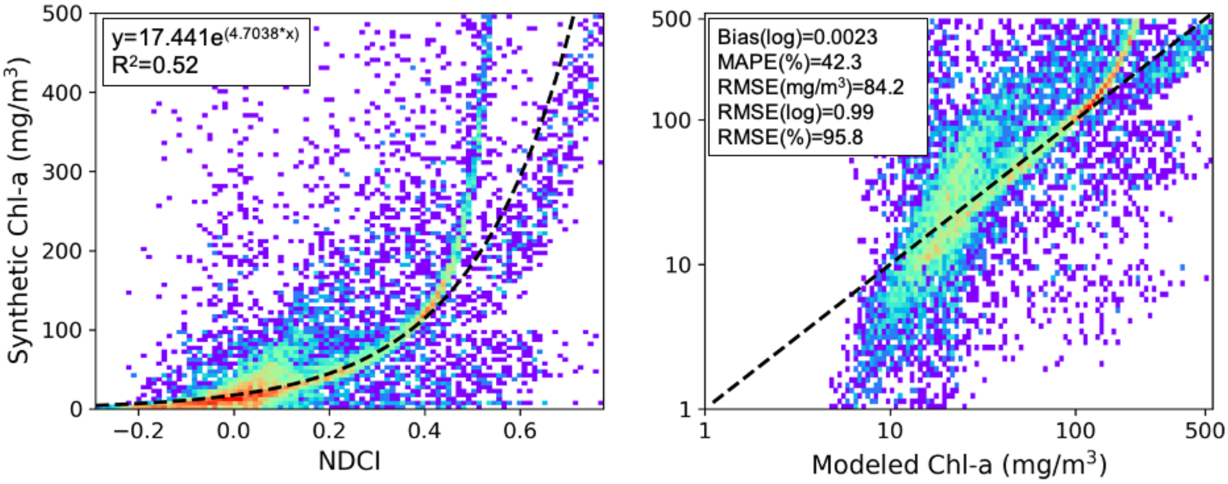


Figure 3.4 (left) Calibration of NDCI using synthetic Rrs and chl-a concentrations. (right) Predictive capability of calibrated NDCI model. Warmer colours indicate regions of higher density of points.

CHAPTER 4: A PHENOLOGY-BASED FORECASTING MODEL FOR LAKES

Abstract

Forecasting of cyanobacteria blooms in lakes is essential to reducing the risk to human and animal health, and operational cost on water treatment, through advanced and early warning. Cyanobacteria forecasting models are typically complex requiring input of biophysical and chemical measurements or DNA sequencing in situ. Satellite imagery presents a unique opportunity to estimate cyanobacteria concentration directly at low cost and over wide spatial and long timescales. This study explores the hypothesis that simple univariate forecasting methods can be applied to reliably forecast cyanobacteria blooms detected using satellite remote sensing. A simple univariate model based on logical decomposition with a moving average and seasonal component was developed to forecast chlorophyll-a concentrations from cyanobacteria and algal blooms in lakes using spatially aggregated satellite remotely sensed data. A small test set of 15 spatially distributed waterbodies was used to assess the model forecast performance on 1-week, 2-week and 4-week forecast horizons using a year-long test timeseries. We found that for a 1-week time horizon, high-risk cyanobacteria blooms could be forecast with 80% accuracy. The 2-week and 4-week forecast accuracy dropped off to 71% and 69%, respectively. Forecast performance was only weakly influenced by lake size, meaning that the spatial-aggregation approach may be valid even for large lakes. Additionally, we found that longer time series reduced the observed forecast error presumably due to better seasonal characterisation. The study is the first to demonstrate that simple univariate models can be used to forecast cyanobacteria and algal blooms in lakes from time-series of satellite remotely sensed data with almost the same reliability as more complex approaches.

4.1 INTRODUCTION

The CyanoLakes web and mobile applications ("Your Weather App for Lakes" – the App) provide weather-like information from satellites for human health risks from cyanobacteria blooms and pollution levels from eutrophication in the world's lakes. The satellite imagery is processed in near real-time (two to three hours after acquisition) providing same-day updates with alerts and notifications. The applications are useful to water utility companies (water managers) providing drinking water and who experience taste, odour and toxin problems as a result of cyanobacteria blooms; and recreational water users who require information on safety for recreational water use.

There is a need to provide forecasts through the App for cyanobacteria concentrations providing early-warning to water managers to enable the implementation of in-lake and in-plant remedial actions, evidenced by the growing literature on the subject (Recknagel et al., 2017; Cayelan et al., 2020; Rousso et al., 2020). Such actions may include adding algicides, activating a mixing system, optimising water treatment for toxin removal, or changing the depth or location of drinking water abstraction points (Carey et al., 2020). Similarly, there is a need for recreational water users to have access to forecasts that enable advanced decision making and can prevent unnecessary exposure to high concentrations of cyanobacteria or poor water quality leading to adverse health effects. The forecast horizon used for the purposes of water management averages at 7-days with up to 30-days or as little as 1-day (Rousso et al., 2020). A near term-forecast horizon of 1-week or less is useful for recreational purposes, similar to how people access and use weather information.

Methods for forecasting cyanobacteria blooms incorporating satellite remote sensing data combine it with meteorological models and hydrodynamic models to create lake-specific forecasts on a weekly or seasonal basis (Recknagel et al., 2018 and references therein). This approach, while comprehensive, is impractical as it cannot be generalised to other lakes easily and requires many often-complex models to be run simultaneously. Further, these models often rely on in situ datasets which negates the benefits of remote sensing data that is collected autonomously and can be processed in near real-time. One study that relies entirely on a spatiotemporal multivariate model derived exclusively from satellite remote sensing data is Myer et al (2020) who derived a model for high-risk cyanobacteria blooms in 102 lakes in Florida. Their approach used spatially aggregated estimates of cyanobacteria cell count from the Ocean and Land Colour Instrument on board Sentinel-3, with additional variables in the form of surface water temperature, precipitation and air temperature and lake geomorphology. Their model could predict the probability of high-risk cyanobacteria blooms with an 82% accuracy for a 1-week forecast horizon. However, they did not forecast the actual concentration of cyanobacteria cells, nor did it address longer time-horizons.

Non-remote sensing methods of forecasting cyanobacteria blooms utilise long timeseries of in situ datasets with variables such as chlorophyll-a, phytoplankton cell counts, Secchi Disk depth measurements, nutrient concentrations, and weather information (e.g. temperature and wind), etc. These data are then fed into either physical models that simulate lake environments, or even the lifecycle of cyanobacteria, or statistical methods such as multi-variate regression analysis, neural networks or other advanced statistical methods to produce forecasts (see Rousso et al., 2020, Echard, 2021 for details). The “process-based” approaches are by nature site-specific and highly complex to initialise and parameterise, while “data-driven” or statistical methods are parameterised only for the dataset or species under consideration and therefore also largely site-specific. Further both approaches require continuous monitoring of a set of variables that are time-consuming and costly to monitor. Alternative approaches based on sequencing the DNA of the microbial community have a high degree of predictive success, but again rely on complex measurements the spatial applicability of which are also untested (e.g. Tromas et al., 2017).

All of these approaches neglect the possibility of utilising simple univariate forecasting methods which are often successfully used for meteorological or commercial applications (e.g. Ji & Peters, 2004; White & Nemani, 2006; Peng & Chu, 2009; Caillault & Bigand, 2018). Such approaches forecast single variables in near-term time horizons with high levels of success and often make use of satellite data. Forecasting from satellite data should be performed not on a per-pixel basis, but on aggregated pixel values for an area, such as a field in case of crop monitoring, or in the case of the present study, a lake (White & Nemani, 2006). Using spatially aggregated single-variables from satellite remote sensing data that has consistent measurement basis, long time-series and near real-time availability seems a favourable approach for near-term forecasting of cyanobacteria blooms.

Further, given that cyanobacteria blooms exhibit strong seasonality in response to light availability and temperature due to their lifecycle, a phenology-based approach taking advantage of seasonal patterns is likely to be successful (e.g. White & Nemani, 2006; García-Mozo, et al., 2008, Peng & Chu, 2009). The bacterial community also has a predictable cyclical change which further drives seasonality and predictability of blooms (Tromas et al., 2017). These logical forecasting approaches (known generally as classical seasonal decomposition) decompose univariable timeseries into error, trend and seasonal (sometimes referred to as ETS) components that allow forecasting based on a logical approach (for detailed methods see Hyndman & Athanasopoulos, 2018; Svetunkov, 2021). Very simply, the trend component is the average trend over the entire timeseries, the seasonal component is the average for a period (e.g. weekly or monthly averages) after subtraction of the trend, and the error component is the anomaly or deviation between the seasonal averages and the observed data. There are many different implementations of classical decomposition technique ranging from simple to complex. Given that the current inoculation of cyanobacteria cells (i.e. today’s value) is the strongest predictor of tomorrow’s value (Swanepoel et al., 2016), simple moving-average techniques are likely to have some success for near-term forecasts.

This paper aims to develop an operational near-term cyanobacteria forecasting approach that can be generalised to thousands of lakes worldwide and is both logical and computationally light-weight suitable for operational use. Or in terms of requirements, this necessitates an approach that is:

- computationally simple and lightweight (fast)

- applicable across very diverse lake systems (generalised)
- does not require model parameterisation (logical)
- does not require training of complex algorithms

These requirements eliminate the possibility of using site-specific process-based models and complex data-driven methods that require training using in situ datasets. Rather a more pragmatic approach based on logical, univariate techniques is required.

In this paper, we test simple univariate forecasting methods and seek to develop a simple logical method for cyanobacteria forecasting using a small test set of 15 lakes that are globally spatially distributed. Further we test two hypotheses, firstly that the forecast performance is sensitive to lake size (larger lakes are more challenging to forecast owing to increased heterogeneity), and secondly that a longer time-series of historical data that is used to characterise seasonal variability leads to improved forecasting ability in the near-term (more data leads to better forecasts). This study is the first of its kind to test the feasibility for using simple univariate methods for forecasting cyanobacteria blooms in the world's lakes based exclusively on data from satellite remote sensing.

4.2 METHODOLOGY

4.2.1 Satellite-derived data

The App uses data from the Copernicus Sentinel-3 Ocean and Land Colour Instruments (OLCI) onboard Sentinel-3A and Sentinel-3B satellite platforms. OLCI provides synoptic coverage of the globe with up to 6 updates per week for any site on earth, at a spatial resolution of 300 m, with 21 spectral bands with high signal-to-noise ratios that are a requirement for water-colour related applications. This allows lakes larger than approx. 1 km^2 to be imaged, which accounts for a large percentage of the world's lake surface area.

Estimation of chlorophyll-a concentration and the detection of cyanobacteria are performed from OLCI data using the maximum peak-height algorithm (Matthews & Odermatt, 2015). The accuracies of cyanobacteria detection have been determined from in situ datasets in Australia and South Africa at greater than 70%, and the error for chlorophyll-a is around 50%, which is acceptable for a log-scaled variable that ranges over four orders of magnitude from 0.1 to 1000 μg/L in lakes. The algorithm provides chlorophyll-a values for cyanobacteria and algal blooms, as well as a binary mask for cyanobacteria detections. Cyanobacteria is detected at concentrations exceeding approx. 5 μg/L due to the inherent sensitivity of the satellite instrument design.

The image data are spatially aggregated using mean or median values for the lake surface area to produce two variables that are forecasted:

1. Chla_cyano – the mean value of chl-a for pixels identified as cyanobacteria for the lake
2. Chla_med – the whole-lake median value of chl-a including both cyanobacteria and non-cyanobacteria pixels

Chla_cyano represents the average biomass of cyanobacteria detected in the lake, although this is not a whole-lake average but only the average for pixels identified as cyanobacteria. Chla_med represents the average algal and / or cyanobacteria biomass for the lake at a moment in time.

In addition to these variables, the App provides two main indices for lake management which are derived based on thresholds from the forecast variables:

1. the cyanobacteria risk levels based on World Health Organisation guidelines; and,
2. the trophic status levels based on OECD guidelines

The cyanobacteria risk level thresholds are defined as:

- Low risk: chla_cyano < 10 µg/L , up to 20 000 cells/mL or 4 µg/L microcystin.[1]
- Medium risk: chla_cyano between 10 and 50 µg/L, up to 100 000 cells/mL or 20 µg/L microcystin
- High risk: chla_cyano between 50 and 100 µg/L, up to 200 000 cells/mL or 40 µg/L microcystin
- Very high risk: chla_cyano > 100 µg/L, more than 200 000 cells/mL or 40 µg/L microcystin

The trophic class thresholds are defined as:

- Low: chla_med < 10 µg/L
- Medium: chla_med between 10 and 20 µg/L
- High: chla_med between 20 and 50 µg/L
- Very high: chla_med larger than 50 µg/L

The forecast accuracy was calculated with regards to the two indices by calculating the percentage agreement between observed and forecast values (i.e. the agreement between the forecast and observed values in the test dataset).

4.2.2 Study areas

This study uses a cross-section of timeseries from fifteen lakes mainly from mid-latitude regions and known to be affected by seasonal or permanent cyanobacteria blooms. The lakes are mostly from South Africa and Australia where the satellite data have been corroborated with in situ datasets, as well as Italy (Trasimeno), the USA (Clear Lake) and China (Taihu). Most of the lakes have been chosen as they are known to be affected by seasonal or permanent cyanobacteria blooms, evident by chla_cyano values (see Table 4.1). This includes lakes such as Clear Lake, Copeton, Hartbeespoort (also affected by floating vegetation), Roodeplaat, Trasimeno, Taihu and Vaal. Several lakes exhibiting very infrequent and low-concentration cyanobacteria blooms and function as controls (e.g. Burragorang, Grahamstown, Midmar and Zeekoevlei). The lakes also vary according to trophic class, which is easily assessed by the average chlorophyll-a value with respect to the thresholds above. Oligotrophic (low productivity) lakes include Burragorang, Grahamstown, Midmar and Trasimeno, while hypertrophic lakes (highly productive) include Zeekoevlei, Roodeplaat, Vaal.

The number of lakes selected (15) is somewhat random but represent a small but large-enough sample to draw conclusions using regression statistics. The lakes vary greatly in size, from just 2,6 km^2 (Zeekoevlei) to a massive 3000 km^2 (Taihu). The timeseries also differ in length and data frequency, with start dates ranging from October 2016 to as recent as July 2019 (Clear Lake), and counts ranging from just 435 (Taihu) to up to 949 (Vaal). Data frequencies differ mainly due to regional cloud cover differences. The differences in time-series length and frequency, as well as size are intentionally chosen to assess relationships between forecast performance and data quantity and target size. All timeseries data were available until 15 November 2021.

