INDEPENDENT COMPONENT ANALYSIS: A TOOL FOR ALGAL BLOOM DETECTION

C. Penne[1], *J. L. Garrett*[2], *T. A. Johansen*[2], *M. Orlandić*[1*]

[1]Department of Electronic Systems
[2]Department of Engineering Cybernetics
Norwegian University of Science and Technology,
Trondheim, Norway

R. Heggebø

Grieg Seafood
Bergen, Norway

ABSTRACT

Independent component analysis decomposition of hyperspectral signals has characteristics ideal for detecting harmful algal blooms in coastal waters. In this proceeding, independent component analysis is used to process images taken by the HYPSO-1 satellite at two different locations: Lake Erie and the Salish Sea. For each location, a set of components are produced and compared to two common water quality indices. The results demonstrate that some of the spatial-spectral features uncovered by independent component analysis resemble common water quality indices but also highlight the need for improved interpretation of the recovered features.

***Index Terms*—** Harmful algal blooms, Aquaculture, Remote sensing, Hyperspectral imaging, unsupervised machine learning

1. INTRODUCTION

Harmful algal blooms (HABs) are a major environmental threat to bodies of water all across Earth. In addition to limiting recreation, fishing, and tourism, they also pose a significant threat to the aquaculture industry. They are difficult to predict and monitor because many different species of algae can be harmful, including some that are harmful only under certain conditions. Moreover, the complexity of currents and tides complicates prediction about how nutrients that feed the blooms move and mix. HABs have caused significant economic damage, including a historic 2016 bloom in Chile that destroyed about a billion USD of fish and shellfish [1].

Nonetheless, it is possible to respond to algal blooms when they occur near certain aquaculture sites. The simplest strategy is to change the feeding pattern at a given site so that the fish stay at a depth below the algal bloom, but this is only viable for the shallowest blooms and can only be sustained for a short duration. If the bloom harms through oxygen depletion, then the water can be oxygenated so that the fish can still breathe. In extreme cases, if there is sufficient advance warning, the fish can be moved to another site. These methods are viable for species such as salmon, but are not feasible to employ on species such as shellfish, which are less mobile.

The unpredictability of algal blooms means that they are one of the biggest threats to aquaculture industry. Since they are not only difficult to predict, but can also be catastrophic, they complicate the insurance of the industry, which could in turn lead to harmful impact on the industry and local communities.

Fig. 1. Independent component analysis is applied to two images from the HYPSO-1 satellite: the Salish Sea, August 30, 2022 (Top), and Lake Erie, August 27, 2022 (bottom). Left is north.

Because of the threat from algal blooms, several programs have been established across the world to monitor for algal blooms. For instance, the Norwegian government administered a series of aquaculture-funded buoys along the coast from about 2000 to 2010 [2]. However, around 2010 the system was decommissioned because it was found to cost more than it contributed to the fish farmers. Thus, the warning system in place during the 2019 HABs in Northern Norway was not sufficiently effective, and lead to high mortality in the farmed salmon populations and heavy losses for the farming companies [3].

Hyperspectral remote sensing is a useful technology for detecting algal blooms near the ocean surface, when there is little cloud coverage. It combines the spatial coverage of re-

*Thanks to the Research Council of Norway and industry partners (Green Platform, grant no. 328724 and ARIEL, grant no. 333229) as well as Norway Grants 2014 – 2021 (ELO-Hyp, contract no. 24/2020) for funding.

mote sensing together with the spectral resolution necessary to characterize algal pigments. However, it produces vast amounts of data which require computational assistance for humans to analyze. The problem of detecting algal blooms has several properties which make it difficult for standard machine learning algorithms, which will be discussed below. Attempts to avoid these difficulties and extract information about algal blooms from hyperspectral images include defining empirically-motivated indices and principal component analysis [4, 5]. In this proceeding, we propose that independent component analysis (ICA) family of algorithms is well-suited to for detecting algal blooms and show examples of how it could be done using radiance-calibrated images from the HYPSO-1 satellite [6].

2. ALGAL BLOOM DETECTION COMPLICATIONS

The detection of algal blooms is a difficult problem for several reasons.

- **Ground Truth** is typically challenging to collect for water bodies because they are dynamic and thus measurements are only valid for a brief time. This presents a particularly large problem for validating satellite data, which generally passes a location no more than a few times a day, and often less.
- **Many algal species** lead to harmful algal blooms. A system designed to detect one species might not be able to detect others.
- **Ecosystem interactions** can modify the behavior of algae. For instance, *C. leadbeateri*, the algae species which caused the 2019 Norway bloom is not always harmful, and is typically present at low levels [3].
- The **spatial extent and dynamics** of algal blooms is not known a priori. Therefore, it is difficult to know where and when to perform water sampling.

3. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis (ICA) refers to a family of methods for identifying signals in high-dimensional data sets [7, 8, 9, 10]. What unites these methods is that they seek signals characterized by a particular statistical distribution which characterizes independence. Independence is characterized by inverting the central limit theorem, which states that the sum of many signals approaches a Gaussian distribution. Inverted, this suggests that if a signal is non-Gaussian then it is not the sum of many signals, and may even have a distinct physical origin.

ICA methods, combined with hyperspectral remote sensing, are appropriate for meeting the challenge of algal blooms. First of all, since they are unsupervised, they are not inhibited

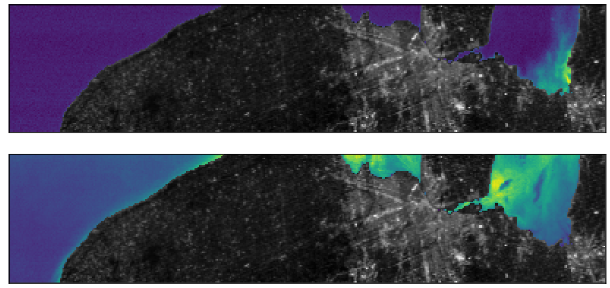

Fig. 2. Lake Erie cyanobacteria index (top) and CDOM index (bottom). The scale goes from blue to yellow.

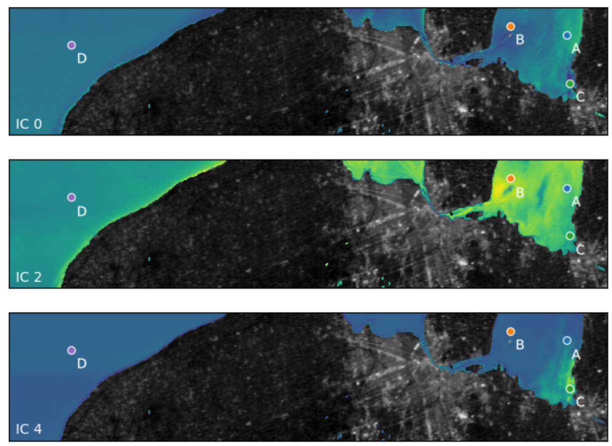

Fig. 3. Independent components from Lake Erie scene. Green and yellow indicates additional signal from the component relative to the mean value, while blue indicates less than the mean.

by limited ground truth, unlike supervisory machine learning algorithms with require significant amounts of labelled data. Second, ICA produces information about the spectra which it detects, in addition to spatial maps. The found spectra could then be compared to a repository of known algal spectra to determine which types of algae are present, and what their condition could be. Finally, ICA applied to hyperspectral images produces a spatial map of where a particular signal occurs, so once an algal bloom is detected, additional measurements on the ground can be used for further inspection.

ICA can complement the band-ratio algorithms which are commonly used as markers for algal blooms. One way to combine ICA with band-ratio algorithms would be to use a band-ratio algorithm to detect a property such as chlorophyll, and then use ICA to look for signatures of additional spectral peaks within the chlorophyll-rich region. Another way would be to develop band-ratios for the most common types of algae, and then subsequently apply ICA to identify any blooms which may have been missed by the first set of algorithms.

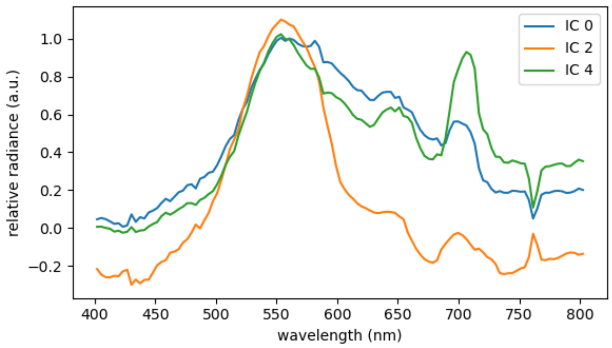

Fig. 4. The mixing spectra from the top (blue), middle (orange), and bottom (green) signals in Fig. 3.

