Erasmus University Rotterdam

Rotterdam School of Management

Master Thesis. Master of Science in Business Analytics & Management

U-Net CNN: Systematically Monitoring HABs in the Netherlands with Sentinel 3 Satellite Imagery

Romain Huet (532091)

Coach: Jason Roos

Co-reader: Sebastian Gabel

Date of submission: 15th of June 2024

Abstract

Deep learning has proven very useful for a broad range of applications, including the field of HAB monitoring. Indeed, by training on satellite images of water shelves models can generate accurate predictions for bloom presence and expansion where in-situ data might otherwise be challenging to obtain for a variety of reasons, with a primary one being the economic cost it entails. While these models are not without their drawbacks, they remain a rather unexplored method of detecting HABs and thus constitute an interesting object of research. However, combining the latter and satellite imagery possesses a range of advantages that if implemented successfully would enable great advancements in terms of scaling monitoring practices towards global coverage while significantly reducing economic costs traditionally associated with the field. This paper explores this possibility by adapting a U-Net CNN originally geared towards biomedical image segmentation and making use of the ESA's Sentinel 3 satellite. It seeks to determine if it is indeed possible to reliably monitor HABs with simple means, and if it is preferable to train a model rather on one lake or multiple. Results are mixed: while the models achieve satisfying results (IoU around 0.9) when tested on a lake that has already been observed during training, the same cannot be said for the second task where the models are made to predict HABs for a new, previously unseen lake. Thus, the paper concludes that while there is promise in the method, significant improvements need to be implemented.

Table of Contents

1) Introduction 03
2) Theoretical Background 08
3) Methodology & Data 12
4) Results 25
5) Discussion 32
6) Appendices 39
7) Acknowledgements 42
8) References 42

1) Introduction

Problem contextualization

Harmful algal blooms (HABs) are one of the grand challenges that humanity is facing in the years to come as it menaces waters around the globe, which are essential parts of the ecosystem and provide sources of nutrition, livelihood and leisure to humans. Unlike most grand challenges however, HABs are one that is easily identifiable, at least in the scope of its consequences: they appear as large patches of green-blue matter floating at the surface of water bodies ranging from ponds to entire seas and can uniformly cover hundreds of hectares of water. Given recent developments in computer vision technology, this is an issue that presents high potential with regards to possible solutions that can be engineered to help mitigate the problem.

For context, HABs are the rapid development of phytoplankton in marine, brackish or freshwater environments. These consist of a number of different species of microscopic algae or bacteria, such as cyanobacteria, dinoflagellates and diatoms (McGowan, 2023). While the latter are not nefarious by nature, their quick multiplication increases the risk that they release toxins into their habitat; if this event occurs, the bloom is then qualified as harmful. Blooms have a range of hazardous effects on the environment, such as contaminating drinking water, and can even be fatal to humans and animals (Ho & Michalak, 2015).

The causes of HABs generally stem from abrupt changes in the environment, such as temperature rises. These cause nutrient abundance and thus provide favourable conditions for the development of blooms. With humanity's intensifying impact on the planet and climate change, these environmental shifts have come to be much more frequent and extensive (Sangiorgi & Villanueva, 2024). An important though rather alarming illustration of this event is the handful of lakes around the world that are suffering from permanent infestations of HABs, such as lake Nieuwe Meer in the Netherlands (Ho et al., 2015). Hence, research presently attributes the recent development of HABs largely as a result of anthropogenic activity.

It is interesting to observe HABs for a variety of research purposes. HABs are a safety concern as they have negative effects on human and animal health, either directly via their inherent toxicity or through their impact on their environment. This can be of interest for a range of

economic actors, from tourism entities to agriculture, as well as governmental and non-governmental organizations specialized in sustainability or health sectors. Furthermore, given that HABs are largely caused by humans, it is also interesting to explore what are some of the activities that have a direct or indirect effect on the blooms' development. Notable examples of these causes are intensive agricultural practices that use fertilizers which permeate their surroundings and create an imbalance in nutrients in natural habitats, as well as any industrial activity that contribute to the production of greenhouse gas emissions affecting global temperatures (Khan et al., 2021).

In light of these examples, it is thus relevant to observe HABs as their current stature represents a clear symptom of human impact on the environment. Having the possibility to reliably and systematically monitor these events provides an outcome variable - a means of quantifying a given activity's impact on the environment - and can enable research to crystallise and illustrate the state of sustainability in a range of economic sectors.

A review of current literature reveals that efforts to observe and monitor HABs is predominantly done through in-situ measurements within specific case studies. However, further exploration yields the discovery of more extensive monitoring projects using data from the growing sector of satellite-based Earth Observation (EO). Indeed, the Finnish institute for the Environment developed their own initiative, TARKKA, using the European Space Agency's (ESA) Copernicus programme for EO-purposed satellite constellations (Dewulf & Mamais, 2022). The methodology they employed consisted in manually assembling satellite images over time of the larger Finnish lakes to create small case studies and observe the development of blooms in a qualitative manner, using surface scum. This already presents several advantages, such as the ability to survey any of the 187,000 lakes that make up the Finnish territory at a high frequency of observations, with a high resolution (down to ten meters per pixel) for virtually no costs (Dewulf & Mamais, 2022).

The potential of Earth Observation thus presents itself as striking. For clarification, EO is the sector that regroups all remote sensing technologies that have for purpose the monitoring of land, marine environments and the atmosphere (europa, 2024). The technology of interest throughout this study lies in satellite-based methods, which gather EO data with the help of measuring instruments included aboard the payload. Indeed, analysing data gathered through these measuring tools can help recognizing HAB formation and development. Indeed, blooms of cyanobacterial origin can be identified spectrally, by measuring the prevalence of

wavelengths that correspond to their major pigments in the given satellite images (Viso-Vasquez et al., 2021). The effectiveness of this method has been extensively reported in literature, with a range of studies that have conducted cross-analyses with in-situ measurements reporting highly statistically significant correlations between the two methods of identifying HABs (Khan et al., 2021).

A major advantage that satellite-based remote sensing presents is the ability to capture data of both high spatial (surface area per pixel) and temporal resolution (rate of return to a same area) over extremely vast territory while remaining highly cost-efficient (Dewulf & Mamais, 2022). Given the vast number of lakes in the Netherlands, using satellite data thus presents an opportunity to reliably and systematically monitor the development of HABs in water shelves without having to physically visit them, or it can at least guide sampling efforts in a more informed manner. This has significant implications for the economic cost of regulating water quality across the country.

Furthermore, given the wide availability of satellite imagery and data that is made freely available by national and supra-national space programs such as the European Space Agency (ESA) and their Copernicus programme of Sentinel satellites, EO is now considered a source of big data. Thus, it is a perfectly suitable environment for the implementation of analytical methods such as machine learning to automate traditional monitoring functions. Indeed, research has already been made in implementing support vector machines and neural networks to complete this task (Khan et al., 2021).

A unique set of techniques have already been developed which facilitate the implementation of machine learning in this field. Within remotely sensed images, there are a number of proxies that have been identified, such as the wavelength of specific pigments that are characteristic of the blooms: Chlorophyll-a and phycocyanin. The first has light absorption peaks at around 433 nm and 686 nm whereas the other has an absorption peak between 550 nm and 650 nm (Khan et al., 2021). Models, in combination with in-situ measurements, can use this information to label pixels that most closely correspond as representing an HAB-infested area. However, there are limitations as to how well these data can be derived from satellite imagery due to disturbances caused by other materials present in the water or by water's own light absorption properties (Khan et al., 2021).

Given these limitations, it can be of interest to make use of a U-Net Convolutional Neural Network (CNN) which can learn from remotely sensed images and ground truth segmentation masks to detect HABs for new images, without having to specifically take in account these scientific and technical details. Indeed, neural networks are able to identify and develop their own features for what constitutes the object they are trained to recognize. Moreover, the U-Net architecture as first described by Ronneberger et al. (2015) enables pixel-wise classification and with its unique inclusion of skip connections achieves consistent and performant results for the given task (segmenting tumours in brain scans). Implementation of such networks has become relatively simple and accessible with popular packages such as pytorch; the only requirement is to possess a computer.

Thus, while much progress has already been achieved in developing analytical methods for this field, there remains significant potential to increment and reinforce these solutions. According to a meta-analysis on the field of remote sensing for monitoring HABs, there are large discrepancies in research with regards to the locations that previous studies have chosen to focus on (Khan et al., 2021). Whereas extensive research that has been conducted for China and the United States, monitoring solutions for areas such as Europe remain limited. Currently, the only solution that can be found for the Netherlands is a tool that monitors potential blooms across European coastal waters which is undergoing development at TU Delft (Villars, 2020). There are no systems that exist to systematically monitor HABs in freshwater, while infestations persist in places such as lake Nieuwe Meer.

Hence, this thesis aims to examine whether using machine learning in combination with satellite-based remote sensing is a reliable way to systematically monitor HABs in Dutch freshwaters. More specifically, the study will assess the predictive capability of a model that is relatively scientifically uninformed and that comes with virtually no economic costs. This leads to the following research questions:

- *Is a simple U-Net CNN implementation technically sophisticated enough to recognize HABs without taking into consideration geoscientific and biological properties? In other words, will the model be able to recognize HABs well enough even with using simple, relatively unprocessed satellite imagery as training data?*
- *Is the model generalizable? Is it preferable to focus model training and application on only one specific body of water or more, for a widely scalable application or a case-wise approach?*

The advanced analytics solution that will be used is a U-Net CNN. This is a specific type of deep learning architecture that is able to recognize patterns in images through analysing pixel information with a segmentation mask as reference, to then reproduce the image with its predicted segmentation of the subject of interest. Motivations for employing this method are varied: the model does not need a very large dataset given its sophistication and the reliability of data augmentation techniques (Ronneberger et al., 2015). Further, it achieves image segmentation, which has the potential to cover large areas at once with great precision, rather than having to generate predictions on a lake-per-lake basis. Lastly, the developers of the architecture claim that it yields significantly better results than any other image segmentation algorithm (Ronneberger et al., 2015). Parenthetically, the U-Net in the context of HAB monitoring has seldom been developed, and the one model that was deployed focuses only on the Korean peninsula (Park et al., 2020). Thus, the idea was to perform the study with the most state-of-the-art tools and assess their ability to generate satisfying results even with little computational and epistemological resources.

