European Journal of Sustainable Development (2025), **14**, 4, 283-294 ISSN: 2239-5938
Doi: 10.14207/ejsd.2025.v14n4p283

Harmful Algal Blooms and Wind-related Variables Association

By Georgia Xanthopoulou[1], Dionysios Kechagias[1], Konstantinos Koukaras[2], Katerina Aligizaki[3], Charalampos Bratsas[4]

ABSTRACT:
Harmful algal blooms (HABs) are rapidly expanding in marine environments, posing serious risks to ecosystems, public health, and water quality. There is a growing need for more accurate forecasting techniques and intelligent systems that can generate accurate forecasts about the presence of HABs and, as a result, evaluate the effects of them on water quality. Direct forecasts of HABs presence are extremely difficult to be implemented due to the nature of the problem. Many studies use chlorophyll concentrations to detect HABs but such an alteration in chlorophyll concentration consists an immediate effect and not the root of the problem. Trying to early detect a forthcoming HAB, by investigating the impact of wind-related variables in the appearance of a HAB event in the marine system of Thermaikos gulf (NW Aegean Sea) with in-situ data, is the purpose of this research. Primary results have shown a causation among them, especially to east and north-west winds, leading to research paths towards to the spatial distribution of the phenomenon. An outcome that demonstrates the potential transporting of the HAB and also could assist in the creation of early warning systems.

Keywords: HABs, wind, causal inference, prediction, early warning

1. Introduction and related work

Photosynthetic organisms such as planktonic microalgae are the life force of aquatic ecosystems. They produce oxygen, fix carbon and form the foundation of food webs. Nevertheless, under some circumstances the abundance of their population can increase to the point where it could be harmful to people as well as other aquatic species. These growths are commonly called ”Harmful Algal Blooms” (HAB). According to the type of algae, it is possible in some cases to spot an algal bloom by its color; this leads to at least a harmful aesthetic impact, affecting e.g. tourism and the local community (Red tide). However, not all algal blooms occur on the surface of the water body having a direct visual impact. Fish and other species may perish as a result of HABs’ reduction of water’s oxygen content, harming the ecosystem. In other cases, microalgae, which are capable of producing polysaccharides and thus form dense layers of mucus in the water column or on the surface, are likely to grow, with adverse effects on many sectors, such as fisheries, aquaculture, tourism, and, where the phenomena are severe, may kill the flora and fauna of the marine area where they occur (Nikolaidis et al., 2008).

[1] Center of Research and Technology Hellas – CERTH/ Information Technology Institute (ITI), Greece
[2] Center of Research and Technology Hellas – CERTH/ Institute of Applied Biosciences (INAB), Greece
[3] Aristotle University of Thessaloniki, Greece
[4] Department of Information and Electronic Engineering of the International Hellenic University, Greece

During a HAB, people can get exposed to toxins from fish and shellfish they may consume, from swimming in or drinking the water, and from the air they breathe (Anderson et al., 1995, Theodorou et al., 2020). It is an indisputable fact that an exposure to toxins usually has no antidote (Grattan et al., 2016). For this reason, it is of paramount importance to have early predictions to achieve prevention and mitigation measures. A topic of particular interest is the identification of the factors/variables that affect the HAB event. HABs may be caused by a variety of variables, but research is still ongoing to determine how these environmental parameters interact to produce a harmful algal bloom. Climate change is an indisputable factor that impacts the HABs (Wells et al., 2015, Gobler et al., 2020, Petoussi et al., 2011), along with the subsequent increase in temperature (Silva et al., 2023, Aboualaalaa et al., 2022). A fact that is also highlighted from the report (IPCC, 2022) of the Intergovernmental Panel on Climate Change (IPCC). Even though HABs are a natural phenomenon, predicting them remains difficult. As one can observe in the simplest case of the recording of the dinoflagellates of the genus *Gambierdiscus* in the Mediterranean, the distribution range of these species was previously thought to occur in tropical and subtropical areas. The worrying aspect of this expansion is that these dinoflagellates are the causative agents of Ciguatera disease, which is caused in humans by the consumption of fish that have accumulated toxins produced by the cells of the genus *Gambierdiscus* (Aligizaki et al. 2008).

Many studies attempt to predict HABs using supervised learning methods, such as classification and support vector machines, while others exploit statistical models. Some others used a more promising class of forecasting models, the hybrid methods that are also used to forecast algal blooms as mentioned by Giddings et al. (2014).

While others, trying to work with issues that occur when dealing with HABs forecasts, like data availability. Nevertheless, hybrid methods are the focus of recent studies; Giddings et al. used a combination of models to reduce the number of false positive events (80% accuracy), while Liu et al. (2022) combined a wavelet transformation and an LSTM network to create a method that can reliably forecast the algal dynamics on multiple time scales. A comparison between hybrid machine learning models is the work of study of Molares et al. (2023). Their study verified that hybrid models provide more accurate results than simple techniques. This verification has been accomplished by implementing the models proposed in other studies since there is no reference data set in the prediction of HAB that allows an objective comparison of the models.