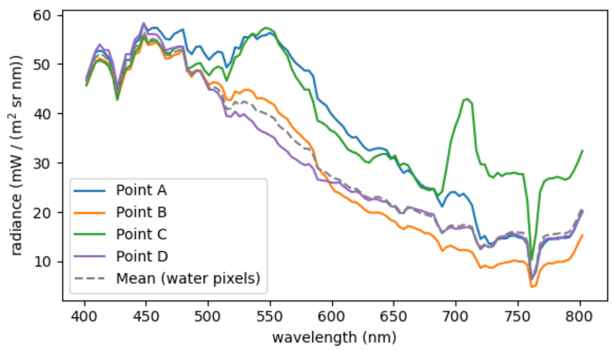

Fig. 5. The top of atmosphere (ToA) radiance spectra of pixels at various relative radiance values selected from the independent components in Fig. 3. The dashed grey line indicates the mean radiance values from unmasked water pixels.

Mathematically, ICA finds an unmixing matrix W, with the same number of bands as the original data, that has the property $S = XW$, where X is the original dataset and S contains the recovered non-Gaussian signals, known as independent components (ICs). The mixing matrix, W^+, is then the pseudo-inverse of the unmixing matrix, such that $SW^+ \approx X$, or, more precisely,

$$XW = SW^+W. \tag{1}$$

Physically, the mixing matrix can be understood as what spectra that come from the different components, and the un-mixing matrix is a tool to extract that component from the background.

4. LAKE ERIE SCENARIO

Because large, well-characterized algal blooms occur yearly on Lake Erie, it was used as the first test of the ICA algorithms. The bloom is dominated by *Microcystis aeruginosa*, which produces hepatotoxin microcystin, a poison. Because the bloom shut down the water supply to the city of Toledo, Ohio in 2014, and has been recurring for decades, it has been considered a model system for studying HABs and how to monitor them [4, 11].

The signals recovered by ICA were compared to standard indices to interpret the components. These indices have been validated through their long-term application to multi-spectral data, but do not utilize the fine spectral resolution of HYPSO-1's sensor. The fluorescence line height (FLH) at 681 nm can reveal the presence of photosynthesizing phytoplankton.

$$\mathrm{FLH} = \Bigg(R(681) - R(665) - (R(709) - R(665)) \times \frac{681 - 665}{709 - 665} \Bigg), \tag{2}$$

where R indicates the radiance at a particular wavelength. However, the *Microcystis aeruginosa* in Lake Erie tend to cause an inversion at the wavelengths relevant to the FLH because it fluoresces little and scatters more light above 700 nm. Thus, the cyanobacteria index (CI), which is commonly used to monitor the intense algal blooms on Lake Erie [4], is defined as:

$$\mathrm{CI} = -\,\mathrm{FLH}, \tag{3}$$

which has been shown to indicate the presence of a *Microcystis aeruginosa* bloom [12]. The second standard index, a colored dissolved organic matter (CDOM) index, based off an algorithm developed for Landsat, is calculated from the ratio of the signal at 560 nm to the signal at 658 nm [13]. Since these are calculated from the radiance rather than reflectance, they only give a general image of the variation within the image, and are not calibrated to particular values.

One variant of ICA, FastICA [7], was applied to the water pixels in the scene. It searched for five components, of which three (IC 0, IC 2, and IC 4) are spatially plotted in Fig. 3. The mixing spectra from these components are plotted in Fig. 4. The ToA radiance spectra from four different points selected from the Lake Erie scene in Fig. 3 are plotted in Fig. 5. The components showed spectral signals which appear reminiscent of algae (Fig. 6).

All the signals showed a strong peak around 550 nm, but the components IC 0 and IC 4 show additional signals above 600 nm. A strong peak around 709 nm appears in both IC 0 and IC 4, but is more notable in the latter. Indeed, the area which they cover shows a signal in the cyanobacteria index. The component in IC 2 roughly corresponds spatially to the CDOM index. One other component showed a strong signal at wavelengths above 750 nm, and was thus likely related to the the shallowness of water and its proximity to land.