[1] 1 µg/L chlorophyll-a equivalent to 2 000 cells/mL or 0.4 µg/L microcystis (World Health Organization, 2002)

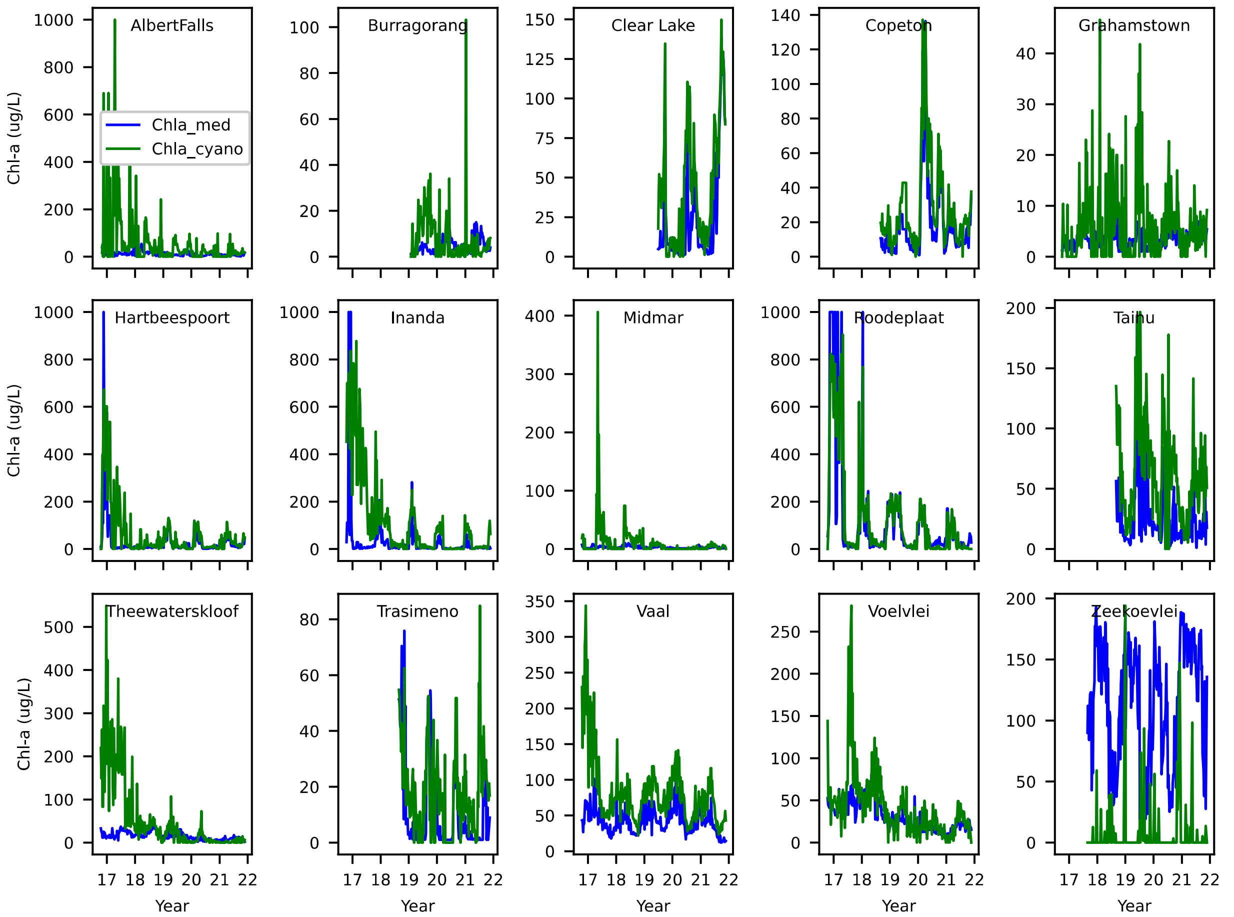


Figure 4.1 Timeseries for chla_med and chla_cyano variables re-sampled to weekly averages between 2017 to 2022. Note intentionally variable timeseries duration, and y-axis scale. Note also that the chla_cyano values are not whole-lake averages but are restricted.

Table 4.1 Description of Lakes for which timeseries were selected.

Name	Lat	Lon	Area (km²)	Chla_med (µg/L)	Chla_cyano (µg/L)	Start	End	Count
AlbertFalls	-29,44	30,40	31	10,1	50,0	2016-10-26	2021-11-15	749
Burragorang	-34,00	150,40	68	4,8	8,3	2019-01-25	2021-11-15	534
Clear Lake	39,04	-122,81	265,6	26,4	46,8	2019-06-27	2021-11-14	585
Copeton	-29,91	150,99	19,9	19,2	28,1	2018-08-27	2021-11-15	623
Grahamstown	-32,74	151,81	27	3,5	6,0	2016-10-02	2021-11-15	722
Hartbeespoort	-25,75	27,86	21,4	26,4	49,1	2016-10-14	2021-11-15	861
Inanda	-29,68	30,85	15,9	19,5	75,1	2016-10-15	2021-11-15	683
Midmar	-29,51	30,18	20,4	2,8	9,8	2016-10-15	2021-11-15	748
Roodeplaat	-25,63	28,36	4,6	92,5	92,9	2016-10-15	2021-11-15	778
Taihu	31,27	120,18	3033,7	25,8	62,4	2018-09-02	2021-11-15	435
Theewaterskloof	-34,03	19,20	33,3	15,1	44,3	2016-10-14	2021-11-15	922
Trasimeno	43,14	12,10	222,3	9,8	18,0	2018-08-23	2021-11-11	577
Vaal	-26,97	28,25	349,6	40,8	75,7	2016-10-15	2021-11-15	949
Voëlvlei	-33,37	19,04	21,5	28,0	37,0	2016-10-14	2021-11-15	913
Zeekoevlei	-34,06	18,51	2,6	115,7	9,0	2017-08-26	2021-11-15	675

4.2.3 Forecasting approach

Forecasts were derived at weekly (7-day), 2-weekly (14-day) and 4 weekly (28-day) time-horizons. These intervals are likely the most useful horizons for near-term forecasting for management and/or recreational purposes when combined with updated near real-time data (Rousso et al., 2020, Carey et al., 2020). The satellite timeseries are sporadic (not a regular interval) due to the presence of cloud and satellite acquisition schedules. The timeseries were aggregated (downscaled) from sporadic daily updates to weekly averages (mean) and backfilled (using prior values not interpolation) where data were incomplete.

The timeseries were split into training and test (hold out) datasets. The test dataset consisted of a 1-year timeseries from November 15, 2020, to November 15, 2021, while the training dataset consisted of all data prior to November 15, 2020. Forecasts were calculated using a rolling-origin technique for a period of one year using the training set to produce seasonal information and the test dataset to test forecast performance. In this manner, only past data were used to determine the seasonal components (lake 'climatology'). Forecast performance was assessed using the root mean square error (RMSE) which was determined as the optimal measure of performance for the type of data under consideration (other methods such as mean average error and mean average percent error were less appropriate for the data type and are therefore not shown). R-squared values (obtained by linear correlation), despite commonly being used to test the performance of lake-related forecasts, should not be used because they cannot measure bias in the forecast result.

Common univariate forecasting methods were used as a benchmark against which other models were assessed. The benchmarks included naïve, moving average, seasonal naïve, exponential smoothing, and trend adjusted exponential smoothing techniques (see Svetunkov, 2021 and Hyndman & Athanasopoulos, 2018). If these simple methods cannot be outperformed by more complex models, there is no justification to use more complex models, provided the performance of the simple models is satisfactory for the application in question (e.g. Peng & Chu, 2009)

The naïve forecast (denoted NA) was determined by:

$$\widehat{y_t} = y_{t-1}$$

where $\widehat{y_t}$ is the forecasted value at time t and y_{t-1} is the observed value of the previous period.

The simple moving average (MA) forecast was determined by:

$$\widehat{y_t} = \frac{1}{m}\sum_{j=1}^{m} \hat{y}_{t-j}$$

where *j* is the period over which the moving average is calculated, and *m* is the window of previous values (a value of 2 was used meaning that the average of the two previous observations was determined as the forecast value).

The seasonal naïve (SNA) forecast was determined by the average value of all previous observations in the training dataset for the period, *m* (in this case the period was a week) in question:

$$\widehat{y_t} = y_{t-m}$$

The simple exponential smoothing (ES) model was:

$$\hat{y}_{t+1} = \hat{\alpha} y_t + (1 - \hat{\alpha})\widehat{y_t}$$

where $\hat{\alpha}$ is the smoothing parameter between 0 and 1, and $\hat{y}_t$ is the forecast value for the previous period. The value of $\hat{\alpha}$ was determined as 0.7 by minimising the RMSE for the entire set of timeseries.

The trend-adjusted exponential smoothing (TAES) model was:

$$\hat{y}_{t+1} = \hat{\alpha} y_t + (1 - \hat{\alpha})\widehat{y_t}$$

$$T_{t+1} = \beta(\hat{y}_{t+1} - \widehat{y_t}) + (1 - \beta)T_t$$

$$TAF_{t+1} = \hat{y}_{t+1} + T_{t+1}$$

where T is the trend, and β is the trend adjustment factor usually ranging from 0 to 1, and TAF means trend-adjusted forecast. A value of zero was used for the initial trend T_t. Varying the values of the smoothing parameters between 0.3 and 0.7 produced only slight changes in the overall RMSE, therefore values of 0.5 were used.

It was essential that any advanced model outperformed or matched the performance of the above baseline models, while also being able to be used for forecasts more than one period ahead (i.e. 2-week and 4-week forecasts). A model based on classical decomposition and on a combination of the methods above was developed by combining the following logic:

1. The best predictor of future cyanobacteria biomass is the current inoculation of cells which is equal to the current value or moving average
2. Cyanobacteria exhibit strong annual seasonality that varies on a roughly weekly timescale, therefore, the value for the current week of year is likely to be similar to the value of the average of all previous years in that week (the lake 'cyanobacteria climatology')
3. There should be some accounting for the anomaly from the seasonal norm (i.e. deviation of the current value from the expected value based on the seasonal signal)

Using the above logic, a so-called moving average seasonal error adjusted (MASEA) model was developed based on a combination of logical decomposition (seasonality) and a moving average approach:

$$\hat{y}_t = \hat{\alpha} * MA + (1 - \hat{\alpha})(s_t + e)$$

$$e = MA - s_{t-1}$$

where s_t is the seasonal average, calculated as the mean of all previous observations for period *t*, and *e* is the seasonal anomaly calculated as the difference between the moving average and the seasonal average in the previous period, and $\hat{\alpha}$ is the weighting factor adjusting the contribution of the anomaly-adjusted seasonal value and the moving average.

A centralised moving average was used to compute the seasonal weekly component. The MASEA model is based on a combined moving average and seasonal approach and enables forecasting at time horizons greater than 1-week (in this case 2- and 4-week horizons). The weighting factor was adjusted to determine the optimal values for 1-week, 2-week and 4-week forecast horizons, and determined as 0.8, 0.7 and 0.6, respectively, with longer forecast-horizons being more heavily weighted towards the value of the seasonal average s_t.

4.3 RESULTS

4.3.1 Performance of baseline models

All of the baseline models, with the exception of seasonal naïve (SN) model, performed similarly for forecasting chla_cyano and chla_med on a weekly forecast horizon, with the exponential smoothing model performing the best overall (Table 4.2, Table 4.3). The SN model was significantly worse than models using the observed value of the previous period, giving weight to the theory that the current cyanobacteria inoculation is the primary forecast variable, with annual seasonality, in general, playing a lesser, but sometimes important role (SNA performed best for 2 or 3 lakes).

Table 4.2 Performance of baseline models measured by RMSE for forecasting mean chlorophyll-a from cyanobacteria on a 1-week horizon

Waterbody	na	ma	sna	es	taes	masea
AlbertFalls	30.6	24.9	54.4	26.1	26.8	25.4
Burragorang	20.3	17.5	12.0	17.8	18.5	16.8
Clear Lake	14.5	14.6	23.7	14.2	14.0	14.4
Copeton	6.7	6.7	17.2	6.4	6.6	7.0
Grahamstown	3.7	3.8	4.0	3.4	3.7	3.7
Hartbeespoort	22.8	22.7	51.8	21.6	22.2	22.0
Inanda	26.0	27.1	100.5	25.8	26.4	29.3
Midmar	3.0	2.6	12.9	2.7	2.7	2.8
Roodeplaat	40.1	44.0	117.5	40.6	42.4	39.6
Taihu	26.6	25.3	24.4	24.2	25.3	24.4
Theewaterskloof	5.6	3.9	51.8	4.6	4.6	5.6
Trasimeno	13.2	12.0	13.8	12.3	12.5	11.8
Vaal	16.6	15.2	27.3	15.0	15.2	14.8
Voëlvlei	9.2	9.2	26.4	8.7	8.9	9.7
Zeekoevlei	24.2	26.4	23.6	24.5	25.0	26.3
Mean RMSE	17.5	17.1	37.4	16.5	17.0	16.9

Table 4.3 Performance of baseline models measured by RMSE for forecasting median chlorophyll-a on a 1-week horizon

Waterbody	na	ma	sna	es	taes	masea
AlbertFalls	5.4	5.7	6.2	5.4	5.7	5.5
Burragorang	3.0	3.0	2.5	2.8	2.9	2.9
Clear Lake	13.1	15.1	28.0	13.9	14.0	15.2
Copeton	5.5	6.2	13.7	5.6	5.8	6.5
Grahamstown	1.6	1.4	1.4	1.4	1.5	1.4
Hartbeespoort	14.3	14.1	31.8	13.4	13.9	14.2
Inanda	11.0	11.2	45.4	10.4	10.8	12.3
Midmar	1.5	1.4	1.1	1.3	1.4	1.3
Roodeplaat	32.1	35.3	112.8	32.2	34.2	30.9
Taihu	18.7	18.1	18.0	17.3	18.0	17.8
Theewaterskloof	2.7	2.6	7.8	2.6	2.7	2.7
Trasimeno	4.7	4.4	7.9	4.4	4.4	4.1
Vaal	10.6	9.4	12.0	9.3	9.4	9.4
Voëlvlei	4.3	4.7	14.2	4.3	4.5	4.8
Zeekoevlei	28.7	28.4	45.9	27.3	27.8	28.0
Mean RMSE	10.5	10.7	23.3	10.1	10.5	10.5

4.3.2 Performance of MASEA model

When compared to the baseline models, the MASEA model performs slightly more poorly than the ES model, but no worse than the other baseline models. Although it would be ideal to have a model that significantly outperforms the baselines, it is not always feasible to achieve this even with very complex and advanced models (e.g. Peng & Chu, 2009). Given that more complex models were not evaluated in this study (owing to the application) a comparison with more complex models was not required.