The motivation for this study is to understand the opportunities and limitations for using freely available satellite data in combination with a simple implementation of deep learning to detect HABs in Dutch freshwater. The primary question is whether it is feasible to leverage a single U-Net CNN model to reliably detect HABs across any inland body of water in the Netherlands, given limited resources. This question considers a presupposed complication that might arise in applying a simple model to predict HABs for many water shelves, rather than for one, similarly to how the U-Net developed by Ronneberger et al. (2015) was trained to recognize tumours on brain scans that all look relatively similar.

Regardless of if the solution is indeed scalable, the first point of interest of the methodology for this research is the process of data collection. The data compiled consists of satellite images depicting various lakes in the USA and their corresponding HAB segmentation mask that are made available by American governmental institutions. Furthermore, a small set of satellite images that depict the Zeeland region were downloaded from the ESA's Copernicus browser for testing the model's performance when generating predictions in the Netherlands. These are to be matched with in-situ measurements that were retrieved from the Bloomin' Algae app.

The objective is to construct two models that have the same architecture but with different training datasets to then compare their performance on three different tasks. Having this set of models and their error metrics for the different tasks should allow for answering the study's

primary question, whether it is possible to leverage one simple U-Net CNN to reliably predict HABs, and if it is scalable for use across the Netherlands.

The overarching problem of water contamination is one that is central to help preserve the environment for future generations. It is of great importance to provide efforts that help to mitigate the issue, as stated by the United Nations in the sixth Sustainable Development Goal (UN, 2024).

Managerial and academic relevance

In the specific context of HABs in the Netherlands, stakeholders that are involved in the issue are the entire Dutch society (in terms of their health), the Dutch government through lake conservation entity, Dutch agricultural entities, Dutch farmers, Dutch tourism entities, as well as Environmental institutions at large. Indeed, the objective outcome of this thesis could provide a monitoring tool for interested stakeholders to inform their decision-making or help researchers gather vast amounts of data, at high spatial and temporal resolutions.

Literature in terms of HABs is extensive, and satellite data has been used to monitor them already in the past. However, there remain some gaps. One meta-analysis on the topic of HAB detection through remote-sensing methods identified the relative lack of machine learning algorithms used to help scale HAB detection (Khan et al., 2021). This same review also found that research was concentrated on specific areas, namely China and the United States, while other continents were overlooked and hence could benefit from being studied more extensively, in the hope of advancing a global understanding of HAB dynamics (Khan et al., 2021).

2) Theoretical Background

Traditional HAB Monitoring Techniques

Since the era of industrialization and intensified farming practices, there has been a global surge in the occurrence of HABs. Researchers have employed traditional techniques to monitor these phenomena, focusing on various bacteria responsible for the blooms. Traditional HAB monitoring involves collecting water samples and analysing their composition. Biomass and biovolume abundance are reported in terms of total phytoplankton, total cyanobacteria, and

individual species abundance (Ho & Michalak, 2014). This is done with the help of proxies such as Chlorophyll-a (chl-a) concentration (Ho & Michalak, 2014). Furthermore, the samples indicate species-specific metrics through DNA analysis. Cell counts, biomass, biovolume, and the presence of surface scum are also used as indicators (Ho & Michalak, 2014; Millie et al., 2009).

However, these traditional methods are labour-intensive, costly, and not suitable for continuous, wide-scale monitoring. Collecting and analysing samples manually over extensive water bodies is impractical and economically unviable (Caballero et al., 2020). This limitation has driven the search for more efficient and comprehensive monitoring methods.

HAB Monitoring and Remote Sensing

Advances in technology have led to the integration of remote sensing techniques with traditional in-situ methods. Remote sensing utilizes the spectral analysis of radiation to identify characteristics of HABs, leveraging the knowledge gained from in-situ techniques (Ho & Michalak, 2014). Metrics such as seston abundance, cyanobacterial abundance, Secchi depth and phycocyanin content can be derived from spectral data to quantify the extent and importance of a bloom (Ho & Michalak, 2014). The latter metrics are then compared against thresholds set by official organisms (such as water authorities) as to what constitutes an HAB – when the bloom is considered toxic. Methods such as these have been successfully implemented in web applications like those developed by the USGS, which provide a general view of HABs across the USA (USGS, 2017).

Indeed, the successful integration of remote sensing into HAB monitoring practices is further corroborated by Caballero et al. (2020) which state that the Sentinel-2 twin satellite mission of the Copernicus program, when combined with in-situ data, offers valuable spatiotemporal information and can map small blooms at a 10-meter spatial resolution, revealing detailed surface patches and heterogeneous distributions. Despite these advancements, limitations persist in the use of new methods. First, there is a large pool of satellites and onboard instruments to choose from with each their different advantages; in the ESA's Copernicus programme, the Sentinel 2 satellite possesses higher spatial resolution (10 meters per pixel) compared to Sentinel-3, however, the Sentinel 3 satellite's OLCI instrument has a higher spectral resolution and can identify a larger array of wavelengths that might offer more precise detection of HABs (Caballero et al., 2020). Parenthetically, though less spatially resolved, the

Sentinel 3 is preferred for the purpose of this study due to the availability of segmentation masks necessary for a U-Net CNN. This is supported by Von Tress et al. (2021) which confirm the low resolution and the inability of satellites to distinguish specific spectral bands of different HAB species. Other papers state that while proxies such as the wavelengths of specific pigments like chl-a and phycocyanin do help in recognizing HABs, though these methods face limitations due to disturbances from other materials in the water and the limited availability of sensors with fine spectral and spatial resolution (Khan et al., 2021; Viso-Vasquez et al., 2021; Sagan et al., 2020)

Thus, while supplementing in-situ monitoring of HABs with remote sensing is already an effective solution, there remains some limitations. The economic factor of the monitoring systems remains unaddressed, as scalability of the applications of these methods in both spatial and temporal qualities would require significant gathering of new in-situ data. Hence, some articles recommend that they should supplement rather than replace traditional monitoring methods (Von Tress et al., 2021). However, new statistical methods that achieve more performant results may be able to shift the balance towards remotely sensed monitoring and thus drive down economic costs significantly.

HAB Monitoring and Machine Learning

Indeed, Machine learning (ML) has emerged as a powerful tool for HAB detection, leveraging advances in algorithms, computing power, and data availability (Sagan et al. 2020). ML methods can automatically learn from data to develop models for detection, estimation, or classification without explicit programming (Sagan et al., 2020). Khan et al. notes that studies have used a wide array of models, such as support vector machines, gradient boosting, random forests, and convolutional neural networks. This has proved to achieve satisfying results in the context of HAB monitoring; for instance, Pahlevan et al. (2020) demonstrated that their Mixture Density Network, a form of neural network, outperformed existing threshold-based HAB detection algorithms, significantly reducing errors in chl-a measurements and improving accuracy. Moreover, Khan et al. (2021) describe that an instance of random forest application could effectively deal with multisource data and was not affected by the outliers, which traditionally are two key considerations for working with HAB monitoring.

However, ML methods face recurrent remote sensing challenges, such as sensitivity to uncertainties in reflectance objects present in satellite imagery and the variability of optical

proxies and their properties across regions (Pahlevan et al., 2020). Additionally, ML models do not necessarily resolve the economic dimension of HAB monitoring as the models still require extensive collection of in-situ data, such as the chl-a measurements, as shown by a range of studies (Khan et al., 2021). Still, there is great potential for the use of ML in this context, given that deployment of these models is not widespread yet. Indeed, Khan et al. (2021) note that though ML methods have proved advantageous in many other applications of remote sensing, they were only used in 5% of the HAB studies reviewed. They also mention that classifying HABs into high algal bloom, low algal bloom, or no bloom has not been attempted, as studies tend to favour regression techniques.

HAB Monitoring and Neural Networks

Neural networks (NN) have already shown great promise in remote sensing applications for HAB detection. Studies have reported improved performance using NNs compared to traditional methods (Pahlevan et al., 2020). Khan et al. (2021) state that empirical-based methods are in fact easy to implement as no prior understanding of water−light interaction is needed, and because these models take into account the characteristics of a specific water body, they tend to provide better results as compared to other spectral based models. For instance, Park et al. (2019) employed NNs to detect red tides around the Korean peninsula, achieving high accuracy when matched with in-situ measurements.

Kim et al. (2019) on the other hand successfully deployed a U-Net model for pixel-specific HAB detection, demonstrating the potential of this approach. However, their study also focused on oceanic red tides, and using low-resolution imagery that remains fixed to the Korean peninsula. This highlights the need for further research into U-Net CNNs for detecting smaller-scale cyanobacterial blooms in inland water shelves using higher spatial resolution data like Sentinel-3 imagery, which, incidentally, covers the entire globe.

Despite their potential, NNs face limitations. Khan et al. state the most prevalent issue, in that all NN applications are developed for a specific water body only and hence are not transferable. Moreover, they are also impacted by atmospheric uncertainties, and rely on large training and validation datasets, which can be difficult to obtain (Pahlevan et al., 2020; Sagan et al., 2020). The networks also come with high computational costs and the risk of overfitting, given that they rely on optical proxies that vary regionally (Sagan et al., 2020). The complexity of implementing neural networks is further corroborated by the fact that deeper, more complex

models need to be constructed to obtain significant results, which enhances computation costs still and risks more overfitting yet due to the increments of trainable parameters and sometimes inadequate training samples (Sagan et al., 2020).

Research motivation

The challenges research still faces on the issue of HAB monitoring can be synthesised as the limitations of scaling traditional remote sensing techniques in terms of their economic cost, the complexity of scaling neural networks geographically and the lack of effective monitoring in the Netherlands. Meanwhile, the use of networks and other ML methods remains promising for HAB monitoring, especially given the ever-wider availability of satellite data. Of these networks, the U-Net is especially promising given that it is a highly performant model that has seldom been attempted in the context of HAB monitoring save in the Korean peninsula. Furthermore, its attractiveness as a solution is reinforced given its ability to perform image segmentation, which would allow precise mapping of the extent of blooms and allow for a simpler geographic scalability than models that would have to observe lakes on a case-per-case basis. The U-Net also does not need extensive datasets to generate usable predictions which shows great potential to significantly drive down the economic costs associated with current means of HAB monitoring. Thus, this study aims to explore the feasibility of training a simple implementation of a U-Net CNN using Sentinel 3 images of American lakes affected by HABs, with the objective of enabling systematic HAB monitoring across waters in the Netherlands.