As already been mentioned, one of the many issues that occur when dealing with HABs forecasts, is the data availability. Most of the times, researchers have to deal with large number of missing values or limitations on the number of variables that are available for process, etc. That is the reason why most of the studies focus on indirect forecasts relevant to the next possible algal bloom event, meaning that they monitor several water characteristics and give an indirect forecast about the time of an algal event peak, without providing any information about the percentage of the peak. Harley et al. (2020) used environmental data to classify shellfish above and below a threshold, achieving an accuracy of approximately 80%. Wind speed and satellite (chlorophyll) data, aiming to classify bloom or no bloom events were used in the study of Silva et al. (2023). The same data combination of chlorophyll images and wind direction was utilized in the conceptual forecasting model in the study of Cusack et al. (2016). As for Kleindist et al. (2014) used historical shellfish harvesting data to predict bloom severity levels aiming to define major storms. Wind circulation was exploited by Raine et al. (2010) to predict algal events according to natural wind circulation in the bays of Ireland. To achieve their predictions, they also used a five-day weather forecast. Gonzales et al. (2014) also used meteorological data to validate their SVM models and identify if there was a bloom or not and forecast the algal occurrence indirectly. Weather variables and more specifically wind speed and direction were also used by Grifoll et al., (2011) in order to early forecast a forthcoming pollution in harbours. According to Grifoll, the hydrodynamics of a receptor are important and highly connected to the problem. A strong interaction between oceanographic, meteorological, and biological variables in relation to the behavior of phytoplankton abundance is also highlighted in the works of Sandoval et al. (2018). They also mention that the understanding of these phenomena requires the understanding of the physiological adaptations of these microalgae, meaning the understanding of which are the variables that are directly related to the HAB at the national, regional and local levels. Moreover, Yang et al. (2024) created two forecasting models that could reconstruct the conditions on the surface highlighting the importance of dynamic factors and specifically the wind. Marine meteorological factors such as wind and atmospheric pressure were used and in the study of Kang et al. (2023). Apart from the data availability issue, many either investigate the spatial or temporal dynamics of the problem and simultaneously neglect the underlying biological root of the problem. By seamlessly incorporating both the spatial and temporal dimensions alongside an understanding of biological causality, a more holistic and explainable forecasting approach emerges. This integration enables the identification of intricate patterns and relationships between environmental variables over time and space, aligning with the underlying biological mechanisms governing phenomena of interest.

Given the fact that many studies have used machine learning or hybrid methods to predict HABs, it is indisputable that they struggle to deal with limited data and the complexity of environmental systems. While various meteorological factors like temperature, rainfall, and wind have been linked to HABs, most existing work focuses on correlations rather than clear explanations. In this study, we investigate the possible effect of wind speed and direction on the appearance of HAB events and show that these blooms are changing over time and space.

2. Data

In situ data of the marine environment are collected in the area of the port of Nea Michaniona, Greece (Thermaikos gulf, NW Aegean Sea). In the area of Nea Michaniola, data come from sensors installed at the outer pier of the port from the Department of Environmental and Hydrology of Central Macedonia Region. The available data are from 14/10/2022 with a total of 79,185 records (to date) and with a collection frequency of every 15'. The data collected are related to physico-chemical characteristics of the water such as water temperature, salinity, etc. as well as meteorological variables such as wind speed, wind direction, humidity, etc.