The 1-week, 2-week and 4-week MASEA forecasts in general, correlate closely with observed values (see Table 4.4, Figure 4.2, Figure 4.3), although there is a noticeable lag in some cases between the observed values and those of the 4-week forecast. The 1-week forecast, with few exceptions, consistently has the lowest RMSE values, which become larger for the 2- and 4-week forecasts, respectively. The disadvantage with using a 4-week forecast that is heavily weighted on a moving average, is that the forecasted values may significantly underestimate larger changes (Figure 4.2, Figure 4.3). This is particularly evident for targets where there is less historical data for computing seasonal averages (e.g. Clear Lake, Copeton, Burragorang). In these cases, the model produces a more conservative 4-week estimate than the actual observed increases or decreases. However, even a conservative 4-week forecast may prove valuable, assuming that there is enough historical data to characterise the seasonal changes.

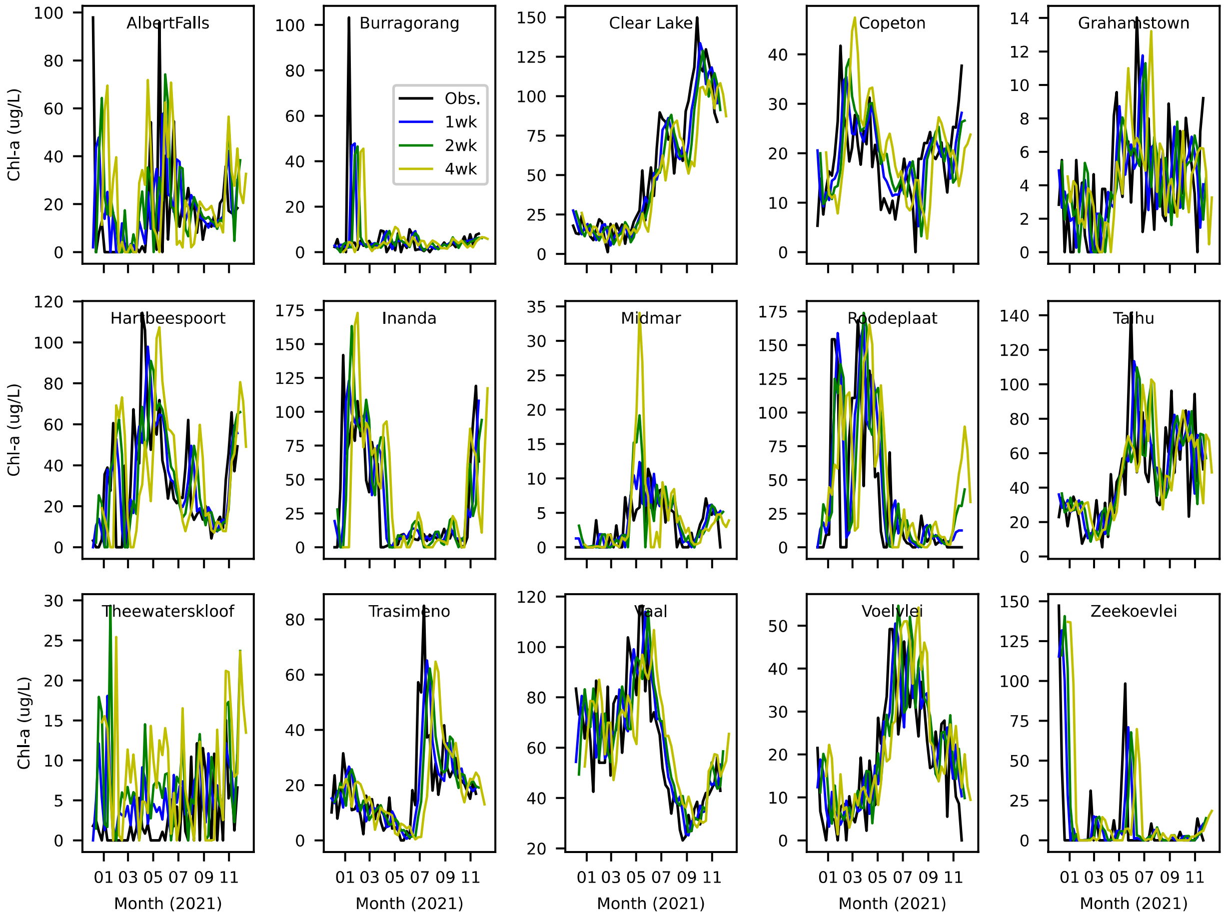


Figure 4.2 Rolling-origin forecasts at 1-week, 2-week and 4-week horizons for chla_cyano

With respect to chla_cyano, the 4-week forecast was overall 5 µg/L less accurate than a 1-week forecast (RMSE of 22.7 vs 16.9 µg/L). Similarly, for chla_med, the 4-week forecast accuracy was overall 3 µg/L less accurate than the 1-week forecast (RMSE of 13.5 vs 10.5 µg/L, respectively). This indicates that the 4-week forecast, although somewhat insensitive to sudden or larger changes, does account for seasonal changes enough for forecast performance not to drop off significantly from a 1-week horizon.

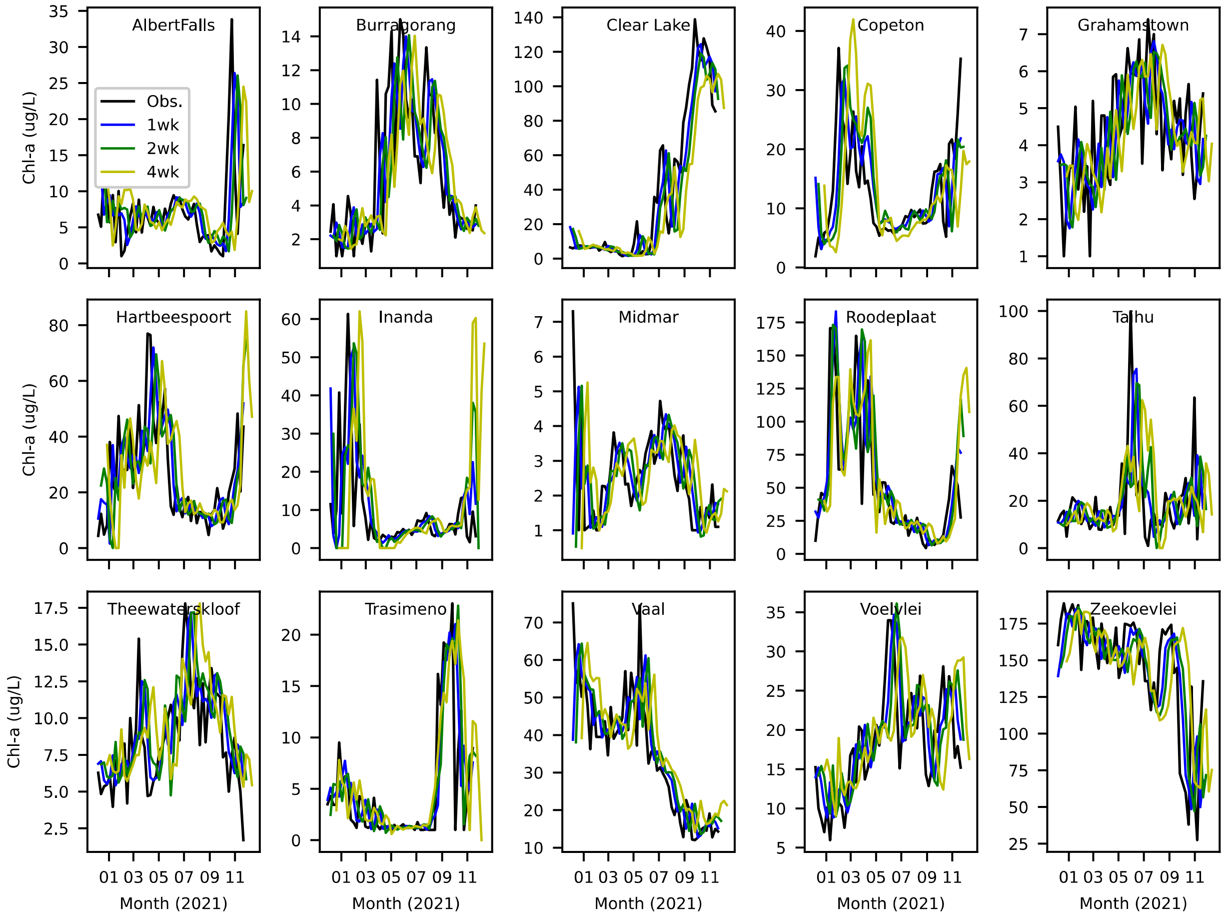


Figure 4.3 Rolling-origin forecasts at 1-week, 2-week and 4-week horizons for chla_med

Table 4.4 Performance of MASEA forecasts for 1-week, 2-week and 4-week horizons for chla_cyano and chla_med (RMSE, units µg/L). Normalised 1-week RMSE values are shown on right. Colour scale indicates lower (green) to higher (red) RMSE values.

	Chla_cyano			Chla_med				chla_cyano	chla_med
Waterbody	**1wk**	**2wk**	**4wk**	**1wk**	**2wk**	**4wk**		**1wk norm.**	**1wk norm.**
AlbertFalls	25.4	21.1	27.0	5.5	6.5	6.2		0.51	0.55
Burragorang	16.8	16.9	16.8	2.9	3.1	3.2		2.03	0.60
Clear Lake	14.4	17.5	23.1	15.2	19.6	24.2		0.31	0.57
Copeton	7.0	8.8	11.8	6.5	8.2	10.9		0.25	0.34
Grahamstown	3.7	3.8	3.4	1.4	1.5	1.4		0.62	0.40
Hartbeespoort	22.0	26.7	34.5	14.2	16.3	17.9		0.45	0.54
Inanda	29.3	37.5	41.0	12.3	14.4	19.0		0.39	0.63
Midmar	2.8	3.7	6.1	1.3	1.0	1.4		0.28	0.48
Roodeplaat	39.6	50.3	56.2	30.9	37.0	35.4		0.43	0.33
Taihu	24.4	24.9	24.5	17.8	18.8	19.6		0.39	0.69
Theewaterskloof	5.6	7.6	8.9	2.7	3.3	3.4		0.13	0.18
Trasimeno	11.8	15.2	19.7	4.1	4.0	4.3		0.66	0.42
Vaal	14.8	15.7	17.9	9.4	8.3	9.9		0.20	0.23
Voëlvlei	9.7	10.9	12.4	4.8	6.1	7.5		0.26	0.17
Zeekoevlei	26.3	33.7	37.7	28.0	31.1	38.1		2.92	0.24
AVERAGE	16.9	19.6	22.7	10.5	11.9	13.5			

4.3.3 Performance for forecasting indices

The 1-week horizon forecast ability for cyanobacteria risk level and trophic state had a 74% and 75% accuracy, respectively (Table 4.5). When considering the range of concentrations of the waters under consideration, the result is encouraging. For the 4-week forecast horizon, the forecasting ability drops off to 66% for CRL and 71% for TS, respectively. The finding that a simple univariate forecasting technique can forecast CRL and TS with more than 50% accuracy 4-weeks in advance lends weight to the approach presented in this study.

Table 4.5 Performance of MASEA forecasts (in percentage agreement between observed and forecasted value) for 1-week, 2-week and 4-week horizons for CRL, TS and high CRL.

	CRL				TS				High CRL			
	1wk	2wk	4wk		1wk	2wk	4wk		N	1wk	2wk	4wk
AlbertFalls	66.7	50.0	50.0		78.4	84	79.2		4	0	0	25
Burragorang	94.1	94.0	89.6		80.4	78	70.8		1	0	0	0
Clear Lake	72.5	72.0	75.0		72.5	76	72.9		24	87.5	91.7	83.3
Copeton	82.4	84.0	75.0		62.7	62	56.2		0			
Grahamstown	94.1	94.0	93.8		100	100	100		0			
Hartbeespoort	54.9	50.0	39.6		54.9	56	52.1		10	60	50	40
Inanda	51.0	48.0	56.2		72.5	74	64.6		15	73.3	60	66.7
Midmar	90.2	90.0	89.6		100	100	100		0			
Roodeplaat	49.0	40.0	37.5		62.7	66	52.1		14	85.7	64.3	71.4
Taihu	70.6	68.0	60.4		51	46	45.8		19	89.5	78.9	73.7
Theewaterskloof	82.4	74.0	58.3		80.4	78	72.9		0			
Trasimeno	72.5	70.0	68.8		88.2	90	87.5		3	33.3	0	0
Vaal	82.4	78.0	70.8		62.7	74	64.6		30	96.7	90	86.7
Voëlvlei	72.5	70.0	64.6		64.7	68	52.1		0			
Zeekoevlei	74.5	72.0	66.7		92.2	94	95.8		3	33.3	0	0
AVERAGE	74.0	70.3	66.4		74.9	76.4	71.1					

4.3.4 Impact of lake size and timeseries length

There was only weak evidence for the hypothesis that forecast accuracy for chla_med is inversely related to lake size given that larger lakes exhibit more heterogeneity (Figure 4.4). In general, forecast accuracy (for chla_med and chla_cyano) represented by normalised RMSE was not significantly affected by the size of the target. In contrast, it was evident that the length of historical time-series, represented by the number of images, had a significant effect on forecasting skill of chla_med ($R^2 = 0.45$, Figure 4.4), with longer timeseries exhibiting lower normalised RMSE values. A similar relationship, although weaker, was also observed for chla_cyano. This confirms the hypothesis that longer time-series leads to increased forecasting skill owing to improved characterisation of the underlying seasonal component used by the MASEA model.