Spatially, the components in IC 0 and IC 4 resemble different portions of the region highlighted by the cyanobacteria index in Fig. 2. However, ICA divides the region into two

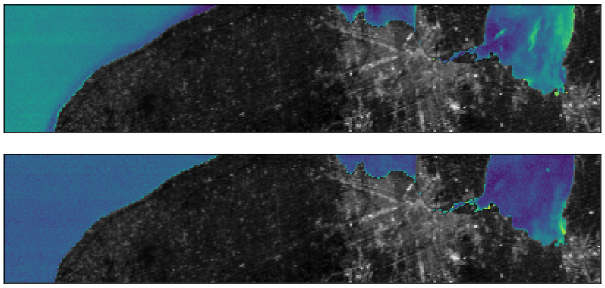

Fig. 6. Spatial variation of an extra peak at 588 nm, plotted as radiance at 588 nm divided by the radiance at 547 nm (top). Spatial variation in the chlorophyll fluorescence peak near 700 nm, plotted as the difference radiance difference between 709 nm and 699 nm. Both features were found through the use of ICA.

portions, one with a higher reflectance around 600 nm and another with larger signal above 700 nm, which could be indicative of a more intense algal bloom [14]. Although the biological origin of this varying signal is not immediately clear, plots of these features do replicate the spatial pattern of ICA (Fig. 6). Thus, this illustrates how ICA can give additional context to the output of extensively tested band-ratio algorithms. The spatial distribution of the second component closely follows that found by the CDOM band ratio algorithm.

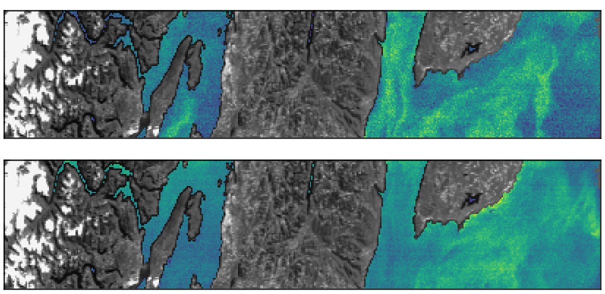

Fig. 7. Salish Sea fluorescence line height and CDOM indices.

5. SALISH SEA SCENARIO

The coasts of the Salish Sea, along the west coast of British Columbia, Canada, are home to many aquaculture sites. There are frequent algal blooms in the region, although blooms as long-lasting or prominent as the bloom on Lake Erie have not yet been recorded here by HYPSO-1. Thus, it presents a chance to test whether ICA can detect less prominent blooms. Of the five ICA components which are found, two appear to be possibly related to algal blooms (Fig. 8), while the peaks in the other three are more difficult to distin-

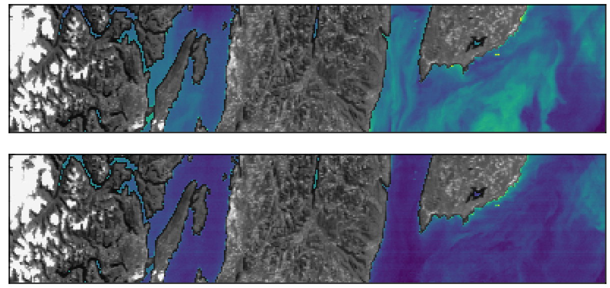

Fig. 8. Spatial representation of the two ICA components that appear similar to the spatial patterns in the fluorescence line height and CDOM indices.

guish from noise. In the first component, which has a similar spatial distribution to the FLH, two peaks can be seen, one of which is near the 685 nm fluorescence peak of chlorophyll. The second component also shows a smooth, clear peak in the radiance.

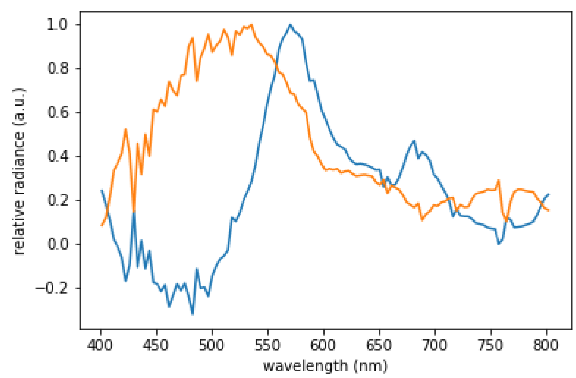

Fig. 9. The ICA spectra which produce spatial patterns similar to the FLH (blue) and the CDOM index (orange) in the Salish Sea scene.

6. DISCUSSION

In this proceeding, we have shown that ICA is able to detect regions of water bodies with unusual spectra and to report what those spectra are. However, the interpretation of those spectra has also been found to be quite difficult in practice. There are two ways then to improve the capacity of identifying algal blooms with ICA. In the first option, ICA itself could be altered to take on an even more interpretable functional form. For example, it could be made to search for band ratios rather than linear transformations. The second option is to focus on the interpretation of the linear coefficients that ICA finds. For example, they could be compared to a spectral database, or a repository of reference spectra could be collected from known algal bloom events.