3) Methodology and Data

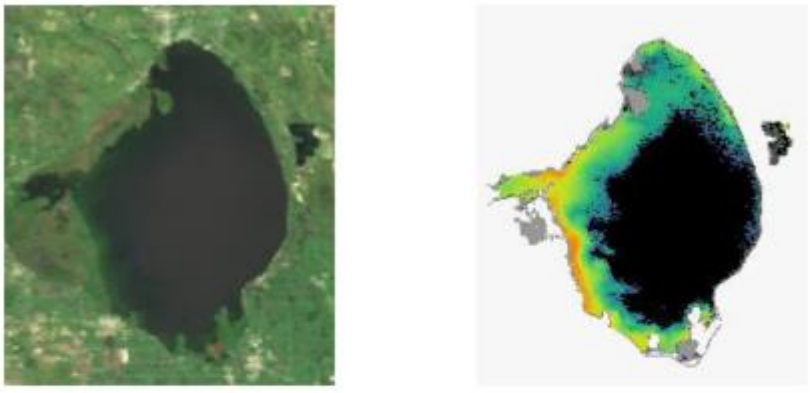
The methodology is somewhat inspired by the research conducted by Pahlevan et al. (2020), who developed a Mixture Density Network and assessed performance by comparing results for two different tasks: predictions generated using the MSI instrument aboard Sentinel 2 and predictions generated using the OLCI instrument from Sentinel 3. In broad terms, the study first prepares two sets of data, one which includes images of one lake only, Okeechobee in 2022, and the other with images of three different lakes, Okeechobee in 2022, and Pontchartrain and Green Bay in 2024, to train two separate neural networks with the same architecture. Subsequently, the objective is to compare the performance of these models on three tasks: predicting HABs for a single lake that was already trained on (Okeechobee), predicting HABs

for a single lake that was previously unseen during training (Albemarle Sound) and the qualitative predicting of HABs for the waters around Anna Jacobapolder in the Zeeland province of the Netherlands. For testing purposes, three additional datasets would have to be compiled. First, and to avoid data leakage, a dataset of images of Lake Okeechobee in 2021, and second, a dataset of Albemarle Sound in 2024; lastly, a set of images of the Zeeland province and the corresponding in-situ measurements. Ultimately, this process would allow to answer the research questions, first if neural networks are sophisticated enough to recognize HABs with little to no geoscientific pre-processing and then if it is preferable to leverage a model for one lake only (i.e. focusing training and application on that lake) or if it is possible to generalize model use to potentially monitor any lake in the Netherlands.

Training data

- **Retrieving data**

The data used throughout the bulk of the project consists of satellite images depicting selected lakes in the USA compiled by the National Centers for Coastal Ocean Science (NCCOS), a subsidiary organization of the National Oceanic and Atmospheric Administration (NOAA). Selected lakes were Lake Okeechobee in Florida, Lake Pontchartrain in Louisiana, Green Bay in Wisconsin and Albemarle Sound in North Carolina.



S3 OLCI Image HAB Segmentation

Figure 1: NCCOS-compiled Sentinel 3 Images of Lake Okeechobee

For Lake Okeechobee, images were downloaded for the entire year of 2022, resulting in a dataset of 366 images. These were originally taken by the Ocean & Land Colour Imager (OLCI) instrument included aboard the payload of the ESA's Sentinel 3 satellite. They represent the lake in its natural state such as is seen by the human eye and have dimensions of 196 in width and 225 pixels in height, with 3 colour channels (red, green and blue). As mentioned previously, resolution for this instrument amounts to one pixel representing an area of 300 by 300 meters, or 90,000 square meters. Furthermore, each image came with a corresponding segmentation mask, labelling the extent of HABs pixel per pixel with a gradient colour scheme to indicate the state of development of algae in that area, with values of blue signifying early bloom of the algae and values shifting towards orange signifying greater longevity of that instance of algae, while water pixels with no algae were black (RGB value = 0) and land pixels were omitted. Lastly, images with the same properties were downloaded for 255 days of 2021, for testing purposes.

Images for Lake Pontchartrain (see Appendix A) and Green Bay (see Appendix B) possessed the same characteristics, also including ground truth segmentation masks, however their size differed: images of the lake in Louisiana covered a wider area, with dimensions of 650 in width and 323 in height, while Green Bay images had dimensions 410 x 537. Further, images from the start of 2024 until the end of May 2024 were the only ones available, thus 83 images were collected for Green Bay and 130 for Lake Pontchartrain.

For Albemarle Sound (see Appendix C), 103 images from the 13th of February to the 9th of June 2024 and their corresponding segmentation masks were collected. These images have dimensions 530 x 673, and display mostly the entire bay, as well as small freshwater lakes such as lakes Mattamuskeet and Phelps.

Region	Dimensions	N	Segmentation masks?
Lake Okeechobee 2022	196 x 225	366	Yes
Lake Okeechobee 2021	196 x 225	255	Yes
Lake Pontchartrain	650 x 323	130	Yes
Green Bay	410 x 537	83	Yes
Albemarle Sound	530 x 673	103	Yes
Anna Jacobapolder	100 x 100	8	No

Table 1: Summary Statistics of Sentinel 3 Images

A few key observations on the aspect of these data, and how they differ between regions were made. First, it is clear that not all satellite images are created equal; while some represent the area of interest perfectly, with no undesirable noise, others have suffered from the conditions of satellite imagery. There are images that appear cut in half by smooth, black and/or white diagonal bands, representing the slight shift in the satellite's swath on that given day which prevented it to cover and map the area in its entirety (see Appendix E). Furthermore, there are days where the satellite's OLCI instrument captured strong sunlight reflection off the water's surface, possibly as a result of the time of day when the swath over the area occurred (see Appendix D). Finally, there are days when clouds obstruct a large portion of the lake and render the recognition of pixels pertaining to the algae class compromised (see Appendix F).

Second, there are some notable differences in the topography of the four regions depicted in the images. Lake Okeechobee in Florida is generally the most consistent, in that it shows a very simple elliptical lake shape, with relatively few clouds, well contrasted colours (lush green vegetation, dark waters) and easily identifiable bloom patterns. In comparison, Green Bay in Michigan is generally cloudier, has a drier land profile, waters of a lighter blue and shows fewer blooms overall. Images of Lake Pontchartrain have the most dissimilar aspect as they display a very large area covering the city of New Orleans, the aforementioned lake above it and a large portion of land around the two, with sparse, smaller bodies of water scattered around. Images of Albemarle Sound strongly resemble the aspect of the latter, with more contrasted green and blue colours. These subtle, differing characteristics are expected to play a significant role in the multi-lake model's ability to generalize the task of identifying HABs across any number of lakes. Indeed, the slight variance in water colours, contrasts between surface water and bloom pixels and land and water body profile should allow the model not to overfit a certain region but learn to recognize more abstract, key features that make up an HAB and that can be found in lakes across the world, regardless of local conditions.

- **Pre-processing the data**

To ensure the smooth operation of the neural network, the images had to undergo a certain amount of pre-processing before being ready for training. First, the target variable, the HAB segmentation image as seen above, was deemed too complex for the scope of this study. Indeed, it represents categorization of different levels of HAB evolution in a continuous form – this would equate to having possibly hundreds of classes for the network to learn to recognize. Thus, the segmentation was drastically simplified to a binary mask, with pixels that are part of

the algae class set to 1 and those that are not to 0. Essentially, this produced an image akin to the original segmentation, only with the coloured pixels all transposed to white while the unaffected waters and land became black (when multiplying the new values by 255 for observational purposes). It is assumed that a performant model would be much easier to produce with this simplified task of simply recognizing if a given pixel represents an HAB or not.

As for the features, the natural images, no processing was executed in terms of modifying pixel values, such as for regularizing brightness, removing cloud cover, water reflection or any other potential processing that can be implemented for satellite imagery. The idea behind this hands-off approach towards the raw images is to observe the ability of the network to adapt and learn how to perform the task even with retaining defects in the images. This would answer the question if it were indeed of primordial importance to include geoscientific pre-processing or if the model is indeed sophisticated enough to overcome these challenges.

Nevertheless, to ensure that the model's training would be optimized and to avoid severe overfitting, resizing, selection criteria and data augmentation techniques were employed. The images were resized using a tiling technique, which cut out 100 x 100 squares (or tiles) of the original image, with the number of tiles based on the truncated integer result of the division of the original image's width by the tile's width (x-axis), times the same result of the same operation for the height (y-axis). For Lake Okeechobee, this resulted in 196/100 = 1 and 225/100 = 2; thus, the resulting 2 tiles were taken by shifting along the y-axis while remaining in the same position on the x-axis.

This technique left the data with several outliers (here considered as outlier tiles). Indeed, some of the tiles were left with no or very few algae pixels. Thus, a selection criterion was introduced to select tiles which possessed at least 20 pixels corresponding to the algae class, according to the segmentation mask. The reason for this criterion is to prevent overfitting: indeed, the model would not learn as efficiently by looking at tiles where there are too few algae pixels, since it would likely optimize weights to fit patterns too specific to that given image.

Furthermore, the resulting images were all augmented (or duplicated) with inducing a random left or right, up or down transposition. This process doubles the number of images to provide more variance and subsequently enhances the robustness of the model as it receives more diverse training data.

A final pre-processing step that was not included in the preparatory script but rather as a function that is applied to every selected batch of images during training is the normalization of the natural images. This scaled all the pixel values down from 0-255 to 0-1, to ensure the well-functioning of the model and limit the importance attributed to random, excessive variance in the RGB values of the input images, which could be caused by some of the key considerations listed above such as abundant sunlight reflection off the water's surface.