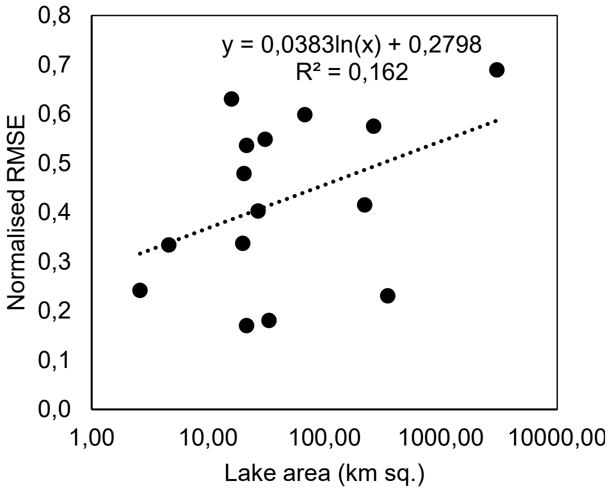


Figure 4.4 Relationship between 1-week forecast accuracy for chla_med and log-scaled lake size (left) and the length of the time-series (historical data) (right).

4.4 DISCUSSION & CONCLUSION

The results demonstrate the feasibility of using simple univariate forecasting methods with spatially aggregated statistical variables derived from satellite remote sensing to forecast future values of chlorophyll-a and cyanobacteria concentrations across a variety of lakes of different size and types, with various lengths of historical data. Here I discuss the accuracy of the forecasts, the factors affecting forecast performance, variations on the MASEA model, and how the approach may be implemented in practice.

4.4.1 Forecast accuracy for median whole-lake chlorophyll-a

On average, the 1-week forecast for chla_med can be estimated with an error of 10.5 µg/L, although the forecast error must be contextualized by the trophic state (or typical range of chlorophyll-a values) for each waterbody. For an oligotrophic system such as Grahamstown with typical values between 1 and 10 µg/L and only occasional exceedances, a forecast error was only 1.4 µg/L. For a hypertrophic system such as Roodeplaat, which regularly exhibits chlorophyll-a concentrations exceeding 100 µg/L, the forecast accuracy was 30.9 µg/L, which also appears adequate. As demonstrated by these examples, the forecast performance is lake-specific, and the values must be considered with reference to the range of values typical for the waterbody under examination. Examining the RMSE normalised for each waterbody by its average chl-a concentration, the overall 1-week forecast accuracy was 42%. This value is more representative of typical forecast performance. Examining the above two examples, accuracies were 40% for Grahamstown and 33% for Roodeplaat, respectively. This value can be contextualised by the typical error on chl-a in situ measurements, which is between 10 and 30%. Forecasting the chlorophyll-a concentration 1-week ahead with a 40% error may meet the minimum requirements for applications such as ecological observation, but it is unclear whether this would meet the requirements for other applications such as water treatment or issuing of recreational advisories.

4.4.2 Forecast accuracy for cyanobacteria

Cyanobacteria (chla_cyano) can be forecast with an average error of 16.9 µg/L with a 1-week forecast horizon, which increases to 22.7 µg/L for a 4-week horizon. Contextualizing this value by normalising the result with the average chla_cyano value for each of the waterbodies, the normalised error can be expressed as 65% for a 1-week forecast. It is relevant to examine the ratio of chla_med to chla_cyano to understand the prevalence of cyanobacteria blooms relative to algal blooms within each of the waterbodies. Waterbodies that exhibited the lowest ratios (e.g. Zeekoevlei, Burragorang, Grahamstown) had significantly larger 1-week forecast errors (292%, 203%, 62%, respectively). By contrast, the lakes with higher concentrations and prevalence of cyanobacteria, such as Vaal, Copeton and Theewaterskloof, had significantly smaller 1-week normalised forecast errors (20%, 25%, 13%). This suggests that forecasting of cyanobacteria is more successful in lakes where cyanobacteria blooms occur more often and at higher concentrations. Forecasting in lakes with few cyanobacteria detections and low concentrations is understandably associated with larger errors.

How do these forecasting performances compare with alternative forecasting approaches? Based on a literature review of forecasting models, the primary measure of performance is the correlation coefficient, or R^2, which is unable to measure bias or provide a quantitative measure of model performance (Roussou et al., 2020). The use of R^2 as the primary measure of model performance seems to be a major oversight with publications in the domain as it precludes quantitative assessment of forecast errors in cells/mL or µg/L for cell counts or toxin concentrations. Nevertheless, based on the data presented in Roussou et al., the R^2 values for a 7-day time horizon for various advanced algorithms reached a maximum value of ± 0.875 with typical values near 0.6 (see Fig 6). The models therefore account for a maximum of 87%, but usually around 60%, of the variability in cyanobacteria cell counts or microcystin toxin concentrations. The only results from a 30-day (4-week) horizon had R^2 near 0.4. It is not immediately clear how the MASEA model compares to these models, however, the 65% mean forecast error for chla_cyano appears comparable to the 60% accounting of variability in these complex models.

The multivariate model presented by Myer et. al (2021) based on remotely sensed data had a 1-week forecast accuracy of 82% (n = 103 lakes) when forecasting whole-lake spatially aggregated high-risk cyanobacteria bloom probabilities (cells > 100 000 cells/mL). They found that there was a high degree of autocorrelation (90%), indicating that the largest factor in predictive success is the previous week's value (which is one of the primary underpinnings of the present study). They also expected their model to remain accurate for a time-horizon of no more than 2 weeks, but no performance was not assessed for longer time horizons. By comparison, the current approach, which is vastly simpler than the one presented by Myer et al, the 1-week CRL forecast accuracy across all cyanobacteria risk level scenarios (low to very high) over a test dataset of 1 year (52 weeks) was 74% (n = 780). When assessing the 1-week forecast accuracy for high or very high risk scenarios (chl-a > 50 µg/L or 100 000 cells/mL), the overall accuracy was 80% (98 of 123 events detected across all lakes), however the lake-specific accuracy varied widely depending upon the waterbody in question (Table 4.5). The 2-week and 4-week accuracy dropped off to 71% and 69%, respectively. Using this metric, the MASEA 1-week forecast performance for high-risk cyanobacteria blooms is comparable to that of Myer et al (2021).

4.4.3 Factors affecting forecast performance

Longer forecast horizons (4 weeks) that depend more on a seasonal characterisation are often only slightly less accurate than a 1-week forecast horizon, depending upon the waterbody in question.

There was only weak evidence to suggest that a spatially aggregated statistical approach may be less suitable for forecasting larger waterbodies. This lends some confidence that the approach may be valid for both large and small waterbodies. There was stronger evidence that longer time-series used to characterize the underlying seasonal component led to increased forecast accuracy. Those lakes for which there were more than 900 historical images (around 4 years of data) had the lowest normalised RMSE. By contrast, those lakes less than 600 images, had the worst or near-worst forecast performance. This implies that 1) the seasonal component should be taken into account when forecasting, and 2) that longer historical time-series lead to improved forecast accuracy.

4.4.4 Variations on the MASEA model

The MASEA model is a simple moving average model that adjusts for the seasonal average and the observed anomaly. Alternatives to the selected model were tested, including a simpler model that included no error adjustment, and a variation using the seasonal probability of a cyanobacteria bloom. However, not all combinations of model possibilities were tested, and it is likely that it could be improved upon with alternative models, or by selecting more advanced approaches that are also computationally simple for operational usage. As we specifically excluded the use of more advanced models in this study, it would be interesting to take advantage of machine learning models in future.

4.4.5 Operational considerations for forecasts

Practically applying a model for forecasting blooms without prior knowledge of the system in question is feasible, however there are some operational limitations. Firstly, a forecast for which there is no recent data due to prolonged cloud cover would effectively revert to a seasonal naïve forecast with considerably less precision. Further, forecasting for waterbodies where there is less or no historical data would revert to a MA model, also with significantly worse predictive ability especially for 2 and 4-week horizons. It would appear from the analysis above, that 900 images (or 5 years of data) provided more optimal seasonal characterisation. This implies that long timeseries need to be analysed (now easily performed using cloud-based computation) for a lake before forecasts become more accurate. One additional challenge is that many lakes are frozen for up to 6 months or more in a calendar year. Whilst this is not a problem per se given that seasonal characterisation is on a week-by-week basis, some account has to be taken not to provide forecasts for frozen lakes as the freeze date may vary significantly from year to year. Lastly, an important limitation is that the forecast is only as accurate as the remotely sensed data, which may be prone to significant errors and occasional anomalies.

CHAPTER 5: CONCLUSIONS & RECOMMENDATIONS

5.1 CONCLUSIONS

- A novel mobile application was developed for the Android and Apple platforms, providing free near real-time updates on cyanobacteria health risks and eutrophication for lakes worldwide from satellite remote sensing data. The mobile app is the only app on the app store provide daily updates for lakes worldwide and represents a significant technological innovation.

- Novel products for Sentinel-2 MSI have been developed for chlorophyll-a detection, bloom detection and land / water separation. The high-resolution imagery provides supplemental information to the medium resolution time series data used for the information provided by the mobile app. Sentinel-2 imagery is available to paid users via the mobile App.

- A simple univariate forecasting approach based on timeseries of satellite data forecast high risk cyanobacteria blooms with 80% accuracy for a 1-week time horizon for a test set of 15 waterbodies. This is the first time that such an approach has been used to forecast cyanobacteria and algal blooms in lakes, and the results show convincing evidence that the approach is robust, computationally simple, and applicable to a range of waterbodies across the globe.

- The mobile App has more than 1500 installs across iOS and Android devices, with the most users in South Africa, followed the US, Sweden, Canada and Europe. The user base is expected to grow significantly in the coming years, especially as further uptake in the US, Europe and Australia is encouraged.

- The project provided internship opportunities for two 3rd year computer science students, providing them valuable experience and training in an operational computer coding environment. In addition, the project has seen the development of valuable, novel technologies that will be exploited by CyanoLakes.

- The App is currently in a state of pre-revenue generation, and up to 13 000 users are required to break even based on the current ratio of paid to unpaid users and pricing model.

- An open API similar to weather APIs was developed allowing information to be retrieved using only a location information (latitude and longitude) and made publicly available

5.2 RECOMMENDATIONS

- The mobile app should be promoted through paid advertising, social media, and conferences to increase user uptake, especially in the US, Europe and Australia in order to increase the number of users, and especially paying users to ensure financial sustainability.

- The app should be maintained and improved after the project to ensure its longevity.

- Ultra-premium features should be developed for paid users, including charts, historical time series of satellite imagery, integration of weather data, and incorporation of more advanced forecasting models to provide hourly / sub daily forecasts, and point-based station monitoring, etc.

- Alternative revenue streams should be explored including advertising and the commercial user sector with ultra-premium features

- The forecast model should be refined, further developed, and applied to 1000 lakes worldwide to increase robustness

- Additional remote sensing data streams from additional sensors should be considered for integration in the mid to long term (3 years +)

REFERENCES

Acharya, T. D., Subedi, A., & Lee, D. H. (2018). Evaluation of water indices for surface water extraction in a Landsat 8 scene of Nepal. *Sensors*, *18*(8), 2580.

Beal, M. R., O'Reilly, B., Hietpas, K. R., & Block, P. (2021). Development of a sub-seasonal cyanobacteria prediction model by leveraging local and global scale predictors. *Harmful Algae*, *108*, 102100.

Caillault, É. P., & Bigand, A. (2018, September). Comparative Study on univariate forecasting methods for meteorological time series. In *2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 2380-2384). IEEE.

Carey, C. C., Woelmer, W. M., Lofton, M. E., Figueiredo, R. J., Bookout, B. J., Corrigan, R. S., ... & Thomas, R. Q. (2021). Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*, 1-14.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., & Li, X. (2016). Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, *8*(4), 354.

Echard, J. S. (2021). A Review of Harmful Algal Bloom Prediction Models for Lakes and Reservoirs.

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, *140*, 23-35.

García-Mozo, H., Chuine, I., Aira, M. J., Belmonte, J., Bermejo, D., de la Guardia, C. D., ... & Galán, C. (2008). Regional phenological models for forecasting the start and peak of the Quercus pollen season in Spain. *agricultural and forest meteorology*, *148*(3), 372-380.

Gower, J., King, S., & Goncalves, P. (2008). Global monitoring of plankton blooms using MERIS MCI. *International Journal of Remote Sensing*, *29*(21), 6209-6216.

Hu, C. (2009). A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment*, *113*(10), 2118-2129.

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 22 November 2021.

Ji, L., & Peters, A. J. (2004). Forecasting vegetation greenness with satellite and climate data. *IEEE Geoscience and Remote Sensing Letters*, *1*(1), 3-6.

Li, H., Qin, C., He, W., Sun, F., & Du, P. (2021). Improved predictive performance of cyanobacterial blooms using a hybrid statistical and deep-learning method. *Environmental Research Letters, **16,** 124045*.

Matthews, M. W., Bernard, S., & Robertson, L. (2012). An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sensing of Environment*, *124*, 637-652.

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.

Mobley, C. D. (1994). *Light and water: radiative transfer in natural waters*. Academic press.

Myer, M. H., Urquhart, E., Schaeffer, B. A., & Johnston, J. M. (2020). Spatio-temporal modeling for forecasting high-risk freshwater cyanobacterial harmful algal blooms in Florida. *Frontiers in Environmental Science*, 202.

Peng, W. Y., & Chu, C. W. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and computer modelling*, *50*(7-8), 1045-1057.

Recknagel, F., Orr, P., Swanepoel, A., Joehnk, K., & Anstee, J. (2018). Operational Forecasting in Ecology by Inferential Models and Remote Sensing. In *Ecological Informatics* (pp. 319-339). Springer, Cham.

Recknagel, Friedrich, et al. "Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling." *Harmful algae* 69 (2017): 18-27.

Rousso, B. Z., Bertone, E., Stewart, R., & Hamilton, D. P. (2020). A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Research*, *182*, 115959.

Svetunkov, A. (2021). Forecasting and analytics with ADAM. https://www.openforecast.org/adam/. Accessed on: 22 November 2021.

Swanepoel, A. (2015). *Early warning system for the prediction of algal-related impacts on drinking water purification* (Doctoral dissertation, North-West University).

Swanepoel, A., Barnard, S., Recknagel, F., & Cao, H. (2016). Evaluation of models generated via hybrid evolutionary algorithms for the prediction of Microcystis concentrations in the Vaal Dam, South Africa. *Water SA*, *42*(2), 243-252.