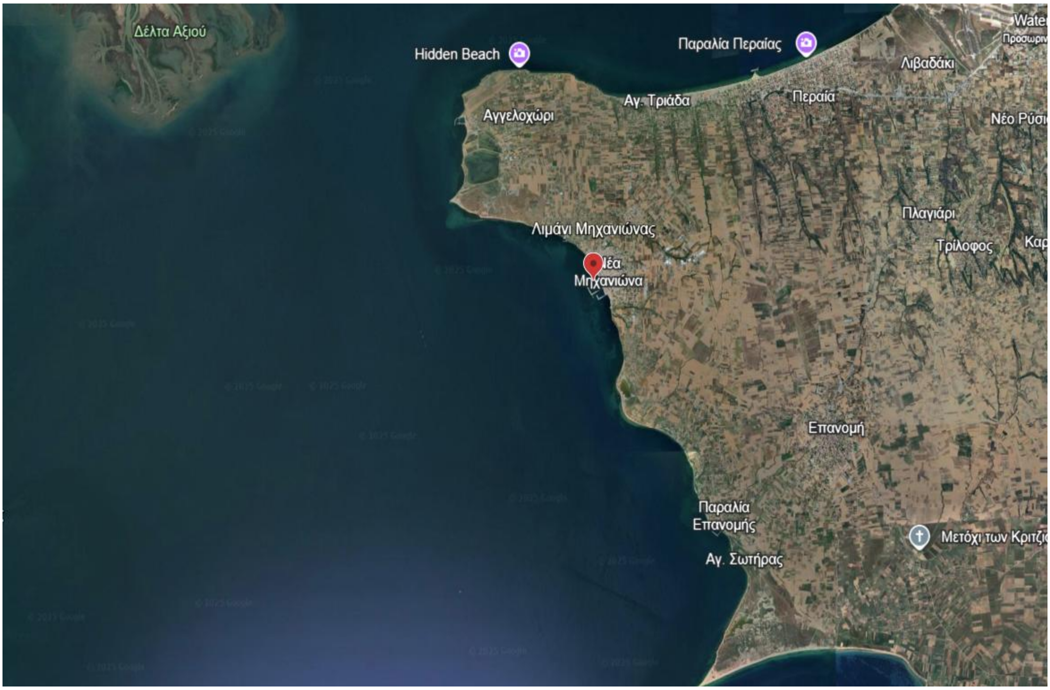

Figure 1: Location of the sensor in Nea Michaniona harbor (Thermaikos gulf, NW Aegean Sea)

Apart from in situ data, for the purpose of this study the Laboratory Unit of Marine Toxic Microalgae of the Aristotle University of Thessaloniki provides data on the presence and abundance of potentially toxic and/or harmful microalgae. In a first phase, the study is processing populations of the dinoflagellate *G.* cf. *hyalina*, which is associated with the production of polysaccharides and the formation of often extensive mucoid aggregations in marine areas, causing a multitude of problems in anthropogenic activities. The data concerning harmful microalgae have mainly a weekly measurement frequency.

The variables that are collected are the following:

Table 1: Variables description

N/O	Description
0	AIR PRESSURE (mbar)
1	Battery Voltage (V)
2	EXO BATTERY (V)
3	Phycoerythrin (μg/l)
4	CHLOROPHYL (μg/l)
5	DO (mg/l)
6	DO_% (%)
7	TDS (mg/l)
8	TSS (mg/l)
9	CONDUCTIVITY_ACT (μS)

10	CONDUCTIVITY_SP (μS)
11	SALINITY (PSU)
12	RELATIVE HUMIDITY (%)
13	AIR TEMP (°C)
14	TURBIDITY (NTU)
15	WATER TEMPERATURE (°C)
16	WIND DIR avg (°)
17	WIND DIR gust (°)
18	WIND SPEED avg (m/s)
19	WIND SPEED gust (m/s)

3. Results

In order to investigate possible relationships between the variables we implemented correlation analysis and causal inference. Correlation analysis is the statistical method that determines if there is a relationship between two variables and how much one variable changes the other. It is expressed as a correlation coefficient ranging from -1 to 1. A value close to 1 indicates a strong positive relation, a value close to -1 a strong negative one. A value close to 0 indicates weak on no correlation. More specifically, we used Pearson correlation coefficient, which measures the strength of the linear relationship between two variables. When correlation coefficient is applied to a sample is represented in equation (1):

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

Where n is sample size, x_i, y_i are the individual sample points indexed with i, $\bar{x}$ the sample mean and analogously for $\bar{y}$.

The results of the correlation analysis are shown in the following picture. As we observe there are linear relationship between phycoerythrin and chlorophyll, between dissolved oxygen (DO) and conductivity, between air temperature and water temperature, etc. In the heatmap, dark red or dark blue shows the strongest relationship between the variables. These colors represent values close to +1(strong positive) or -1(strong negative) respectively.

Even though there are correlations between variables, there is not a direct linear correlation to the variable of interest, the abundance of *G.* cf. *hyalina*. Causal inference is the process of identifying and quantifying the causal effect of one variable on another. Unlike correlation, causation suggests a cause-and-effect relationship between two variables. In causal inference, the first step is to formulate a hypothesis and then test it with statistical methods. To explore the causality in this study we implemented a **Directed Acyclic Graph (DAG),** meaning a directed graph with no directed cycles, in which nodes represent the variables and the edges directed from one vertex to another, such that following those directions will never form a closed loop. The DAG that was constructed based on the knowledge of the domain, along with correlation analysis results, as seen in Figure 3.