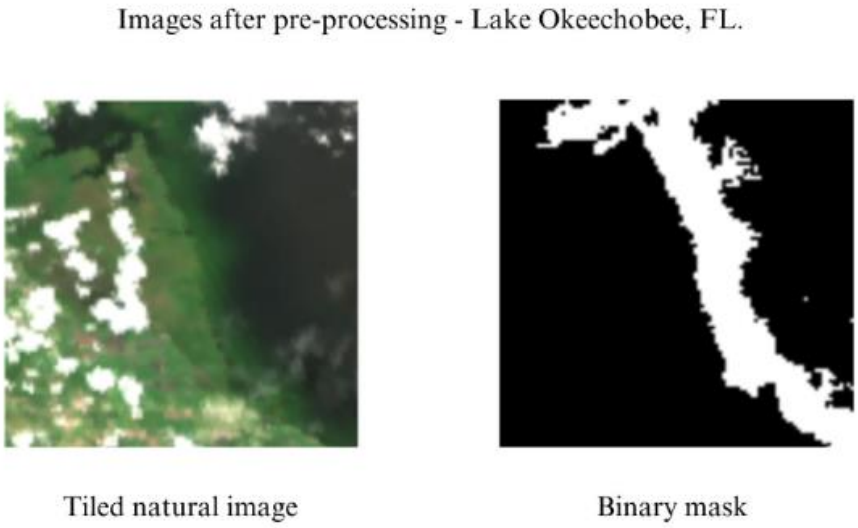


Figure 2: Lake Okeechobee Images after Pre-processing

- **Compiling the dataset**

With all pre-processing completed, the dataset was ultimately compiled. For the model trained uniquely on Lake Okeechobee, this consisted of 996 images and 996 of their binary mask counterparts. These were then split into training and testing sets using a 75/25 ratio, which amounted to 747 images and their masks randomly selected to compose the training set and 249 images and their masks forming the test set.

For the model trained on multiple lakes, namely Lake Okeechobee, Lake Pontchartrain and Green Bay, further subdivision was undertaken. To attempt to bring balance in the number of images per lake, the 80 first and 80 last images of Lake Okeechobee were dropped, especially

considering that they represent the early spring and winter months of 2022 and presented few HABs overall. Thus, the dataset for the multi-lake model contained 206 images (and their masks) of Lake Okeechobee, 130 images of Lake Pontchartrain and 83 of Green Bay. This totalled to 419 images, which were then processed and augmented. The final dataset resulted in 812 images, which were randomly split into training and testing sets with a 75/25 ratio: 609 images and their masks in the training set and 203 in the testing set.

The test datasets also underwent some of these modifications, which resulted in 273 images for the 2021 Lake Okeechobee set and 740 images for the Albemarle Sound set.

Regions	N	Training/Testing?
Okeechobee 2022	996	Training
Multi-lake	812	Training
Okeechobee 2021	273	Testing
Albemarle Sound	740	Testing
Anna Jacobapolder	8	Testing

Table 2: Summary of Compiled Datasets

Netherlands data

Data for the Anna Jacobapolder area situated in Zeeland were also retrieved (see Appendix G), with the purpose of assessing the feasibility of scaling the use of the models trained on American waters to Dutch water shelves. Thus, in-situ measurements (courtesy of Pr. Mackay of the "Bloomin' Algae" app) that indicated when the area faced HABs were matched with images from the Sentinel 3 satellite's OLCI instrument (retrieved from the Copernicus browser) based on corresponding dates. Dates for the measurements ranged over the late summer of 2023, from August to September. The eight retrieved images were of the same quality as the tiled natural images mentioned in the previous section; they also had dimensions 100 x 100 and had three channels (RGB). In their aspect, they resembled more so Lake Okeechobee with the clear contrasts between water, land and HAB areas.

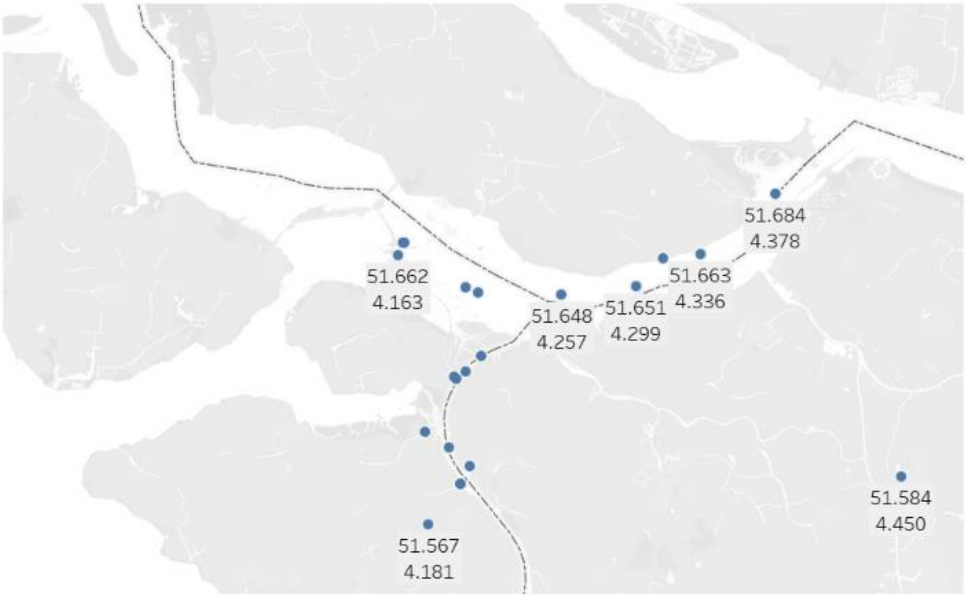


Figure 3: Locations of the In-situ Measurements of HABs in Zeeland

Data Leakage

Data leakage is not a very complicated issue to manage in the context of this study. Given the wide availability of data for a variety of lakes as well as for a range of different years, training is set to be conducted on sets of images that are not the same as the ones used for the two tasks designed to assess performance. Furthermore, for testing the models at various points in its training, the initial data sets are split into a training and testing set using a 75/25 ratio split. The above considerations ensure that any data leakage is prevented throughout the study.

U-Net Convolutional Neural Network

- **Architecture**

The analytics approach that was adopted for the purpose of this study is a convolutional neural network with a U-Net architecture to perform image segmentation. This is a supervised application of deep learning, a field of machine learning which creates artificial neural networks that attempt to model the functioning of human neurons. This is achieved by iteratively computing nodes that each learn to recognize a specific pattern or feature of the input and are then used to recognize higher-level entities as part of a specified task. The model is made somewhat easy to implement through Python's torch library.

The U-Net is a modified version of a convolutional neural network that was first developed by Ronneberger et al. in 2015 to help resolve the task of identifying pixels corresponding to tumours on brain MRI scans. Their innovation consisted in the characteristic U shape of the diagram they composed to explain their method: the model takes an image input which is first contracted in the encoder path, then subsequently expanded and developed into an outcome in the decoder path. More specifically, throughout contraction (or encoding) the model down-samples the image and learns its most salient features, from more detailed, low-level ones at the start of the path to coarser, higher-level ones at its lowest point (or bottleneck), to then up-sample and construct its predicted segmentation with the help of skip connections, which are feature maps saved at each layer of the encoder path, in the expansion or decoder path. The specific model applied for the purpose of this paper is an adaptation of a model that was developed for segmenting brain MRI scans and recognizing tumours throughout student Tiril Mageli's coursework at Imperial College London (Mageli, 2024).

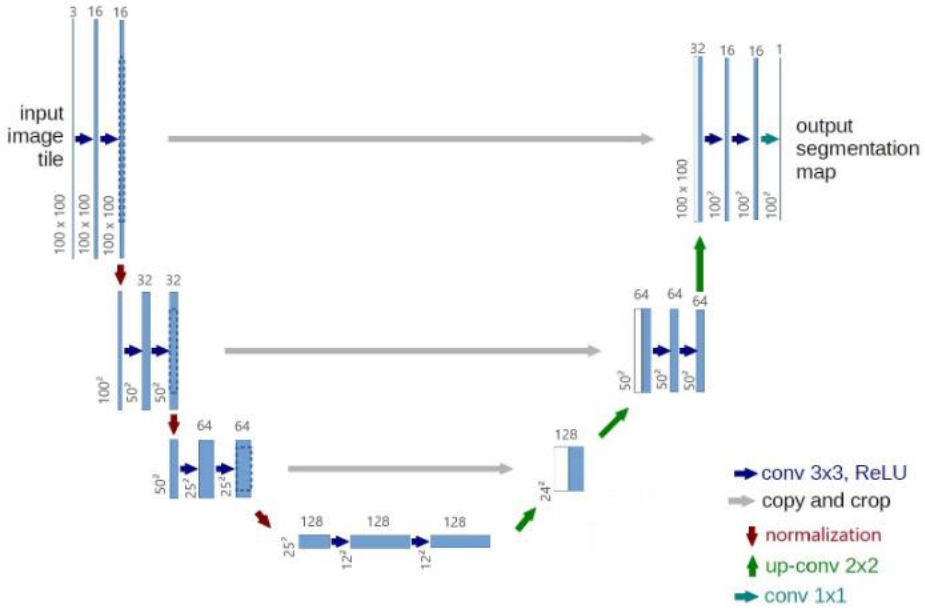


Figure 4: Diagram of the Adapted U-Net

The model consists in 8 different layers, with four convolution blocks that form the encoder path and four deconvolution blocks that form the decoder path. It is initiated with a forward pass, starting at the first convolution block and its 16 filters. Within this block, the filters slide over the input and apply convolutional operations, which output a single value per observed region: the nodes that, put together, form a feature map. The nodes are subsequently transformed by batch normalization and then a rectified linear unit (ReLU) activation function.

This sequence is repeated twice, first to extract low-level features such as edges and corners and then, given an increase of the stride parameter, to allow larger-scale processing of the observed features to decipher higher-level patterns. This results in an output of 16 feature maps. The model then moves on to a second convolution block which doubles the number of filters and subsequently outputted feature maps to 32, then to third and fourth blocks which double the number of filters and outputted feature maps each time to arrive at the bottleneck of the model with a total of 128 feature maps. At this point, given parameter changes, the initial dimensions of the input image have shrunk down to 5 x 5 for each feature map. The bottleneck is then followed by four deconvolution blocks which up-sample the feature maps by combining the n-1 output with the corresponding skip connections to produce the predicted HAB segmentation mask, going from 128 feature maps to the single output mask.

The different convolution and deconvolution blocks possess a range of parameters to perform their convolution operations. These parameters consist of the kernel size, stride and padding. The kernel size represents the number of pixels that form the dimensions of the convolution kernel, or filter (such as a small window which slides over the image). Thus, a kernel size of 3x3 suggests that the filter is a 3 x 3 matrix. Parenthetically, this means that there are 9 randomly initialized weights assigned to that filter, plus a bias term, which are used in the computation of a node. Stride indicates this kernel's movement in number of pixels from columns left to right, and once the given row has been fully observed and a node computed at each position, downwards. Here, the first convolution block has a stride of 1, which indicates that the filter moves one pixel at a time, ensuring that each region of the input is observed in detail. This is sometimes increased to 2, which means that one pixel is skipped over, and the output dimensions of the feature maps are subsequently halved, as twice as less nodes are computed. Padding is the outline of artificial, zero-value pixels added as an outline to the image for the border pixels to have the opportunity to be at the centre of the convolution kernel as it is shifted across the image. Padding is set to 1 in the first block, which means that there is an outline of one artificial pixel around the input image and ensures outputted feature maps keep the same dimensions as the input (provided the stride equals 1).