Tromas, N., Fortin, N., Bedrani, L., Terrat, Y., Cardoso, P., Bird, D., ... & Shapiro, B. J. (2017). Characterising and predicting cyanobacterial blooms in an 8-year amplicon sequencing time course. *The ISME journal*, *11*(8), 1746-1763.

White, M. A., & Nemani, R. R. (2006). Real-time monitoring and short-term forecasting of land surface phenology. *Remote Sensing of Environment*, *104*(1), 43-49.

APPENDIX A CAPACITY BUILDING & DISSEMINATION

Date	Description	Location	Person / Organisation
19th May 2020	2020 LIFT Webinar Series: Source Water Quality	Online	Aaron Fisher WERF
April 19-23, 2021	12th National Monitoring Conference (USA)	Online	North American Lake Management Society
May 2021	Internship	CyanoLakes Offices	Ian Edwards
17 August 2021	Mobile App Launch Event	Online	Hosted by the WRC
October 2021	World Water Tech North America	Online	Featured start-up at the event
November 2021	Internship	CyanoLakes Offices	Mukundi Chitamba
November 2021	GEO Week 2021	Online	Hosted by GEO
November 2022	Science to Policy: From EO to Legislation Webinar and Podcast	Online	International Water Association
November 2022	Sydney Water Innovation Festival (Urban Plunge Tech Summit)	Sydney, Australia	Sydney Water
November 2022	In Person Conference Presentation and Expo	Minneapolis	North American Lakes Association

Publications and Conference Proceedings

Matthews, M. W. (2022). The CyanoLakes mobile app: Weather-like information for lakes from satellite imagery. *Water Wheel*, *21*(4), 32-34.

William Matthews, M. (2022). Near-term forecasting of cyanobacteria and harmful algal blooms in lakes using simple univariate methods with satellite remote sensing data. *Inland Waters*, 1-29.

Matthews, M.W & Kravitz, J.A. (2021). Cyanobacteria chlorophyll-a from Sentinel-2. Available from: https://custom-scripts.sentinel-hub.com/sentinel-2/cyanobacteria_chla_ndci_l1c

Matthews, M.W & Kravitz, J.A. (2021). The Maximum Peak Height Bloom Index. Available from: https://custom-scripts.sentinel-hub.com/sentinel-2/maximum_peak_height_bloom_index

Kravitz, J., Matthews, M., Lain, L., Fawcett, S., & Bernard, S. (2021). Potential for high fidelity global mapping of common inland water quality products at high spatial and temporal resolutions based on a synthetic data and machine learning approach. *Frontiers in Environmental Science*, *9*, 587660.

Matthews, 2022. A Mobile App for Cyanobacteria Health Risk Monitoring and Forecasting Based on Satellite Imagery. In Person Presentation at the North American Lakes Management Society, Minneapolis, November 14-17, 2022.

Abstract

Cyanobacteria blooms represent a considerable public health risk, and operational concern for water treatment, across the United States and globally. Rising temperatures and increasing nutrient concentrations through anthropogenic pollutants are increasing the incidence and severity of harmful algal blooms globally. There is a need for rapid, synoptic monitoring driven by cutting-edge algorithms with AI, combined with mobile technologies to bring near real-time information to water treatment operators, catchment managers and the public to ensure human health concerns are adequately addressed. This talk demonstrates a novel mobile application, and novel methods for cyanobacteria detection and near real-time forecasting, to produce weather-like information for lakes across the United States and globally. The utility of the App for water companies and the public is demonstrated through case studies. A novel forecasting approach based on a logical decomposition and univariate forecasting methods is demonstrated that enables cyanobacteria risk levels can be forecast at 1-week, and up to 4-week time horizons with up to 75% accuracy. Novel algorithms applied to high-resolution satellite imagery (Sentinel-2 multi-spectral imagery) are presented for chlorophyll-a detection in cyanobacteria blooms, and high-resolution imagery is exploited to visualize bloom progression in near real-time. The App is demonstrated for forecasting with case studies for Clear Lake in California, known for severe annual cyanobacteria outbreaks.

Knowledge Dissemination Activities

Social media	CyanoLakes Twitter Account (1000+ followers) CyanoLakes Mailing List (350+ subscribers) CyanoLakes Facebook Account
Advertising	Google AdWords Campaigns
Commercial Pilot Users	Rand Water City of Cape Town Umgeni Water

Internship & Scholarship Opportunity

CyanoLakes (www.cyanolakes.com) is a small company using satellite imagery to help water utilities monitor harmful algal blooms in their water storages. CyanoLakes, funded by the Water Research Commission (www.wrc.org.za), is offering an internship and scholarship opportunity to eligible 3rd year computer science students at the University of Cape Town. Only 3rd year students should apply.

The internship will provide the opportunity to work in a small team with a technology stack including python, Django web framework, JavaScript and html (in a docker environment) and git version control software. The week-long internship, funded by the Water Research Commission, will include a scholarship to the value of R50 000 that will be used to settle any outstanding tuition fees, with the remainder being paid to the student for living expenses. The aim of the internship is to provide the student with the opportunity to learn and grow their skills and capabilities.

Requirements:
You must be a 3rd year BSc Computer Science student currently registered at the University of Cape Town with excellent academic performance. The student must demonstrate a level of proficiency in the listed technologies and have a willingness to learn. The student must be available to work at the CyanoLakes office from 3rd to the 7th of May 2021 (1st vacation) in the southern suburbs using their own means of transport. COVID protocols (sanitizing and mask wearing) will apply. Preference will be given to students from low-income households.

How to apply:

Applicants will be required to complete a series of skills assessments that will count towards their application and an online form. To access the application form, please email sandra@cyanolakes.com from your UCT email account.

The closing date for all applications is 15 April 2021.

Only the successful candidate will be contacted on or before the 23rd April 2021.

In case that no suitable candidate can be identified, CyanoLakes reserves the right to not offer the internship or scholarship.

MSc / PhD in Satellite Remote Sensing Phenology and Forecasting

CyanoLakes (Pty) Ltd. is seeking a Masters or Doctoral candidate to conduct research on a Water Research Commission (WRC) funded project: The Earth Observation National Eutrophication Monitoring Programme Phase II. The candidate will conduct research under the co-supervision of Dr. Mark Matthews at CyanoLakes (Pty) and must be registered for a MSc or PhD program at a South African University.

Background and Topic:
Satellite remote sensing data provides the ideal dataset to derive land surface phenology related to the timing and seasonality of biological events. Cyanobacteria blooms in lakes are a global concern of increasing severity. Your task will be to develop a phenology-based predictive model using remotely sensed data for forecasting cyanobacteria and algal blooms in lakes and reservoirs using satellite remote sensing data. You will develop methods to forecast cyanobacteria blooms based on predictive models and climatology. Your methods will be applied in the CyanoLakes mobile application.

Eligibility:
You must have a bachelors or honours degree. Your background must include statistics, computer science or programming, and applied physical environmental sciences. Candidates in the climate and related sciences are encouraged to apply. You will be required to read and write code in Python Programming Language and contribute to CyanoLakes' online code repositories. The candidate must demonstrate a high degree of proficiency in the required skills. Knowledge of water remote sensing is not a requirement. Only shortlisted candidates will be contacted.

Funding:
Funding will be conditional upon successful application to the The Water Research, Development, and Innovation (RDI) Roadmap Postgraduate Student Support Call. Support will be 2 years for a MSc, and 3 years for a PhD (up to R194 000 per year).

Start Date:
You should be registered on or before 15 February 2020.

How to apply:
To apply please forward a brief CV, certified academic transcripts and a letter of motivation to mark@cyanolakes.com.

APPENDIX B DISCLAIMER & TERMS OF USE

Disclaimer

This Application provides information on the risk to human health from cyanobacteria blooms in lakes and reservoirs derived exclusively from satellite remote sensing imagery. It does not provide any information related to other important water quality indicators such as dissolved metals, other microorganisms or contaminants. The information provided through this Application should be used in conjunction with other available water quality information before determining whether or not a waterbody is safe for recreational or other use.

Recreational advisories are provided only with respect to cyanobacteria and algae concentrations. The cyanobacteria cell counts, and potential *microcystin* toxin concentrations provided are based on accepted conversions using World Health Organisation standards and guidelines.

Use of the information provided is entirely at your own risk and this Application is provided without warranty. CyanoLakes (Pty) Ltd., its employees and affiliates are not liable for any loss or damage resulting from decisions made on the basis of the information provided.

All information should be considered preliminary until verified by field observations or measurements. Always follow advisories issued by your local authority.

By accessing and using this Application, you agree that you have read, understood and agree to be bound by the Terms of Use available from https://www.cyanolakes.com/terms-of-use/.

Terms of Use

https://www.cyanolakes.com/terms-of-use/

These are the Terms of Use (“Terms”) which you (“the User”) are required to accept when subscribing to and using the CyanoLakes Application (“the Application”) operated by CyanoLakes (Pty) Ltd. (“CyanoLakes”, “we”, “us”). By accessing and using the CyanoLakes App, you acknowledge that you have read, understood, and agree to be bound by these Terms and the Privacy Policy. If you do not accept these Terms, you are not authorized to use the Application. In order to improve the services we provide, we reserve the right to modify or update these Terms, which will become effective at the time of update.

Privacy

You may opt to grant CyanoLakes access to your current location in order to retrieve nearby waterbodies. By using the App, you agree that CyanoLakes may collect and use this information as set forth in our Privacy Policy at www.cyanolakes.com/privacy-policy.

Subscriptions

You may choose to subscribe (“Subscription”) to access premium features in the Application. You will be billed in advance on a recurring basis for the duration of your choice (monthly or annually).

Your Subscription will be charged to your Apple or Google Play account at confirmation of purchase. Your Subscription will automatically renew at the end of your Subscription period unless cancelled at least 24 hours prior to the end of the current period. Your account will be charged for renewal within 24 hours prior to the end of the current period.

You can manage your subscription or turn off auto-renewal at any time from your iTunes or Google account settings. Cancellation of the active subscription period is not allowed.

All subscriptions may be final and non-refundable. By accepting these terms, you agree that CyanoLakes in not required to provide a refund for any reason.

CyanoLakes may change, modify or remove features of the Subscription at any time, with or without notice to you. This includes changing the source of satellite imagery provided through the app as a result of a change in pricing or the terms of satellite imagery providers. In this instance, you agree that CyanoLakes will not be liable to you or any third party.

CyanoLakes, at its sole discretion and at any time, may modify the fees charged for Subscriptions. Subscription fee changes will require you to opt-in to the new fee to continue using the Subscription at the end of the then-current Subscription period.

Intellectual property

The CyanoLakes Application is owned and operated by CyanoLakes. The visual interfaces, graphics, design, compilation, information, computer code, products, software, services and all other elements of the Application are protected by international copyright and intellectual property law.

All material contained on the Application are the property of CyanoLakes and/or third-party licensors. All suggestions, enhancement requests, feedback, recommendations or other input provided by the User shall be owned by CyanoLakes.

The User grants CyanoLakes, a worldwide, perpetual, irrevocable, royalty-free license to use and incorporate said suggestions, enhancements, feedback and/or recommendations into the Application. Unless authorized by CyanoLakes in writing, you may not sell, license, distribute, copy, modify, publicly perform or display, transmit, publish, edit, adapt, create derivative works from, or otherwise make unauthorized use of the materials.

Limitation of liability

CyanoLakes, its directors, employees or agents shall not be liable for any damages or loss whatsoever, including without limitation any direct, indirect, special, incidental, consequential or punitive damages or loss, howsoever arising (whether in an action arising out of contract, statute, delict or otherwise) related to the use of, or the inability to access or use the Application or any information, or inaccuracies, defects, errors, whether typographical or otherwise, omissions, out of date information or otherwise. Excluded consequential and indirect loss and damage will include but not be limited to loss of profits, loss of goodwill, and wasted expenditure.

Disclaimer and Warranties

Use of the Application is at your sole risk. The Application is provided “as is” and without warranties of any kind (express or implied, including, but not limited to, fitness for a particular purpose).

CyanoLakes does not warrant that a) the information will function uninterrupted, or be available at a specific time or location, b) any errors or defects present in the Application will be corrected, c) the information provided through the Application will meet your requirements.

The Application relies upon information collected by satellites. The latency (delivery time) and availability (coverage) of information are affected by natural weather phenomenon, and may be interrupted as a result of maintenance, satellite transmission failure, instrument failure, internet connectivity failure, computer malfunctions, cyber-attacks, power failures or Acts of God. The User indemnifies and holds CyanoLakes and its employees harmless against any loss or damages that may result from such interruptions.

The User acknowledges that variability beyond the control of CyanoLakes or its employees, related to, but not limited to, natural phenomenon, satellite instrument performance or calibration, may result in errors being present in the information provided. The User agrees that the information provided is preliminary in nature and must be verified through observations or measurements made on the ground. The User acknowledges that CyanoLakes and its employees will not be liable for any loss or damages they may suffer as a result of its reliance on the data obtained from CyanoLakes.

This Application provides information on the risk to human health from cyanobacteria blooms in lakes and reservoirs. It does not provide any information related to other important water quality indicators such as dissolved metals, other microorganisms or contaminants. The User acknowledges that the recreational advisories provided through the Application are provided only with respect to cyanobacteria and algae concentrations, and no other important water quality variables. The cyanobacteria cell counts and potential microcystin toxin concentrations provided through the Application are based on accepted conversions using World Health Organisation standards and guidelines.

The User will therefore be responsible to use the information provided through this Application in conjunction with other available water quality information, that may or may not be available, before determining whether or not a waterbody is safe for recreational or other use.

Termination
You agree that CyanoLakes may terminate your access to the Application, at our sole discretion, without cause or notice, and without penalty, at any time. CyanoLakes may also at any time discontinue providing access to the Application, or any part thereof, with or without notice to you. You agree that termination of your access to the Application or any portion thereof may be affected without prior notice, and you agree that CyanoLakes will not be liable to you or any third party for any such termination. Upon termination, your right to use the Application will immediately cease. If you wish to terminate your account, you may uninstall the Application from your device, or discontinue using the Application.

Update of these Terms
CyanoLakes may update these Terms, from time to time, due to changes in the services we offer or other circumstances. The User agrees to familiarise themselves with these Terms, as revised from time to time, and your continued use of the Application shall indicate your acceptance of any revised Terms.

Governing Law
Any claim related to these Terms will be governed by and construed and interpreted in accordance with the laws of the Republic of South Africa.