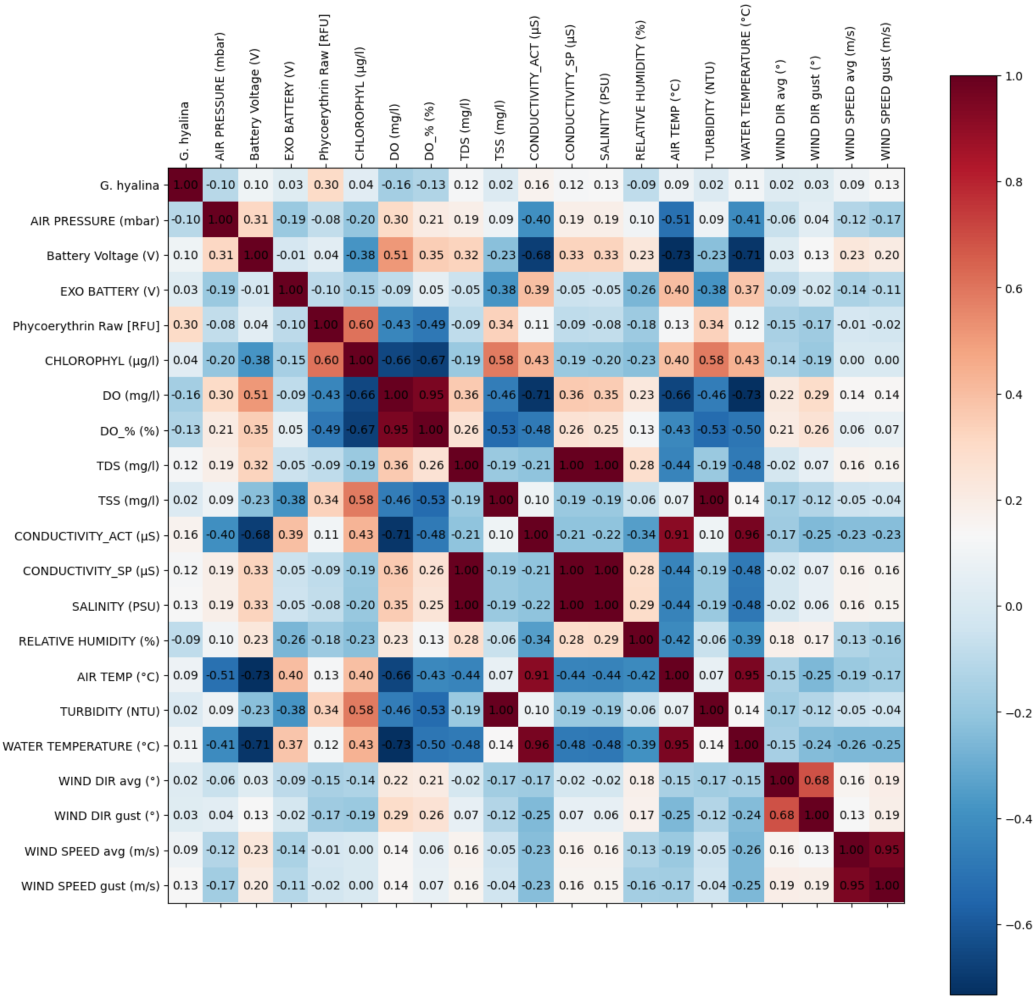

Figure 2: Correlation analysis heatmap

After the construction of the DAG and the use of it to define the causal model, we determine which variables (estimands) need to be controlled to estimate the causal effect, the appearance of *G.* cf. *hyalina*. The results indicate that the phenomenon appears to be influenced by wind speed and as a result, it could be spread in nearby locations. Particularly, the estimation of the causal inference showed in the following expression:

$$\frac{d}{d\left[Wind\ speed\ avg\ \left(\frac{m}{s}\right)\right]}(E[G.hyalina]) \tag{2}$$