After convolution, the individual outputs of the sliding convolution kernel, the nodes, are normalized using batch normalization, which provides a means of regularization and helps stabilize and accelerate training. This operation is followed with an activation function: the Rectified Linear Unit (ReLU) (1), which applies an element-wise operation, replacing all

negative values with zero while keeping positive values unchanged. It also introduces non-linearity into the model to capture more complex patterns and prevents the vanishing or exploding gradient problem.

(1) $\mathrm{ReLU}(x) = \max(0, x)$

In broad terms, these previous paragraphs have summarized what constitutes the forward pass of the U-Net neural network used in this study, whose architecture had to be coded in manually. After completion of this first pass, the model performs a backward pass; here, the loss function (the error), a penalty for wrongly classified pixels, is computed which informs the update of a set of randomly initialized weights that were assigned to each filter of each layer – indeed, the weights are updated with the objective to minimize the error. This starts at the output, the final deconvolution block, and, using a backpropagation algorithm, computes the gradients, or the contribution of each weight towards the error, and subsequently updates these weights based on the resulting gradient and a predefined learning rate. The chosen optimization algorithm was the Adam one, with a learning rate set to 0.001. To compute loss, the Binary Cross-Entropy with logits method was chosen due to it being well adapted to binary classification tasks.

Further parameters that were manually set were the training iterations (10,000) and the batch size (16). During an iteration, the model performs a forward pass to compute the predictions for each image in the batch, computes the loss, and then performs a backward pass to update the model's weights. Using batches of this size allows for a simple computer to handle the complex task and leads to more stable and reliable gradient estimates compared to using a single data point (stochastic gradient descent) or the entire dataset (batch gradient descent).

Parameter	Value
Layers	4
# of Filters	16
Batch Size	16
Activation Function	ReLU
Loss Function	BCEwithlogits
Optimizer	Adam
Learning Rate	0.001
Training Iterations	10,000

Table 3: Summary of Network Parameters

This same methodology was used for both single-lake and multi-lake models. Based on the stated parameters and training data numbers, we can compute the epochs, or the number of iterations required to observe each instance of the training dataset, that constitute each model (2, 3). For the single-lake model, the number of epochs equals:

(2) 747 / 16 = **47** (rounded to the higher nearest integer)

For the multi-lake model, the number of epochs equals:

(3) 609 /16 = **39** (rounded to the higher nearest integer)

- **Evaluating and interpreting predictions**

To understand the performance of the model and benchmark the two models against each other, several methods, using both qualitative and quantitative means, were retained. A first technique to assess the results, given the graphic, visual nature of the output of the model, is to print out the *predicted segmentation masks* and compare the accuracy of the predicted algae pixels with regards to the ground truth segmentation. This provides a very accessible and salient understanding over performance as it appears extremely clear if the model is generating accurate predictions – if the algae pixels match between the two masks, then it is a satisfactory result. However, while it can provide this initial, solid first impression, it is not sufficient to use only this method, especially when trying to evaluate which model performs best.

Thus, a range of test metrics that are adapted to image segmentation tasks were employed. A first metric that remains very simple and similar in nature to the visual comparison is computing the *pixel accuracy*, which essentially provides a percentage of how many pixels were correctly classified in the predicted segmentation mask. This is calculated as the total number of matching pixels between the predicted and actual segmentation masks for a given picture, divided by the total number of pixels in the image, thus taking in account all classification terms. The mathematical representation is as follows:

(4) $$PA = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FN_i)}$$

Another metric that is commonly found to assess the performance of a model in image segmentation tasks is the *Jaccard coefficient*, or the *Intersection over Union* (IoU) coefficient. This attempts to capture the similarity between two datasets, which here are the ground truth and predicted segmentation masks. It is computed by dividing the intersection of predicted positive pixels and true positives with the union of both predicted and true positive pixels (positive here references that the pixel represents an HAB and equals 1). This provides a more accurate measure than the pixel accuracy as it does not take in account true negatives which makes it more robust when it comes to imbalanced datasets. An imbalanced data set in the scope of this study would be one where there are images with very small portions of HABs compared to the size of the image. The mathematical representation of the IoU is as follows:

$$\text{(5)}\quad \text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}}$$

Lastly, the *Dice coefficient* or *F1 score* was also used as a further test metric to benchmark the two models against each other and compare performance. Akin to the IoU, it measures the overlap between the predicted segmentation and ground truth masks but with a different formulation. It doubles the number of intersected pixels between the two masks and divides the result by the sum of all pixels from each mask. The metric was included as it also gives a great idea of overlap, while balancing the importance of false positives and negatives given the multiplication by 2. It is also very useful for imbalanced datasets where the object of interest is much smaller compared to the background, as it emphasizes the overlap more than pixel accuracy. The formula is as follows:

$$\text{(6)}\quad \text{DSC} = \frac{2\cdot\text{TP}}{2\cdot\text{TP+FP+FN}}$$

The loss function for the network was computed using the Binary Cross-Entropy with Logits method. It measures the difference between two probability distributions: the predicted probabilities and the actual binary labels. The BCE loss for a single prediction is defined as:

$$\text{(7)}\quad \text{BCEWithLogitsLoss}(x, y) = \log(1 + \exp(-|x|)) + \max(0, x) - x \cdot y$$

In practice, the model outputs raw scores, the logits, rather than probabilities. These logits can take any real value and are converted to probabilities by applying a sigmoid function, which in python is combined into a single step. The loss allows the model to learn as it provides an

objective to minimize (low values mean better performance) and thus informs the update of the weights associated with each filter.

4) Results

Training the models over 10,000 iterations took roughly 3,300 seconds each. Subsequently, results were obtained for the two tasks that were detailed in the introduction to the methodology. All of the figures were rounded to the nearest third decimal.

First, the two models were tested by generating predictions for new images of an already observed lake, Lake Okeechobee. The first model was trained exclusively on images of this lake, while the second model was trained on images of this lake as well as two others, Lake Pontchartrain and Green Bay. Any issues with regards to data leakage were circumvented by downloading the 2021 set of images depicting the lake in Florida and conducting testing on that dataset. The purpose of this task is to understand if it is important to contain training and application on a case-by-case basis. If results differ significantly between the two models, then it can be said that there is a strong location factor during the networks' training for the purpose of recognizing HABs.

Secondly, the two models were made to generate HAB predictions for an entirely new, unseen lake during training: Albemarle Sound in North Carolina. The purpose of this task is to assess and compare the ability of the two models to generalize to any other body of water in the world. If predictions for the new lake demonstrate accuracy this would mean that the models are not too location-dependent and can indeed be scaled to a wide range of applications.