Contact
If you have any questions regarding these Terms, please email support@cyanolakes.com.

Last Update
14 May 2021

Privacy Policy
https://www.cyanolakes.com/privacy-policy/

CyanoLakes (Pty) Ltd. built the CyanoLakes app as a Commercial app. This SERVICE is provided by CyanoLakes (Pty) Ltd. and is intended for use as is. This page is used to inform visitors regarding our policies with the collection, use, and disclosure of Personal Information if anyone decided to use our Service. If you choose to use our Service, then you agree to the collection and use of information in relation to this

policy. The Personal Information that we collect is used for providing and improving the Service. We will not use or share your information with anyone except as described in this Privacy Policy.
The terms used in this Privacy Policy have the same meanings as in our Terms and Conditions, which is accessible at CyanoLakes unless otherwise defined in this Privacy Policy.

Information Collection and Use
For a better experience, while using our Service, we may require you to provide us with certain personally identifiable information. The information that we request will be retained by us and used as described in this privacy policy. The app does use third party services that may collect information used to identify you. Link to privacy policy of third party service providers used by the app
Google Play Services

Log Data
We want to inform you that whenever you use our Service, in a case of an error in the app we collect data and information (through third party products) on your phone called Log Data. This Log Data may include information such as your device Internet Protocol ("IP") address, device name, operating system version, the configuration of the app when utilizing our Service, the time and date of your use of the Service, and other statistics.

Cookies
Cookies are files with a small amount of data that are commonly used as anonymous unique identifiers. These are sent to your browser from the websites that you visit and are stored on your device's internal memory. This Service does not use these "cookies" explicitly. However, the app may use third party code and libraries that use "cookies" to collect information and improve their services. You have the option to either accept or refuse these cookies and know when a cookie is being sent to your device. If you choose to refuse our cookies, you may not be able to use some portions of this Service.

Service Providers
We may employ third-party companies and individuals due to the following reasons:
To facilitate our Service;
To provide the Service on our behalf;
To perform Service-related services; or
To assist us in analysing how our Service is used. We want to inform users of this Service that these third parties have access to your Personal Information. The reason is to perform the tasks assigned to them on our behalf. However, they are obligated not to disclose or use the information for any other purpose.

Security
We value your trust in providing us your Personal Information, thus we are striving to use commercially acceptable means of protecting it. But remember that no method of transmission over the internet, or method of electronic storage is 100% secure and reliable, and we cannot guarantee its absolute security.

Links to Other Sites
This Service may contain links to other sites. If you click on a third-party link, you will be directed to that site. Note that these external sites are not operated by us. Therefore, we strongly advise you to review the Privacy Policy of these websites. We have no control over and assume no responsibility for the content, privacy policies, or practices of any third-party sites or services.

Children's Privacy

These Services do not address anyone under the age of 13. We do not knowingly collect personally identifiable information from children under 13. In the case we discover that a child under 13 has provided us with personal information, we immediately delete this from our servers. If you are a parent or guardian and you are aware that your child has provided us with personal information, please contact us so that we will be able to do necessary actions.

Changes to This Privacy Policy

We may update our Privacy Policy from time to time. Thus, you are advised to review this page periodically for any changes. We will notify you of any changes by posting the new Privacy Policy on this page. This policy is effective as of 2020-10-26.

Contact Us

If you have any questions or suggestions about our Privacy Policy, do not hesitate to contact us at support@cyanolakes.com.

This privacy policy page was created at privacypolicytemplate.net and modified/generated by App Privacy Policy Generator

APPENDIX C FAQS

https://www.cyanolakes.com/faqs/

How do I get started?
When you first open the App begin by adding a lake (see How do I add a lake?). You may add up to four lakes in your free account. By subscribing you can add an unlimited number of lakes access high-definition satellite imagery and receive notifications of updates for your lakes (see What features are included in a premium subscription?).

How do I view information for different lakes?
To view information for different lakes swipe left and right on the home screen. Tap the list icon on the bottom right of the home screen to view a summary list of your lakes.

How do I view a list of my lakes?
To view a list of your lakes, tap on the list icon on the bottom right corner of the home screen.

How do I add a lake?
Adding an existing lake:
On the landing screen tap on the list icon in the bottom right-hand corner. This will open the You Lakes screen where a list of your lakes is displayed. Then tap the plus (+) icon in the bottom right-hand corner.
This will open the Add a Lake screen. You must allow the app to access your location before you can add a lake (see How do I enable location services?). The App will use your location to fetch nearby lakes that are already in the database. Using your fingers zoom in and out and move the map to the lake you would like to add. Existing lakes will appear as pins on the map. If you don't see any pins tap Refresh Map to check if any have already been added in the area. To add an existing lake, tap on the pin and then tap \Add This Lake\.
The lake will then be added to your list. If you already know the name of the lake you can search for it by tapping on the magnification glass icon.
Adding a new lake:
If you would like to add a lake that does not already have a pin on it select Add New Lake. Then on the map zoom in and tap near the centre of the waterbody you would like to add. The App will try to add the new waterbody. You may receive an error message if the lake is too large or small or if it could not be added. If you get a message which says the lake already exists select \Refresh Map\ and then add the lake by selecting the pin first.

How do I remove a lake?
To remove a lake, go to the list view. Then swipe left on the lake you want to remove and tap delete.

How do I enable location services?
Apple devices:
Go to Setting Privacy Location Services and select While using the app.
Devices running Android:
Go to Setting Location App permissions and select Allow only while using the app.

I got an error message when trying to add my lake. Why can't I add my lake?
There are several reasons why you may not be able to add a lake. The lake may be too small or too large.
By default, lagoons and estuaries affected by tides can't be added. Lakes connected to rivers and canals may also cause problems due to their irregular shapes. It may also be that when adding a new lake, you did not touch the screen inside a waterbody or that the lake does not exist in the Open Street Map database.

How do I update the information?

The App will update by itself or every time you open it. To force an update close and then re-open the App (see also How frequently is the information updated?).

How is the cyanobacteria risk level determined?
The cyanobacteria risk level is calculated using the average value of chlorophyll-a for water identified as cyanobacteria by the satellite. The risk level for the waterbody is the highest risk class in which three or more detections are made in the image pixels. This means that the average chlorophyll-a value for the waterbody may not correspond to the cyanobacteria risk level. For example, a bloom where the large majority of pixels have a chl-a value near 20 μg/L will have an average chlorophyll-a value of 20 μg/L (medium-risk). However, because there are a few pixels identified as cyanobacteria that have chl-a values larger than 100 μg/L the risk level will be very high. Subscribers can view the latest satellite imagery to determine which areas are at a high-risk level. The risk classes are based on the World Health Organisation (1999 2003) guidelines.

Low: < 10 μg/L chlorophyll-a < 20 000 cyanobacteria sp. cells/ml <4 μg/L microcystin toxin concentration.

Medium: 10-50 μg/L chlorophyll-a 20 000 to 100 000 cyanobacteria sp. cells/ml 4 to 20 μg/L microcystin toxin concentration.

High: 50-100 μg/L chlorophyll-a 100 000 to 200 000 cyanobacteria sp. cells/ml 20 to 40 μg/L microcystin toxin concentration.

Very high: > 100 μg/L chlorophyll-a > 200 000 cyanobacteria sp. cells/ml > 40 μg/L microcystin toxin concentration.

The US Environmental Protection Agency ambient water quality guidelines for recreational water use (2020) recommend a microcystin toxin concentration of less than 8 μg/L equivalent to 20 μg/L chl-a for recreational use.

What does the cyanobacteria risk level mean?
The application provides four risk levels to human health from cyanobacteria blooms based on World Health Organization guidelines.

Low risk
No cyanobacteria have been detected in the water. There is a low risk of adverse health effects from cyanobacteria. But since cyanobacteria can exist below the detection limit of the satellite it does not guarantee that cyanobacteria are not present. Most recreational activities will be safe provided additional water quality information is first consulted and that any warnings issued by government or local authorities are followed. Routine sampling on the ground should still be performed to determine if cyanobacteria are present.

Medium risk
Cyanobacteria are present in the water but at medium concentration levels which means there is a small risk of adverse health effects from cyanobacteria. High risk individuals should avoid all recreational activities. The water should be observed carefully for any areas of high concentrations that may form as a result of wind. Most recreational activities will be safe for uncompromised persons although full contact recreational activities should be avoided (e.g. swimming). Again, additional water quality information should first be consulted, and any warnings issued by government or local authorities must be followed.

High risk

Cyanobacteria are present in the water at high concentration levels. There is a high risk of adverse health effects from cyanobacteria. It is recommended that all recreational activities be suspended until water conditions improve. The map should be consulted to determined which areas are affected and to determine whether recreational activities can be continued in mild or low risk areas (note may not be current). Depending on the prevailing wind dangerously high concentrations of cyanobacteria can form within minutes to hours. Therefore, the water should be carefully monitored.

Very high risk

Cyanobacteria are present in the water at very high concentration levels. There is a very high risk of adverse health effects from cyanobacteria. All recreational activities should be suspended until water conditions improve. By consulting the map areas of moderate and high risk can be avoided in which case recreational activities can be continued in medium or low risk areas. Cyanobacteria scums are most likely present, and these can contain dangerously high concentrations of cyanotoxins which may cause poisoning of animals and adverse symptoms for recreational users.

What risk does cyanobacteria pose to human health?

Cyanobacteria also called blue-green algae produce various toxic compounds that are as poisonous as cobra venom. These “cyanotoxins” may be present when cyanobacteria occur at high or low concentrations in the water but are most dangerous when cyanobacteria accumulate as scum on the water surface.

Many wild and domestic animals have died as a result of drinking water from a dam or river containing cyanotoxins. In humans there are also many documented cases of human deaths resulting from cyanobacteria poisoning around the world although these are quite rare. This is because people do not generally drink water containing untreated or raw algae.

However, humans often come into contact with cyanobacteria through partial or full body contact with water. In such cases water may be swallowed and various symptoms of exposure to compounds causing rashes nausea and vomiting and sometimes symptoms associated with cyanotoxin poisoning may occur.

Cyanotoxins may still be present in drinking water even after conventional treatment. This is because conventional water treatment does not remove cyanotoxins. If concentrations of cyanotoxins are high this may be deadly and result in acute poisonings of humans. At low concentrations some symptoms may be present only however long-term exposure can lead to cancer (mainly of the liver) or neurodegenerative diseases (such as Amyotrophic lateral sclerosis or ALS).

Cyanotoxins have been shown to cause ALS in cases where fruit bats that were eaten after they had ingested cyanobacteria toxins in their food source. Cyanotoxins can also accumulate in crops (fruit vegetables, etc.) that have been irrigated with water containing cyanobacteria and may cause adverse health effects (although this is difficult to detect).

Thus, cyanobacteria pose a high risk to human health by causing acute poisonings various adverse short-term symptoms and long-term health effects.

How are the recreational advisories determined?

The recreational advisories are provided with respect to partial and full contact water sports. Partial-contact includes activities such as fishing dog-walking or canoeing. Full contact includes activities such as swimming skiing or diving. The recreational advisories are based on the cyanobacteria health risk and nutrient pollution levels. Full-contact recreational activities are only recommended at a low cyanobacteria risk level. Partial-contact activities are not advised at high or very high-risk levels. Because the advisories provided through the App are derived using average values for a lake there may still be areas that are suitable for recreational activities. By subscribing you can view the latest satellite imagery to help you locate areas that may be suitable. Always follow warnings issued by your local authority.

How is the chlorophyll-a concentration determined?

The chlorophyll-a value displayed on the centre of the home screen is the average chlorophyll-a value determined for the entire waterbody from satellite remote sensing. The minimum maximum and standard deviation is also provided. Chlorophyll-a is the best proxy for cyanobacteria and algal biomass or density.
A chl-a value between 1 and 10 µg/L means that the water will be mostly clear.
A chl-a value between 10 and 20 µg/L means that the water will be slightly green or brown in colour.
A chl-a value between 20 and 50 µg/L means that the water will be green or brown and turbid.
A chl-a value larger than 50 µg/L means the water will be opaque as a result of a cyanobacteria or algal bloom.
Subscribers can view the latest chlorophyll-a maps from satellite imagery.

How is the pollution level determined?
The nutrient pollution levels correspond directly to the OECD trophic status classes for freshwaters. The pollution level is calculated using the average value of chlorophyll-a for the waterbody. Therefore, different areas of the waterbody may be categorised at different pollution levels Subscribers can view the latest satellite imagery for chlorophyll-a to determine which areas are at lower pollution levels.

Low: < 10 µg/L chlorophyll-a oligotrophic (low nutrients).
Medium: 10-20 µg/L chlorophyll-a mesotrophic (medium nutrients).
High: 20-50 µg/L chlorophyll-a eutrophic (plentiful nutrients).
Very high: > 50 µg/L chlorophyll-a hypertrophic (high concentration of nutrients).

How is the cyanobacteria cell count determined?
The cyanobacteria cell counts are derived from the average value of chlorophyll-a from water pixels identified as cyanobacteria. The conversion from chlorophyll-a to cells follows the World Health Organisation recommended conversion of 1 µg/L chl-a to 2 000 cell/ml (WHO 2003). The cell count may not correspond to the risk level because of the averaging process. For example, a bloom where the large majority of pixels have a chl-a value near 20 µg/L will result in a cyanobacteria cell count near 40 000 cells/l calculated using the average value (20 µg/L). However, because of small area with chl-a values larger than 100 µg/L the risk level will be very high.

How is the microcystin toxin concentration determined?
The microcystin toxin concentrations are derived from the average value of chlorophyll-a from water pixels identified as cyanobacteria. The conversion from chlorophyll-a to microcystin follows the World Health Organisation recommended conversion of 1 µg/L chl-a to 0.4 µg/L microcystin (WHO 2003). The microcystin concentration may not correspond to the risk level because of the averaging process. For example, a bloom where the large majority of pixels have a chl-a value near 20 µg/L will result in a microcystin concentration near 8 µg/L calculated using the average value (20 µg/L). However, because of small area with chl-a values larger than 100 µg/L the risk level will be very high. Subscribers can view the latest satellite imagery to determine which areas are high risk.