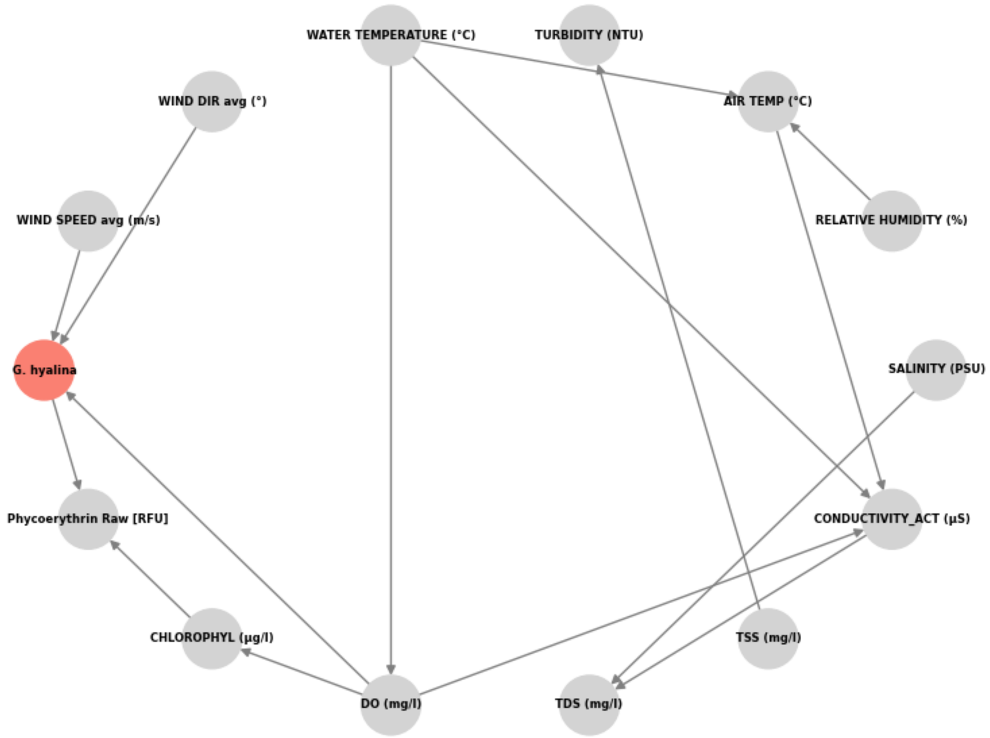

Figure 3: Directed acyclic graph

Once the estimand is defined (a.k.a. wind speed), we selected a suitable statistical method to estimate it. We used linear regression and the results showed that the estimated causal effect is 1.53, meaning that if the "*treatment*" (wind speed) is changed by 1 unit, the "*outcome*" (*G.* cf. *hyalina*) will be increased by 1.53 units on average due to the causal effect. Moreover, in order to validate the results we performed a refutation test; specifically we added a random cause to ensure that the estimated effect is robust. The refutation showed a minimal change in the estimated effect as presented in Table 2, which indicates that the result is robust and reliable.

Table 2: Causal inference – Refutation test results

Type of effect	Value
Estimated effect	1.53
Effect after refutation	1.52

Knowing that wind speed affects the appearance of the phenomenon, the hypothesis that wind direction plays also a significant role is something to be investigated. Firstly, in Figure 4 we can observe the peaks of abundance of *G.* cf. *hyalina*, the variable of interest, from October 2022 to December 2024. As we observe, there is no periodicity in the data and as we have already mentioned that, it is influenced by wind speed. In Figure 5, we can observe that wind speed increases shortly, preceding spikes in *G.* cf. *hyalina* concentrations, while during the peaks there is a decrease in wind velocity.