Model 1

Metric	Task 1	Task 2
Pixel Accuracy	0.951	0.869
IoU	0.941	0.963
Dice	0.387	0.053

Table 4: Model 1 Results

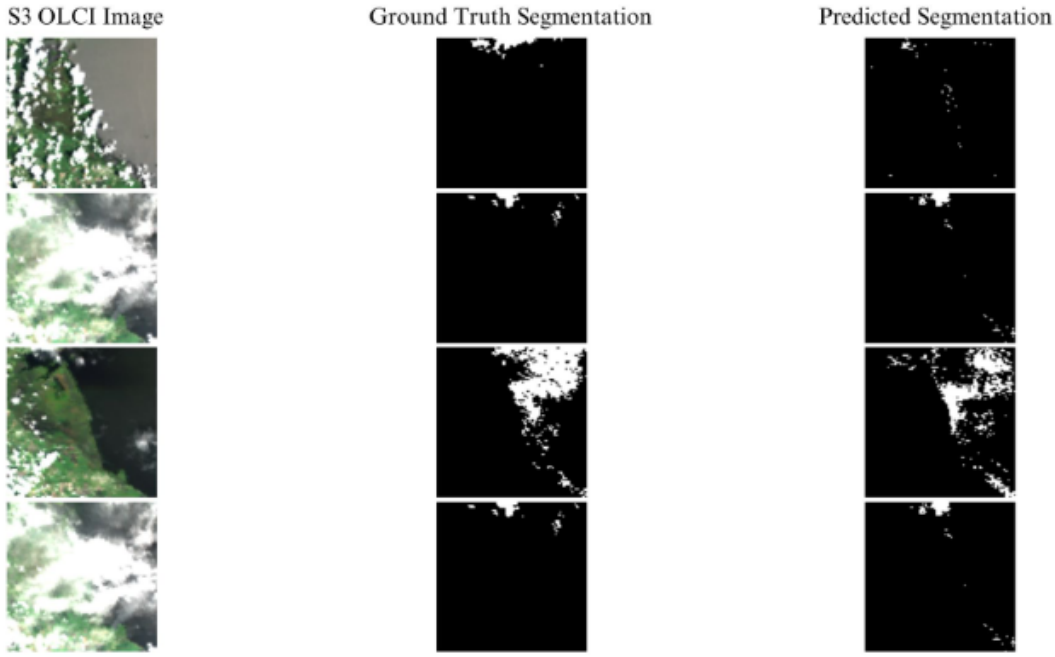


Figure 5: Model 1 Predictions Comparison, Task 1

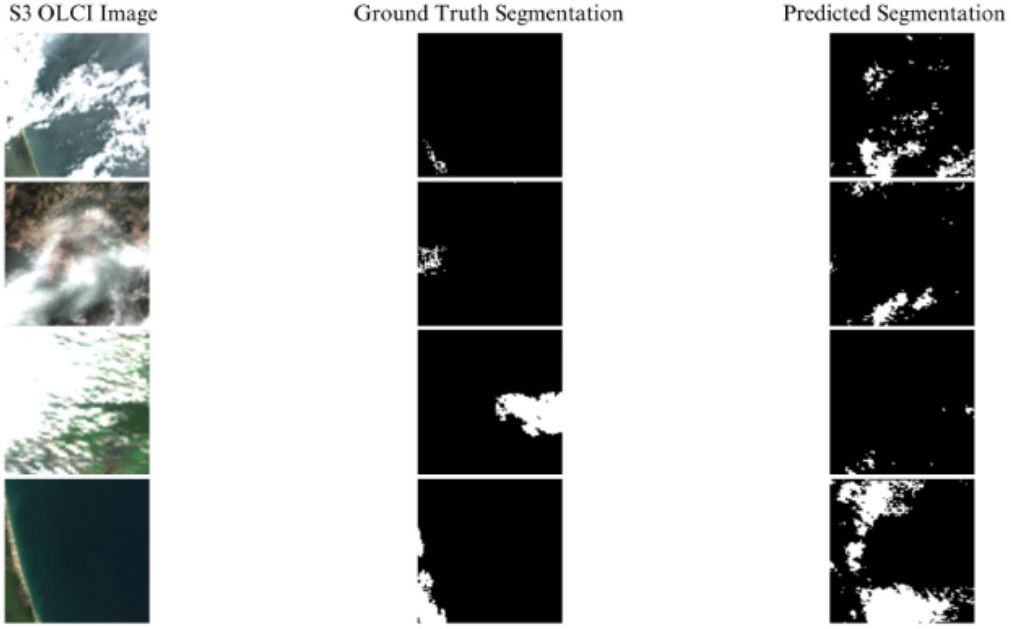


Figure 6: Model 1 Predictions Comparison, Task 2

Qualitative interpretation

This comparison between the natural colour images, the ground truth segmentation and the predicted segmentation masks allows for a clear visualisation of the model's performance. For the images of Lake Okeechobee in 2021, the model is able to generate quite accurate predictions of the location of HABs. This is demonstrated by the significant overlap between the two segmentation masks, where the column in the middle shows the actual repartition of HABs on the observed section of the lake, while the column on the right is the model's prediction. This level of overlap looks promising for the deployment of U-Net neural networks to recognize HABs; however, these predictions are only for the same lake, where accurate predictions are already being generated by the NCCOS.

For Albemarle Sound in the first half of 2024, predictions appear much less performant. Indeed, while HABs are being predicted, they do not appear to overlap correctly with their appearance on the ground truth segmentation masks. This is not promising for the scalability of the given model, as it seems to be largely location dependent, especially given that it was trained only on Lake Okeechobee. Potential reasons for this are developed in the discussion section of this paper.

Test metrics

- **Pixel Accuracy**: 0.951, 0.869

This high pixel accuracy achieved for the first task suggests that the model correctly classifies 95.02% of pixels in the test images. It indicates that the model is generally very good at predicting the right class for each pixel. For the second task, performance is unsurprisingly lower as images of the North Carolina lake were never seen by the model prior to testing. However, the metric remains high (86.92% of pixels correctly classified) which is promising and indicates the potential for scalable application.

For further context, this metric is typically very high for segmentation models, as it includes both the correctly classified background and foreground pixels. However, it can sometimes be misleading if there is a significant class imbalance (i.e., a lot more background than foreground, which is the case as illustrated by figures x, y, z). Thus, we look at the next metric, the Jaccard coefficient or IoU.

- **IoU (Intersection over Union)**: 0.941, 0.963

For the first task, an IoU score of 0.941 indicates that the model's predictions are highly accurate in terms of spatial overlap with the ground truth, i.e. there are many similarities between the predicted and the ground truth masks. This is indeed a strong performance metric, which suggests that the model is effective in segmenting the images accurately. Furthermore, the model's average IoU was surprisingly even higher for the second task than for the first one, with a coefficient of 0.963 for the overlap between ground truth and segmentation masks, which again appears very promising for the potential to scale application of the network.

For context, compared to the pixel accuracy, the IoU provides a more balanced view of the model's performance by focusing on the overlap between the predicted and ground truth segments rather than the correct classification of each pixel.

- **Dice Coefficient**: 0.397, 0.053

The Dice coefficient (or F1 Score) of 0.397, while lower than the pixel accuracy and IoU, still shows reasonable performance. The Dice coefficient is particularly sensitive to class imbalance and penalizes false negatives more than false positives. A score of 0.397 indicates room for improvement, especially in minimizing false negatives. Furthermore, the score being lower suggests that while the overall pixel-wise accuracy is high, the model might struggle more with capturing the exact boundaries of the segmented regions or with correctly segmenting smaller, less dense clusters of algae.

With regards to the second task, the model's performance is drastically lower than what was indicated by the first two metrics, with a struggling 0.052. This means that while the model is predicting an appropriate number of pixels as being of the HAB class, they tend to never be in the correct area compared to the ground truth segmentation. This calls for great improvements to the model, which are developed in the fifth section of this paper, or can be interpreted as a rather unmovable barrier for a model trained on only one lake in terms of its scalability to a variety of other lakes.

Model 2

Metric	Task 1	Task 2
Pixel Accuracy	0.943	0.951
IoU	0.936	0.958
Dice	0.351	0.021

Table 5: Model 2 Results

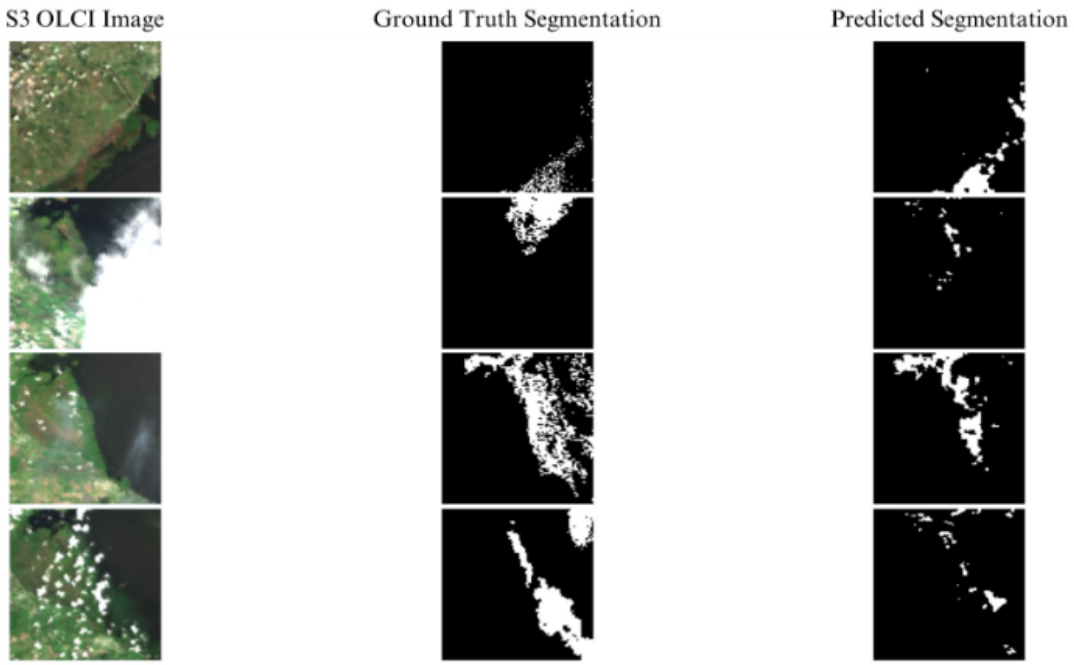


Figure 7: Model 2 Predictions Comparison, Task 1

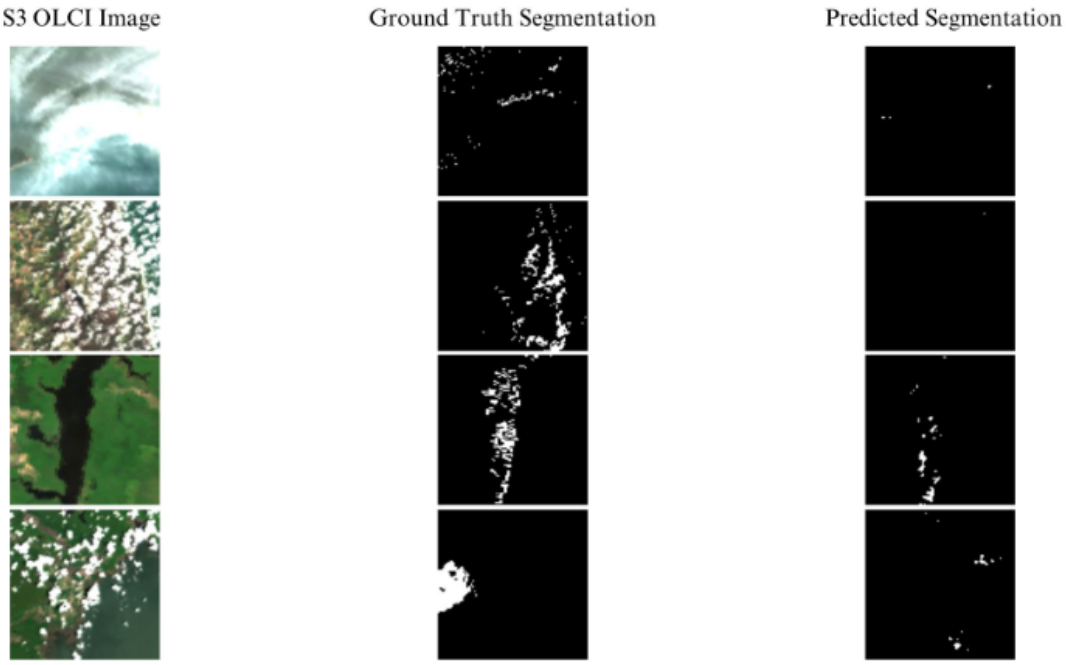


Figure 8: Model 2 Predictions Comparison, Task 2

Qualitative interpretation

For the images of Lake Okeechobee in 2021, the model trained on multiple lakes including the observed one is once more able to generate quite accurate predictions of the location of HABs. This level of overlap looks again promising for the scalability of the neural network to generate predictions for other lakes it was not trained on.

However, akin to the previous model, for Albemarle Sound, predictions appear much less performant. Indeed, while HABs are being predicted, they do not appear to overlap correctly with their appearance on the ground truth segmentation masks.