What does all the other statistical information mean?
Cyanobacteria risk level – the risk to human health from cyanobacteria blooms
Potential microcystin toxin – the potential concentration of microcystin toxin based on a simple conversion from chlorophyll-a (assumes that microcystin producing species are present)
Average cyanobacteria cell count – the average count of cyanobacteria cells derived using a conversion from chlorophyll-a for cyanobacteria
Cyanobacteria coverage – the percentage of the waterbody where the water is dominated by cyanobacteria species
Cyanobacteria scum coverage – the percentage of the waterbody where the water is dominated by cyanobacteria species at a concentration above 50 µg/L chlorophyll-a
Average cyanobacteria chl-a – the average concentration of chlorophyll-a for water dominated by cyanobacteria species
Average algae chl-a – the average concentration of chlorophyll-a for water dominated by algal species
Vegetation coverage – the percentage of the waterbody covered by floating vegetation

St. Dev. Chl-a – the standard deviation of chlorophyll-a for the entire waterbody

How frequently is the information updated?
Depending on your location the information is updated 5 to 6 days per week, however cloud ice or snow cover may result in fewer updates.

Why has the information not been updated recently?
If the information has not been updated within the last week it almost certainly due to cloud ice or snow covering your lake.

How do I know if the water is safe to use?
Knowing if water is safe to use is determined by many different variables. This Application provides information ONLY related to cyanobacteria and algal blooms. It does not provide any information related to other important water quality indicators such as microbiological organisms (other bacteria and viruses) heavy metals pathogens and diseases inorganic and organic chemicals or other contaminants. Therefore, the recreational recommendations provided by the App are NOT comprehensive. The information provided in the App MUST be used in conjunction with other available information before determining if the water is safe to use. Always follow warnings issued by your local authorities.

What is the source of the information?
The information provided via the Application is derived exclusively from satellite imagery. We use multiple sources of satellite data from the EU Copernicus Program and NASA.

The information doesn't agree with what I can observe why?
In cases where the information does not seem to correspond with your own measurements first check the date that the information was updated. There may have been a change in conditions since the satellite last past overhead. There can be many other reasons: a difference in time a difference in place or a difference in how you took your measurements.

How often are the satellite images updated?
The high-definition satellite imagery is updated between 2 and 3 times per week depending on your location as well as cloud ice and snow cover.

Why is the satellite imagery blanked out?
The satellite imagery may occasionally be blacked or whited out. This is caused by cloud cover or coverage gaps. If this problem occurs, it will resolve itself after a cloud-free image of your lake is acquired which may take a few days or longer depending on your location and the time of year.

Why isn't the satellite imagery loading?
The satellite imagery is downloaded from the internet in real-time. The speed it will render will depend on the speed of your internet connection. A slow connection may load imagery very slowly or not appear to load imagery at all.

Why is the resolution of the satellite imagery worse than Google Maps?
The satellite imagery from Google is ultra-high-resolution imagery and is not current. We provide the latest imagery at a slightly lower spatial resolution than Google. The lower-resolution imagery provided through the App generates value-added products (such as chlorophyll-a) with higher accuracy than ultra-high-resolution imagery reducing errors.

What features are included in a premium subscription?
By subscribing you will have access to the following premium features:

- Access the latest high-definition satellite imagery for your lakes including maps to enhance monitoring cyanobacteria and algal blooms aquatic weeds and vegetation and water colour;
- Receive cyanobacteria risk level alerts and updates from satellite in your notifications centre;
- Unlock adding an unlimited number of waterbodies;
- Additional premium features that may be released in future.

What satellite imagery is available when I subscribe?
The following satellite imagery is available with a 10 m pixel size (spatial resolution):
- Chlorophyll-a – the chlorophyll-a concentration between 1 and 1000 µg/L – used for monitoring cyanobacteria and algal blooms;
- True Colour – a representation of the water colour as seen by the human eye – used for monitoring water transparency water colour and blooms;
- Bloom index – (blue-green colour map showing blooms in bright red) – used for monitoring cyanobacteria or algal blooms;
- Enhanced Vegetation Index – shows vegetation in bright green – used for monitoring cyanobacteria surface scum floating aquatic weeds and vegetation;
- Normalised Difference Vegetation Index (NDVI) – blue-red colour map showing vegetation in bright red – used for monitoring cyanobacteria surface scum floating aquatic weeds and vegetation.

What notifications are available when I subscribe?
After subscribing you can turn on notifications for the following events:
- high or very high-risk events – receive notifications for high or very high cyanobacteria blooms;
- medium risk events – receive notifications for medium risk cyanobacteria blooms;
- all updates – receive notifications of all updates.

How many lakes can I add when I subscribe?
You can add an unlimited number of lakes when you subscribe.

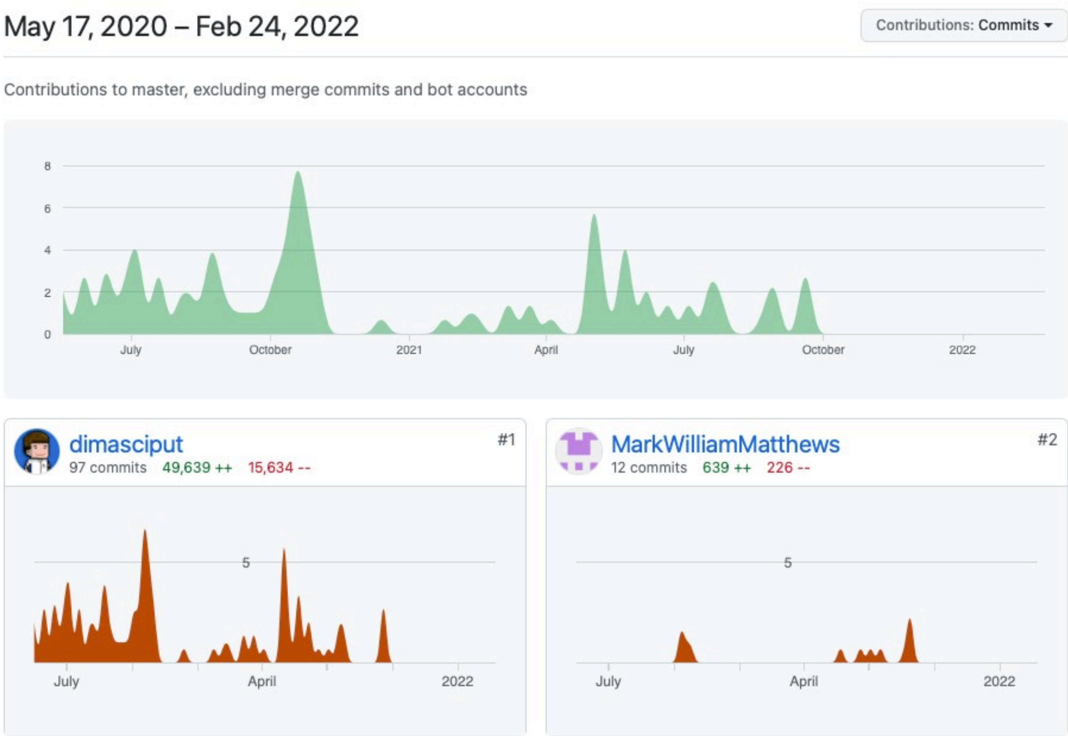
How do I turn on notifications?
To turn on notifications tap the settings icon on the top right corner of the home screen. Then turn on notifications for the events of your choice by using the slider.

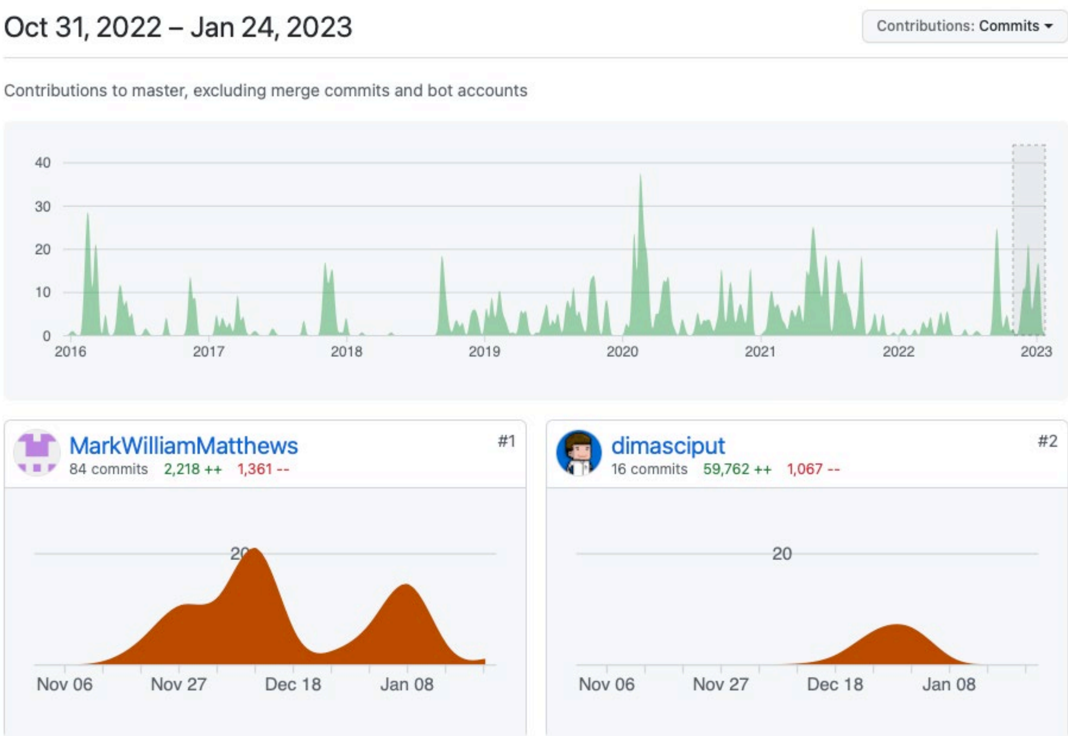
How do I cancel my subscription?
You can manage your subscription or turn off auto-renewal at any time from your Apple or Google account. Cancellation of the active subscription period is not allowed.

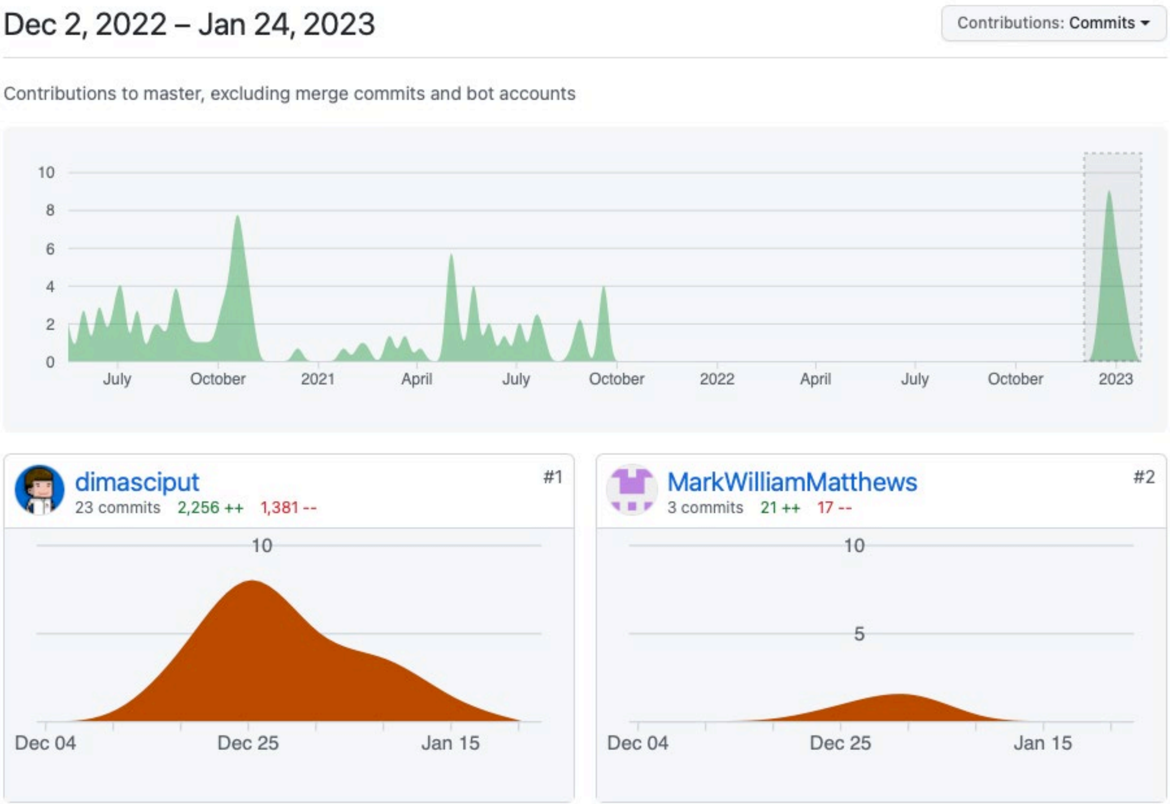
I am not happy with my purchase can I get a refund?
Subscriptions are non-refundable for the current period. You can turn off auto-renewal to cancel the next subscription period.