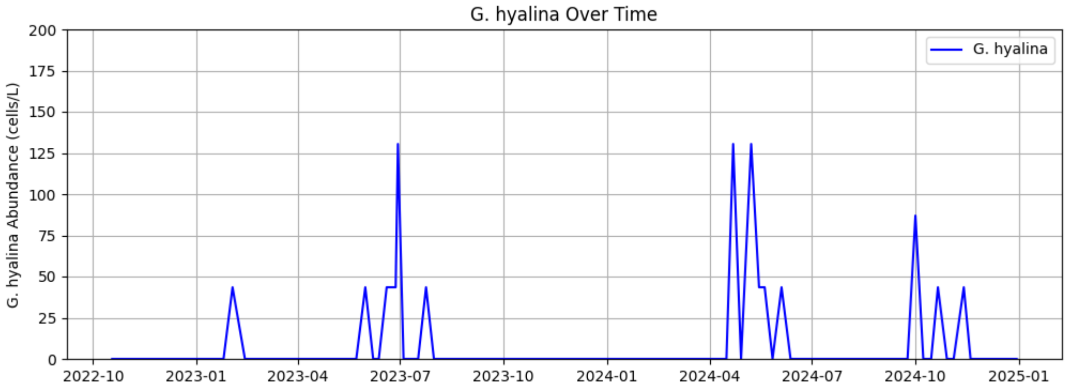

Figure 4: Gonyaulax – hyalina abudance

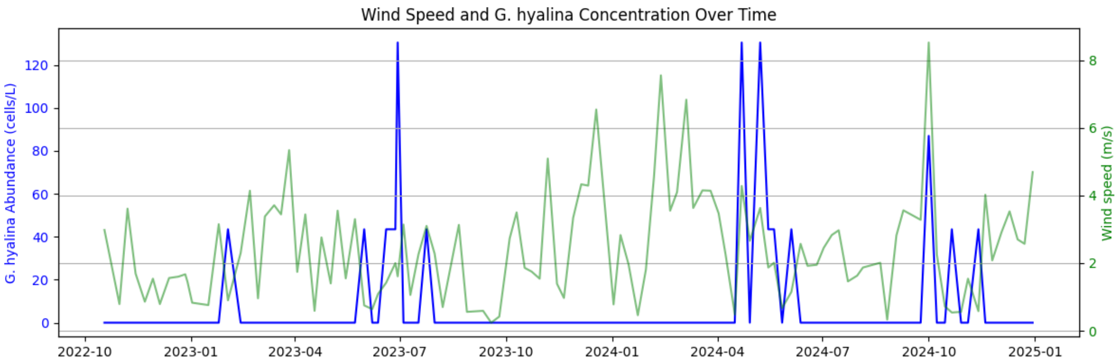

Figure 5: Gonyaulax hyaline abudance and wind speed time series

The observed patterns suggest that the initial position of the phenomenon is not fixed but instead shifts according to prevailing wind conditions, leading to noticeable spatial changes over time. Regarding the relationship of *G.* cf. *hyalina* and wind direction, Figure 6 is a wind rose plot that shows that *G.* cf. *hyalina* concentrations vary according to wind direction. When wind is blowing from Eastern direction, *G.* cf. *hyalina* population has the highest concentration, while blowing from North or North-West direction, has also high concentration but lower than when blowing from the East. Winds from other directions (West, South-West, South-East) were associated with much lower or negligible concentrations of this species.

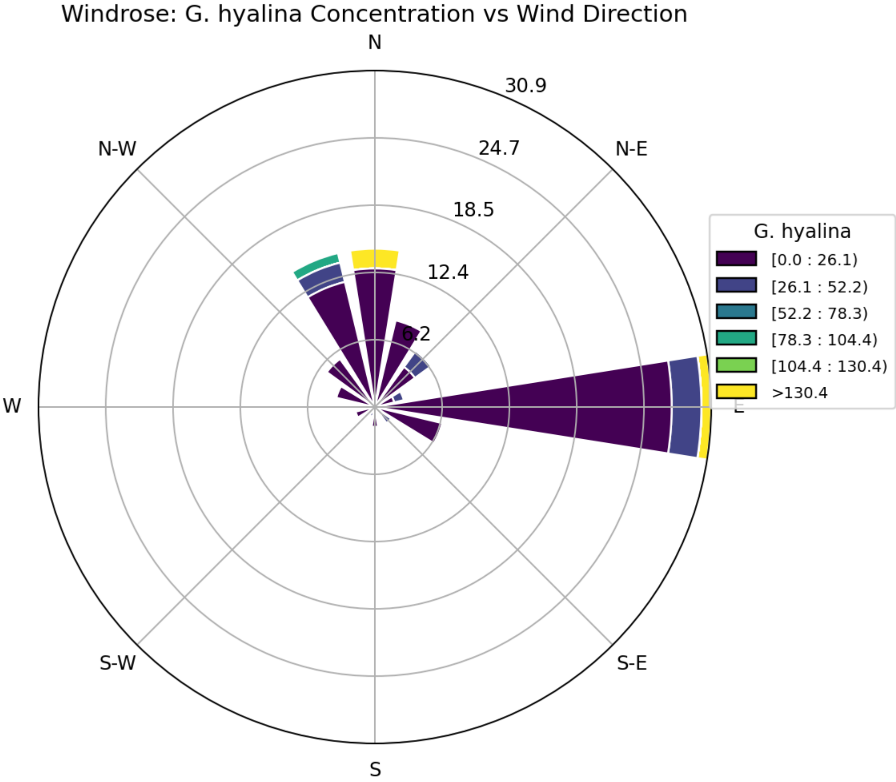

Figure 6: Windrose plot

4. Conclusion

One of the many issues that occur when dealing with HABs forecasts, is data availability, complex ecological interactions and the indirect approach of the most forecasting studies. In this study, there is a HAB research by focusing on *G.* cf. *hyalina*, a species of interest whose sporadic appearances and elevated concentrations are difficult to predict using traditional methods.