Even with these improvements ICA faces several challenges. First of all, signals from different sources can get mixed together. For example, the increased near infrared (NIR) signal from shallow water is often present in at least one of the ICA components, and can interfere with the detection of algal blooms near the shore. Second, with the standard ICA, it is not clear whether the mixing or un-mixing weights should be inspected. Here, we have found that the mixing weights show the absorption peaks and valleys more clearly, but it is not clear that this is generally the case.

7. CONCLUSION

In this proceeding, it has been demonstrated that ICA is able to detect spatial-spectra features in hyperspectral images. It excels in searching for spectral details which simple band-ratio algorithms do not detect. However, like principal component analysis, it also suffers from the mixing of signals, and while the mixing weights do provide insight into the detected signal, additional interpretation is still required. Therefore, ICA could function as a complement to more standard band-ratio algorithms in a HAB monitoring program, to look for sub-portions of the blooms that were found and to search the scene for any remaining undetected blooms.

Acknowledgements

We thank Emil Lindfors, John Constantino, Mike Anderson, Trond Kathenes, and Glaucia Fragoso for helpful conversations.

8. REFERENCES

[1] D Anderson, J Rensel, J Forster, and S Hart, "Harmful algae blooms: assessing Chile's historic HAB events of 2016," *Global Aquaculture Alliance Report*, 2016.

[2] G Johnsen and E Sakshaug, "Monitoring of harmful algal blooms along the Norwegian coast using bio-optical methods," *South African Journal of Marine Science*, vol. 22, no. 1, pp. 309–321, 2000.

[3] U John, L Šupraha, S Gran-Stadniczenko, C Bunse, A Cembella, W Eikrem, J Janouškovec, K Klemm, N Kühne, L Naustvoll, et al., "Spatial and biological oceanographic insights into the massive fish-killing bloom of the haptophyte Chrysochromulina leadbeateri in northern Norway," *Harmful Algae*, vol. 118, pp. 102287, 2022.

[4] R P Stumpf, T T Wynne, D B Baker, and G L Fahnenstiel, "Interannual variability of cyanobacterial blooms in Lake Erie," *PLOS ONE*, vol. 7, no. 8, pp. 1–11, 08 2012.

[5] J D Ortiz, D M Avouris, S J Schiller, J C Luvall, J D Lekki, R P Tokars, R C Anderson, R Shuchman, M Sayers, and R Becker, "Evaluating visible derivative spectroscopy by varimax-rotated, principal component analysis of aerial hyperspectral images from the western basin of lake erie," *Journal of great lakes research*, vol. 45, no. 3, pp. 522–535, 2019.

[6] S Bakken, M B Henriksen, R Birkeland, D D Langer, A E Oudijk, S Berg, Y Pursley, J L Garrett, F Gran-Jansen, E Honoré-Livermore, et al., "HYPSO-1 cubesat: First images and in-orbit characterization," *Remote Sensing*, vol. 15, no. 3, pp. 755, 2023.

[7] A Hyvärinen and E Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[8] P Ablin, J F Cardoso, and A Gramfort, "Faster independent component analysis by preconditioning with Hessian approximations," *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4040–4049, 2018.

[9] D Lupu, I Necoara, J L Garrett, and T A Johansen, "Stochastic higher-order independent component analysis for hyperspectral dimensionality reduction," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 1184–1194, 2022.

[10] H Li, S Yu, and J C Príncipe, "Deep deterministic independent component analysis for hyperspectral unmixing," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3878–3882.

[11] J C Ho and A M Michalak, "Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie," *Journal of Great Lakes Research*, vol. 41, no. 2, pp. 317–325, 2015.

[12] T Wynne, R Stumpf, M Tomlinson, R Warner, P Tester, J Dyble, and G Fahnenstiel, "Relating spectral shape to cyanobacterial blooms in the laurentian great lakes," *International Journal of Remote Sensing*, vol. 29, no. 12, pp. 3665–3672, 2008.

[13] J Chen, W N Zhu, Y Q Tian, and Q Yu, "Estimation of colored dissolved organic matter from Landsat-8 imagery for complex inland water: case study of Lake Huron," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, pp. 2201–2212, 2017.

[14] J Gower, S King, G Borstad, and L Brown, "Detection of intense plankton blooms using the 709 nm band of the meris imaging spectrometer," *International Journal of Remote Sensing*, vol. 26, no. 9, pp. 2005–2012, 2005.