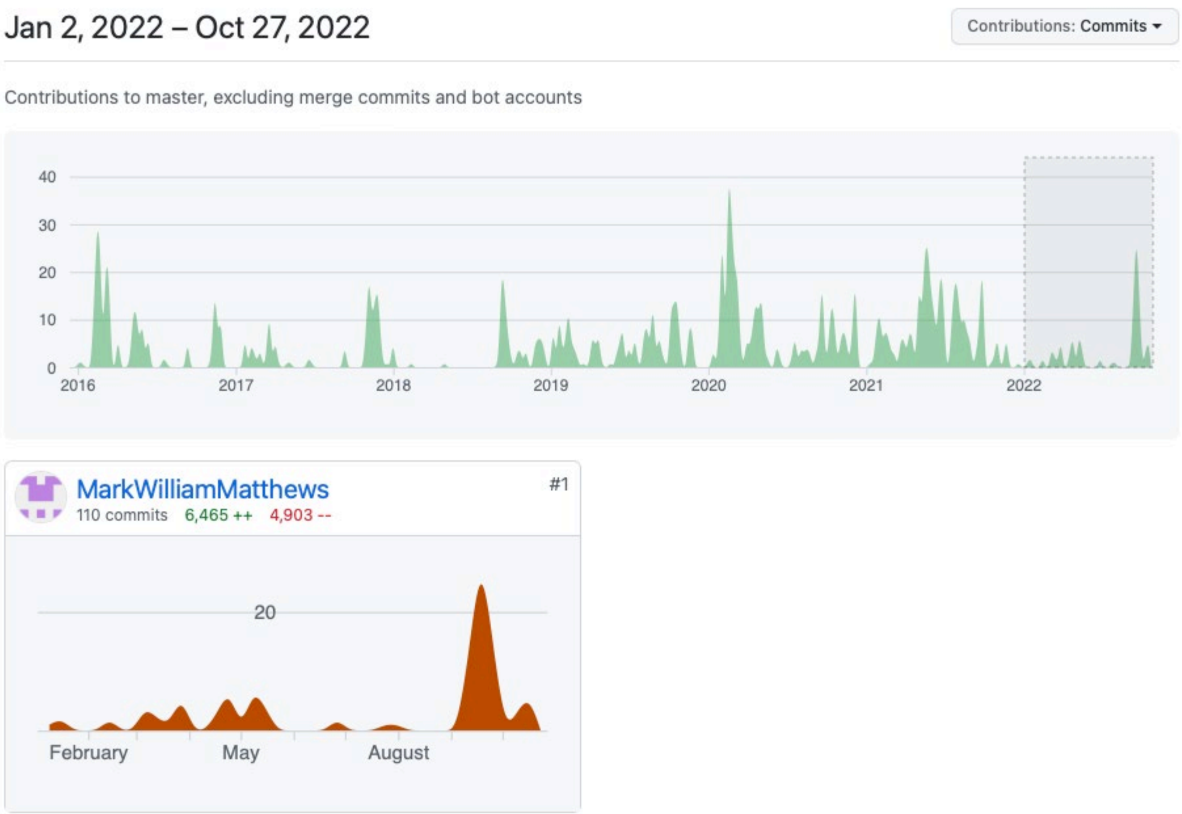
How do I log in using my organisational account?
From the home screen tap on the settings icon on the top right corner. Then tap Log in here under Already have an account for your organisation? Enter your organisation's credentials and tap log in. After you have logged in you will have access to Premium features.

APPENDIX D DEVELOPMENT LOG


May 17, 2020 – Feb 24, 2022
Contributions: Commits
Contributions to master, excluding merge commits and bot accounts
0
2
4
6
8
July
October
2021
April
July
October
2022
dimasciput
97 commits 49,639 ++ 15,634 --
#1
5
July
April
2022
MarkWilliamMatthews
12 commits 639 ++ 226 --
#2
5
July
April
2022


Oct 31, 2022 – Jan 24, 2023
Contributions: Commits
Contributions to master, excluding merge commits and bot accounts
0
10
20
30
40
2016
2017
2018
2019
2020
2021
2022
2023
MarkWilliamMatthews
84 commits 2,218 ++ 1,361 --
#1
Nov 06
Nov 27
Dec 18
Jan 08
dimasciput
16 commits 59,762 ++ 1,067 --
#2
20
Nov 06
Nov 27
Dec 18
Jan 08


Dec 2, 2022 – Jan 24, 2023
Contributions: Commits
Contributions to master, excluding merge commits and bot accounts
10
8
6
4
2
0
July
October
2021
April
July
October
2022
April
July
October
2023
dimasciput
#1
23 commits 2,256 ++ 1,381 --
10
Dec 04
Dec 25
Jan 15
MarkWilliamMatthews
#2
3 commits 21 ++ 17 --
10
5
Dec 04
Dec 25
Jan 15


Jan 2, 2022 – Oct 27, 2022
Contributions: Commits
Contributions to master, excluding merge commits and bot accounts
40
30
20
10
0
2016
2017
2018
2019
2020
2021
2022
MarkWilliamMatthews
#1
110 commits 6,465 ++ 4,903 --
20
February
May
August

APPENDIX E HIGH RESOLUTION ALGORITHMS

The following algorithm(s) are available on the Sentinel-hub Custom Scripts Repository, which is hosted on Github: https://custom-scripts.sentinel-hub.com

Water / Land Masking

```
// Author: Mohor Gartner
var MNDWI_threshold=0.42; //testing shows recommended 0.42 for Sentinel-2 and Landsat 8. For
the scene in article [1] it was 0.8.
var NDWI_threshold=0.4; //testing shows recommended 0.4 for Sentinel-2 and Landsat 8. For the
scene in article [1] it was 0.5.
var filter_UABS=true;
var filter_SSI=false;

function wbi(r,g,b,nir,swir1,swir2) {
        //water surface
        let ws=0;
        //try as it might fail for some pixel
        try {
                //calc indices
                //[4][5][1][8][2][3]
                var ndvi=(nir-r)/(nir+r),mndwi=(g-swir1)/(g+swir1),ndwi=(g-
nir)/(g+nir),ndwi_leaves=(nir-swir1)/(nir+swir1),aweish=b+2.5*g-1.5*(nir+swir1)-
0.25*swir2,aweinsh=4*(g-swir1)-(0.25*nir+2.75*swir1);
                //[10][11][12]
                var dbsi=((swir1-g)/(swir1+g))-
ndvi,wii=Math.pow(nir,2)/r,wri=(g+r)/(nir+swir1),puwi=5.83*g-6.57*r-30.32*nir+2.25,uwi=(g-
1.1*r-5.2*nir+0.4)/Math.abs(g-1.1*r-5.2*nir),usi=0.25*(g/r)-0.57*(nir/g)-0.83*(b/g)+1;
                //DEFINE WB
                if
(mndwi>MNDWI_threshold||ndwi>NDWI_threshold||aweinsh>0.1879||aweish>0.1112||ndvi<-
0.2||ndwi_leaves>1) {ws=1;}
                //filter urban areas [3] and bare soil [10]
                if (filter_UABS && ws==1) {
                        if ((aweinsh<=-0.03)||(dbsi>0)) {ws=0;}
                }
                //filter shadows and snow/ice
                if (filter_SSI && ws==1) {
                        //SHADOWS[3]
                        if ((aweish<=0.1112&&ndvi>-0.2)){ws=0;}
                        if ((aweinsh<0.5&&ndvi>-0.2)){ws=0;} //or 0.1897
                        if (((aweinsh<0||aweish<=0||ndvi>-0.1))){ws=0;}
                        //SNOW AREAS[6][7][8]
                        if
((((g>=0.319)?((mndwi>0.2)?((nir>0.15)?((b>0.18)?1:0):0):0):0))){ws=0;}
                        if (g>0.319){ws=0;}
                        //WII,WRI[11]
```

```
                    if (wii>0.04||wri<2){ws=0;}
                    //PUWI,UWI,USI[12]
                    if (puwi<0||uwi<0||usi<=-1){ws=0;}
                    //spectrum based[13]
                    if (mndwi<aweinsh){ws=0;}
                    if (ndwi-aweinsh>0.5){ws=0;}
            }
        }catch(err){ws=0;}
        return ws;
}

let trueColor = [3*B04,3*B03,3*B02];
let water = wbi(B04,B03,B02,B08,B11,B12);
if (water==0) {
  return trueColor }
else {
    return [0, 0, 1.0]
}
```

CyanoLakes Bloom Index

```
// Kratitz, J.A., Matthews, M.W., 2020
// Water body detection - Credit : Mohor Gartner
var MNDWI_threshold=0.42; //testing shows recommended 0.42 for Sentinel-2 and Landsat 8. For
the scene in article [1] it was 0.8.
var NDWI_threshold=0.4; //testing shows recommended 0.4 for Sentinel-2 and Landsat 8. For the
scene in article [1] it was 0.5.
var filter_UABS=true;
function wbi(r,g,b,nir,swir1,swir2) {
        //water surface
        let ws=0;
        //try as it might fail for some pixel
        try {
                //calc indices
                //[4][5][1][8][2][3]
                var ndvi=(nir-r)/(nir+r),mndwi=(g-swir1)/(g+swir1),ndwi=(g-
nir)/(g+nir),ndwi_leaves=(nir-swir1)/(nir+swir1),aweish=b+2.5*g-1.5*(nir+swir1)-
0.25*swir2,aweinsh=4*(g-swir1)-(0.25*nir+2.75*swir1);
                //[10][11][12]
                var dbsi=((swir1-g)/(swir1+g))-
ndvi,wii=Math.pow(nir,2)/r,wri=(g+r)/(nir+swir1),puwi=5.83*g-6.57*r-30.32*nir+2.25,uwi=(g-
1.1*r-5.2*nir+0.4)/Math.abs(g-1.1*r-5.2*nir),usi=0.25*(g/r)-0.57*(nir/g)-0.83*(b/g)+1;
                //DEFINE WB
                if
(mndwi>MNDWI_threshold||ndwi>NDWI_threshold||aweinsh>0.1879||aweish>0.1112||ndvi<-
0.2||ndwi_leaves>1) {ws=1;}
                //filter urban areas [3] and bare soil [10]
```

```
            if (filter_UABS && ws==1) {
                    if ((aweinsh<=-0.03)||(dbsi>0)) {ws=0;}
            }
        }catch(err){ws=0;}
        return ws;
}
let water = wbi(B04,B03,B02,B08,B11,B12);

// Baseline subtractions
function MCI (a,b,c) {return (b-a-(c-a)*(705-665)/(740-665))};
function FAI (a,b,c) {return (b-a-(c-a)*(740-665)/(842-665))};

// True color representation
var trueColor = [3*B04,3*B03,3*B02];

// Switches to FAI if FAI > MCI
function Switch (a,b) {
  if (a > b) {
  return a
  } else {
  return b
  }
}

// Bloom index
var bloom_index = Switch(MCI(B04,B05,B06),FAI(B04,B06,B08));

// HighlightCompressVisualizer
// This compresses bloom index values over 0.05 corresponding to floating algae
// 0.05 will return ~ 0.9
const compressviz = new HighlightCompressVisualizer(0, 0.05);

// Visualise on color scale between 0 and 1 (after compression)
var viz = ColorGradientVisualizer.createBlueRed(minVal=0, maxVal=1);

//Compressed index
compressed_index = compressviz.process(bloom_index);

return water==0 ? trueColor : viz.process(compressed_index);
```

Chlorophyll-a

```
// Kravitz, J.A., Matthews, M.W. (2020)
// Water body detection - Credit : Mohor Gartner
var MNDWI_threshold=0.42; //testing shows recommended 0.42 for Sentinel-2 and Landsat 8. For
the scene in article [1] it was 0.8.
```

```
var NDWI_threshold=0.4; //testing shows recommended 0.4 for Sentinel-2 and Landsat 8. For the
scene in article [1] it was 0.5.
var filter_UABS=true;

function wbi(r,g,b,nir,swir1,swir2) {
        //water surface
        let ws=0;
        //try as it might fail for some pixel
        try {
                //calc indices
                //[4][5][1][8][2][3]
                var ndvi=(nir-r)/(nir+r),mndwi=(g-swir1)/(g+swir1),ndwi=(g-
nir)/(g+nir),ndwi_leaves=(nir-swir1)/(nir+swir1),aweish=b+2.5*g-1.5*(nir+swir1)-
0.25*swir2,aweinsh=4*(g-swir1)-(0.25*nir+2.75*swir1);
                //[10][11][12]
                var dbsi=((swir1-g)/(swir1+g))-
ndvi,wii=Math.pow(nir,2)/r,wri=(g+r)/(nir+swir1),puwi=5.83*g-6.57*r-30.32*nir+2.25,uwi=(g-
1.1*r-5.2*nir+0.4)/Math.abs(g-1.1*r-5.2*nir),usi=0.25*(g/r)-0.57*(nir/g)-0.83*(b/g)+1;
                //DEFINE WB
                if
(mndwi>MNDWI_threshold||ndwi>NDWI_threshold||aweinsh>0.1879||aweish>0.1112||ndvi<-
0.2||ndwi_leaves>1) {ws=1;}
                //filter urban areas [3] and bare soil [10]
                if (filter_UABS && ws==1) {
                        if ((aweinsh<=-0.03)||(dbsi>0)) {ws=0;}
                }
        }catch(err){ws=0;}
        return ws;
}
let water = wbi(B04,B03,B02,B08,B11,B12);

// Floating vegetation
function FAI (a,b,c) {return (b-a-(c-a)*(783-665)/(865-665))};
let FAIv = FAI(B04,B07,B8A);

// Chlorophyll-a
function NDCI (a,b) {return (b-a)/(b+a)};
let NDCIv = NDCI(B04,B05);
let chl = 826.57 * NDCIv**3 - 176.43 * NDCIv**2 + 19 * NDCIv + 4.071; // From simulated data

// Ture colour
let trueColor = [3*B04,3*B03,3*B02];

// Render colour map
if (water==0) {
  return trueColor;
} else if (FAIv>0.08){
```

```
  return [233/255,72/255,21/255];
} else if (chl<0.5){
  return [0,0,1.0];
} else if (chl<1){
  return [0,0,1.0];
} else if (chl<2.5){
  return [0,59/255,1];
} else if (chl<3.5){
  return [0,98/255,1];
} else if (chl<5){
  return [15/255,113/255,141/255];
} else if (chl<7){
  return [14/255,141/255,120/255];
} else if (chl<8){
  return [13/255,141/255,103/255];
} else if (chl<10){
  return [30/255,226/255,28/255];
} else if (chl<14){
  return [42/255,226/255,28/255];
} else if (chl<18){
  return [68/255,226/255,28/255];
} else if (chl<20){
  return [68/255,226/255,28/255];
} else if (chl<24){
  return [134/255,247/255,0];
} else if (chl<28){
  return [140/255,247/255,0];
} else if (chl<30){
  return [205/255,237/255,0];
} else if (chl<38){
  return [208/255,240/255,0];
} else if (chl<45){
  return [208/255,240/255,0];
} else if (chl<50){
  return [251/255,210/255,3/255];
} else if (chl<75){
  return [248/255,207/255,2/255];
} else if (chl<90){
  return [134/255,247/255,0];
} else if (chl<24){
  return [245/255,205/255,1/255];
} else if (chl<100){
  return [245/255,164/255,9/255];
} else if (chl<150){
  return [240/255,159/255,8/255];
} else if (chl<250){
  return [237/255,157/255,7/255];
```

```
} else if (chl<300){
  return [239/255,118/255,15/255];
} else if (chl<350){
  return [239/255,101/255,15/255];
} else if (chl<450){
  return [239/255,100/255,14/255];
} else if (chl<500){
  return [233/255,72/255,21/255];
} else return [233/255,72/255,21/255];
```