Most of the times, researchers have to deal with large number of missing values or limitations on the number of variables that are available for process, etc. In the present study, we utilized the most complete dataset available, nevertheless, future efforts should incorporate advanced imputation techniques (e.g., multiple imputation, deep learning approaches) and data fusion strategies that integrate satellite, in-situ, and historical datasets. Another limitation, even though HABs forecasts are studied in an abundance of papers, is the lack of studies on the pathway of toxin accumulation in shellfish and the geomorphology of the marine area where the events take place. In this study, by applying a causal inference approach, we identified a robust causal relationship between wind-related variables and *G.* cf. *hyalina* concentrations. Specifically, our analysis revealed that both wind speed and wind direction significantly influence the temporal and spatial patterns of *G.* cf. *hyalina* occurrence. Although linear regression offers interpretability in estimating causal effects within a DAG-based framework, it may not capture nonlinear interactions inherent in ecological systems. Future studies could apply non-parametric estimators (e.g., generalized additive models, causal forests) or causal discovery algorithms (e.g., PC or FCI) to learn causal structure and improve robustness. In addition, temporal overlays demonstrated that spikes in abundance often follow periods of intensified wind activity, while windrose analysis further indicated that winds from the East and Northwest directions are particularly associated with high *G.* cf. *hyalina* levels. While this study highlights wind-driven spatial patterns of *G. cf. hyalina*, it does not explicitly model physiological or ecological mechanisms behind bloom formation. Future work should integrate key biological variables, including nutrient availability, light conditions, and species-specific life history traits such as cyst germination and growth thresholds. Moreover, it is in our purposes to incorporate spatial-temporal simulations (e.g., wind-driven diffusion models) to better capture the dynamic propagation of *G. cf. hyalina* blooms. Another indisputable fact is that the study is restricted to *G.* cf. *hyaline* within a specific region. This fact is limiting the generalizability to other HAB species or ecosystems. This focus was dictated by the availability of consistent and well-documented historical observations. Nonetheless, future research should aim to expand monitoring efforts to include multiple species and leverage satellite data.

All the above findings demonstrate the potential transporting of the *G.* cf. *hyalina*, leading to the conclusion that predictive models could assist in early-warning systems and strategies. Future research should translate these insights into a real-time monitoring and forecasting system by integrating environmental sensors, predictive modeling, and decision-support interfaces. These tools should be rigorously assessed for forecasting accuracy, cost-effectiveness, and usability to support aquaculture and coastal management under HAB threat conditions. Lastly, the ability to generate a public data set with a complete time series not only would allow for novel approaches to HAB prediction but also would provide an objective comparison of the models investigated in other studies (Molares et al., 2023).

Acknowledgment: This work was supported by the Project SeaGuard (ID: 1221ΠΚΜ001ΚΑΠ23) which is financed by the Prefecture of central Macedonia/Greece

References

Aboualaalaa, H., el kbiach, M. L. B., Rijal Leblad, B., Hervé, F., Hormat-Allah, A., Baudy, L., Ennaskhi, I., Hammi, I., Ibghi, M., Elmortaji, H., Abadie, E., Rolland, J. L., Amzil, Z., & Laabir, M. (2022). Development of harmful algal blooms species responsible for lipophilic and amnesic shellfish poisoning intoxications in southwestern Mediterranean coastal waters. In *Toxicon* (Vol. 219). Elsevier Ltd. https://doi.org/10.1016/j.toxicon.2022.09.002

Aligizaki K., Nikolaidis G. & Fraga S., (2008). Is *Gambierdiscus* expanding in new areas? *Harmful Algal News,* 36: 6-7.

Anderson, D. M. (1995). Toxic red tides and harmful algal blooms: A practical challenge in coastal oceanography. *Reviews of Geophysics*, *33*(2 S), 1189–1200. https://doi.org/10.1029/95RG00440

Cusack, C., Dabrowski, T., Lyons, K., Berry, A., Westbrook, G., Salas, R., Duffy, C., Nolan, G., & Silke, J. (2016). Harmful algal bloom forecast system for SW Ireland. Part II: Are operational oceanographic models useful in a HAB warning system. *Harmful Algae*, *53*, 86–101. https://doi.org/10.1016/j.hal.2015.11.013

Giddings, S. N., Maccready, P., Hickey, B. M., Banas, N. S., Davis, K. A., Siedlecki, S. A., Trainer, V. L., Kudela, R. M., Pelland, N. A., & Connolly, T. P. (2014). Hindcasts of potential harmful algal bloom transport pathways on the Pacific Northwest coast. *Journal of Geophysical Research: Oceans*, *119*(4), 2439–2461. https://doi.org/10.1002/2013JC009622

Gobler, C. J. (2020). Climate Change and Harmful Algal Blooms: Insights and perspective. *Harmful Algae*, *91*. https://doi.org/10.1016/j.hal.2019.101731

González Vilas, L., Spyrakos, E., Torres Palenzuela, J. M., & Pazos, Y. (2014). Support Vector Machine-based method for predicting Pseudo-nitzschia spp. blooms in coastal waters (Galician rias, NW Spain). *Progress in Oceanography*, *124*, 66–77. https://doi.org/10.1016/j.pocean.2014.03.003

Grattan, L. M., Holobaugh, S., & Morris, J. G. (2016). Harmful algal blooms and public health. In *Harmful Algae* (Vol. 57, pp. 2–8). Elsevier B.V. https://doi.org/10.1016/j.hal.2016.05.003

Grifoll, M., Jordà, G., Espino, M., Romo, J., & García-Sotillo, M. (2011). A management system for accidental water pollution risk in a harbour: The Barcelona case study. *Journal of Marine Systems*, *88*(1), 60–73. https://doi.org/10.1016/j.jmarsys.2011.02.014