Test metrics

- **Pixel Accuracy**: 0.943, 0.951

This high pixel accuracy achieved for the first task suggests that the model correctly classifies 94.3% of pixels in the test images. It once more indicates that the model is generally very good at predicting the right class for each pixel. For the second task, performance is again unsurprisingly lower as images of the North Carolina lake were never seen by the model prior to testing. However, the metric remains high (95.10% of pixels correctly classified) which is better than the first model's result on that task and gives promise to the possibility of scaling a network based on the variety of topographies it is trained on.

- **IoU (Intersection over Union)**: 0.936, 0.958

For the first task, an IoU score of 0.936 indicates that there are many similarities between the predicted and the ground truth masks. This is indeed a strong performance metric, which suggests that the model is effective in segmenting the images accurately. The model's average IoU was once again higher for the second task than for the first one, with a coefficient of 0.958 for the overlap between ground truth and segmentation masks.

Reasons for this odd occurrence are that the model, for the unseen lake, might over-generalize and predict large regions as algae blooms, leading to a higher intersection with the ground truth (i.e. large but few errors). Conversely, the errors in the known lake might be more minute due to the complex structure of algae blooms and the accumulation of small errors across many regions reduces the IoU. Ultimately, the difference in IoU results across both tasks is not vast,

though this explanation should provide context as to why performance on the second one remains high.

- **Dice Coefficient**: 0.351, 0.021

The Dice coefficient of 0.351 for the second model is also lower than the pixel accuracy and IoU. Again, it indicates room for improvement, especially in minimizing false negatives, which translates to capturing the exact boundaries of the areas with pixels corresponding to the HAB class.

With regards to the second task, the model's performance is even lower than the first model's with a striking coefficient of 0.021. This means that pixels classified as pertaining to the HAB class tend to never be in the correct area compared to the ground truth segmentation. This result comes as a surprise given that the model was trained on a wider variety of lakes and that it might have thus been assumed to be more robust with regards to overfitting location. Further interpretation is included in the following discussion.

Zeeland Qualitative Test

Given the results to the second task, it is safe to assume that generalization of the models' application is not viable in their current state. However, given that the scope of the study includes the Netherlands, it remains important to observe the models' performance on the country. Since segmentation masks are not available for the Netherlands, the task must be adapted: in-situ measurements are meant to cross-check that there is in fact an HAB where the models are generating predictions. Thus, the results of this task fall into a qualitative category, as only the interpretation of the outputted masks of the models can be conducted.

The period August-September of 2023 demonstrated a significant issue of HABs in the waters around Anna Jacobapolder in the Zeeland province, according to the Bloomin' Algae records. Both models predicted that there were in fact HABs in the presented images, which is a satisfactory result; however, the model trained uniquely on Lake Okeechobee appears to outperform the multi-lake model by a landslide. Indeed, a simple interpretation with the naked eye shows that the lake at the centre of the image is entirely infested with HABs due to its brownish colour that is highly contrasted by the bluer waters around. This is aptly recognized

by the first model, though the latter is largely overpredicting, while the multi-lake model struggles to identify this, as it predicts small patches of HABs in odd corners.

Figure 9: Model 1 Predictions Comparison, Task 3

Figure 10: Model 2 Predictions Comparison, Task 3

5) Discussion

Answers

With the results now retrieved, it is possible to conceptualize some answers to the research questions that have guided this paper. The first consisted in observing if the U-Net neural network was sophisticated enough in it of itself to be able to recognize HABs with little to no geoscientific processing applied to the images prior to model training and testing. It is clear given the results to the first task that this is indeed feasible for the networks, as rather accurate predictions for the location of HABs were generated for new, previously unseen images of

Lake Okeechobee. This is a satisfactory and promising result, as it confirms that the model is able to perform well even given the strong variability of satellite imagery in terms of cloud cover, sunlight reflection and shifting colour schemes across seasons and weather conditions. For the second task, predictions from both models appear to be more random and are therefore less satisfactory; however, the output can still be considered with some value as there remain instances of accurately predicted HAB pixels. This indicates that the model is able to understand general and location-invariant features that characterize HABs, though this assertion holds only to a limited extent.

The second research question sought to know if it was possible to generalize the model's application to any water shelf in the Netherlands, and subsequently if it was preferable to focus model training and application in a location-specific way or a broader, multi-location approach. Given the results to the second task, which was to generate predictions for the previously unseen waters of Albemarle Sound, it is currently impossible to validate the feasibility of scaling the application of either single or multi-location models. Indeed, both the visual and statistical results indicate that each model is unable to classify pixels corresponding to the HAB class correctly, especially when looking at their Dice coefficients, which represents the degree of overlap between ground truth and predicted segmentations, respectively 0.053 and 0.021. Thus, with the current solution, it is preferable to limit training and application to a single location to achieve significant results. Possible avenues to explore that could shift this answer in favour of scalability are detailed in the limitations.

With the research questions answered, a new question emerges: why is it that the first model appears to significantly outperform the second one on most tests? The hypothesis for the multi-lake model was that given the variety in geological features it was exposed to during training, it would learn to recognize more generalizable features that characterize HABs across regions. However, it appears that this assumption did not hold, as predictions for the second and third tasks were quite random. This reinforces the idea that networks perform best given enhanced homogeneity in the observed images, as they seem to base their predictions strongly on a specific topographical profile.

Limitations

Given the limited scope of this study, there is a range of limitations that impede both the effective recognition of HAB pixels and subsequently the scalability of the networks'

application to a wider pool of lakes. First and foremost, the minimal extent of geoscientific pre-processing considerably hindered the ability of the models to consistently provide significant predictions. Indeed, it is abundantly clear that the variability in terms of weather conditions, such as the density of the cloud cover, the time of day and resulting sunlight reflection, or the region's topographical profile, such as the soil that composes the lakebed, the vegetation cover and a wide array of other parameters still, significantly affect the model's ability to consistently provide accurate predictions over time and space. This is clearly visible through the predictions generated for the second task, where the models do not seem to understand that Albemarle Sound is separated from the ocean by a terrestrial line. In fact, it appears that the freshwater is considered as land cover, which given its dark green, brownish appearance and sparse distribution that is highly in contrast with the bluer and homogeneous ocean, is sensical. Thus, the variation and resulting confusion of colours and water shelf shape and distribution which is due to the lack of geoscientific pre-processing constitutes a strong limitation to the models' ability to systematically recognize HABs.

A further limitation to this study is the spatial resolution of the Sentinel 3 OLCI images. The Copernicus program's satellite offers a resolution of 300 by 300 meters per pixel, which is rather suitable for observing large water shelves such as the lakes found in the USA. However, this presents some disadvantages; first, the level of detail is not very granular, as one pixel represents 90,000 square meters. Thus, lower-level patterns and features that make up the texture of HABs are somewhat blurred and effective recognition may be compromised as a result. Furthermore, this is essentially a colossal barrier towards scaling model application given that a wide range of water shelves would be represented by only a few pixels; for reference, Kralingse Plas in Rotterdam appears as one pixel through Sentinel 3 imagery.

A further limitation concerns the availability of ground truth segmentation to enable training of the U-Net neural network. The NCCOS appears to be the only source that provides accurate representations of HAB spread, and their operations are limited to a few select regions. Given that the U-Net requires the segmentation masks as a target variable, their limited availability hinders the potential to create extensive datasets that would represent a wide array of region profiles, to hypothetically prevent the overfitting that occurs in the results of this study.

The U-Net architecture deployed throughout this research may constitute a further limitation. Indeed, the model implemented remained rather simple to enable the available hardware to execute training effectively. The lack of complexity could prevent filters from accurately

deciphering unique HAB patterns that may generalize better and allow scaling of the models' application. This lack of complexity stems from the four layers, and the relatively low number of filters at the input layer: in original article by Ronneberger et al. (2015) that detailed the U-Net, the input layer started with 64 filters, while the current implementation starts with 16.

The tiling technique employed might also hinder the models' ability to effectively learn to recognize HAB patterns. Indeed, the tiling was executed somewhat randomly as a form of data augmentation, to generate extra variance in the dataset and hypothetically ensure the models learn patterns that are specific to HABs rather than those of the observed location. However, given the wide range of topographical profiles detailed previously, this might enhance confusion of features and thus limit the models' performance and ability to learn significant patterns that represent HABs.

A final limitation concerns the inflexibility of the U-Net regarding the image dimensions it accepts in its input tensors. A tensor, in *pytorch*, is an object that contains the image and filters which is passed through the model. Indeed, the U-Net works with a fixed input size, and thus images need to be homogenized. This limitation is the reason for the implementation of the aforementioned tiling technique – however, with no size limitations, it would be easier to methodically choose the image framing and thus have more consistency in the data.

Recommendations

Mending to the first obvious limitation would be a great opportunity for improving performance. Indeed, adding a range of geo-scientific pre-processing steps, such as removing the cloud cover, countering water reflection of sunlight, using images that suppress land cover and other means would homogenize images to a certain extent and allow for the models to be less location dependent.

Further, a more informed selection of the snapshot should be given further consideration. It is quite clear that predictions for Lake Okeechobee were generally the most successful, especially from the first model which was trained exclusively using images that represented it. Key characteristics of the lake were that it had a clear contrast of colour, shape and texture between land cover, water surface and HAB areas. The tiles that were developed omitted the smooth elliptical shape of the lake, however they still represented a clear and obvious demarcation of where the lake was situated and where the HAB pixels could thus be located. When comparing

to Albemarle Sound, the tiles appear much more confusing in terms of what we are presented with; indeed, it is occasionally hard to differentiate between water shelves and land cover, especially given that some of the tiles also include the nearby ocean that is much more contrasted in colour and shape than the inland reservoirs. The topography also varies vastly between tiles, as they present small lakes further inland, as well as larger inlets and the bay that adjoins the ocean.

While the model that was trained on a variety of locations was faced with the aforementioned scenarios, it obviously could not generalize features for HAB recognition, as it performed somewhat worse than the first model. Thus, it is an avenue for future research to thoughtfully consider the selection of the snapshot of the water shelves of interest. It could be beneficial to ensure that the images represent mostly the water shelf and not the land cover around it, and that they appear generally consistent, as attempting to acclimate the model to various forms of noise was obviously not successful. For example, the tiling for Albemarle Sound could have been guided to focus on the two inland lakes, Phelps and Mattamuskeet. With these presumed simpler tiles as a test set, model performance could potentially be claimed to generalize better.