Harley, J. R., Lanphier, K., Kennedy, E., Whitehead, C., & Bidlack, A. (2020). Random forest classification to determine environmental drivers and forecast paralytic shellfish toxins in Southeast Alaska with high temporal resolution. *Harmful Algae*, *99*. https://doi.org/10.1016/j.hal.2020.101918

Intergovernmental Panel on Climate Change (IPCC). (2022). The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009157964

Kang, D., Kim, B. K., Jung, S. W., Baek, S. H., Choi, J. Y., Cho, H. Y., Lee, S. J., & Kim, H. (2023). Development and Application of an Integrated System for the Detection and Prediction of Harmful Algal Blooms in Korea. *Journal of Marine Science and Engineering*, *11*(12). https://doi.org/10.3390/jmse11122207

Kleindinst, J. L., Anderson, D. M., McGillicuddy, D. J., Stumpf, R. P., Fisher, K. M., Couture, D. A., Hickey, J. M., & Nash, C. (2014). Categorizing the severity of paralytic shellfish poisoning outbreaks in the Gulf of Maine for forecasting and management. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *103*, 277–287. https://doi.org/10.1016/j.dsr2.2013.03.027

Liu, M., He, J., Huang, Y., Tang, T., Hu, J., & Xiao, X. (2022). Algal bloom forecasting with time-frequency analysis: A hybrid deep learning approach. *Water Research*, *219*. https://doi.org/10.1016/j.watres.2022.118591

Molares-Ulloa, A., Rivero, D., Ruiz, J. G., Fernandez-Blanco, E., & de-la-Fuente-Valentín, L. (2023). Hybrid machine learning techniques in the management of harmful algal blooms impact. *Computers and Electronics in Agriculture*, *211*. https://doi.org/10.1016/j.compag.2023.107988

Nikolaidis, A., Aligizaki, K., Koukaras, K., & Moschandreou, K. (2008). Mucilage phenomena in the North Aegean Sea, Greece: another harmful effect of dinoflagellates*?* Proceedings of the 12th International Conference on Harmful Algae.

Petoussi, M. A., Koukaras, K., & Kalogerakis, N. (2011). A simple mathematical model for real-time prediction of harmful algal blooms in Greece. Proceedings of the 12th International Conference on Environmental Science and Technology.

Raine, R., McDermott, G., Silke, J., Lyons, K., Nolan, G., & Cusack, C. (2010). A simple short range model for the prediction of harmful algal events in the bays of southwestern Ireland. *Journal of Marine Systems*, *83*(3–4), 150–157. https://doi.org/10.1016/j.jmarsys.2010.05.001

Sandoval, M., Parada, C., & Torres, R. (2018). Proposal of an integrated system for forecasting harmful algal blooms (HAB) in Chile. *Latin American Journal of Aquatic Research*, *46*(2), 424–451. https://doi.org/10.3856/vol46-issue2-fulltext-18

Silva, E., Counillon, F., Brajard, J., Pettersson, L. H., & Naustvoll, L. (2023). Forecasting harmful algae blooms: Application to Dinophysis acuminata in northern Norway. *Harmful Algae*, *126*. https://doi.org/10.1016/j.hal.2023.102442

Theodorou, J. A., Moutopoulos, D. K., & Tzovenis, I. (2020). Semi-quantitative risk assessment of Mediterranean mussel (*Mytilus galloprovincialis* L.) harvesting bans due to harmful algal bloom (HAB) incidents in Greece. *Aquaculture Economics and Management*, *24*(3), 273–293. https://doi.org/10.1080/13657305.2019.1708994

Wang, Y., Xu, C., Lin, Q., Xiao, W., Huang, B., Lu, W., Chen, N., & Chen, J. (2024). Modeling of algal blooms: Advances, applications and prospects. In *Ocean and Coastal Management* (Vol. 255). Elsevier Ltd. https://doi.org/10.1016/j.ocecoaman.2024.107250

Wells, M. L., Trainer, V. L., Smayda, T. J., Karlson, B. S. O., Trick, C. G., Kudela, R. M., Ishikawa, A., Bernard, S., Wulff, A., Anderson, D. M., & Cochlan, W. P. (2015). Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae*, *49*, 68–93. https://doi.org/10.1016/j.hal.2015.07.009

Yang, G. G., Wang, Q., Feng, J., He, L., Li, R., Lu, W., Liao, E., & Lai, Z. (2024). Can three-dimensional nitrate structure be reconstructed from surface information with artificial intelligence? — A proof-of-concept study. *Science of the Total Environment*, *924*. https://doi.org/10.1016/j.scitotenv.2024.171365