A third recommendation towards improving the scalability of a model that is trained to recognize HABs in freshwater, is to employ satellite imagery with a higher spatial resolution. Indeed, for the Netherlands, lakes are generally extremely smaller than those observed by the NCCOS: retrieving the same images with the Sentinel 3 satellite's OLCI instrument yields about six pixels for lake Nieuwe Meer, and 1 for Kralingse Plas. Thus, for further improvements to model performance, it would be interesting to download images from the Sentinel 2 satellite which has a resolution of 10 meters per pixel, or thirty times the precision of Sentinel 3 images. Images from the new satellite could be retrieved for the same times and region delimitations as the NCCOS' Sentinel 3 images to still have access to the segmentation masks. Furthermore, this would allow implementing cloud cover corrections directly when downloading from the Copernicus browser, as it is an available feature for Sentinel 2 imagery, and an unfortunately prevalent in the Netherlands.

Parenthetically, this raises some technical considerations since resolutions between the new satellite's images and the segmentation masks are different: a 100x100 image for Sentinel 3 will cover a much larger area than the same for a Sentinel 2 image (3000 x 3000 S2 image for the same area as a 100 x 100 S3 image). It is to be explored if the model could handle having much larger input images than the actual ground truth mask, and if it would be able to learn

from having large size discrepancies. Otherwise, it is to be explored if the segmentation mask can be converted to the size of the raw image without losing excessive accuracy over which pixels represent HABs.

Relatedly, for application to the Netherlands and given the smaller average water shelf size there, it would also be of interest to drop the U-Net architecture in favour of a traditional CNN. Training data could then be retrieved by compiling Sentinel 2 images from the Copernicus browser and matching them with in-situ measurements which are made available by the different *'waterschaps'* (water departments) and the Bloomin' Algae app.

Given the current inability to generalize, it would be of interest to explore a region per region application protocol. This would ensure topographical profiles remain relatively similar and would likely yield more significant results.

Further data pre-processing could also be explored. Given that the data generally has significant class imbalance, applying techniques such as class weighting, oversampling, or reinforced data augmentation could ensure that the model learns to recognize the minority class, the HAB pixels, more robustly.

Changes to the U-Net model itself also offer strong potential for improvements. Experimenting with different hyperparameters, such as the learning rate, batch size, number of filters, or network architecture modifications with deeper layers could further improve results. Lastly, implementing post-processing techniques such as Conditional Random Fields (CRFs) might help in refining the segmentation boundaries and thus improve performance (Prasad, 2019).

Conclusion

This paper has been an exploration of the potential of using the U-Net as a new, inexpensive means of systematically monitoring HABs in the Netherlands. Given the obvious complications when scaling the application of the two models to previously unseen water shelves, it cannot be said that the U-Net CNN is sophisticated enough in it of itself to perform the ambitious task. Moreover, the proposed solution as a new tool for monitoring HABs is not very desirable given that the data it was trained on is more granular, more viable in detecting HABs and is already established as a pipeline to monitor blooms daily in the select locations, although its scope remains limited to the USA.

However, some findings were made throughout this study that continue to offer promise in the use of the U-Net as a viable, scalable and inexpensive means of monitoring HABs in the Netherlands. Indeed, it appears that having the utmost consistency in training data empowers the network to generate the most consistent and accurate predictions when deployed across a range of different regions. Thus, giving thoughtful consideration to the aspect of the images that are being trained and tested on could significantly improve the model' results. Were the images all clearly depicting the given water shelf with minimal land pixels included in the frame and the colour contrasts across regions normalized to a certain extent, the model would likely overcome the confusion it may experience in recognizing meaningful features that characterize HABs across images.

Ultimately, the objective of developing a performant and inexpensively scalable HAB monitoring tool continues to represent a valuable opportunity for the stakeholders involved, and its necessity will grow as the availability of good-quality water becomes increasingly at stake in the Netherlands and elsewhere. Therefore, efforts should be pursued to refine the proposed solution.

6) Appendices

Appendix A: Lake Pontchartrain, LA.

Appendix B: Green Bay, WI.

Appendix C: Albemarle Sound, NC.

Appendix D: Sunlight Reflection Issue

Appendix E: Corrupt Image Issue

Appendix F: Cloud Cover Issue

Appendix G: Anna Jacobapolder, Zeeland

Appendix H: All code for the project was published to a public GitHub repository. Please send an email to the author to request access.

7) Acknowledgements

Special thanks to Tiril Mageli for sharing her U-Net notebook, Sélim Behloul for sharing his satellite data pre-processing steps and insights on the project, Pr. Miquel Lurling from Wageningen University for providing some context on the state of HAB research in Europe and lastly Dr. Ellie Mackay from the UK Centre for Ecology and Hydrology and the entire Bloomin' Algae team for providing in-situ measurements of HABs collected through their app.

8) References

Caballero, I., Fernández, R. & Escalante, O.M. (2020). *New capabilities of Sentinel-2A/B satellites combined with in situ data for monitoring small harmful algal blooms in complex coastal waters*. Sci Rep 10, 8743. Available at: New capabilities of Sentinel-2A/B satellites combined with in situ data for monitoring small harmful algal blooms in complex coastal waters | Scientific Reports (nature.com)

Dewulf, L. & Mamais, L. (2022). *Water Quality in* Finland. Available at: Water-quality-in-Finland-Final.pdf (earsc.org)

ESA (2024). *Copernicus Programme*. Available at: Copernicus Programme

Europa (2024) *What is Earth Observation?* EU Agency for the Space Programme. Available at: (europa.eu)](https://www.euspa.europa.eu/european-space/eu-space-programme/what-earth-observation#:~:text=Earth Observation (EO) refers to the use of,to gather imaging data about the Earth's characteristics.)

Sangiorgi, F. & Villanueva, L. (2024). *Harmful Algal Blooms: have they really become more frequent?* NIOZ. Available at: Harmful Algal Blooms: have they really become more frequent? - NIOZ

Ho, J.C. & Michalak, A.M. (2015). *Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie*. J. Great Lakes. Accessed at: (PDF) Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie (researchgate.net)

Khan, R. M., Salehi, B., Mahdianpari, M., Mohammdimanesh, F., Mountrakis, G. & Quackenbush, L. J. (2021). *A Meta-Analysis on Harmful Algal Bloom Detection and Monitoring: A Remote Sensing Perspective*. Available at: Remote Sensing | Free Full-Text | A Meta-Analysis on Harmful Algal Bloom (HAB) Detection and Monitoring: A Remote Sensing Perspective (mdpi.com)

Kim, S. M., Shin, J., Baek, S. & Ryu, J. (2019). *U-Net Convolutional Neural Network Model for Deep Red Tide Learning Using GOCI*. Journal of Coastal Research, 302–309. Available at: U-Net Convolutional Neural Network Model for Deep Red Tide Learning Using GOCI | Journal of Coastal Research (allenpress.com)

Lee, M. S., Park, K. A., Chae, J., Park, J. E., Lee, J. S., & Lee, J. H. (2019). *Red tide detection using deep learning and high-spatial resolution optical satellite imagery*. International Journal of Remote Sensing, 41(15), 5838–5860. Available at: https://doi.org/10.1080/01431161.2019.1706011

Mageli, T. (2024). *Coursework*. Imperial College London.

McGowan, S. (2023). *Harmful Algal Blooms*. Biological and Environmental Hazards, Risks and Disasters. Available at: Harmful algal blooms - ScienceDirect

Millie, D.F., Fahnenstiel, G.L., Bressie, J.D., Pigg, R.J., Rediske, R.R. & Klarer, D.M. (2009). *Late-summer phytoplankton in western Lake Erie (Laurentian Great Lakes): bloom distributions, toxicity, and environmental influences*. Aquat. Ecol.43 (4), 915–934. Available at: Late-summer phytoplankton in western Lake Erie (Laurentian Great Lakes): bloom distributions, toxicity, and environmental influences | Aquatic Ecology (springer.com)

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M. K., Ondrusek, M., Oppelt, N. & Stumpf, R. P. (2020). *Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach*. Remote Sensing of Environment, Volume 240. Available at: Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach - ScienceDirect

Prasad, A; (2019). *Conditional Random Fields Explained*. Towardsdatasceince.com. Accessed at: Conditional Random Fields Explained | by Aditya Prasad | Towards Data Science

Ronneberger, O., Fischer, P. & Brox, T. (2015). *U-Net : Convolutional Networks for Biomedical Image Segmentation*. University of Freiburg. Available at: U-Net: Convolutional Networks for Biomedical Image Segmentation (arxiv.org)

Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S. & Adams, C. (2020). *Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing.* Earth-Science Reviews, Volume 205. Available at: Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing - ScienceDirect

Schaeffer, B. A., Bailey, S. W., Conmy, R. N., Galvin, M., Ignatius, A. R., Johnston, J. M., Keith, D. J., Lunetta, R. S., Parmar, R., Stumpf, R. P., Urquhart, E. A., Werdell, P. J. & Wolfe, K. (2018). *Mobile device application for monitoring cyanobacteria harmful algal blooms using Sentinel-3 satellite Ocean and Land Colour Instruments*. Environmental Modelling &

Software, Volume 109, pp 93-103. Available at: Mobile device application for monitoring cyanobacteria harmful algal blooms using Sentinel-3 satellite Ocean and Land Colour Instruments - ScienceDirect

UN (2024). *Goal 6: Ensure access to water and sanitation for all.* Available at: Water and Sanitation - United Nations Sustainable Development

USGS (2017). *Satellite Imagery Can Track Harmful Algal Blooms*. Available at: Satellite Imagery Can Track Harmful Algal Blooms | U.S. Geological Survey (usgs.gov)

Villars, N. (2020). *Real-time algae monitoring*. Available at: http://www.coastalwiki.org/wiki/Real-time_algae_monitoring

Viso-Vazquez, M., Acuna-Alonso, C., Rodrigez-Somosa, J. L. & Alvarez Bermudez, X. (2021). *Remote Detection of Cyanobacterial Blooms and Chlorophyll-a Analysis in Eutrophic Reservoir Using Sentinel-2.* Available at: (PDF) Remote Detection of Cyanobacterial Blooms and Chlorophyll-a Analysis in a Eutrophic Reservoir Using Sentinel-2 (researchgate.net)

Von Tress, N., Nelson, N. & Young, S. (2021). *Supporting Cyanobacterial Bloom Monitoring with Satellite Imagery*. NC State Extension Publications. Available at: Supporting Cyanobacterial Bloom Monitoring with Satellite Imagery | NC State Extension Publications (ncsu.edu)